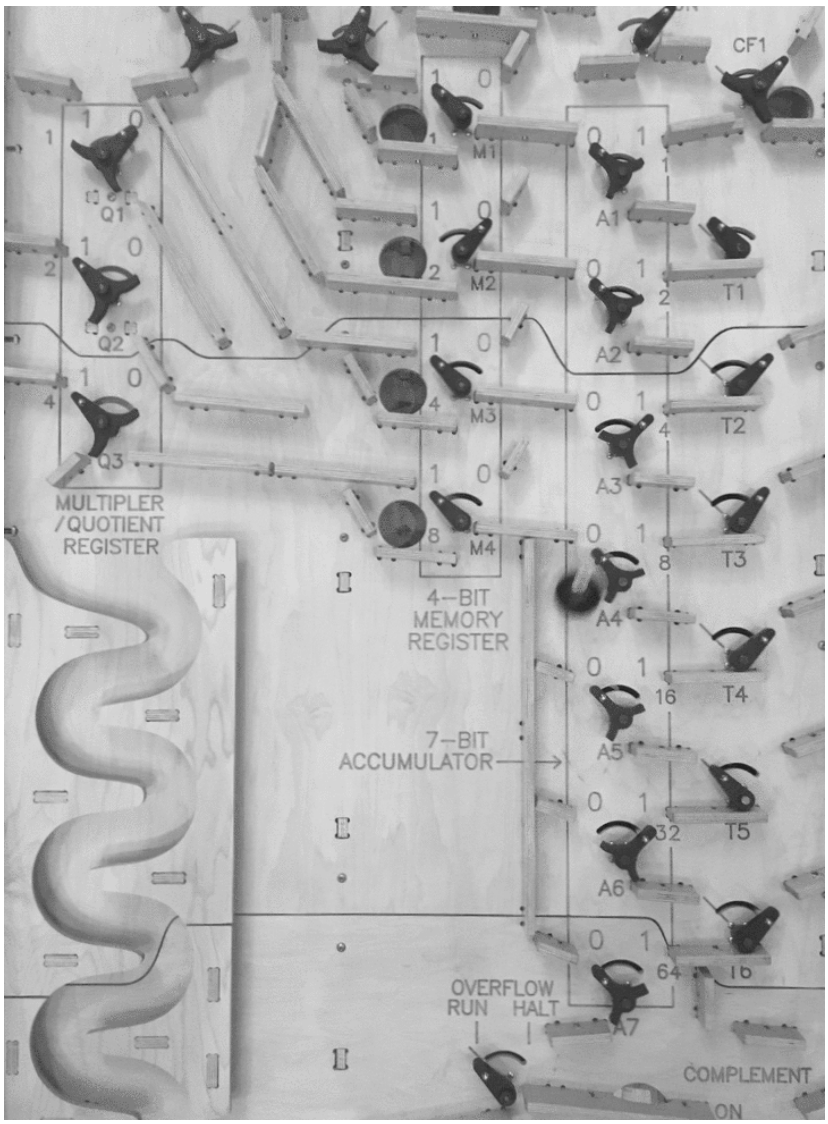
Measuring the Impact of Public Policy on Socioeconomic Disparities in Health

Bastian Ravesteijn



ISBN: 9789051708073

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

Cover image: This is a picture of the Digi Comp II at the Massachusetts Institute of Technology, photographed by the author.

This book is no. 656 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Measuring the Impact of Public Policy on Socioeconomic Disparities in Health

Het meten van de effecten van beleid op sociaaleconomische gezondheidsongelijkheid

Proefschrift

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

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

donderdag 19 mei 2016 om 15:30 uur

door

Bastian Ravesteijn
geboren te *Haarlem*

Erasmus University Rotterdam

Promotiecommissie

Promotor:

Prof.dr. E.K.A. van Doorslaer

Overige leden:

Prof.dr. M. Lindeboom

Prof.dr. O.A. O'Donnell

Prof.dr. H.D. Webbink

Copromotor:

Dr. J.L.W. van Kippersluis

Contents

Chapter 1

Introduction

Health is an important determinant of human well-being. On average, individuals on the lower rungs of the societal ladder have worse health and shorter lives compared with those at the top. This dissertation describes several pathways through which these socioeconomic disparities in health come to be, and estimates the effects of public policies aimed to reduce these disparities. I investigate the role of occupational conditions, retirement policies, educational policies, and health insurance in shaping differences in health.

Health disparities are shaped by both choice and circumstance. While a moral case can be made in favor of government policies correcting differences in circumstance, it is less obvious that public policy should compensate health losses that result from freely chosen unhealthy behavior. The focus of this dissertation is on the disparities that are caused by factors that are exogenous to individuals. However, it is not always possible to identify and target individuals with adverse circumstances. Targeting policy at unhealthy individuals could both help those who are unhealthy at no fault of their own, but it also indirectly rewards unhealthy behavior.

To understand the causes of socioeconomic disparities in health, consider the opportunities and constraints that people face when making choices that affect health. Individuals may choose to to invest in current and later health through healthy behaviors (Grossman, 1972a). Conversely, they can voluntarily choose to "lose" health by engaging in unhealthy behaviors for which the benefits to utility are greater than the losses. For example, a healthy diet and exercise can be viewed as investment in health, while smoking or drinking can be seen as ways to give up health in exchange for the utility gains of consuming these unhealthy goods (Becker and Murphy, 1988; Cutler and Glaeser, 2005).

Apart from the way in which individuals directly derive utility from health, it can be viewed as a form of human capital in the production process (Galama and Van Kippersluis, 2015). In chapters 2 and 3 of this dissertation my co-authors and I show that

healthy workers are more productive and that good health increases earnings, both contemporaneously and in the future. Like any other form of capital, health capital is a stock that depreciates over time but can be increased by investment. However, investment in health capital is not necessarily positive: individuals can divest in health when they choose a (physically or psychosocially) "hard" occupation which is harmful to health, in return for greater earnings (Viscusi, 1978b).

Even though health is affected by choices, individuals face different constraints when making these choices. Every person is faced with endowments: factors that limit the choice set and that are determined exogenously. Endowments can take many forms: they may be genetic, result from conditions during gestation (Scholte et al., 2015), during child birth or throughout (early) childhood (Van den Berg et al., 2006; Currie, 2009; Conti et al., 2010). Intelligence, cognitive ability and personality traits result from these endowments (Borghans et al., 2008) and place limits on an individual's choice set: they limit the combinations of consumption and health over the life cycle that one can obtain by making behavioral choices. Chapters 4 and 5 discuss how educational policies can reduce the importance of these endowments in determining later-life outcomes.

One can distinguish between deterministic endowments, such as being born blind or missing a limb at birth, and endowments that shape the ex ante probability distribution of negative and positive shocks to health. For example, a genetic disorder (which can be viewed as an endowment) increases the probability of onset of a disease, even if the disease has not manifested itself. In countries with functioning capital markets, individuals can insure themselves against the financial risk of a negative health shock. However, in systems where insurance premiums can be made conditional on risk profiles, insurance protects against ex ante risk, but it does not necessarily reduce disparities due to ex ante imbalances in endowments if individuals with higher risks face higher insurance premiums. Chapter 6 shows how an increase out-of-pocket expenditures for mental health health care affects access to care for subpopulations with different degrees of vulnerability.

Empirically, socioeconomic status is strongly and positively associated with health (Mackenbach et al., 2008; Meara et al., 2008). This holds for one's own educational attainment, earnings, and occupational class, and for a variety of measures of morbidity, mortality, and self-assessed health. Parental background is a strong predictor of later life health outcomes as well, suggesting that—exogenously determined—circumstance plays an important role in determining these disparities in health. Whether a society is willing to accept disparities that result from differences in endowments, is subject to normative considerations.

A Rawlsian perspective on health disparities highlights the role of exogenous deter-

minants of health that are not controlled by the individual, such as endowments and shocks. These exogenous determinants can for example be related to genetic or cultural differences, or plain bad luck. A Rawlsian social welfare function—which equates social welfare with the utility of the worst-off in society—favors levelling the initial playing field at birth (i.e. endowments, following Roemer, 2009). In contrast, there is no clear rationale for public policy to offset health disparities that derive from voluntarily giving up health to obtain other benefits (i.e. effort, as argued by Cawley and Ruhm, 2011).

The chapters of this dissertation evaluate the degree to which a variety of different policies have been or can be successful in reducing health disparities due to exogenous ("unfair" according to Fleurbaey and Schokkaert, 2009) factors. In doing so, I explicitly take into account behavioral responses and possibly unintended consequences of these policies.

1.1 Occupation and health

In the second chapter, together with my co-authors I outline a theoretical utility maximization model to shed light on the choices and exogenous factors that together determine health. While the association between socioeconomic status and health is well-established, less is known about the mechanisms underlying this relationship. One possible pathway is that parental background influences educational decisions, which in turn influence career opportunities throughout the lifetime. We investigate the relationship between occupational characteristics and health in the Netherlands and confirm a pervasive association between occupational class and both subjective and objective (e.g. mortality) health measures. We then survey the literature and find that while the strong association is found in other studies as well, causal effect estimates are scarce.

In the third chapter, we construct measures of occupational stress using two German data sources to investigate the effect of specific occupational characteristics on health. However, selection into certain occupations is probably not random: if physically strong individuals select into physically demanding occupations, a simple comparison is not informative about the causal effect of occupation on health. This would make it impossible to formulate policy implications to give people long and healthy careers. One reason for the lack of causal estimates in the literature is that few cases exist where occupational selection is exogenous. We estimate a dynamic panel data model to eliminate as much selection bias as possible given our rich panel data, but in absence of exogenous variation in occupational choice. We link the resulting estimates and corresponding identifying assumptions to a dynamic model of occupational selection and health investment to interpret the results.

1.2 Schooling and health

Investigation of the transmission mechanisms of causal effects in the economics literature—such as the degree to which the health effect of parental SES is transmitted by schooling—has often relied on the comparison of regression coefficients before and after an intervening control variable was introduced. However, even though nowadays most economists are aware of the problems due to endogeneity of the regressor of interest, the consequences of controlling for an endogenous intervening control variable have not been well-understood. In the fourth chapter, I show that the estimators of direct and indirect effects of the regressor of interest are—potentially severely—biased if we control for endogenous "bad" intervening control variables. I mathematically derive the bias of these estimators and provide an illustration of the bad controls problem based on the relationship between parental SES, schooling and health. I show that using exogenous variation in schooling can solve the bad controls problem using two UK compulsory schooling reforms in the middle of the twentieth century . Additionally, I simulate a data-generating process and compare different methods to illustrate the bias of the bad controls method.

In the fifth chapter of this dissertation, we investigate the importance of parental background in determining health and health disparities. Since children do not control their parental socioeconomic status (SES), some individuals may have less opportunities at birth. One important policy lever to improve equality of opportunity is education: education can diminish the adverse effects of bad endowments, which could correct disparities that resulted from a weak family background. We evaluate the effect of the Finnish compulsory schooling reform in the 1970s which meant that the tracking age—the age at which students are tracked into different educational streams—was extended from 11 to 16. We investigate how the effect of the reform varies by parental background and look at educational attainment, measures of health care utilization, and the ultimate health indicator: mortality. Since several factors may have coincided with the reform, we use the fact that some regions were reformed later than others—which allows us to account for region and cohort effects—in a differences-in-differences setup.

The chapters in this dissertation shed light on the various causes, consequences, and remedies of socioeconomic disparities in health. For each particular research question, we use the appropriate econometric tools given the data at hand. We evaluate the strengths and weaknesses of these methods, welcoming the insights from the so-called credibility revolution in economics that emphasizes the importance of internal validity in the scientific economic literature. This forces us to be upfront about the identifying assumptions in each paper, improving transparency and limiting the role of technical

assumptions in driving the results. The role of the economic researcher then becomes more modest, which is what it should be: to estimate the implications of different policy proposals, while leaving it to the rest of society to choose between these proposals by weighting the benefits to some against the losses to others.

Chapter 2

The Contribution of Occupation to Health Inequality

Chapter Abstract

Health is distributed unequally by occupation. Workers on a lower rung of the occupational ladder report worse health, have a higher probability of disability and die earlier than workers higher up the occupational hierarchy. Using a theoretical framework that unveils some of the potential mechanisms underlying these disparities, three core insights emerge: (i) there is selection into occupation on the basis of initial wealth, education, and health, (ii) there will be behavioural responses to adverse working conditions, which can have compensating or reinforcing effects on health, and (iii) workplace conditions increase health inequalities if workers with initially low socioeconomic status choose harmful occupations and don't offset detrimental health effects. We provide empirical illustrations of these insights using data for the Netherlands and assess the evidence available in the economics literature.

2.1 Introduction

Health and mortality are distributed unequally by occupation (Mackenbach et al., 2008). For example, Smith et al. (1998) show that in the UK, those in the highest occupational classes had a 70 percent lower mortality rate over a 21-year period than those in the bottom occupational class. Not only mortality rates differ: Case and Deaton (2005a) find that in the US those employed in manual occupations self-report

This chapter is based on Ravesteijn, B., H. van Kippersluis, and E. van Doorslaer (2013). The contribution of occupation to health inequality. In Research on economic inequality, Volume 21, pp. 311–332. Emerald Group Publishing Limited.

lower health than those who work in professional occupations, and that their health declines more rapidly with age. Galama and Van Kippersluis (2015) similarly find that among working Dutch males, health differences across manual and non-manual workers widen until around early retirement age and decline thereafter. Does this evidence mean that lower skilled occupations exert a higher health toll and thereby contribute to observed health disparities by socioeconomic status (SES)?

In this chapter we review the literature on whether occupation affects health and, if so, to what extent this contributes to socioeconomic inequalities in health. The discussion is illustrated with descriptive evidence from the Netherlands, which provides an interesting context in which there is a heated ongoing debate on whether certain 'hard' occupations should be exempted from a proposed rise in the statutory retirement age, and where it only recently became possible to link occupational information to administrative registers containing mortality and disability figures. Our assessment of the literature is guided by a theoretical framework, which indicates that it is vital to understand the association between occupation and health along three dimensions.

First, it is unclear whether the strong and persistent association between occupation and health derives from a direct, causal effect of occupation on health. Alternatively, the association between occupation and health could stem from health enabling and/or limiting factors that induce individuals to self-select into certain types of occupation. A third possibility is that predetermined characteristics such as education or initial endowments affect both occupation and health, implying that health inequalities by occupation are simply a reflection of 'deeper' societal inequalities across socioeconomic groups. Even if a causal effect of occupation on health is established, it is essential to understand which occupational characteristics are most important in producing these health differentials. Is it simply the manual aspect of the job in terms of hard physical labour, or are health differences mostly caused by the psychosocial stressors, such as low job control and high work load? The answers to these questions have implications for policies that aim to prolong working lives and provide foundations for possible differentiation of the statutory retirement age on the basis of individual occupational histories.

Second, individual choices are not made in isolation. Decisions regarding occupational choice are made simultaneously with decisions regarding health investment and consumption. Hence, workers may (partially) offset the occupation-related damage to their health by investing in health, or may add to this health risk by engaging in unhealthy types of consumption.

Third, given a heavily constrained occupational choice set, workers with worse endowments may choose to "sell" part of their health by engaging in harmful occupations, in return for higher earnings. If this health risk is not fully compensated, it implies that

occupation could exacerbate socioeconomic inequalities in health. If workers with poor endowments are more likely to choose such harmful occupations and are not fully compensated financially, there is scope for compensating labour market policies to offset health inequalities resulting from "hard" occupations.

Our review suggests that the evidence on the causal impact of occupation on health – in contrast to evidence on the association – is fairly thin. Identification of causal effects is particularly hampered by the sheer difficulty of finding suitable sources of exogenous variation in occupational and working conditions. Recent contributions do suggest that, while health differences across occupations largely reflect health-based selection, at least some part of the effect runs through physical working conditions affecting health outcomes, thereby exacerbating socioeconomic inequalities in health.

The paper is organized as follows. In the next section we document patterns of occupation, health and longevity in the Netherlands. In the third section we describe the basics of an economic model of health behaviour across the life cycle which can help us understand occupational differences in health. In the penultimate section we review the scientific evidence guided by the core insights derived from the theory. The final section concludes.

2.2 The association between occupation and health in the Netherlands

The association between occupation and self-reported health, as well as other measures of morbidity, has been widely documented, both internationally and for the Netherlands (Cavelaars et al., 1998; Mackenbach et al., 2008). There is some evidence of an association between occupation and mortality (Kunst et al., 1990), but less on the association between occupation and the onset of disability (e.g. Currie and Madrian, 1999). We add to this with evidence from the Netherlands, made possible through the recent linking of surveys and administrative registers by Statistics Netherlands.

We use the Dutch Permanent Survey of Living Conditions (Dutch acronym POLS) which has been linked to the registers since 1997. POLS is a repeated cross section survey which includes questions on measures of self-assessed health (SAH) and a measure of type of occupation based on the Dutch Standard Classification of Occupations (SBC, 1992). We use observations of individuals in the years 1997 through 2006, 115,888 of whom report an occupational title. The earliest cross-sections are much larger, with the first four years accounting for 70 percent of observations.

We distinguish between five major occupational groups: elementary, low-level, mid-level, high-level and university-level occupations (cbssbc2012). The 1992 SBC occu-

pational classification is based on the skill level that is required for each occupation.[1] Examples of elementary occupations include conveyor belt workers or cleaners; low-level occupations include lumberjacks, miners and construction workers; mid-level occupations include foremen, building contractors or mechanics; high-level occupations include primary school teachers, mid-level managers, and head nurses; university-level occupations include medical doctors, accountants, and architects.[2] We investigate the association between occupational type and three measures of health. We find that workers in a "higher" type of occupation on average report better health, are less likely to become disabled and live longer than workers in lower ranked occupations.

2.2.1 Occupation and self-assessed health

In the survey years 1997 to 2001, respondents were asked to assess their health on a five-point scale from poor to very good. We restrict the sample to working-age individuals between 20 and 65 years old, which is the current legal retirement age. Figure 2.1 shows the health self-reports, uncontrolled for age and gender, for workers in five occupational groups, for those on disability benefits, and for those not working. Only a very small proportion of people on disability benefits reports good health. People on disability benefits report worse health than others who are not in paid employment. The figure shows that health status monotonically improves with higher levels of occupation: 81 percent of elementary workers report good or very good health as opposed to 90 percent of those in high-level and university-level occupations

Figure 2.2 shows the proportion in good or very good health of individuals in the five occupational groups at different ages. It is striking that already at age 25 marked differences in health are observed across occupational groups. Since occupation is unlikely to have had much of an effect on health already at that age, this strongly suggests that there is health-related selection into occupations. The health disparities between occupational groups increase with age, suggesting rapid health deterioration among workers in the lower occupational groups, yet it should be kept in mind that these are not life cycle profiles and hence could reflect cohort effects, selective promotion between occupational groups, selective mortality and other sources of confounding. The occupational health gradient by age is somewhat steeper for men than for women (not shown).

[1] The idea here is to present associations of an occupational measure and health outcomes. While we acknowledge that this classification is likely to pick up not just the effects of occupation, but additionally the effects of education and other individual characteristics, similar issues would plague other classifications of occupation into e.g. blue and white collar jobs.

[2] Due to changes in the coding procedure, 17,897 observations of occupational titles had to be excluded from the analyses

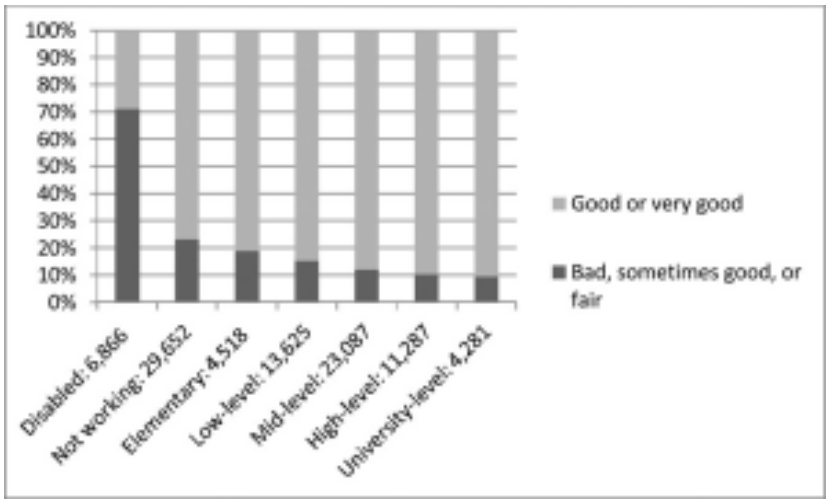


Figure 2.1: Occupation and self-assessed health.

Notes: Self-assessed health by disability, employment and occupation. Source: POLS 1997–2006.

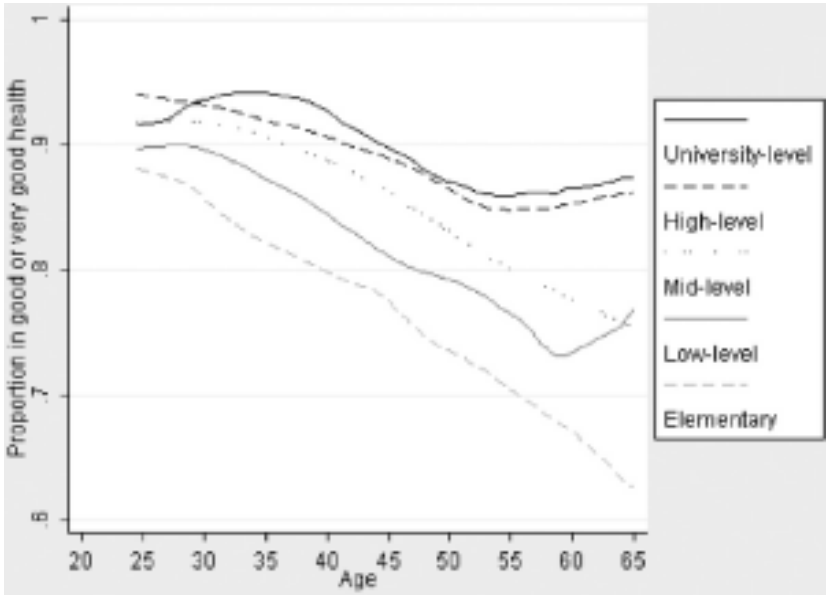


Figure 2.2: Proportion of workers in good or very good health by occupation and age (Netherlands POLS surveys 1997–2006)

2.2.2 Occupation and disability

In the Netherlands, people who become unable to work because of health reasons qualify for a disability benefit. Using a linkage with longitudinal data from the Social Statistics File (Dutch acronym SSB), we can follow up all workers aged 20 to 65 observed in the POLS survey data to identify those moving into disability (until 2006). This allows analysis of the duration until exit out of a disability-free spell and into disability of all individuals who reported to be working at the time of the POLS survey. We take into account the left-truncation resulting from the fact that we start to observe individuals at different ages and we take into account right-censoring because of death, reaching the retirement age of 65 or the fact that we observe disability only until 2006. We estimate a Cox proportional hazard model of the duration until exit to disability. The base category is elementary work and the likelihood of exiting into disability monotonically decreases with level of occupation. Taking into account the 95 percent confidence intervals, individuals in low-level occupations are between 17 and 33 percent less likely to exit into disability than elementary workers at any age. Individuals in mid-level occupations are between 38 and 51 percent less likely to exit to disability. For individuals in high-level occupations this is between 45 and 57 percent and for university-level occupations this is between 56 and 70 percent.

Figure 2.3 shows the proportion of the working population that survives disability-free until the age of 65. At the age of 55, already 18 percent of elementary workers have exited to disability while only 7 percent of university-level workers are receiving disability benefits. Also very striking is the monotonicity of the gradient: with every step down the occupational ladder, the risk of disability increases significantly. The survivor function flattens after the age of sixty, most likely reflecting the increased likelihood of exiting the labour force into early retirement instead of disability.

2.2.3 Occupation and mortality

Occupational disparities in health are also reflected by differential survival. While these have been documented for many countries, hitherto they could not be examined for the general Dutch population because of absence of mortality data by occupation. Linkage of the POLS survey to the Cause of Death registry (Dutch acronym DO) enables analysis of the duration until death of individuals by occupational status. As for disability, our duration analysis accounts for left-truncation and we also account for right censoring because we observe mortality only until 2010.

Again, we estimate a Cox proportional hazard model of the duration until exit due to death. At any age, individuals receiving disability benefits are more than twice as likely to die in the period of observation compared to individuals who were

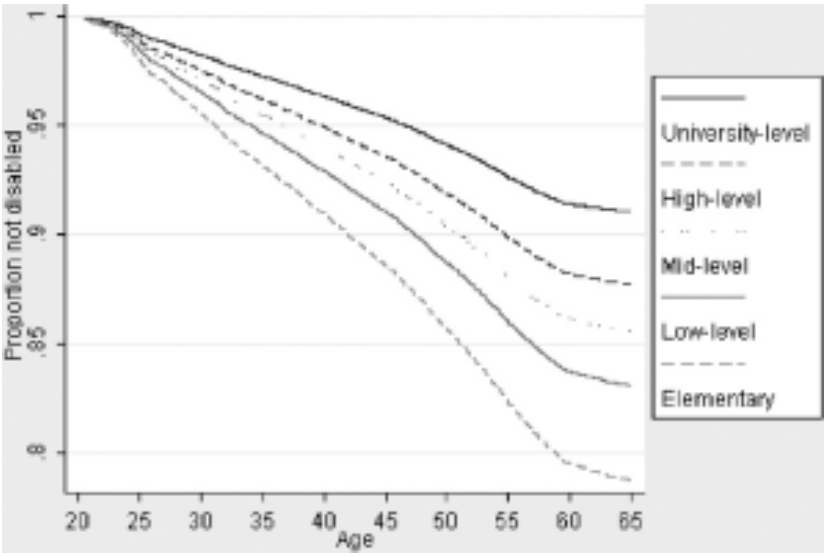


Figure 2.3: Survival until exit to disability by occupation.

Notes: Cox proportional hazard model. Source: POLS and SSB follow-up until 2006.

in an elementary occupation when they were observed in the POLS survey. In our sample, taking into account the 95 percent confidence intervals, individuals in low-level occupations are between 7 and 23 percent, those in mid-level occupations are between 20 and 32 percent, those in high-level occupations are between 32 and 44 percent, and university-level occupation even between 45 and 57 percent less likely to die at any given age compared to elementary workers.

Figure 2.4 shows the estimated survival curves by occupation. It can be seen that, at the age of 65, more than 20 percent of individuals who were on disability benefits have passed away, around 13 percent of elementary workers, and around 7 percent of university-level workers. In the coming years, the legal retirement age in the Netherlands will be increased to 67. At that age, and not taking into account any increase in life expectancy, 16 percent of elementary workers will have passed away, as opposed to 8 percent of individuals in university-level occupations. It is clear that both the probability of reaching retirement age, as well as the survival chances beyond that age, decrease monotonically with lower occupations.

All in all, the analysis confirms that in the Netherlands – as in many other industrialized countries – patterns of morbidity, disability and mortality differ by occupation. Health and survival prospects generally improve monotonically when moving up the occupational ladder. This raises the question of the extent to which these disparities reflect selection of healthier individuals into higher occupations, as opposed to these occupations offering a health advantage. In the case that the differences do reflect a causal effect of occupation on health, it is still of crucial importance to know to

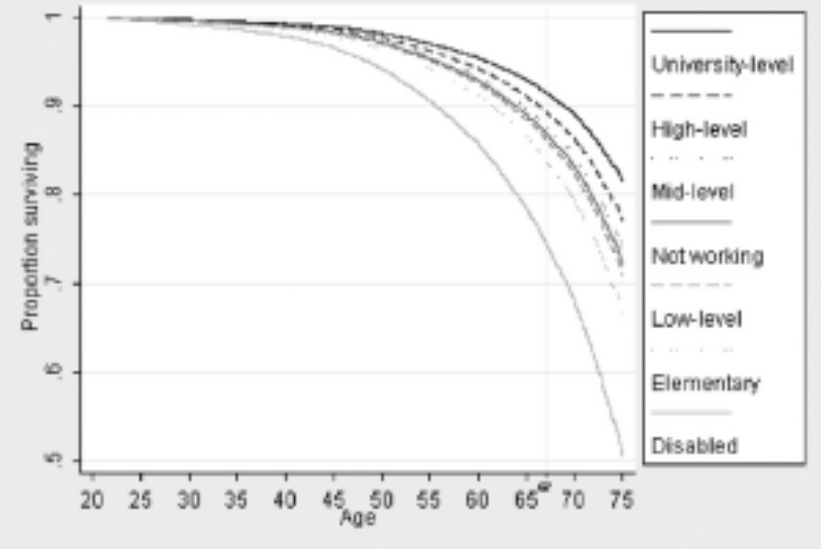


Figure 2.4: Survival until exit to death by employment and occupation.

Notes: Cox proportional hazard model. Source: POLS and cause-of-death follow-up until 2010.

what extent this derives directly from the physical or psychosocial working conditions or rather follows from different life styles caused by occupation (e.g. through peer effects).

2.3 Theoretical framework

In this section, we present a theoretical framework based upon Grossman (1972a); Case and Deaton (2005a); Galama and van Kippersluis (2013); Galama and Van Kippersluis (2015), which represents choices of occupation, health investment, and consumption levels. It should help us to better understand (a) the empirical patterns observed in the previous section and (b) the evidence obtained so far in the wider literature.

2.3.1 Model formulation

The individual maximizes discounted lifetime utility, which depends on consumption c and health h in each period, by choosing a level of consumption, of health investment m, and of physical and psychosocial occupational attributes in vector $\boldsymbol{o}$, given his information set It which includes all state variables at time t.

$$max_{\{c_{t+j},m_{t+j},\boldsymbol{o}_{t+j}\}_{j=0}^{T-t}} E\left[\sum_{j=0}^{T-t}\beta^j u(c_{t+j},h_{t+j})|I_t\right] \tag{2.3.1}$$

The individual faces a health and a budget constraint. In each period, health is determined by permanent health p, the biological aging rate a, the history of idiosyncratic health shocks η, and the history of health investments and occupational choices.

$$H_{t+j} = f(p, a_1, ..., a_{t+j}, \eta_1, ..., \eta_{t+j}, m_1, ..., m_{t+j-1}, \boldsymbol{o}_1, ..., \boldsymbol{o}_{t+j-1}) \tag{2.3.2}$$

Permanent, time-invariant health is a function of endowments $p = g(e)$ and reflects characteristics and circumstances that are stable over time, such as genetic predisposition for certain illnesses. Health deteriorates with age and with exposure to harmful occupational characteristics, but can be improved by health investments. Initial health is viewed as the health level at the beginning of the working career, and it is determined by permanent health, health depreciation due to age, and the effects of all past health shocks on current health. The effect of the history of health shocks is typically smaller than the sum of these health shocks.

Expenditures on consumption and health investment (at prices p_c and p_m) may not exceed total earnings. Wages w depend on endowments, current health, and on the current level of harmful workplace conditions, which can be chosen in each period. Current health and endowments determine the maximum wage (the 'wage frontier') the individual can attain. The wage can be increased by undertaking jobs with harmful workplace conditions that have a deleterious impact on health, leading to health compensating wage differentials. There is no initial wealth in this simple model, endowments take the form of human capital.

$$\sum_{k=1}^{T}(p_c c_k + p_m m_k) \leq \sum_{k=1}^{T} w(\boldsymbol{o}_k | h_k, e) \tag{2.3.3}$$

The model emphasizes that the realization of lifetime utility depends on endowments, effort and institutions. Endowments are characteristics of the individual that are not chosen by the individual, like gender, race or genetic predisposition for learning, athletic ability or disease. But individuals are also endowed with their family background. For example, later-life outcomes may be determined by the level of education and income of the parents or by their ability to raise children. Second, lifetime utility is determined by individual effort: e.g. investment in health is costly but this type of effort is rewarded in terms of earnings potential and future health. Similarly, working may be harmful to health, but work is rewarded with earnings. Third, institutions determine how initial endowments and effort are rewarded in terms of utility and they determine the parameters of the constraints. For example, remedial teaching programs may reduce the importance of endowments while payroll taxes reduce the rewards of endowments and effort in terms of earnings.

2.3.2 Equilibrium conditions

If we make the simplifying assumption that the time preference rate is equal to the interest rate, then, in each period, consumption and health investment are each chosen to equate the respective marginal benefit and marginal cost represented by the conditions

$$\frac{\partial u}{\partial c} = \lambda p_c \tag{2.3.4}$$

$$\frac{\partial h}{\partial m}\left[\frac{\partial u}{\partial h} + \lambda \frac{\partial w}{\partial h}\right] = \lambda p_m \tag{2.3.5}$$

where λ is the 'value' (or shadow price) of lifetime wealth and the LHS of 2.3.5 distinguishes the direct utility and indirect production benefits of health investment. These benefits include the future returns to current health investment through earnings and the discounted marginal utility of health in future periods.

The optimal level of harmful occupational characteristics is determined by

$$\lambda \frac{\partial w}{\partial o} = -\frac{\partial h}{\partial o}\left[\frac{\partial u}{\partial h} + \lambda \frac{\partial w}{\partial h}\right] \tag{2.3.6}$$

where the LHS represents the marginal return to entering an occupation that pays a wage premium in compensation for health hazards imposed and the RHS is the marginal health cost weighted by its direct and indirect consequences. The marginal costs include the effect of current occupational damage on future health and consequently on future earnings. Hence, there is an instantaneous wage benefit of harmful occupational attributes, yet future wages and future utility will be lower due to lower future health.

2.4 Core insights and empirical evidence

The economic framework yields several insights that are relevant for the interpretation of empirical evidence on occupation and health. In this section we discuss three core insights and the extent to which the empirical evidence in the economics literature is in line with them.

2.4.1 Selection and the estimation of causal effects

Individuals select into types of occupation on the basis of endowments, education, and health. Initial endowments influence one's health, but also one's wage prospects, since the marginal benefits of harmful occupational characteristics depend on endowments. Ravesteijn et al. (2013) describe that in the Netherlands, those in "higher ranked"

occupations are generally better educated. It is therefore important to account for all factors that simultaneously influence both occupational choice and health outcomes if we want to obtain estimates of the causal effects of occupation on health. Moreover, not only do occupational characteristics affect health, but health may influence occupational choices. Health, which is subject to shocks, co-determines the wage rate, and the marginal benefits of engaging in harmful occupations. In simple terms, health determines the type of occupation one is able and willing to perform. Both unobserved heterogeneity and reverse causality prevent us from making statements about a causal effect of occupation on health on the basis of simple associations, such as those identified in US data by Case and Deaton (2005a); Cutler, Lleras-Muney, and Vogl (Cutler et al.); Morefield et al. (2012a).

Several recent papers have made an attempt to go beyond the description of the occupation-health association and to estimate the health effects of occupation. Using an estimator that partially allows for correlation of regressors with time invariant unobservables, Gupta and Kristensen (2008) find that working in a satisfactory job environment improves self-reported health and reduces limitations in activities of daily living (ADL) in Denmark, France and Spain. Using data from 15 European countries, Lucifora and Cottini (2013) find that adverse working conditions negatively affect mental health, with the largest effect due to working at very high speed and tight deadlines, low job autonomy, and being involved in complex tasks. They seek to address the potential problem of endogeneity of job conditions by instrumenting working conditions with (i) occupational health and safety regulation by country, and (ii) job control defined by industry and occupation. Both instruments are based upon the idea that institutions or competition at an aggregate level are plausibly exogenous to the working conditions in a particular firm. Estimates exploiting these instruments show even larger effects of working conditions on mental health. Using Danish panel data with detailed information on physical and psychosocial workplace conditions, as well as physical and mental health, and allowing for correlated unobservable determinants of health, lifestyles and work conditions, Cottini and Ghinetti (2012) find that bad working conditions reduce especially mental health.[3]

Using US data and controlling for initial and lagged health using a random effects specification to allow for time-invariant and time-varying factors that may affect both health and occupation, Fletcher et al. (2011a) find that physically demanding work has a strong negative effect on self-rated health for white women but not for men and non-white women, and that harsh environmental working conditions have a strong negative effect for young men but not for young women.

[3]Limitations of the study are that it uses only two waves of the panel, fixed effects are not taken into account and there are no exclusion restrictions, with non-linearity relied on for identification.

On the basis of analysis of German panel data, and exploring several empirical panel data approaches, Ravesteijn et al. (2014) claim that 60 percent of the association between self-rated health and both physical workplace conditions and low control possibilities at work is due to selection effects, while the remaining 40 percent reflects the causal effect on health.

2.4.2 Compensating behaviours and clustering of occupational attributes

A second core insight from the theory is that choices are not made in isolation. An individual simultaneously selects his occupation and investment in health. Health behaviour may be adjusted in response to occupational choice, since the marginal benefits of health investment depend on occupation, through the terms $\frac{\partial h_{t+j}}{\partial m_t}$, and $\frac{\partial w_{t+j}}{h_{t+j}}$ from equation 2.3.5. Note that the behavioural adjustment is different from the selection issue in terms of timing. Ravesteijn et al. (2013) describe that workers in lower occupations are more likely to smoke and less likely to engage in physical exercise. If these smoking differences in adolescence reflect different preferences across individuals which then lead to different occupational choices, this is considered selection. In contrast, if workers in certain occupations initiate smoking to cope with stress on the job or due to peer effects among colleagues, then these are considered behavioural adjustments. Therefore, the total health effect of a change in occupation is the sum of the direct effect of the occupational characteristics on health, and the indirect effect on health behaviour. In the absence of estimates of a full-fledged structural model – which inevitably requires strong functional form assumptions and very detailed data – or an experimental setting, it is extremely difficult to obtain an unbiased estimate of the direct effect since we only observe the total effect after the behavioural response.

Choo and Denny (2006b) claim on the basis of Canadian data that even though the health effect of manual work is reduced by around 10 percent after controlling for health behaviours, there seems to be a direct effect of occupation on health. Cutler, Lleras-Muney, and Vogl (Cutler et al.) similarly observe from US data that controlling for health behaviours reduces, but does not eliminate, occupational differences in health. Kelly et al. (2014) claim to be the first to estimate the causal impact of initial occupation on health behaviours. They find that entering the labour market as a blue collar worker raises the probability of obesity by 4 percent and of smoking by 3 percent. At least part of the observed health differences across occupational groups would appear to arise from differences in health behaviour differences.

Each occupation is characterized by psychosocial and physical workplace conditions. It is important to distinguish between the contributions of each of these attributes to

Table 2.1: Occupational class and job attributes.

	Physically demanding	Inconvenient work postures	Repetitive work	Job control	Under time pressure
Elementary	58%	39%	70%	63%	49%
Low	51%	40%	63%	68%	54%
Mid	39%	39%	57%	81%	70%
High	16%	19%	39%	81%	72%
University	9%	18%	41%	94%	86%

Notes: Percentage of respondents sometimes or regularly exposed to respective job attribute versus no exposure. Source: POLS.

health outcomes. Occupations with heavy manual work duties may simultaneously be characterized by low job control. Researchers should be careful in attributing certain health effects of an occupation to a specific attribute if this characteristic is simultaneously associated with other characteristics within occupational types.

In the Dutch POLS survey, individuals are asked to describe the characteristics of their jobs, including physical exertion, inconvenient work postures, repetitive work movements, and psychosocial aspects, such as time pressure and control over daily work activities. Table 2.1 shows clear gradients in both physical and psychosocial job characteristics favouring the higher occupations. The only exception is working under time pressure, which is reported more often among the higher-ranked occupations. There appears to be clustering of different job characteristics that are potentially damaging to health in certain occupations.

Karasek Jr (1979) defines job strain as the interaction effect between decision latitude and job demands. Decision latitude refers to the degree to which a worker can influence his workplace situation (job control). Job demands are simply stressors on the job. Karasek argues that the combination of low decision latitude and high job demands is particularly bad for health.

According to the effort-reward imbalance model (Siegrist, 1996), it is imbalance between job demands in terms of physical and psychosocial effort, on the one hand, and recognition/reward in terms of wages, esteem, job stability and career opportunities, on the other, that leads to a negative impact on health. Bosma et al. (1998) compare the job demand-control model of Karasek with the effort-reward imbalance model of Siegrist using data from the Whitehall II study.[4] They confirm that the imbalance between personal efforts and rewards predicts higher risk of coronary heart disease. They also find that low job control in itself is strongly related to heart disease, while job strain or more generally high job demands are not.

The Whitehall I and II studies by Marmot et al. (1984, 1991a, 1997) brought the

[4]We refer to Bakker and Demerouti (2007) for a complete overview and review of the two models and proposed alternatives.

concepts of "rank" and "social status" into the theory on psychosocial job demands. These studies showed that male British civil servants in low employment grades suffered from higher mortality and morbidity rates than their colleagues in high employment grades, despite a great deal of homogeneity across the six employment grades. After ten years, mortality for the highest employment grade was about one third of the mortality rate of the lowest for a wide range of causes of death. Marmot and colleagues argue that differences in health outcomes between civil servants of higher and lower rank are due primarily to seniority in the employment hierarchy – rank. They show that low job control seems important, yet they reject the Karasek hypothesis that high demand at work plays an important role.

Case and Paxson (2011) confirm that entry grade and current occupational grade in Whitehall are significantly related to self-assessed health, yet they show that the associations are eliminated after controlling for future occupational grades. This suggests that occupational grade may be more of a marker than a cause of poor health. Using first-difference models, they find no association between current civil service grade and future self-assessed health. In contrast, they do find a significant association between current self-assessed health and future civil service grade. These findings support the health selection hypothesis, rather than the hypothesis that social position in adulthood influences changes in health status.[5]

Fletcher et al. (2011a) combine information on physical requirements of work and environmental work conditions identified in the in the Dictionary of Occupational Titles (DOT) with U.S. panel data and find that both physical demands and harsh environmental conditions harm self-reported health. Fletcher (2012) uses the DOT to construct a 'bad job' factor including ten adverse job conditions, including mostly physical job hazards and perceived reward. While he finds that starting a career in a 'bad job' is detrimental for later life self-reported health, he is unable to disentangle the individual contributions of the ten job conditions. Ravesteijn et al. (2014) find that the effects of late-career exposure of one year to high physical strain and low job control are equivalent to the health deterioration from ageing 16 and 6 months, respectively. They find no evidence of a causal effect of psychosocial stress on health.

In sum, there is little doubt that physical demands of a job matter for health. There is more controversy surrounding the question whether and how psychosocial aspects of the job matter for health.

[5] A recent study by Anderson and Marmot (2012) does find evidence that promotion on the job reduces the risk of heart disease.

2.4.3 Occupational characteristics and socioeconomic inequalities in health

The third insight from the theoretical model is that occupational choices may contribute to socioeconomic inequalities in health. Agents with low endowments, for example in terms of parental background and intelligence, have lower permanent health and lower earnings for each choice of occupation. They may choose to "sell" part of their health capital – by choosing harmful occupational attributes – to obtain an earnings bonus (Smith, 1776; Viscusi, 1978b, 1979). Hence, the choice for a compensating wage differential may be induced partly by initial endowments, and therefore represents a heavily constrained choice.

This phenomenon could occur through multiple channels. First, since the marginal utility of consumption is higher for low levels of consumption, marginal increases in consumption are more attractive for agents with low endowments. Second, the marginal benefit of engaging in harmful occupational characteristics increases in the value of lifetime wealth, λ, as can be seen in equation 2.3.6. The value of lifetime wealth is generally assumed to decrease with more wealth (e.g. Galama and Van Kippersluis, 2015), such that those with low endowments will have a higher value of lifetime wealth and hence higher marginal benefits for harmful workplace conditions. Finally, Case and Deaton (2005a) argue that the marginal benefit of harmful job characteristics decrease with endowments and education, that is $\frac{\partial^2 w}{\partial o \partial e} < 0$, since "professors, unlike construction workers, [...], get no increase in earnings by wearing out their bodies more rapidly". This would imply an additional reason for those with fewer endowments to accept an unhealthy job.

The compensating wage premium historically was associated with a potentially dangerous occupation requiring hard physical labour, e.g. mining, yet a modern-era variant is perhaps an investment banker who depletes his health capital through high levels of job stress and long working hours in exchange for a relatively high wage. Hence, the sign of $\frac{\partial^2 w}{\partial o \partial e}$ may depend on the actual type of workplace condition, e.g. plausibly negative for physical demands, yet potentially positive for psychosocial stressors on the job.

The gradient between socioeconomic status and health becomes steeper if workers with a lower socioeconomic status in terms of education, parental background, and financial wealth are more likely to choose occupations with harmful workplace conditions, and if they don't fully offset these negative effects on health by making health investments. This implies that workplace conditions could directly contribute to health inequalities.

In the epidemiological literature there have been numerous attempts to assess the

contribution of occupation to health inequalities by socioeconomic status. Typically, the bivariate association between health and one particular socioeconomic indicator is estimated, after which occupation or occupational characteristics are added to the regression to assess how the association changes (e.g. Marmot et al., 1997; Borg and Kristensen, 2000). These studies suggest a large role for occupation: for example, Borg and Kristensen estimate that as much as 59% of socioeconomic disparities can be attributed to occupation and the work environment. Yet, simply adding occupation to a regression of health outcomes on other socioeconomic indicators does not provide a reliable estimate of the contribution of occupation to socioeconomic inequalities in health. Since occupation is 'intervening' in the relationship between some other socioeconomic indicators and health, and occupation is clearly endogenous, this provides a particular form of the "bad controls" problem (Angrist and Pischke, 2008), that is elaborated upon in this context by Ravesteijn (2015). One requires sources of exogenous variation in both the socioeconomic indicator (the regressor of interest) and in occupation (the intervening control variable) in order to identify the contribution of the latter to socioeconomic-related health inequality, which is obviously a tall order.

2.5 Conclusion

There is a strong and persistent association between occupation and health. We have verified that occupation is correlated with self-assessed health, risk of disability and longevity in the Netherlands. For example, those in professional occupations are on average 63 percent less likely to enter into disability, and 51 percent less likely to die at any given point in time, when compared to those in elementary occupations.

While this evidence is often interpreted as indicating unfair health inequalities, a fully informed normative policy response requires knowledge of why health is related to occupation. We have sketched a theoretical model of choices with respect to occupation and health. Three core insights emerge: (i) there is selection into occupation on basis of initial endowments and health, (ii) there will be behavioural responses to adverse working conditions, which can be compensating in the form of health investment, or worsening in terms of unhealthy behaviours, and (iii) workplace conditions increase health inequalities if workers with initially low socioeconomic status choose harmful occupations and do not offset detrimental health effects.

Guided by these insights, four main conclusions emerge from our review of the evidence. First, while health inequalities across occupational largely reflect health-related selection into occupations, recent evidence suggests that there also exists a causal effect of occupation on health, that mainly derives from physical work conditions. We believe future research should go beyond the unravelling of (partial) associations

to unveil the underlying causal health effects of job conditions. While the panel-data approaches are promising, an alternative would be to look for exogenous variation in occupational characteristics that might be obtained from changes in regulation, unanticipated firm downsizing, or plant closures.

Second, observed health outcomes across occupational groups may be the result from a behavioural response to changing job conditions rather than the direct effects of occupational characteristics. It is known that health behaviours differ across occupational groups – in the Netherlands close to 50 percent of workers in elementary occupations smoke versus just over 20 percent in university-level occupations. There is some evidence that at least part of these differences in health behaviours is caused by (first) occupation (Kelly et al., 2014). Future research should focus on disentangling the direct effects of occupation and the indirect effects resulting from behavioural response, which is crucial for policy purposes.

Third, research should attempt to separate the contributions of the various physical and psychosocial aspects of work places. While there is convincing evidence both from the epidemiological literature (e.g. Bernard et al., 1997) and the economics literature (Fletcher et al., 2011a; Ravesteijn et al., 2014) that physical characteristics of occupations impact on health, there is more controversy on the impact of psychosocial aspects. Since adverse physical and psychosocial job conditions tend to be clustered, much remains to be learned on their relative contributions.

Finally, more effort should be devoted to establishing the cumulative effects of occupation on health, and the extent to which occupation contributes to the association between (childhood) socioeconomic status and health. This is the most challenging, but potentially also the most rewarding item on the research agenda. Improved understanding of the interrelation between (childhood) SES, occupation, and health is not only helpful for addressing SES-health inequalities, but is equally important in guiding policy choices regarding the regulation of occupational safety and regarding premature labour market exits for health reasons.

Chapter 3

The Wear and Tear on Health: What Is the Role of Occupation?

Chapter Abstract

Health is well known to show a clear gradient by occupation. While it may appear evident that occupation affects health, there are multiple sources of selection that preclude the strong association to be interpreted as exclusively deriving from a causal effect of occupation on health. Despite abundant literature documenting the association, quantification of the relative importance of selection into occupation and the effect of occupation on health is scarce. We link job characteristics to German panel data spanning 29 years to characterize occupations by their physical and psychosocial burden. Employing a dynamic model to control for factors that simultaneously affect health and selection into occupation, we find that selection into occupation accounts for at least 60 percent of the association between health and both physical strain and job control, while selection accounts for nearly 100 percent of the association between psychosocial workload and health. The residual effect of occupational characteristics such as physical strain and low job control is negative and increases with age. The effects of late-career exposure of one year to high physical strain and low job control are equivalent to the health deterioration from ageing 16 and 6 months, respectively.

3.1 Introduction

It is well known that average health and life expectancy display a clear gradient by occupation (e.g. Marmot et al., 1991b). Manual workers in the US, for example, are 50

This chapter is based on Ravesteijn, B., H. Van Kippersluis, and E. Van Doorslaer (2014). The wear and tear on health: What is the role of occupation? SOEPpaper 15256.

percent more likely to die within a given year than workers in managerial, professional and executive occupations (Cutler et al., 2008). The mortality rate for manual workers in Europe is higher than for non-manual workers throughout the age distribution, and this gap has widened over time (Mackenbach et al., 2003). For the Netherlands, Ravesteijn et al. (2013) find a strong gradient in self-assessed health by occupational class—particularly at an older age—and note that 20 percent of elementary workers[1] at the age of 60 have exited the workforce into disability, as opposed to eight percent of workers in occupations that require academic training.

While the occupation/health gradient is widely documented, less is known about the underlying mechanisms that generate it. Occupation may exert a causal effect on health, but the strong correlation may also stem from reverse causality, with health constraining occupational choice. Moreover, individuals in different occupational groups can differ in other—observed and unobserved—"third factors" that influence health. For example, manual workers are less educated or possess different genetic predispositions than non-manual workers. Both reverse causality and third factors may lead to selection effects: people with good health prospects are selected into certain types of occupations. As a result, the magnitude of the association between occupation and health is likely to be (much) higher than the magnitude of the causal effect of occupation on health.

Identification of health effects of occupations is policy relevant for at least two reasons. From a fairness perspective, health disparities that result from a heavily constrained occupational choice set may be more objectionable than differences in health due to free choice behaviour, such as smoking or drinking. From a productivity perspective, it is vital for both regulators and employers to know which specific occupational characteristics are most harmful to health. For example, occupations with harmful ergonomic workplace conditions may simultaneously be characterized by low control possibilities at work, which may exert an independent effect on health. Consequently, separating out these effects is crucial for better-targeted efforts at reducing sickness absenteeism and disability by adjusting specific labor conditions.

Many studies have documented strong associations between occupational characteristics and health (see e.g. Kunst et al., 1999; Goodman, 1999), but few have attempted to obtain estimates of a causal effect, and those that do often focus on specific occupations or specific types of exposure to unhealthy circumstances.[2] The relationship between occupation and health has received surprisingly little attention in the economics literature, but interest in the topic has grown in recent years. Case and Deaton (2005b) show that the self-reported health of manual workers is lower and declines

[1] Elementary occupations consist of simple and routine tasks which mainly require the use of hand-held tools and often some physical effort.

[2] For example, Bongers et al. (1990) study back pain among helicopter pilots.

more rapidly with age than that of non-manual workers. Choo and Denny (2006a) report similar patterns for Canadian workers while controlling for a more extensive set of lifestyle factors and suggest that manual work has an independent effect on health over and above any differences in lifestyle across occupations. Using the longitudinal Panel Study of Income Dynamics (PSID), Morefield et al. (2012b) estimate that five years of blue-collar employment predicts a four to five percent increase in the probability of moving from good health to poor health.[3]

The most comprehensive attempt to estimate the health impact of occupation is Fletcher et al. (2011b) who combine information on the physical requirements of work and environmental conditions taken from the Dictionary of Occupational Titles (DOT) with occupational information in the PSID. Their aim is to estimate the health impact of five-year exposure to physical and environmental conditions as the DOT lacks information on psychosocial stressors. Controlling for first-observed health and five-period lagged health in their empirical model, they estimate negative health effects of physical requirements and environmental conditions. They acknowledge that the potential endogeneity of occupation and occupational change does not allow for a causal interpretation of their random effects estimates. Their data also do not permit to disentangle the contributions of physical and psychosocial occupational stressors.

In sum, the literature so far has failed to establish to what extent the strong association between occupation and health reflects a causal effect of occupation on health, and to what extent it reflects the selection of unhealthy individuals into occupations with harmful job characteristics. In part this is due to the difficulty of finding credible sources of exogenous variation in occupation. Previous studies have also had only limited success in disentangling the health effects of different types of occupational stressors.

Our contribution to the literature is threefold. First, we derive an empirical specification that is grounded in a theoretical model of occupation and health over the life cycle. The explicit link between the theory and the empirical specification (i) identifies the sources of health-related selection into occupation, (ii) shows how our econometric estimators relate to the structural parameters, and (iii) details the conditions under which our dynamic panel data estimates allow for a causal interpretation. These insights provide a theoretical foundation for the dynamic panel data estimation.

[3]Apart from current occupation, a worker's entire occupational history is likely to affect current health. Thus, Fletcher and Sindelar (2009) use father's occupation during childhood and the proportion of blue-collar workers in the state as instrumental variables for first occupation and find that a blue-collar first occupation negatively affects self-assessed health. Kelly et al. (2012) question the statistical relevance of the two instrumental variables used in Fletcher and Sindelar (2009) and instead propose methods developed by Lewbel (2012) and Altonji et al. (2005) to investigate the causal effect of first occupation on health. They find that entering the labor market as a blue-collar worker raises the probabilities of obesity and smoking by four and three percent, respectively, which indicates that the effect of occupation on health may—at least in part—be transmitted through lifestyles.

Second, we estimate an empirical model, on German longitudinal data, that can account for various sources of selection: (i) unobserved time-invariant variables due to the inclusion of individual fixed effects, (ii) time-varying observed variables such as age and wave dummies, and (iii) time-varying unobserved shocks that exponentially die out through the inclusion of the lagged dependent variable. We argue that with panel data spanning 29 years and in the absence of credible sources of exogenous variation in occupation, our model provides the most promising estimates of the relative importance of selection into occupation versus effects of occupation on health.

Third, we show that blue-collar occupations are both more physically demanding and more often characterized by low job control. Previous studies have often characterized occupation with a binary indicator of manual versus non-manual occupation or have focused only on the physical aspects of occupation. This approach has left the contributions of the various ergonomic and psychosocial stressors unseparated and made clear policy conclusions difficult to draw. By linking German data on occupational stressors to individual-level longitudinal data, we are able to unravel the health effects of job characteristics in greater detail.

Our findings suggest that at least 60 percent of the association between physical demands at work and self-reported health stems from the selection of individuals with worse health (prospects) into occupations with high physical demands, and that the same holds for the degree of job control. This leaves at most 40 percent that could potentially stem from the causal effect of both stressors on health. In contrast, our estimate of the effect of psychosocial workload is close to zero. Under the, admittedly stringent, assumptions laid out in the theoretical framework, we estimate that the average effect of one year exposure to a one standard deviation increase in the degree of physical strain (e.g., working as a toolmaker instead of as a teacher) is comparable to the effect of ageing nine months, and the effect increases with age. A lower degree of control over daily activities at work (e.g., a secretary versus a librarian) is harmful to health at older ages but not at younger ages. We estimate that exposure to a one standard deviation increase in handling heavy burdens between the ages of 50 and 54 leads to a health deterioration that is comparable to ageing 16 months. In other words, exposure to physical strain increases the biological health deterioration rate by 130 percent. The estimated effect of exposure to low job control between the ages of 50 and 54 is comparable to ageing 6 months, or an increase of the biological aging rate by 50 percent.

3.2 Occupation and health over the life cycle

In the economics literature, health is treated as a durable capital stock that depreciates with age and can be increased with investment (Grossman, 1972b). The age-related health depreciation rate is exogenous, but an individual can invest in his health by purchasing preventive and curative medical care. The effect of behaviour on health can be positive or negative. Occupational choice can be understood as a form of health disinvestment/erosion: an individual chooses an occupation that is characterized by a set of potentially harmful occupational stressors (Case and Deaton, 2005b; Galama and van Kippersluis, 2010). Occupations with more harmful characteristics may yield higher earnings than other less harmful occupations in the choice set of the individual, which is known as the compensating wage differential (Smith, 1974; Viscusi, 1978a). The additional earnings may be used to partially offset the detrimental effect of work on health by investing in health or to increase consumption. This economic paradigm is useful for distinguishing between the sources of health-related selection into occupation.

Our empirical investigation is based on a theoretical model of an individual maximising the expected present value of lifetime utility, which is derived from consumption c and health h, by choosing levels of consumption c, occupational stressors in vector $\mathbf{o}$, and health investment m. Each occupation is characterized by physical and psychosocial occupational stressors that tend to be clustered, i.e., occupations with low psychosocial workload are often characterized by high physical demands. Future utility is discounted at discount rate β. The information set $\mathcal{I}$ includes endowments e and permanent health h_p, all state and choice variables up to time t, and all future values of the ageing rate, but not future unanticipated health shocks η.

$$max_{\{c_{t+j},\mathbf{o}_{t+j},m_{t+j}\}_{j=0}^{T-t}} E\left[\sum_{j=0}^{T-t}\beta^j u(c_{t+j},h_{t+j})|\mathcal{I}_t\right] \tag{3.2.1}$$

The health production function depends on (i) characteristics and circumstances that remain constant over time that are embodied by permanent health $h_p = f(e)$, which is a function of endowments and reflects all circumstances and personal characteristics that remain constant over the life cycle; (ii) anticipated health deterioration due to ageing a; (iii) a vector of (physical and psychosocial) occupational characteristics $\mathbf{o}$;[4] (iv) medical investment m; and (v) exogenous health shocks η. The effect of occupational characteristics on health, $\boldsymbol{\gamma}_o$, is nonpositive, and $0 \leq \theta \leq 1$ reflects diminishing marginal benefits to health investment. The effects of occupational stressors, health investments and shocks are assumed to decay at the same rate ϕ, which lies between 0

[4]In section 3.3 we will link this to the seminal work by Karasek (1979) on occupational stressors.

and 1. Total lifetime T is exogenous and known to the individual.

$$h_{t+j} = h_p + \sum_{k=2}^{t+j} \left(a_k + \phi^{t+j-k} (\boldsymbol{\gamma}_{\mathbf{o}}' \mathbf{o}_{k-1} + \gamma_m m_{k-1}^{\theta} + \eta_k) \right) \tag{3.2.2}$$

Expenditures on consumption and health investment, at prices p_c and p_m, respectively, should not exceed the net value of wage earnings. The individual can lend and borrow at real interest rate r, but he must repay any remaining debt at the end of his life. Wage w is a function of (i) current occupational choice $\mathbf{o}$, (ii) current health h, and (iii) endowments e.

$$s.t. \sum_{k=1}^{T} (p_c c_k + p_m m_k) \leq \sum_{k=1}^{T} (1+r)^{k-1} w(\mathbf{o}_k, h_k; e) \tag{3.2.3}$$

Consumption, health investment and occupational choice are chosen by equating marginal benefit with marginal cost. The marginal utility of consumption is equal to the shadow price of income λ multiplied by the price of consumption.

$$\frac{\partial u_t}{\partial c_t} = \lambda p_c \tag{3.2.4}$$

For each occupational attribute o_l in vector $\mathbf{o}$, the marginal benefit of occupational stress is represented by the product of λ and the instantaneous wage premium. The marginal cost includes the marginal deterioration of health in all future periods multiplied by (i) the discounted marginal utility of future health and (ii) the product of λ and the present value of the marginal wage returns to future health.

$$\lambda \frac{\partial w_t}{\partial o_{t,l}} = -\sum_{j=1}^{T-t-1} \frac{\partial h_{t+j}}{\partial o_{tl}} \left[\beta^j \frac{\partial u_{t+j}}{\partial h_{t+j}} + \lambda \left(\frac{1}{1+r} \right)^j \frac{\partial w_{t+j}}{\partial h_{t+j}} \right] \quad \forall l \tag{3.2.5}$$

Health investment is the 'mirror image' of occupational choice. The marginal benefit (the product of the marginal effect of health investment on health and both the discounted marginal utility of health and the marginal wage returns to health in all future periods) is equated with marginal cost (the product of the shadow price of income and the price of medical care).

$$\sum_{j=1}^{T-t-1} \frac{\partial h_{t+j}}{\partial m_t} \left[\beta^j \frac{\partial u_{t+j}}{\partial h_{t+j}} + \lambda \left(\frac{1}{1+r} \right)^j \frac{\partial w_{t+j}}{\partial h_{t+j}} \right] = \lambda p_m \tag{3.2.6}$$

The theoretical framework shows how an individual takes the future consequences of his decisions into account while deciding on the optimal levels of harmful occupational stressors. Three insights from the theory are particularly noteworthy. First, both time-invariant initial endowments e—in the form of, for example, physical ability, intelligence

or taste for adventure—and time-varying factors such as health shocks η—e.g. a car accident or the onset of a disease— may influence *both* occupational choice *and* health status through (i) the marginal utility of health, (ii) the marginal wage returns to health, and (iii) the shadow price of income λ. This finding indicates that workers may select themselves into certain types of occupations depending on exogenous factors that directly influence health. Observed health differences across occupational classes should therefore not be interpreted as evidence of a causal effect of occupation on health.

Second, health-related selection into occupation is not only exogenously determined by endowments and shocks: individuals choose their levels of health investment. Health investment may be correlated with occupational choice because (i) exogenous factors influence both health and occupational choice and (ii) workers may choose to offset occupation-related health damage by investing in health (e.g., a bricklayer may seek physiotherapeutic treatment for his back pain, or a manager may take yoga classes to help handle psychological stress).

Third, the relationship between work and health may change over the life cycle for three reasons. First, as equation 3.2.6 illustrates, the expected wage returns on health investment decrease as the individual approaches retirement age, which implies that individuals have fewer incentives to offset occupational damage to health by medical investment.[5] Second, $\boldsymbol{\gamma_o}$ may change over the lifetime, for example if health at older ages is more susceptible to wear and tear at the workplace. Third, the marginal effect of health repair may decrease with age to such an extent that full health repair is no longer feasible at older ages.[6]

In sum, the theory imposes the following conditions on our empirical identification strategy: it should (i) account for factors that can influence selection into type of occupation and may also be related to health, (ii) allow for behavioural adjustments that affect health may coincide with occupational choice, and (iii) accommodate the changing relationship between occupation and health over the life cycle. This requires individual-level data that includes information on health, observes occupational stressors for working individuals, and repeated measures of these variables over an extensive period of time.

[5]However, a model that endogenizes length of life as a function of health can explain an increase in medical investment at older ages.

[6]Our model does not incorporate real-world labor market rigidities, but such rigidities may also prevent individuals from switching occupations at older ages to optimize their exposure to occupational stressors.

3.3 The German Socioeconomic Panel

The German Socioeconomic Panel (SOEP) is a representative longitudinal household survey that started in 1984. We use data from the 29 subsequent annual waves. Respondents are followed over multiple waves, but the panel is unbalanced as many respondents enter the sample after 1984 or leave the sample before 2012. The sample is restricted to 222,726 person-wave observations for which we observe occupation in the previous year, educational attainment, and health in the previous and in the current year. It includes individuals between 16 and 65 years old and employed in the previous period. Sample sizes per wave range between 4,702 in 1989 and 11,798 in 2003. Figure 3.1 shows that we observe 31,216 individuals for at least one period and that 10,577 individuals are observed for at least nine periods, a restricted sample that we will use in a robustness check.

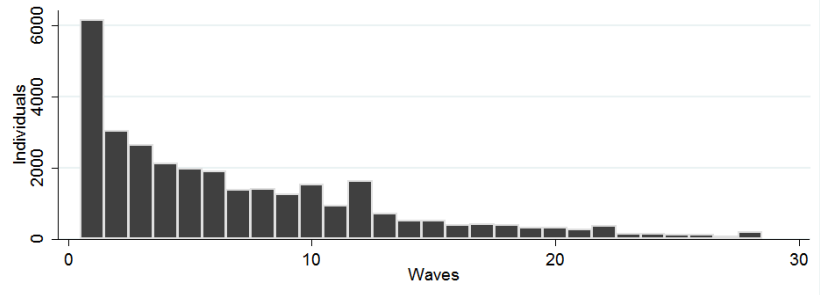


Figure 3.1: Number of individuals by number of observed waves.

Notes: Each bar shows the number of individuals by the total number of (not necessarily consecutive) observed waves in which the individual was employed in the previous period and between 16 and 65 years old. Source: SOEP.

3.3.1 Health

The SOEP has two general health measures: We use health satisfaction (HSAT), measured on an integer scale from 0 to 10, as our main outcome variable. Self-assessed health (SAH), measured on a five-point scale ranging from bad to very good, is used for a robustness check because it was only included since 1992 (and missing in 1993).

3.3.2 Occupation

Occupational titles were coded into the International Standard Classification of Occupations of the OECD (ISCO-88). This gives us 306 occupational titles that were grouped into nine major occupational groups ranked by the OECD classifications,

white-collar workers include legislators, senior officials, managers, professionals, technicians, associate professionals, and clerks. We define blue-collar workers as service workers and shop and market sales workers, skilled agricultural and fishery workers, craft and related trades workers, plant and machine operators, assemblers, and workers in elementary occupations. These definitions are consistent with the distinction between manual and non-manual work of Case and Deaton (2005b), but the blue-/white-collar terminology better reflects the fact that these occupations differ both in terms of physical strain and psychosocial demands. This classification gives a total of 119,456 person-wave observations for white-collar occupations and 103,270 observations for blue-collar occupations.

Figure 3.2 shows that on average, blue-collar workers report better health at younger ages, whereas the opposite is true after the age of 28. HSAT decreases for both blue-collar and white-collar workers over most of the age range but increases after the age of 57. One should keep in mind that these patterns only reflect the HSAT ratings of those who are employed. At older ages, unhealthy workers exit out of employment, whereas healthy workers remain employed which explains the upward slope after the age of 60. Consistent with Case and Deaton (2005b), we find that predicted health decline associated with age in the pooled sample is much stronger among blue-collar than white-collar workers. This begs the question why we observe this.

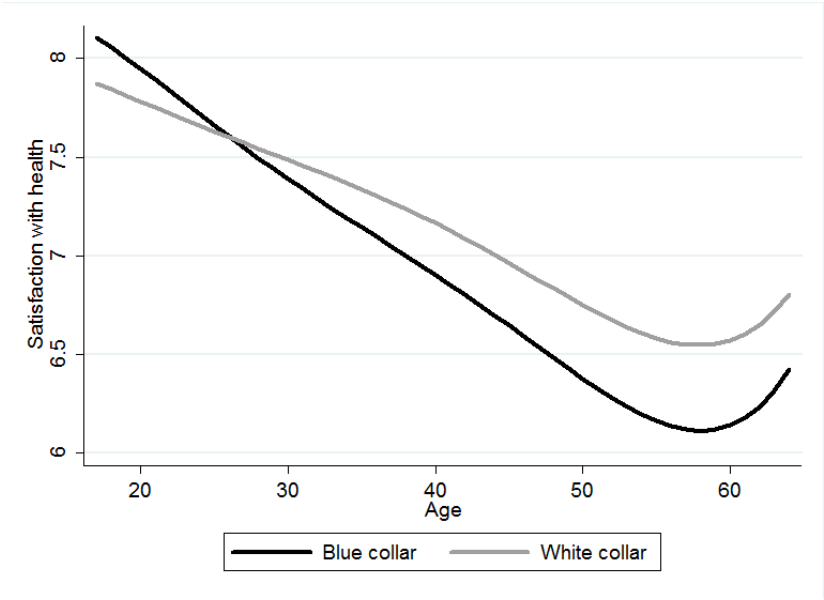


Figure 3.2: Health for blue- and white-collar workers.

Notes: Predicted satisfaction with health for blue- and white-collar workers over the life cycle. Source: SOEP.

Panel A of table 3.1 shows that on average blue-collar workers report worse health

Table 3.1: Summary statistics for the German Socioeconomic Panel.

	HSAT	Age	Female	Schooling	Observations
A. Baseline sample					
All workers	6.94	41.39	.45	12.10	222,726
	(2.05)	(11.47)		(2.71)	
White collar	7.01	42.51	.51	13.34	119,456
	(1.99)	(11.07)		(2.83)	
Blue collar	6.86	40.10	.35	10.66	103,270
	(2.11)	(11.78)		(1.65)	
B. Individuals who were employed in at least nine annual waves					
All workers	6.92	42.16	.43	12.14	151,752
	(2.00)	(10.45)		(2.70)	
White collar	6.98	43.07	.51	13.34	68,585
	(1.96)	(10.12)		(2.82)	
Blue collar	6.85	41.05	.33	10.68	83,167
	(2.05)	(10.74)		(1.61)	

Notes: HSAT, age, female proportion, years of schooling and monthly labor earnings in the German Socioeconomic Panel. Each wave is viewed as a separate observation. Standard deviations are in parentheses. Source: SOEP.

(6.86) than white-collar workers (7.01).[7] Blue-collar workers are slightly younger and less likely to be female, and have an average of three years of schooling less than white-collar workers.

Panel B of table 3.1 shows descriptive statistics for a restricted sample that we will use in a robustness check. The average age in the restricted sample of individuals who were observed in at least nine waves is approximately one year higher than in the full sample. Average health and the proportion of women are slightly lower. The full and restricted samples are similar in terms of education and (blue-collar) employment.

3.3.3 Occupational stressors

While the distinction between blue-collar and white-collar occupations helps us to characterize health differences across broad occupational groups, it does not allow us to identify which occupational stressors associated with blue-collar occupations matter the most. In addition to *physical strain*, we follow the seminal work by Karasek (1979) and distinguish between two psychosocial stressors that could affect health. First, *job*

[7]Health worsens from the top to the bottom of the OECD occupational ladder: 23 percent of legislators, senior officials and managers rate their health with a five or less, as opposed to 31 percent of elementary workers, and 49 percent of legislators, senior officials and managers rate their health with at least an eight, as opposed to 42 percent of elementary workers. This pattern is monotonic across the nine ranked major OECD occupational groups.

control is defined by Karasek as decision authority and intellectual discretion: the individual's potential control over his tasks and his conduct during the working day. Second, *psychosocial workload* refers to the degree to which someone is required to work very fast, hard, or to accomplish large amounts of work and whether he or she is short of time.[8]

We have linked information on occupational stressors from the German Qualification and Career Survey (GQCS) to 306 different ISCO 88 occupational titles in the SOEP.[9] In the GQCS individuals are asked about exposure to occupational stressors and can answer with 'frequently', 'sometimes', 'rarely', or 'never' (Hall et al., 2010). Each occupational title is assigned three measures of occupational stress, and each of these measures is constructed on the basis of one variable to allow for an intuitive interpretation of its coefficient. Robustness checks with composite measure of occupational stressors yield similar results.

First, *physical strain* is measured by the degree of heavy lifting that is required for a job. This variable is constructed as an exposure score: the percentage of individuals in each of the 306 occupations who reported sometimes or frequently carrying heavy stocks of more than 20 kg (men) or 10 kg (women) in the GQCS. 35 percent of all individuals surveyed in the GQCS report that they sometimes or frequently carry heavy stocks. Subsequently, each working individual in the SOEP was assigned the exposure score in the GQCS that pertains to his current occupation. Second, we measure *job control* as the percentage of workers in a given occupation who reported in GQCS that their work was frequently or sometimes stipulated in the minutest details. 46 percent of respondents reported that this was frequently or sometimes the case. Third, we measure *psychosocial workload* as the percentage of respondents in a particular occupation that frequently work under great deadline pressure, which is what 56 percent of all respondents in the GQCS reported. The cutoff value for each of the occupational stressors was determined by taking the value that is closest to the median. Varying the cutoff values did not alter our results.

Although in the analysis we use the mapping of occupational stressors into *all* 306 individual job titles, for illustrative purposes table 3.2 shows the exposure of the nine major OECD occupational groups in the SOEP to the three occupational stressors in the GQCS. Two important observations can be made. First, blue-collar occupations are characterized not only by higher physical strain compared with white-collar occupations but also by lower job control. This illustrates the importance of disentangling the health

[8]Karasek additively combines four measures of decision authority and four measures of intellectual discretion (including a measure of the required skill level) into an aggregate decision latitude scale. He constructs a job demands scale by combining seven measures on not having enough time and working hard and fast.

[9]606 person-year observations were dropped from the sample because 34 of the total of 306 occupational titles in the SOEP could not be linked to the GQCS.

Table 3.2: Occupational stressors across the major ISCO occupational groups.

	High physical strain	Low job control	High psychosocial workload	Observations
Legislators, senior officials and managers	30	28	69	15,263
Professionals	13	29	62	40,079
Technicians and associate professionals	31	46	57	55,650
Clerks	19	56	51	30,819
Service workers and shop/market sales workers	47	50	44	28,269
Skilled agricultural and fishery workers	79	44	44	3,523
Craft and related workers	65	59	60	46,960
Plant and machine operators and assemblers	54	65	52	22,892
Elementary occupations	55	49	36	14,439

Notes: The numbers reflect average percentages for exposure to high physical strain, low job control and high psychosocial workload aggregated by major ISCO 88 occupational group, based on measures of occupational stress for each of the 306 observed ISCO 88 occupational codes. The number of observations refers to the number of person-wave observations in our sample and standard errors are reported in parentheses. White-collar occupations are above the dashed line, and blue-collar occupations are below the dashed line. Source: SOEP, GQCS.

effects of separate occupational stressors. Psychosocial workload is somewhat higher for white-collar occupations.

Second, there is ample variation in occupational characteristics even within the major occupational groups. Even though blue-collar workers are generally more likely to work under more demanding ergonomic conditions and have lower job control compared with their white-collar counterparts, this may not necessarily be the case for many specific occupations. A simple division into blue-collar or white-collar occupations therefore hides the considerable heterogeneity within these groups and the clustering of occupational stressors. We will first look at the blue-collar/white-collar distinction before investigating the effects of the three occupational stressors on health.

3.4 Estimation of the effect of occupational stressors on health

3.4.1 Model specification

We aim to estimate the structural parameter $\boldsymbol{\gamma_o}$ in equation 3.2.2, which refers to the health effects of exposure to occupational stressors $\mathbf{o}$ in the previous year. Note that the one-period lag of the health production function (equation 3.2.2), which includes permanent health h_p, the health effects of ageing a, health investment m and shocks η,

is:

$$h_{t+j-1} = h_p + \sum_{k=2}^{t+j-1} \left(a_k + \phi^{t+j-1-k}(\boldsymbol{\gamma}'_{\mathbf{o}}\mathbf{o}_{k-1} + \gamma_m m^{\theta}_{k-1} + \eta_k) \right) \tag{3.4.1}$$

Substituting equation 3.4.1 into equation 3.2.2, we obtain:

$$h_{t+j} = (1-\phi)\left(h_p + \sum_{k=1}^{t+j-1}(a_k)\right) + a_{t+j} + \boldsymbol{\gamma}'_{\mathbf{o}}\mathbf{o}_{t+j-1} + \gamma_m m^{\theta}_{t+j-1} + \phi h_{t+j-1} + \eta_{t+j} \tag{3.4.2}$$

Switching to individual notation and demeaning the covariates to eliminate the time-invariant factors, we obtain a fixed effects within estimator:

$$h_{i,t+j} - \bar{h}_i = \phi(h_{i,t+j-1} - \bar{h}_i) + \boldsymbol{\gamma}'_{\mathbf{o}}(\mathbf{o}_{i,t+j-1} - \bar{\mathbf{o}}_i) + \boldsymbol{\delta}'(\mathbf{x}_{i,t+j} - \bar{\mathbf{x}}_i) + \varepsilon_{i,t+j} \tag{3.4.3}$$

This specification controls for various sources of selection. First, any unobserved heterogeneity that is constant over time and may be correlated with occupation (such as permanent health h_p in equation 3.2.2) is eliminated: $(1-\phi)h_p - (1-\phi)\bar{h}_p = 0$. Moreover, we go beyond the traditional fixed effects estimator by controlling for lagged health. As detailed in the theoretical framework, this purges the occupational effect from all time-varying unobserved shocks, to the extent that the impact of these unobserved shocks decays exponentially at the same rate over time (see 3.2.2). This decay rate is the interpretation of the coefficient of the lagged dependent variable in an individual fixed effects model.[10] The coefficient ϕ of the demeaned one-period lag of health can be interpreted as the decay parameter through which occupational choice $\mathbf{o}$, health investment m, and unanticipated shocks η in period t-2 and earlier periods affect current health.

$\mathbf{x}$ is a vector of control variables consisting of a fifth order age polynomial and wave dummies to control for a common time trend. A less flexible linear approximation of the age effect would bias our estimates of $\boldsymbol{\gamma}_{\mathbf{o}}$ if health deteriorates more rapidly at older ages, or if workers at older ages would be more or less likely to be exposed to certain occupational stressors.

The error term is $\varepsilon_{i,t+j} = \gamma_m(m^{\theta}_{i,t+j-1} - \bar{m}^{\theta}_i) + \eta_{i,t+j} - \bar{\eta}_i$, which implies two things. First, the ordinary least squares estimator of the coefficient of the lagged dependent variable ϕ is biased because $h_{i,t+j-1}$ is correlated with $\bar{\eta}_i$, and $\bar{h}_i$ is correlated with $\eta_{i,t+j}$. Importantly, however, the estimator is consistent for large T (Nickell, 1981; Bond, 2002). Plausibly, the 29 waves of the SOEP panel satisfy this criterion and we

[10] Only time-varying shocks that are correlated with occupational choice and that do not decay at the average rate may lead to a bias in the estimated effects of occupation. We cannot rule out this source of selection.

check the robustness of our results to restricting the sample to individuals who were observed in nine or more waves.

Second, our estimates should be interpreted as the net effect of occupational stressors, including health investment responses to occupational choice. This net effect is the sum of $\boldsymbol{\gamma_o}$ and an additional term resulting from a possible correlation between occupational choice and unobserved contemporaneous health investment. While this may seem restrictive, we argue that this is a relevant parameter of interest for policymakers because it captures both the direct effects of occupation and the indirect effects through health investment responses to occupation.

3.4.2 Estimation details

Our estimates are based on all within-individual deviations of occupational stressors and health from their observed averages. Within our sample, 13,150 out of 22,526 individuals have switched occupations, for a total of 31,727 times. The number of upward and downward switches in terms of physical and psychosocial demands balance out, and a large proportion of older workers are exposed to high occupational stress.[11]

Our sample selects individuals who were working in the previous year. Even for individuals not working in the current period, we can still estimate the effect of occupation in the previous period on their current health. This implies that sample selection on the outcome variable of the type that warrants a Heckman two-step approach is not an issue here. Obviously, our results should be interpreted as the treatment effect on the working population—and not on children, students, the unemployed, the disabled, and retirees—but that seems to be the policy-relevant effect.

One potential issue could be attrition due to mortality or non-response. Health-related attrition—if present—will lead to a bias toward zero of our estimators, if individuals with the highest vulnerability to occupation-related health deterioration are more likely to suffer from attrition. We find that the likelihood of attrition is at most one percent higher for blue-collar workers than for white-collar workers in our sample. Attrition bias is therefore unlikely to be of major importance to our estimates, and if anything should lead to a slight underestimation of the true effect of occupational

[11] 22 percent of individuals working in two subsequent years in their early twenties experienced a year-to-year job switch that resulted in changes in the level of each of the three occupational stressors. Job switches that result in changes in stressors peak at 24 percent between the age of 25 and 26, and year-to-year switches in occupational stressors still amount to 14 percent of all workers who are in their early sixties. For physical strain, there are slightly more year-to-year switches to lower levels (8.96 percent) than to higher levels (8.55 percent). 8.90 percent of switchers move to an occupation with higher job control while 8.63 percent switch to an occupation with lower control. Slightly more workers switch to occupations with a higher workload (8.94 percent) than with a lower workload (8.57 percent). These small differences are stable across all ages. A substantial proportion of workers in their early sixties are exposed to above-median levels of the occupational stressors: 40 percent for physical strain, 37 percent for low job control, 55 percent for high workload.

stressors.

Self-reported health, as measured on a five-point ordinal scale from poor to excellent, has been shown to be a strong predictor of mortality and morbidity (e.g. Idler and Benyamini, 1997; Mackenbach et al., 2002). We use HSAT (on a 0-10 integer scale) as a proxy for health, which exhibits more variation than the five-point SAH measure which we use in a robustness check. Ferrer-i Carbonell and Frijters (2004) and Frijters et al. (2005) show that for the variable that measures satisfaction with life on a ten-point scale, assuming ordinality or cardinality makes little difference, such that a linear specification is acceptable. Reporting heterogeneity arising from different subgroups reporting the same objective health status differently (Lindeboom and van Doorslaer, 2004) is eliminated by the individual fixed effect, to the extent that reporting heterogeneity is time-invariant.

3.5 Results

3.5.1 Main results

Table 3.3 shows the main results for six different models, where we first present results for a dichotomous indicator for blue-/white-collar occupations (columns 1 to 3) and then for occupation as characterized by three occupational stressors (columns 4 to 6). To understand the order of magnitude of the coefficients, note that the average health deterioration of growing one year older (obtained from an individual fixed effects regression of satisfaction with health on age) is -.0616 (.0008) in our sample.

1. Health effects of blue-collar work. The bivariate association in column 1 between satisfaction with health and blue- or white-collar occupation in the previous year confirms that blue-collar workers are in worse health and that the size of this health gap is similar to the average effect of ageing 29 months, which is a sizable and economically meaningful difference. Column 2 shows the results for the model described by equation 3.4.3. Much of the association appears to be driven by health-related selection into blue-collar occupations because the estimate of the effect is -.0487(.0160) compared with -.1483 (.0087) in column 1. When taken at face value, the health effect of exposure to a blue-collar occupation in the previous year is comparable to the average health effect of ageing nine months.

We add an interaction between age and blue-collar work in column 3 to investigate whether the effect of blue-collar employment differs with age. The coefficient in the first row of column 3 refers to the hypothetical effect of blue-collar employment at the age of zero. The coefficients of the interaction term in the second row indicate that blue-collar employment is harmful to health and that this effect increases with age.

Table 3.3: Results.

	Associations for blue/white collar	FE & LDV for blue/ white collar	FE & LDV for blue/white collar and age interactions	Associations for stressors	FE & LDV for stressors	FE & LDV for stressors and age interactions
	(1)	(2)	(3)	(4)	(5)	(6)
Blue-collar at t-1	-.1483***	-.0487**	.1662***			
	(.0087)	(.0160)	(.0456)			
Age × blue collar at t-1			-.0056***			
			(.0011)			
Physical strain at t-1				-.2102***	-.0830***	.4163***
				(.0051)	(.0032)	(.0909)
Job control at t-1				-.0558*	-.0238	.3656**
				(.0289)	(.0434)	(.1411)
Psychosocial workload at t-1				.5978***	.0074	-.1203
				(.0333)	(.0460)	(.15349)
Age × physical strain at t-1						-.0130***
						(.0022)
Age × job control at t-1						-.0098***
						(.0035)
Age × psychosocial workload at t-1						.0014
						(.0037)
Health at t-1		.1122***	.1120***		.1120***	.1117***
		(.0030)	(.0030)		(.0030)	(.0030)
Individual FE, fifth order age polynomial and wave dummies	✗	✓	✓	✗	✓	✓
Observations	222,726	222,726	222,726	222,726	222,726	222,726
R^2	.0013	.5627	.5628	.0026	.5630	.5631

Notes: Main results for satisfaction with health. FE refers to fixed effects estimation, and LDV refers to the inclusion of the lagged dependent variable. Panel-robust standard errors are in parentheses. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level. Fixed effects specifications are obtained by subtracting individual averages for each regressor. The reference category for columns 1 to 3 is working in a white-collar occupation. All results are similar in terms of size and significance for models with lower-order age polynomials. Intercepts not shown. Source: SOEP, GQCS.

2. Health effects of occupational stressors. Column 4 breaks down occupation into three dimensions of occupational stressors: physical strain, job control, and psychosocial workload in the preceding year. As expected, physical strain and low job control are associated with worse health, whereas psychosocial workload is positively associated with health.

From our theoretical model, we expect health-related selection into occupation to partially drive these associations. Column 5 therefore shows estimates of the effects of these three occupational stressors according to the specification in equation 3.4.3, which controls for selection into occupation on the basis of time-invariant and time-varying factors. These results imply that approximately 61 percent of the negative association between physical strain and health can be explained by selection. Our point estimate (-.0830) suggests that a one standard deviation increase in the distribution of physical strain (e.g. being employed as a toolmaker instead of a teacher) leads to a next-year health deterioration that is comparable to ageing 16 months. Similarly, the coefficient of job control drops by 58 percent although it is no longer significant, while the estimated effect of the psychosocial workload in column 5 is close to zero, suggesting that the positive association with health may even be entirely driven by selection.

However, column 6 shows that the effects on health of low job control and handling heavy burdens vary with age. The predicted health deterioration due to a one standard deviation (.26) increase in handling heavy burdens is equal to zero at age 32. At the age of 50 the point estimate of the effect of moving up the distribution of physical strain by one standard deviation is comparable to ageing 12 months: $.26(.4163 - \times.0130 \times 50) = -.0607$. Low job control has a negative effect after age 37: being in a job with a one standard deviation lower job control (e.g. from librarian to secretary) at age 50 leads to a predicted health deterioration comparable to the effect of ageing 4 months: $.17(.3656 - .0098 \times 50) = -.0211$. We conclude that the effects of physical strain and job control are age-dependent. The coefficient of the interaction between psychosocial workload and age is not significantly different from zero, possibly because workload is only important for certain personality types or for a subset of occupations.

The effect of occupational stressors potentially differs by gender because the type of stressors experienced and the vulnerability to certain stressors is likely to be different for men and women. The direction and statistical significance of the estimates for the subsample of men are similar to separate estimates. However, estimated effects of physical strain and job control for women do not seem to increase with age. A regression without the age interactions gives a significant point estimate of -.1082 (s.e. .0500) for the effect of physical strain on women's health.

3.5.2 Cumulative effects

Cumulative health effects can be obtained from the estimated coefficient of the lagged dependent variable ϕ in equation 3.4.3. By assumption, ϕ is the uniform exponential decay rate at which past health investment, occupational stressors, and shocks affect current health in equation 3.2.2. The point estimates of ϕ in table 3.3 suggest that roughly ten percent of the occupation-related health deterioration in period t-2 persists in period t. Using the point estimates in column 6 of table 3.3, the point estimate of health deterioration at the age of 55 caused by a one standard deviation increase in the physical strain between ages 50 to 54 is $\sum_{k=50}^{54} .1117^{54-k}.26(.4163-.0130\times k) = -.0831$, which is comparable to the average health effect of ageing nearly 16 months. Likewise, the point estimate of the effect of working in occupations with a one standard deviation lower degree of job control between the ages of 50 and 54 is -.0311, which is comparable to the effects of ageing 6 months.

3.5.3 Vulnerability to occupational stressors over the life cycle

Health effects may be nonlinear in age. Figure 3.3 shows the results of two regression models that are similar to the specifications of columns 3 and 6 in table 3.3 but include additional interactions between the occupational variables and age up to the fifth power. Panel A shows the estimated next-period effect of blue-collar versus white-collar occupations at different ages.[12] The estimated next-period negative effect of blue-collar employment is statistically significant from the age of 43 onwards.[13]

Panels B, C, and D of figure 3.3 refer to estimates obtained from the model in column 6 of table 3.3 with added interactions between each of the three occupational stressors and age up to the fifth power. Physical strain has a significant negative effect on workers aged 42 and older.

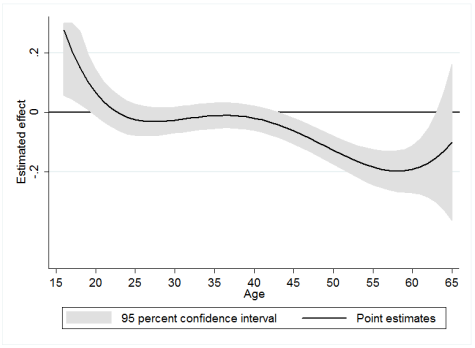
Panel C shows that the negative effect of low job control is significant for workers aged 48 years and older. The estimated negative effects in panels A, B, and C are strongest around age 60. Surprisingly, we find a minor positive effect of high psychosocial work load around the age of 50 as shown in panel D.

3.5.4 Robustness checks

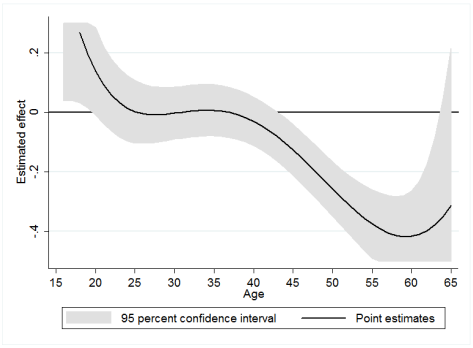
We examine the robustness of our main findings to (i) alternative measures of occupational stressors, (ii) using SAH instead of HSAT as our outcome measure, (iii)

[12] Confidence intervals are computed using the delta method (Oehlert, 1992).

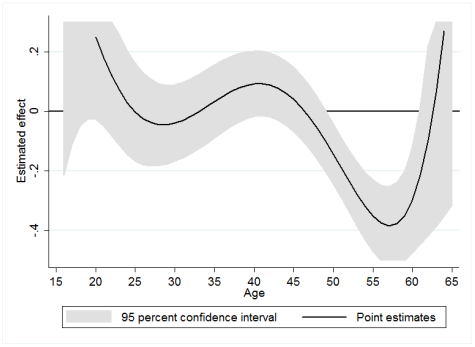
[13] Caution is warranted when interpreting the estimates at the lower and upper end of the age distribution because of the lower number of observations at young and old ages and the polynomial functional form. Less than three percent of observations occur at ages below 20, and only two percent of observations occur at ages over 60.



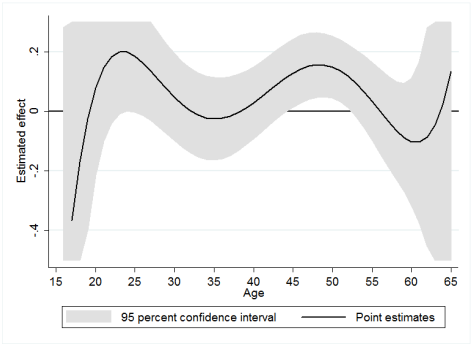
(a) Effect of blue-collar work.



(b) Effect of physical strain.



(c) Effect of low job control



(d) Effect of high work load

Figure 3.3: The effects of occupational stressors over the life cycle.

Notes: 95 percent confidence intervals of the coefficients of the occupational stressors, computed using the delta method. Panel 3.3a refers to the coefficient of a binary variable, the other panels refer to the health effects of occupations with 100 instead of 0 percent exposure to each respective stressor. Source: SOEP, GQCS.

controlling for education-specific health deterioration by age, (iv) limiting the sample to individuals who are observed in nine or more waves, and (v) two additional ways of specifying our dynamic panel data model.

The results are robust to using other measures of occupational stressors and health. Column 1 in table 3.4 includes composite measures of physical strain, job control and psychosocial workload, and shows similar results in terms of effect sizes of a standard deviation increase in the occupational stressors. We obtain composite measures of the three occupational stressors by averaging percentages of high exposure to (i) physical strain as measured by heavy lifting, working crouched down, and working standing up; (ii) job control as measured by work being stipulated in the minutest details, being allowed to plan and schedule work by oneself, and being able to influence the amount of work one has to do; and (iii) psychosocial workload as measured by working under great deadline pressure, reaching the limits of one's capacities, and working very quick. Furthermore, analyses using the Finnish Job Exposure Matrix instead of the GQCS provides results that are similar in terms of order of magnitude and significance (see Ravesteijn et al., 2014).

SAH—measured on a five-point scale—was only included in the SOEP questionnaire in 1992 and from 1994 onwards. Column 2 of table 3.4 substitutes SAH for HSAT and the coefficient estimates reveal similar signs and sizes to our main findings in column 6 of table 3.3, supporting our conclusion that physical strain and low job control are harmful to health at older ages.

Individuals in different occupations may have different biological ageing rates. We have assumed uniform ageing effects in the preceding analyses. If the health of blue-collar workers declines more rapidly regardless of their occupation, our results overestimate the harmful effects of physical strain. In column 3 of table 3.4, we allow for different rates of ageing by interacting each of the levels of educational attainment with a fifth-degree age polynomial and our estimates are similar to our findings in table 3.3.

The estimator of the coefficient of the lagged dependent variable is consistent if the number of time periods in the sample goes to infinity. Our sample spans 29 years and is unbalanced because it includes individuals who are observed for fewer waves. We repeat our analysis for a subsample of 10,373 individuals who have been employed for at least nine of the 29 years to counter the downward bias of the estimator of the lagged dependent variable that plagues short panels (Bond, 2002). The number of person-wave observations drops from 222,726 in our baseline sample to 151,752 in column 4 of table 3.4. The coefficients of the (age-interacted) occupational stressors are similar to those in our baseline specification. However, the coefficient of lagged health is now larger, suggesting that past health investment, occupational stress, and health shocks are more persistent than they were in the full-sample analysis. We conclude that our

Table 3.4: Robustness.

	Composite stressors	SAH	Control for education-specific ageing trends	Only individuals with $T > 8$	FE	LDV
	(1)	(2)	(3)	(4)	(5)	(6)
Physical strain at t-1	.2693***	.1775***	.4032***	.3473***	.4693***	.1324**
	(.0943)	(.0506)	(.0940)	(.1010)	(.0921)	(.0529)
Job control at t-1	.9661***	.2906***	.3584**	.4439***	.3945***	.3953***
	(.2641)	(.0760)	(.1454)	(.1605)	(.1427)	(.0918)
Psychosocial workload at t-1	.2355	.0755	.1125	.4175**	.1564	.1326
	(.2152)	(.0816)	(.1563)	(.1787)	(.1553)	(.1032)
Age × physical strain at t-1	-.0088***	-.0046***	-.0127***	-.0110***	-.0145***	-.0071***
	(.0023)	(.0012)	(.0023)	(.0024)	(.0022)	(.0013)
Age × job control at t-1	-.0028***	-.0059***	-.0095***	-.0111***	-.0107***	-.0130***
	(.0065)	(.0018)	(.0036)	(.0038)	(.0035)	(.0022)
Age × psychosocial workload at t-1	.0044	-.0027	.0014	-.0072*	-.0020	.0035
	(.0052)	(.0020)	(.0038)	(.0043)	(.0038)	(.0024)
Health at t-1	.1117***	.0723***	.1110***	.1613***		.5464***
	(.0030)	(.0034)	(.0030)	(.0034)		(.0021)
Fifth order age polynomial and wave dummies	✓	✓	✓	✓	✓	✓
Fifth order age polynomial interacted with education	✗	✗	✓	✗	✗	✗
Individual FE	✓	✓	✓	✓	✓	✗
Education and gender	✗	✗	✗	✗	✗	✓
Observations	222,726	162,595	222,726	151,752	222,726	222,726
R^2	.5631	.5906	.5635	.5220	.5576	.3389

Notes: Robustness checks for satisfaction with health. FE refers to fixed effects estimation, and LDV refers to the inclusion of the lagged dependent variable. Panel-robust standard errors are in parentheses. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level. Fixed effects specifications are obtained by subtracting individual averages for each regressor. The fourth column refers to sample of individuals who are observed in at least nine—not necessarily consecutive—waves. Intercepts not shown. Source: SOEP, GQCS.

estimates of the effects of occupational stressors are robust across specifications but that an analysis of the full sample leads to underestimation of the coefficient of lagged health. We may have underestimated the cumulative effects of occupational history by underestimating ϕ, and the predictions regarding the cumulative effects provide—in absolute terms—a lower bound on the health effects, which indicates that the true health effects may in fact be even larger.

Angrist and Pischke (2009) have voiced concerns about the violation of strict exogeneity in fixed effects dynamic models, particularly by utilising short panels. They propose checking robustness by separately estimating both a fixed effects and a lagged dependent variable model. Column 5 of table 3.4 presents results from a fixed effects model without a lagged dependent variable.[14] The point estimates in column 5 suggest a somewhat stronger effect of physical strain and low job control at older ages than the baseline specification. However, these estimates may be the result of a bias caused by past events that affected health and occupational choice that are not accounted for by the lagged dependent variable, which is omitted in this specification.

In a model in which we control for a lagged dependent variable, but not for individual-specific fixed effects, the estimator of the decay parameter ϕ in equation 3.4.3 is biased toward one because h_{t-1} contains h_p (see equation 3.4.1), which has a coefficient of one and no longer drops out if we do not subtract $\bar{h}$. We can therefore no longer distinguish between the elements in h_{t-1} that are transitory and the elements that are constant over time, which explains the bias of the estimator of ϕ toward one. In this specification, we therefore overestimate the impact of past events on current health, and we only partly control for unobserved time-invariant heterogeneity.[15] To proxy for time-invariant unobserved factors otherwise picked up by the fixed effect, we control for years of schooling and gender. Our estimates are now mostly driven by variation between individuals. The coefficients of the interaction between age and occupational stressors in column 6 of table 3.4 are similar to our earlier results. Note that the coefficients in the first three rows of the table lack a clear interpretation because of the age interaction and should therefore not be interpreted. Overall, our main conclusions do not change when estimating models that include either individual-specific fixed effects or a lagged dependent variable, which is reassuring.

Other methods have been proposed to consistently estimate $\boldsymbol{\gamma_o}$ in equation 3.4.3 in short panels, of which the so-called Arellano-Bond estimator (Arellano and Bover,

[14] With respect to equation 3.4.3, the error term would now include the deviations of the effects of health investment, occupational stressors, and health shocks before period t-1 from their individual averages. If a past health shock would have a negative effect on current health and lead to higher occupational stress in the previous period, we would overestimate the effect of occupational stressors because this situation leads to additional correlation between $\mathbf{o}$ and the error term.

[15] By not subtracting averages in equation 3.4.3, the error term now includes $(1-\phi)h_p$, which may be correlated with lagged health and occupational characteristics.

1995; Blundell and Bond, 1998) is the most prominent. The Arellano-Bond estimator is based on the first-difference estimator. The most important assumption is that second and further lags of health are uncorrelated with the first differences of the error term and can be used as instrumental variables for $h_{t-1}-h_{t-2}$. The Arellano-Bond test for autocorrelation rejects this assumption in our case, which is not surprising because using lagged values as instruments is difficult to justify in the case of health: chronic illnesses or the introduction of a new medical drug may progressively affect health over time, which leads to second- or higher-order serial correlation in the differenced error term and violation of the exogeneity assumption. In attempting to overcome this problem, more lags of the regressors were included in the model, and further lags of regressors and instruments were used to purge the error term from autocorrelation. However, we still find higher-order autocorrelation in these models, rejecting the validity of the instruments.[16]

3.6 Conclusion

The strong association between occupation and health is widely documented. Our results confirm that German blue-collar workers report worse health than white-collar workers, and that the size of this health gap is similar to the effect of ageing 29 months. However, because of various sources of selection into occupation, the association does not necessarily reflect the causal effect of occupation on health.

In this paper, we make three contributions. First, by proposing a dynamic theoretical model as the foundation of our empirical specification, we highlight the various sources of selection into occupation, and we make explicit under which conditions the coefficient of occupational hazards can be interpreted as causal.

Second, we estimate the empirical equation deriving from the theory using a detailed German longitudinal dataset over many time periods (29 years). In this equation, we account for the sources of selection: (i) unobserved time-invariant variables due to the inclusion of individual fixed effects, (ii) time-varying observed variables such as age and wave dummies, and (iii) time-varying unobserved shocks that exponentially die

[16]Limiting the number of waves can give us the false illusion that serial correlation of the error term is not a problem simply because of the low power of the test. Blundell and Bond (1998) and Michaud and Van Soest (2008) use short panels of six waves and "use up" even more waves due to the inclusion of lagged values of the dependent variable. The autocorrelation tests in these studies do not reject the assumption of no autocorrelation in the error term, which may be the result of limited test power based on the small number of waves. If we include one- and two-period lags of the dependent variable, we find no second-order autocorrelation. However, we find autocorrelation of the third-order, which still violates the Arellano-Bond assumptions. Including third or fourth lags seems to shift the order of autocorrelation downward rather than to solve the problem. The Sargan test may not be informative because it assumes that at least one instrument is exogenous, which is an assumption we are not willing to make.

out through the inclusion of the lagged dependent variable. In doing so, we are able to quantify a lower bound on the selection effect, and an upper bound on the causal effect of occupation on health.[17] Moreover, our results generalize across the entire labor force, which is in contrast to local effect estimates based on a particular reform that affected only part of the employed population.

Third we estimate separate health effects of physical strain and Karasek's two dimensions of psychosocial occupational stress from the German Qualification and Career Survey. However, because we do not observe individual levels of health investment, we are unable to disentangle the effects of such occupational stressors and any health investment made in response to occupational choice. This is a policy-relevant effect, since it can be interpreted as the sum of the direct effect of occupation and the indirect effect of any behavioral response to occupational choice.

Our main finding is that selection of individuals with poor health (prospects) into occupations with heavy physical demands and low job control, accounts for 61 and 58 percent, respectively, of the observed association with health. The association between psychosocial workload and health seems to be entirely driven by selection. While we cannot rule out other sources of selection, we interpret the residual, conditional differences in health as deriving from the occupational characteristics itself. Using this interpretation, we find that both high physical occupational demands and low job control have negative effects on health and that these effects increase with age. The effect of exposure to a one standard deviation increase in the degree of handling heavy burdens (e.g., working as a toolmaker instead of a teacher) between age 50 to 54 is comparable to ageing 16 months. Low job control is harmful to health but only after age 48. The effect of exposure to a one standard deviation decrease in the degree of job control (e.g., working as a secretary instead of a librarian) between age 50 to 54 is comparable to ageing 6 months.

Occupational health and safety policies, career development programs, and retirement policies should be based on the knowledge that selection into occupation explains a substantial fraction of health disparities across occupational groups. It is important to emphasize that ignoring this selection will lead to serious overestimation of the benefits of improving occupational characteristics. Our results cannot rule out, and even strongly suggest, that exposure to physical strain and low job control is harmful to health at older ages. That would mean that successfully shielding workers—especially older workers —from such conditions will prevent accelerated health deterioration and

[17] There could be other sources of selection, like time-varying shocks that are related to occupational choice that do not exponentially decay over time or that have different decay rates. While we cannot rule out that some time-varying unobserved factors operate in the opposite direction, the sign of the bias when excluding fixed effects or the lagged dependent variable renders the case that the true selection effect is larger more plausible.

therefore also illness-related absenteeism and labor force exit due to work disability. Our findings therefore have great relevance for the many OECD countries that are aiming to extend working life careers in order to keep disability and pension systems sustainable.

Chapter 4

When a Bad Control Variable Turns Good: Is the Effect of Parental Socioeconomic Status on Health Transmitted by Schooling?

Chapter Abstract

It is common practice in the economics literature as well as in other disciplines to include an intervening (bad) control variable—which itself is an outcome of the treatment variable—in an OLS regression model. I show that the estimators of direct and indirect treatment effects are asymptotically biased if the intervening control variable is correlated with the error term.

The problem of endogenous intervening controls can be solved in the special case where (i) we have a valid instrumental variable for the bad control variable, and (ii) the local average treatment effect for compliers with the instrument is equal to the average treatment effect. Using these insights to revisit the relationship between parental socioeconomic status, schooling, and health in the UK, the bad controls method suggests that 23 percent of the total effect of parental socioeconomic status is transmitted by schooling. However, after eliminating the bad controls bias by instrumenting educational attainment by the 1947 and 1972 UK compulsory schooling reforms, the estimate of the indirect effect is close to zero and not significant.

I conclude that caution is warranted when interpreting a regression coefficient if other covariates are themselves outcomes of that particular variable. Not only do inter-

This chapter is based on Ravesteijn, B. (2015). When a bad control variable turns good: Is the effect of parental socioeconomic status on health transmitted by schooling? Mimeo.

vening control variables block causal pathways—which is sometimes what we want—, but their inclusion also introduces asymptotic bias which produces results which lack a meaningful interpretation.

4.1 Introduction

Causal effects may be transmitted by intervening factors. For example, to what extent does the causal effect of parental socioeconomic status (SES) on health run through educational attainment? The bad controls method involves comparison of the results from two OLS regressions. First, the dependent variable (e.g. health) is regressed on the regressor of interest (e.g. parental SES) to estimate the total effect. In figure 4.1 the total effect is defined as $\beta_p+\gamma\beta_s$. In the second regression model, the dependent variable is regressed on both the regressor of interest and the intervening control variable (e.g. schooling) to estimate the direct effect β_p. The difference between the total effect and the direct effect is usually interpreted as the indirect effect $\gamma\beta_s$: the part of the effect of the regressor of interest that is transmitted by the intervening control variable. This supposedly tells us whether educational policies can play a role in reducing health disparities by parental background.

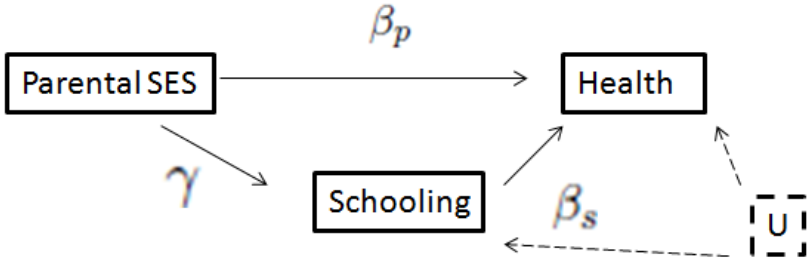


Figure 4.1: Endogenous intervening control variable.

In the economics literature on the relationship between SES and health, this approach is common practice. For example, Cutler and Lleras-Muney (2010) analyze the National Health Interview Survey (NHIS), and conclude that income, health insurance, and family background can account for about 30 percent of the education-health gradient. They find that knowledge and measures of cognitive ability explain an additional 30 percent, while social networks account for another 10 percent. Other notable examples of the bad controls method in the economics literature are Case et al. (2002), who report a strong direct impact of parental education (regressor of interest) on child's health (outcome variable) while controlling for family income (intervening control variable), and Kenkel (1991), Deaton and Paxson (2001), Currie et al. (2007), Cutler and Glaeser (2005), and Goldman and Smith (2002) who employ similar strategies to investigate the mechanisms underlying the association between socioeconomic status and health.

This paper shows under which circumstances this common practice fails. It builds upon a discussion by Angrist and Pischke (2008), who distinguish between *good controls* and *bad controls*. Angrist and Pischke define good controls as *preceding* control variables that are measured before the regressor of interest was determined. Bad control variables are themselves caused by the regressor of interest and *intervene* in the relationship between the regressor of interest and the outcome variable. The "bad controls method" yields biased estimators of both (i) direct and indirect causal effects and (ii) partial associations conditional on intervening variables when the intervening control variable is correlated with the error term.

I show that when the regressor of interest is exogenous, the bias of the bad controls estimator of the indirect effect is equal to the bias of the intervening control variable, inflated by the effect of the regressor of interest on the intervening control variable. The sign and size of the bias term $(\mathbb{E}[\widehat{\beta_s}] - \beta_s)\gamma$ are generally unknown, such that these results lack a clear interpretation. The bad controls bias is equally problematic in studies which acknowledge endogeneity of the regressor of interest: any changes in the association between the regressor of interest and the outcome variable could either be the result of the intervening control truly explaining part of the association, or because of the bad controls bias. In the latter case, the results are not informative about real-world relationships, and often there is no way to be certain. It is important to note that these problems only occur with intervening control variables, controlling for preceding control variables yields a well-understood partial association.

Unbiased estimators of direct and indirect effects can be obtained using exogenous variation in the intervening control variable, for example in an experimental setting or by using an instrumental variable approach in cases where the local average treatment effect (LATE) is close to the average treatment effect (ATE). If LATE differs from ATE however, the estimator of the average indirect treatment effect is biased by γ times the difference between LATE and ATE, and the bias of the direct effect is the same in size but of the opposite sign. Additionally, these indirect and direct effects lack the interpretation of local effects because the coefficient of the regressor of interest is not based on compliers. Using separate instrumental variables for both the regressor of interest and the intervening control variable (see e.g. Frölich and Huber, 2015) leads to similar problems unless both groups of compliers coincide.

The common mistake of using bad controls is perhaps due to a lack of clarity in the previous literature and this is reflected in econometric textbooks, which leads to perpetuation of this bad habit. The widely used introductory econometrics textbook by Stock and Watson (2010) states that for omitted variable bias to arise, at least one of the included regressors must be correlated with the omitted variable and the omitted variable must be a determinant of the dependent variable. Although these two

conditions hold true for any intervening variable that appears in the causal pathway between the regressor of interest and the outcome, omission does not lead to bias if we are interested in the total effect of the regressor of interest. It is in fact inclusion of these variables which introduces bias.

Even in cases where researchers appear to be aware that we may not want to block causal pathways to estimate total effects, many seem unaware of the additional problem that is caused by the correlation between the intervening control variable and the error term. In another popular text book on econometrics, Wooldridge (2012) writes that we should not be controlling for differences in the intervening control variable unless we want to test for some sort of indirect effect of the regressor of interest. This appears to incorrectly validate the use of bad controls in case one is in fact interested in direct and indirect effects.[1]

I illustrate the problem using two compulsory schooling reforms in the United Kingdom raised the minimum school-leaving age first to 15 and then to 16. Because a large portion of students left school at the minimum school leaving age, the reforms affected the behavior of a much larger share of students than similar reforms in other countries. Instrumenting educational attainment by either of the UK compulsory schooling reforms within a fuzzy regression discontinuity framework leads to a LATE that is close to ATE due to the large proportion of compliers with the reforms (Oreopoulos, 2006; Clark and Royer, 2013). I show that the indirect health effect of parental SES through schooling is in fact close to zero at the margin that was affected by the reform, in contrast to the large indirect effect estimates according to the bad controls method.

4.2 Direct and indirect causal effects

Throughout this paper, health h is determined by parental SES p, schooling s, and unobserved factors contained in the error term ε

$$h = \beta_p p + \beta_s s + \varepsilon \tag{4.2.1}$$

Furthermore, schooling is a function of an exogenous component s^*, parental SES (randomly distributed) and an error term η:

$$s = \gamma p + s^* + \eta \tag{4.2.2}$$

[1]The discussion by Angrist and Pischke has been widely misinterpreted, with researchers acknowledging that bad controls block causal pathways, but seemingly unaware of the bias due to the endogeneity of intervening control variables. See e.g. Burke (2012); Bakija (2013); Dolan et al. (2013); Del Valle (2013).

As before, β_p is the direct effect of parental SES that is not transmitted by schooling, $\gamma\beta_s$ is the indirect effect of parental SES that runs through schooling, and $\beta_p + \gamma\beta_s$ is the total effect. The remainder of this section show how direct and indirect can be estimated

4.2.1 Estimation of direct and indirect effects in the absence of endogeneity

To clarify the bad controls problem, first consider the hypothetical situation where both the regressor of interest and the intervening control variable are uncorrelated with the error term in equation 4.2.1, or $\mathbb{E}[p, \varepsilon] = \mathbb{E}[s, \varepsilon] = 0$. If this is the case, the bad controls problem does not exist. OLS regression of h on both p and s in a single regression model yields:[2]

$$\mathbb{E}[\hat{\beta}_s] = \beta_s \tag{4.2.3}$$

$$\mathbb{E}[\hat{\beta}_p] = \beta_p \tag{4.2.4}$$

The estimators of the direct effect of parental SES and the effect of schooling are unbiased in the absence of further omitted variable bias or measurement error. The key assumption—which will be relaxed in the next subsection—is that the intervening control variable is not correlated with the error term.

In two separate univariate regression models where h is regressed on either s or p, the expected values of the estimators are:

$$\mathbb{E}[\hat{\beta}_s] = \beta_s + \beta_p\gamma\mathbb{E}\left[\frac{\sum_{i=1}^{n} p_i^2}{\sum_{i=1}^{n}(\gamma p_i + \eta_i)^2}\right] \tag{4.2.5}$$

$$\mathbb{E}[\widehat{\beta_p + \beta_s}\gamma] = \beta_p + \beta_s\gamma \tag{4.2.6}$$

Where i refers to each individual and n is the sample size. The estimator of β_s (the coefficient of schooling) yields an estimate of the total (unconditional) association between schooling and health. This is equal to β_s plus the omitted variable bias due to the omission of parental SES, which ranges between the ratio of the health effect of parental SES over the effect of parental SES on schooling ($\beta_p\gamma^{-1}$) when all variation in schooling is caused by parental SES (and therefore $Var(\eta) = 0$), and approaches zero when η (the variance of the part of schooling that is not a function of parental SES) goes to infinity when the variance of parental SES is kept fixed. The univariate regression of health on parental SES leads to an unbiased estimator of the *total* effect of parental SES on health, which is equal to the sum of the direct and the indirect

[2]See appendix A for derivations.

effect.

Under exogeneity of both the regressor of interest and the intervening control variable, the omission of an *intervening* control variable that is influenced by the regressor of interest leads to estimation of the total effect of the regressor of interest, while the omission of a *preceding* control variable—that influences both another regressor and the outcome variable—leads to well-understood omitted variable bias.

4.2.2 Omitted variables and intervening controls

Now consider the case in which $\mathbb{E}[\varepsilon, \eta] \neq 0$, when unobserved factors affect both schooling and health (e.g. serious illness during childhood which affects educational attainment and adult health but not parental SES). The expected values of the estimators from the OLS regression model with both regressors are:

$$\mathbb{E}[\hat{\beta}_s] = \beta_s + \mathbb{E}\left[\frac{(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} \eta_i \varepsilon_i)}{(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} (\gamma p_i + \eta_i)^2) - (\sum_{i=1}^{n} p_i(\gamma p_i + \eta_i))^2}\right] \tag{4.2.7}$$

$$\mathbb{E}[\hat{\beta}_p] = \beta_p - \gamma\mathbb{E}\left[\frac{(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} \eta_i \varepsilon_i)}{(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} (\gamma p_i + \eta_i)^2) - (\sum_{i=1}^{n} p_i(\gamma p_i + \eta_i))^2}\right] \tag{4.2.8}$$

As expected, the estimator of the effect of schooling in equation 4.2.7 suffers from omitted variable bias which is similar in nature to the bias term in equation 4.2.5, although it has a different form because the exact relationship between the omitted variable(s) with respect to schooling and health is not specified.

Equation 4.2.8 shows the main result of this paper: the bias term of equation 4.2.7 also shows up in the expectation of the estimator of the direct effect of parental SES, where it is multiplied by the association between the regressor of interest and the intervening control γ and then subtracted from the true value of the parameter. While the omission of a preceding control variable leads to the well-understood bias in equation 4.2.5, the bias in equation 4.2.8 is caused by the inclusion of an intervening control variable. This result also holds if the regressor of interest is not exogenous and instead we are interested in partial correlations (see appendix A.6 for derivations).

In most real-world situations, it is likely that there are unobserved factors that influence both schooling and health *after* the level of parental SES is determined, which means that we cannot determine the size or even the sign of the bias term. As a result, the bad controls method yields biased estimators of direct and indirect effects, and the size of the bias depends on (i) the association between the regressor of interest and the intervening control and (ii) the bias of the estimator of the effect of the intervening control variable.

The estimator of the total effect of parental SES on health in the univariate regres-

sion remains unbiased as in equation 4.2.6, irrespective of endogeneity of variables that transmit the total effect.

4.2.3 Measurement error and intervening controls

In addition to omitted variables, endogeneity of the intervening control variable can be caused by measurement error, for example if schooling is measured in years instead of specific levels. Random measurement error leads to attenuation bias towards zero of the estimator of the coefficient of the mismeasured variable. This in turn causes the estimator of the direct effect of the regressor of interest to be biased towards the total effect. Suppose that schooling is measured with zero mean measurement error w: $\tilde{s} = s + w$. Then the expected values of the estimators in the regression model with both regressors are:

$$\mathbb{E}[\hat{\beta}_s] = \beta_s - \beta_s \mathbb{E}\left[\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n w_i^2)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i + w_i)^2) - (\sum_{i=1}^n p_i(s_i + w_i))^2}\right] \tag{4.2.9}$$

$$\mathbb{E}[\hat{\beta}_p] = \beta_p + \beta_s \gamma \mathbb{E}\left[\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n w_i^2)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i + w_i)^2) - (\sum_{i=1}^n p_i(s_i + w_i))^2}\right] \tag{4.2.10}$$

Expression 4.2.9 gives us the well-known result that the estimator of the effect of the intervening control variable (schooling) in equation 4.2.9 is bounded between β_s (in the absence of variation in the measurement error when $Var(w) = 0$ and the numerator of the expression in brackets goes to zero) and 0 (if the variance of w goes to infinity and the numerator of the bracketed term approaches the denominator in the limit).

Equation 4.2.10 shows that the bias term of the estimator of the direct effect of parental SES is $-\gamma$ multiplied by the bias of the estimator of the effect of schooling in equation 4.2.9, which means that it is bounded between the direct effect β_p (in the absence of measurement error) and the total effect $\beta_p + \beta_s\gamma$. This demonstrates that in case of random measurement error of the intervening control variable, the estimator of the direct effect of the regressor of interest is biased towards the total effect.

4.2.4 Instrumenting the intervening control variable if LATE is equal to ATE

Estimation based on a valid instrumental variable (IV) for the intervening control variable purges the bias from the estimator of the effect of the intervening control variable. This approach would still block the part of the effect that is transmitted by the intervening control variable, but it would remove the additional bias caused by its endogeneity.

IV estimation yields an estimated local average treatment effect (LATE), defined

as the average treatment effect for individuals who comply with the instrument in a potential outcomes framework (Imbens and Angrist, 1994). In the presence of heterogeneous effects, LATE may differ from the average treatment effect (ATE), and consequently the bias of the estimator of the indirect effect is equal to the product of γ and the difference between LATE and ATE.

The coefficient of the regressor of interest reflects the direct average treatment effect for the total population, and not only for compliers. It is usually impossible to identify compliers due to the fundamental problem of causal inference: we observe realized outcomes and not potential outcomes. Therefore we are usually unable to limit our sample to include only compliers with the instrument for the intervening control, which would allow us to give a LATE interpretation to direct and indirect effect. The difference between LATE and ATE is usually unobserved, and our identification strategy suffers from a similar problem as when the intervening control variable was not instrumented. In expectation, the estimator of the direct effect is:

$$\mathbb{E}[\hat{\beta}_p] = \beta_p - \gamma[LATE_{\beta_s} - ATE_{\beta_s}] \tag{4.2.11}$$

These can neither be interpreted as LATE nor as ATE of the regressor of interest because the coefficient of the regressor of interest relates to the ATE, while the coefficient of the intervening control variable relates to LATE. Similarly, using two instrumental variables to instrument both the regressor of the regressor of interest and the intervening control variable is problematic if the two groups of compliers do not coincide.

LATE will be equal to ATE in special cases, for example if compliance is random or if the instrument is particularly strong. I argue that the latter condition holds for the empirical setting in the next section, and therefore we can obtain unbiased estimators of direct and indirect effects in this particular case. One should keep in mind however, that LATE and ATE do not coincide in most applications, and it is not possible to obtain consistent estimators of the direct and indirect effects if this is the case.

4.3 Parental SES, schooling, and health in the UK.

Parental background is an important determinant of health during adulthood (Currie, 2009, 2011). Educational policies offer an important policy lever to improve equality of opportunity, and could be used to offset some of the health disparities by parental background. This raises the question to what extent the gradient between parental background and health is transmitted by educational attainment. A prerequisite for the transmission hypothesis to hold true, is that schooling itself has a causal effect on health. If the association between educational attainment and health is exclusively

spurious, the proportion of the gradient between parental background and health that is transmitted by schooling is zero. In this section, I first estimate direct and indirect effects using the bad control method. Next, I show that once we account for the endogeneity of schooling (the intervening control variable), it becomes clear that the results from the bad control method are misleading and that the indirect effect running through schooling at this margin is in fact close to zero.

According to the bad controls method, we regress health (a higher value indicates worse health to be consistent with the literature) on parental SES and educational attainment and interpret the coefficient of parental SES as a direct effect. However, in accordance with the derivations in the previous section, the resulting estimator of the direct effect of parental SES on health is biased upwards if (i) parental SES is positively correlated with educational attainment, and (ii) educational attainment is negatively correlated with the error term. If parental SES improves health, we underestimate the direct effect of parental SES in absolute terms, and—depending on the size of the unobserved bias term and the association between parental SES and schooling—the expected value of our estimator could even have the opposite sign. By definition, and given that a higher health value indicates worse health, the bad controls estimator of the indirect effect of parental SES that runs through educational attainment is then biased downwards.

To remove the possible bias of the estimator of the direct and indirect effects of parental background due to the endogeneity of schooling, I use exogenous variation in years of schooling that was caused by the introduction of compulsory schooling laws in Britain. This setting is particularly well-suited for two reasons. First, the reform affected a large share of the population, and therefore the resulting LATE estimates are much closer to population effects than those based on reforms that affected a much smaller proportion of the population as was argued by Oreopoulos (2006); Devereux and Hart (2010) for the compulsory schooling reform in 1947 and by Clark and Royer (2013) for both the 1947 and 1972 reforms. Second, Clark and Royer (2013) use both reforms to estimate the effect of schooling on health and report IV estimates that are much lower than the OLS estimates. This indicates that OLS estimators are severely biased and that the estimates of direct and indirect effects based on the bad controls method are highly misleading.

4.3.1 The 1947 and 1972 UK compulsory schooling reforms

The minimum school leaving age in the United Kingdom was raised from 14 to 15 in April of 1947, and then again from 15 to 16 in September of 1972, affecting those born after April 1, 1933 and September 1, 1947, respectively . Evaluating the 1947 reform and controlling for birth cohort trends before and after the reform, Harmon and Walker

Table 4.1: Descriptive statistics.

	1947	1972
Age	40.4	50.6
	(7.8)	(4.7)
Proportion health fair or bad	.32	.38
Female	.51	.51
Age left full-time schooling	16.9	16.7
	(4.0)	(4.6)
Father in white-collar occupation	.30	.28
Observations	90,721	18,925

Notes: Sample statistics for the two subsamples, standard errors in parentheses. Source: GHS.

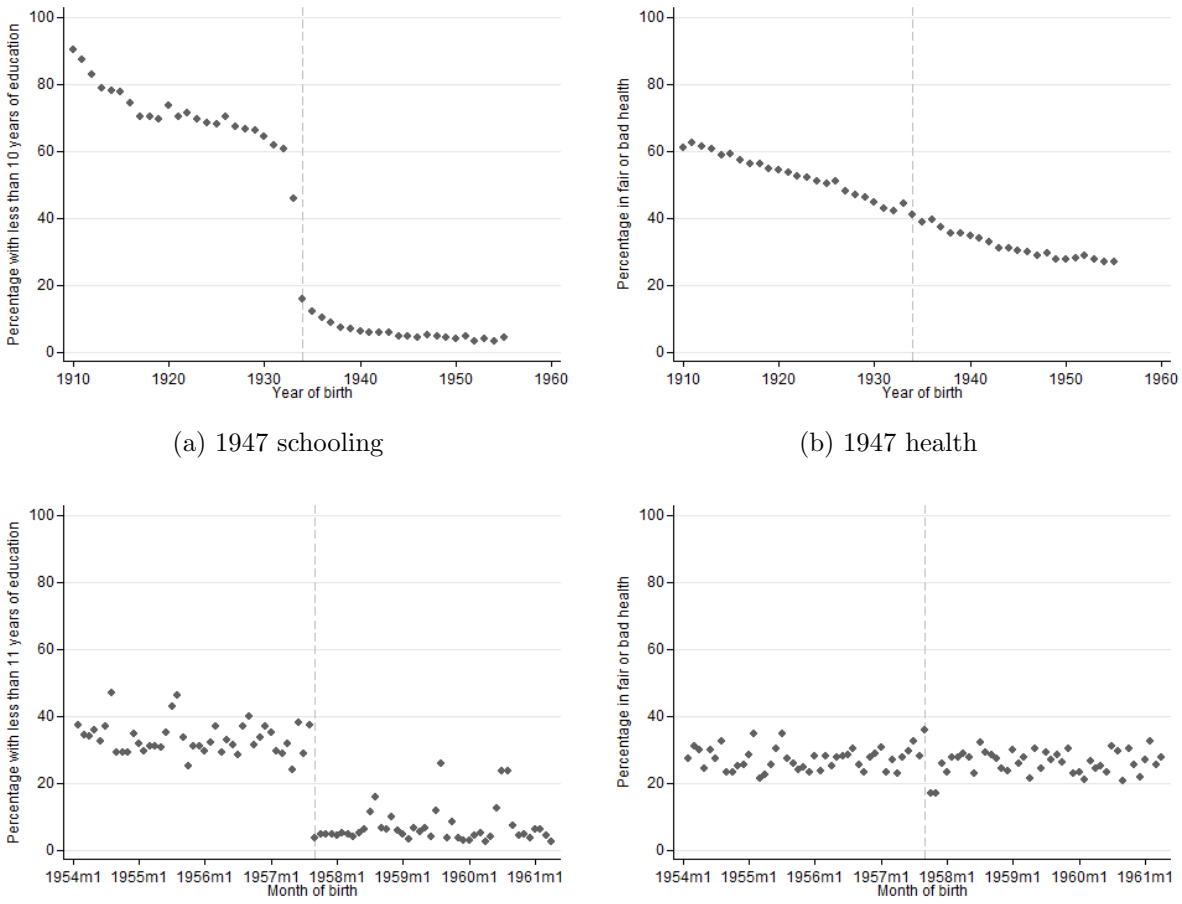
(1995) and Oreopoulos (2006) find that an additional year of schooling raises earnings by about 15 percent, which is somewhat larger than conventional OLS estimates. In later work, Clark and Royer (2013) observe exact date of birth and can therefore control for birth month effects before and after both the 1947 and 1972 reforms. They find that the health returns to schooling generated by these compulsory schooling reforms were small and that the strong association between schooling and health at this particular margin is due to third factors driving both variables.

4.3.2 Data and operationalization

The British General Household survey contains information on parental SES, year of birth, educational attainment, and a subjective health measure in the waves between 1978 and 1992. Up to 1988, the GHS records father's occupational class for individuals up to age 49. From 1989 to 1992, reporting father's occupation was extended to age 59. 1986 is the first GHS survey year for which exact date of birth is made available to researchers. Because of the missing information on exact month of birth and parental SES for older cohorts, my evaluation of the 1947 reform is based on birth-year-to-birth-year variation between 1978 and 1992. For the evaluation of the 1972 reform, I use the 1986 to 1992 waves which contain exact month of birth information.

Table 4.1 shows that due to the wider range of birth cohorts, the inclusion of earlier survey waves, and the exclusion of later survey waves, the average age for the 1947 sample is ten years lower than for 1972. Accordingly, the proportion in bad health is higher for the 1972 sample. The proportion of women, schooling, and the proportion of fathers in a white-collar occupation is similar in both samples.

Following Oreopoulos (2006), I compute year of birth by subtracting age at the time of the interview from the year of birth. Figure 4.2a shows that before the 1947 reform, more than 60 percent left school before the age of 15, and this percentage dropped to less than 15 percent after the reform. The proportion in fair or bad health in figure



(a) 1947 schooling

(b) 1947 health

(c) 1972 schooling

(d) 1972 health

Figure 4.2: —The 1947 and 1972 reforms.

Source: GHS.

4.2b decreases steadily over the birth cohorts, without apparent evidence of an effect of the reform.

Figure 4.2c shows that 33 percent of individuals who were born in the months before September 1957 completed 10 or fewer years of schooling. The figure shows a sharp drop in school leavers before the age of sixteen for the birth cohorts that were affected by the 1972 reform: after the reform only 7 percent completed at most 10 years of schooling, which indicates a 26 percent drop with respect to the pre-reform period. 27 percent of individuals in the sample report fair or bad health, with the remaining 73 percent reporting good health. As figure 4.2d shows, there is no clear jump in the proportion of the sample in fair or bad health in the cohorts before and after the reform.

4.3.3 Estimation methods

The 1947 reform affected the schooling decisions of the majority of students, which lends credence to the generalizability of the results. Availability of month of birth instead of only year of birth improves precision, but the group of compliers is smaller. I evaluate both reforms to illustrate the bad controls problem and to obtain causal estimates for the indirect effect of parental SES through schooling. I first evaluate the 1947 reform, based on the (Oreopoulos, 2006) method, and next I replicate the results by (Clark and Royer, 2013) for the 1972 reform. I use these results to estimate the indirect effect of parental SES on health that is transmitted by schooling, and compare the outcomes to the bad controls estimates.

The key identifying assumption is that the conditional expectation of the outcome is smooth through the 1934 and September 1957 birth cohort thresholds, respectively. If this assumption holds, any discontinuities in the outcome variable at the threshold can be attributed to the causal effect of the compulsory schooling reform.

I drop the notation from the previous section to be consistent with Clark and Royer and estimate the following first stage equation:

$$E_{ict} = \gamma_0 + \gamma_1 D_{ic} + f(R_{ic}, D_{ic}) + \boldsymbol{X}_{ict}^{'}\gamma_2 + \gamma_3 F_{ic} + \nu_{ict} \tag{4.3.1}$$

Where the educational attainment of individual i in yearly or monthly birth cohort c surveyed in month t is regressed on a constant, the reform dummy D for individuals who were born after March 1993 or August 1957, a function of birth cohort R which is explained below, a vector of control variables X, and father's usual job F.

For the 1947 reform, in accordance with the Oreopoulos specification, cohorts are measured in years, f is a fourth-order birth year polynomial, and X contains gender, interview month and year dummies, and a fourth-order age polynomial.

Following Clark and Royer for the 1972 specification, f is a local linear function of birth month which is normalized at zero for September 1957 and interacted with the reform dummy to allow for differential cohort trends before and after the reform, and X includes age in months and age in months squared, and dummies for birth month, birth year, and gender.

The reduced form equation is identical to 4.3.1 with the exception of the outcome variable, which is now a binary variable indicating fair or bad health:

$$H_{ict} = \delta_0 + \delta_1 D_{ic} + h(R_{ic}) + \boldsymbol{X}_{ict}^{'}\delta_2 + \delta_3 F_{ic} + \xi_{ict} \tag{4.3.2}$$

The following outcome equation describes the relationship between the health out-

come and schooling:

$$H_{ict} = \beta_0 + \beta_1 E_{ic} + h(R_{ic}) + \boldsymbol{X}_{ict}^{\iota}\beta_2 + \beta_3 F_{ic} + \mu_{ict} \tag{4.3.3}$$

β_1 in equation 4.3.3 is the ratio of the coefficient of the reform dummy in equation 4.3.2 to the first stage coefficient γ_1 in equation 4.3.1, using two stage least squares estimation.

Using the terminology of the previous section, β_3 is the direct effect of the regressor of interest (father's occupation) on the outcome variable (health), γ_3 is the effect of the regressor of interest on the intervening control variable (educational attainment), β_1 is the effect of the intervening control variable on the outcome variable, and the product of γ_3 and β_1 is the indirect effect of the regressor of interest on the outcome that is transmitted by the intervening control variable.

4.3.4 Results

Panel A of table 4.2 replicates the analysis by Oreopoulos (2006) with health as the outcome variable and including father's occupation as the regressor of interest. Column (1) shows that 31 percent of those born in the ten years leading up to the reform report fair or bad health. Column (2) shows that having a father in a white-collar occupation reduced the likelihood of being in bad health by 8 percentage points, or 24 percent. Column (3) shows the results from the bad controls method, which suggests that 23 percent of the effect of father's occupation on health is transmitted by schooling. Based on these results, one could erroneously conclude that policies to reduce socioeconomic disparities in educational attainment would also lower health disparities.

Column (4) presents results from an appropriate two-stage least squares regression—accounting for educational differences in health outcomes that are explained by selection—based on equations 4.3.1 and 4.3.2. These results show that in fact none of the effect of father's occupation on health is transmitted by schooling, and that the bad controls method leads to the wrong conclusion.

To gain insight in the role of the bias, notice that the direct effect—the difference of the coefficients of father's occupation in columns (2) and (3)—is $-.0767+.0590 = .0176$. Now consider the association between father's occupation and schooling in column (5). If we multiply this estimate of γ in figure 4.1 and equation 4.2.2 by the direct effect β_s in equation 4.2.1 which was shown to result from bias, we get $2.166 \times -.0081 = .0175$. This empirically confirms that the bad controls estimator of the direct effect is biased if the intervening control variable is endogenous. As shown in equation 4.2.8, the size of the bias is equal to minus the product of (i) the bias in the estimator of the effect of the intervening control variable and (ii) the association between the regressor of interest

Table 4.2: Bias of the bad controls estimator and results from the UK compulsory schooling reforms.

	(1) Mean Health	(2) OLS Health	(3) OLS Health	(4) IV Health	(5) OLS Schooling
A. 1947					
		Bandwidth = 1910 to 1955, N = 90,721			
In bad health	.3145***				
Father white-collar occupation		-.0767***	-.0590***	-.0768	2.166***
		(.0033)	(.0034)	(.0561)	(.0273)
Schooling			-.0081***	.0001	
			(.0004)	(.0258)	
% transmitted by schooling			23 %	0 %	
B. 1972					
		Bandwidth = 43 months, N = 18,925			
In bad health	.2730***				
Father white-collar occupation		-.0563***	-.0506***	-.0564	.1733***
		(.0187)	(.0185)	(.0919)	(.0221)
Schooling			-.0332**	.0068	
			(.0131)	(.0473)	
% transmitted by schooling			10 %	0%	

Notes: Column (1) shows average health in the sample Models (2)-(5) of panel A control for a fourth-order birth year polynomial, gender, interview month and year dummies, and a fourth-order age polynomial. With the exception of column (1), which shows average health in the twelve birth months right before the reform, the models in panel B control for a linear function of month of birth interacted with the reform, gender, interview month and year dummies and age and age squared, and are limited to having 10 or 11 years of schooling. To avoid contamination by the reform, columns (2), (3), and (5) are limited to the pre-reform cohorts. Source: GHS.

and the intervening control variable γ.

The results are not driven by the zero point estimate of the effect of schooling; indirect effects are correctly estimated if the coefficient of the intervening control variable is nonzero.[3] The estimated indirect effect, i.e. the difference between (i) the respective coefficients of the regressor of interest in the specification without the intervening control estimate and (ii) the specification with the unbiased estimator is by definition equal to the product of (i) the association between the regressor of interest and the intervening control and (ii) the causal estimate of the effect of the intervening control. Consequently, the bias of the estimator of the indirect effect is γ times the bias of the estimator of the coefficient of the bad control variable.

Panel B of table 4.2 confirms these findings for the 1972 reform for which we observe month of birth but which affected a smaller share of students than the 1947 reform. As explained in section 4.3.3 this panel replicates Clark and Royer (2013) and additionally includes father's occupation as the regressor of interest. Column (1) indicates that 27 percent of individuals who were born in the year leading up to the reform report bad or fair health. Column 2 shows that having a father in a white-collar occupation lowers the probability of being in bad health by 6 percent.

The bad controls method in column (3) suggests that 10 percent of the effect of father's occupation is transmitted by schooling. However, column (4) indicates that this resulted from the endogeneity of schooling. Once we appropriately account for nonrandom selection into level of schooling, the estimated indirect effect drops to zero. As in panel A, the difference between total effect of father's occupation in column (2) and the direct effect according to the biased bad controls estimator in column (3) ($-.0563 + .0506 = -.0058$) is equal to the product of the association between father's occupation and schooling in column (5) and the problematic coefficient of schooling in column (3) ($.1733 \times -.0332 = -.0058$). As before, the bias of the bad controls estimator of the indirect effect is equal to the bias of the estimator of the effect of the intervening control variable multiplied by minus the association between the regressor of interest and the intervening control.

4.4 Discussion and conclusion

The use of bad controls is widespread in the economics literature: Case et al. (2002) report a strong impact of parental education (regressor of interest) on child's health (outcome variable) while controlling for family income (intervening control variable); Case et al. (2005) report that the impact of parental education and SES on child's health and SES in adulthood is insignificant when controlling for the child's education,

[3] See the simulations in appendix B for an illustration.

SES and health at younger ages; Deaton and Paxson (2001) report that the effect of education on mortality disappears after controlling for income; Currie et al. (2007) report that mother's education has a stronger effect on the subjective health of the child than father's education after controlling for the education of both parents; Kenkel (1991) report that the effect of education on smoking, drinking and exercise is slightly smaller after controlling for knowledge about the effect of these behaviors on health; Cutler and Glaeser (2005) report that the education effects on smoking, drinking and obesity are large after controlling for income; Cutler et al. (2006) state that the fetal origins hypothesis is supported by the strong correlation between health in adulthood and birth weight, even when controlling for socioeconomic status; Goldman and Smith (2002) report that men and ethnic minorities are more likely to experience a worsening health while they control for education and health behaviors; Schuster et al. (2012) report that harmful health behaviors and bad health outcomes are more common among black children and Latino children than among white children. Adjustment for socioeconomic status and the child's school substantially reduces most of these differences; Lantz et al. (1998) report that the estimated effect of education on mortality becomes insignificant after controlling for income.[4] If the intervening controls in these studies are correlated with the error term, the estimators of the (direct) effects of the regressors of interest in these models are biased as well.

The bad controls problem affects the instrumental variable literature as well. In his widely-used graduate-level econometrics text book, Wooldridge (2010) discusses the exclusion restriction in the work by Angrist (1990) on the effect of serving in the Vietnam war on the earnings of men. Wooldridge correctly observes that the exclusion restriction for using the draft lottery as an instrument for veteran status is unlikely to hold because the draft lottery number also affected earnings through education, as described by Angrist and Krueger (1992). Wooldridge recommends controlling for education in the earnings equation. However, this solution potentially makes matters worse: Now, education is a bad control variable in the reduced-form equation and this could substantially bias the estimator the coefficient of the draft lottery, which is the numerator of the two-stage least squares estimator of the effect of veteran status on health.

I find that schooling is a bad control variable when we are interested in the direct

[4]Additionally in the health literature, Marmot (1994) reports that adjusting for coronary risk factors explains about 25 percent of the social gradient in CHD; Borg and Kristensen (2000) report that women have a higher risk of having poor self-reported health then men while they control for social class; Abraído-Lanza et al. (2005) report that the differences between Latinos and non-Latino whites in the odds of current smoking, alcohol intake, and BMI were exacerbated and the difference in exercise is mitigated slightly when age, education, and family income were included in the models; Baron and Kenny (1986), MacKinnon et al. (2002) and MacKinnon et al. (2007) describe the common practice of controlling for intervening variables in the psychology literature. In sociology, the use pervasive of these methods can be traced back to Duncan (1966) and Alwin and Hauser (1975).

or total effect of father's occupation on health. Instrumenting the intervening control variable when LATE is equal to ATE—which is arguably the case when the proportion of compliers is very large as with the UK compulsory schooling reforms—leads to unbiased estimators of direct and indirect effects. Unbeknown to the researcher, the bias can be large: the bad controls method suggested that 23 percent of the effect of father's occupation on health for individuals who were born in the first half of the 20th century is transmitted by schooling, while the appropriate method show that the indirect effect is in fact close to zero.

The analyses relying on the compulsory schooling reforms are informative about the role of schooling at the margin that was affected by the reform. Based on these results, we can only draw conclusions about the direct effect through schooling between ages 14 and 16.

We have seen that preceding control variables are good controls, and that intervening control variables are bad controls. When preceding control variables are correlated with omitted variables that influence both the regressor of interest and the outcome variable, they additionally remove part of the bias that results from the omitted variables and do not contaminate the estimators of the coefficients of the other variables in the regression model. The endogeneity of a bad control contaminates the estimator of the coefficient of the regressor of interest while this is not a problem for a good control. The resulting estimates are generally uninformative about the transmission mechanisms of causal effects, and neither can we conclude anything about the part of an association that is explained by intervening variables. If a partial association is the best we can produce, we should estimate a separate model for each regressor of interest separately, including only good controls. Consequently, it is usually bad practice to interpret more than one coefficient from a single regression model.

Chapter 5

Late Tracking, Intergenerational Mobility, and Human Capital: The Impact of the Finnish Comprehensive School Reform

5.1 Introduction

In many western countries, human capital outcomes in terms of education and health are positively associated with parental socioeconomic status (e.g. Case et al., 2002). Educational policies often aim to promote equal opportunities for all children, and one of the educational policy levers is the choice of the tracking age into differing-ability classrooms (Hanushek et al., 2006). In this paper we exploit a Finnish comprehensive schooling reform, which was motivated by a desire to promote more equal educational opportunities irrespective of social background (Kerr et al., 2013), to test whether delayed tracking has affected human capital disparities with respect to family socioeconomic status.

The literature suggests that educational institutions do seem to matter: Hanushek et al. (2006) use a cross-country comparison of internationally comparable student test scores in developed countries, and report that later tracking has positive effects throughout the grade distribution, and reduces inequality by benefiting weak students more than strong students. Duflo et al. (2011) find that in developing countries, teach-

This chapter is based on Ravesteijn, B., M. Avendano Pabon, P. Martikainen, E. van Doorslaer, H. van Kippersluis, H. Vessari (2015). Late tracking, intergenerational mobility, and human capital: the impact of the finnish comprehensive school reform. Mimeo.

ers have incentives to teach at the level of their top students. This implies that any positive peer effects of late tracking for weak students may be offset by the negative effect of a too high level of instruction.

Some previous studies have explicitly focused on educational reforms, in the UK and in some Scandinavian countries. Jones et al. (2014) relate childhood circumstances to later-life health outcomes using the UK National Child Development Study (NCDS), and explicitly frame their empirical analysis into the inequality of opportunity framework (e.g. Roemer, 2009). They exploit the fact that members of their 1958 cohort were exposed to different educational regimes—a comprehensive schooling reform in the 1960s affected only some part of the cohort—to study the effect of different educational policies on inequality of opportunity in health. One limitation of this approach is that it is not possible to control for systematic differences in the outcome between regions which were exposed to the reform and regions which had not been exposed yet. This means that one could mistakenly attribute regional differences to the reform. Their results are mixed: the worst-off groups in society are better off in terms of some health outcomes under the new comprehensive schooling system, while they are worse off in terms of other health outcomes.

In Sweden, Meghir et al. (2013) have exploited a schooling reform in the 1950s and early 1960s that increased the compulsory years of schooling by one or two years depending on the municipality of residence and which also raised the tracking age. They find a positive effect on military test scores, with the strongest gains for boys from low-income families and of low ability. Moreover, they find that the reform increased the completed years of schooling of boys with low-educated fathers by almost three and a half months. While the authors do find a decline in mortality for low-ability individuals, they conclude that the reform only had a minor negative effect on the number of sick days of men with low-educated fathers, and they find no effect on hospitalizations.

A series of papers by Pekkarinen and co-authors has used the Finnish compulsory schooling of 1972-77 reform to examine a number of non-health outcomes. They find that it has led to an increase in the gender differences in educational attainment, to a decrease of the gender wage gap in adult income by four percentage points (Pekkarinen, 2008), to a decrease in the intergenerational income elasticity (Pekkarinen et al., 2009) and to an increase in the military test scores of boys from parents with low levels of education (Kerr et al., 2013).

In this paper we exploit the same Finnish schooling reform that was gradually implemented in six geographical regions between 1972 and 1977. The variation in educational policy over time and across regions allows us to cleanly identify the effect of the reform. The reform effectively had no impact on the average years of schooling, which allows identifying the effect of the increase in the tracking age on human capital

outcomes. We observe an eleven percent sample of the Finnish population which includes an unusually rich set of information on childhood circumstance, educational attainment, earnings, medicine use, hospital visits, and mortality. Combining the richness of the data with a clean identification strategy, this paper is able to investigate whether educational policies can influence socioeconomic disparities in human capital (that is, educational attainment and health).

Compared to the series of papers by Pekkarinen and co-authors on the Finnish reform, we are the first to focus on disparities in health and education with respect to household income during childhood. Our contribution over Meghir et al. (2013) and Jones et al. (2014) is that we are able to identify the effect of a reform that changed the tracking age controlling for ex ante differences between treated and untreated regions and for time trends. In contrast, Jones et al. compare individuals in two different schooling systems at a given point in time, without being able to account for unobserved heterogeneity. The Swedish reform does not permit disentangling the effect of increased years of schooling and tracking. Moreover, the Finnish reform was gradually implemented across six regions, resulting in six clusters for computing the standard errors of our estimates. This gives us an advantage in terms of statistical precision compared to the Swedish case, where treatment was simultaneously assigned to a set of municipalities at a given moment in time. This is potentially problematic if standard errors are clustered at the level of assignment of the treatment (Bertrand et al., 2004).

We find evidence that delaying the educational tracking age in Finland in the 1970s had a persistent impact on socioeconomic disparities in human capital throughout the life-cycle. The reform generally reduced disparities with respect to household income during childhood in terms of educational attainment and in terms of longevity. Yet, we find that the tide of the Finnish compulsory schooling reform did not lift all boats. While the reform reduced socioeconomic disparities, the gains for children from poor families seem to have come at the cost of some losses for children from affluent families.

This paper is organized as follows. In the next section, we outline the theoretical mechanisms by which delayed tracking could affect human capital decisions and mortality. In section 3 we discuss the comprehensive schooling reform, and in section 4 the data. Section 5 discusses the empirical methodology, section 6 the results, after which we conclude in section 7.

5.2 Socioeconomic disparities and the optimal tracking age

Cunha et al. (2006) outline a model of human capital formation where the effect of an early life intervention is multiplied by interactions with later life investments: the intervention sets off a chain of positive events throughout the life course. Despite empirical evidence in favor of this hypothesis for the affected individuals (e.g. Campbell et al., 2014), general equilibrium effects could differ in the presence of spillover effects. In the case of educational policy, since school grades are often rank-based, interventions that particularly benefit students with a disadvantaged background may "push downwards" the grades of their classmates, thereby adversely affecting their employment prospects. This type of reasoning does not solely apply to grades. In the remainder of this section, we outline two further mechanisms through which the effect of late tracking, as in the Finnish schooling reform, can have positive effects on the development and performance of some students, and negative effects for others: through peer effects and the "distance" between the level of the student and the level of instruction.

We envision a model in which student performance at age t is determined by (i) ability, assumed to be largely constant over time, (ii) childhood circumstance, which has a strong effect at younger ages that gets weaker as the child gets older, (iii) peer effects, which are increasing in the performance of the student's classmates (Sacerdote, 2011), and (iv) the distance between past student performance and the level of instruction, which is determined by the average or median student performance. Up to the tracking age, high- and low-performing students are in the same classroom in comprehensive school with a one-size-fits-all curriculum. At the tracking age, they are tracked into either an academic or a vocational track, based on academic performance.

Given these four determinants of student performance, a student at a younger age may not have revealed his true adult potential (Sacerdote, 2011; Brunello and Checchi, 2007; Malamud and Pop-Eleches, 2011). Instead, the student's performance is codetermined by childhood circumstance which becomes less important as the student gets older (Dustmann, 2004). Hence, tracking at a young age based on student performance depends greatly on childhood circumstance, and may lead to mistracking: as the adult potential of a student manifests itself in late childhood, it may become apparent that the student is in a track that is either too demanding or too easy. For a student who is undertracked, his/her peers may drag down the performance and the level of instruction may be too low. By the end of their schooling career, they have academically suffered from interaction with weaker peers and a low level of instruction. A student who is overtracked benefits from the interaction with his/her higher ability peers, but may suffer from the advanced level of instruction and may fail the exams.

Switching tracks midway involves switching costs – the costs of adjusting to a new learning environment.

The upside of early tracking, as in the old Finnish schooling system, is that students – at least those who were not mistracked – are taught at an appropriate level which better matches their performance. Students in the high-level track are not held back by their low-achieving peers. However, students in the low-level track are not able to benefit from high-achieving classmates.

Late tracking, as in the new Finnish schooling system, means that students with high and low performance remain in the same classroom until an older age when their performance may reflect their ability more accurately, such that mistracking is less likely. While low-performing students may benefit from interacting with their high-performing peers, high-performing students may suffer from interacting with weaker students. Moreover, both high-achieving and low-achieving classmates are taught at the same level of instruction, which potentially hampers their development.

The preceding discussion shows that we expect heterogeneous effects of the Finnish compulsory schooling reform, depending on ability and childhood circumstance. The empirical test here is to investigate which of the theoretical mechanisms – the likelihood of mistracking, peer effects, and the distance between the level of instruction and the level of the student – dominate by determining the sign of the treatment effect. While we do not observe childhood ability, we do observe parental socioeconomic status, and we will allow the treatment effect of the reform to vary with parental socioeconomic status of the child.

5.3 The Finnish comprehensive school reform

The Finnish comprehensive school reform was gradually implemented between 1972 and 1977. Finnish children start primary school at the age of 7 (see figure 5.1) but before the reform, children were separated by academic ability into one of two different tracks at age 11. Tracking was based on an entrance examination, school grades and teacher assessment. The high track prepared for upper secondary education and university while the low track prepared for civic school and vocational school. After the comprehensive school reform, tracking was postponed until the age of 16. It meant that children of different academic ability were held together in the same classes throughout the 9 years of comprehensive education.

The reform was gradually implemented from Northern to Southern Finland in six implementation regions in the six years between 1972 and 1977. It was somewhat controversial because it raised concerns that education in the new comprehensive school would be of lower quality than in the old general secondary school. Opposition was

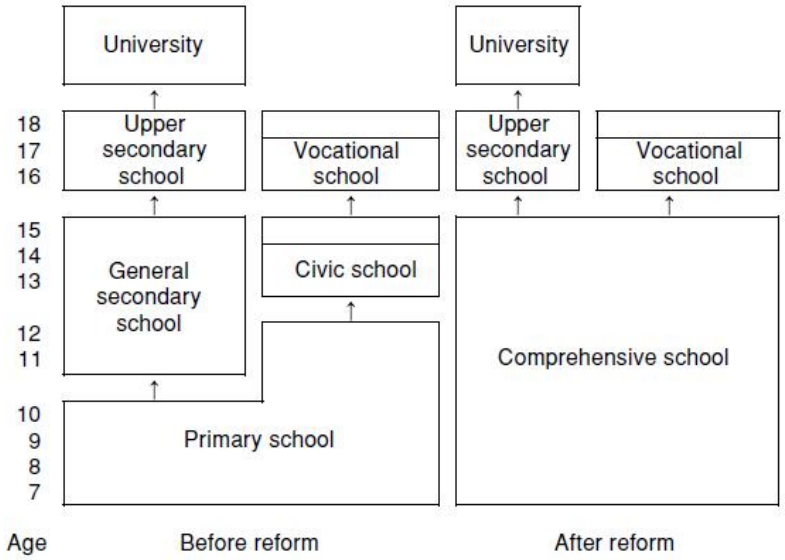


Figure 5.1: Finnish school system before and after the reform.

Source: Pekkarinen (2009)

Table 5.1: Implementation of the reform

		Region					
		1972	1973	1974	1975	1976	1977
Birth cohort	1960	C	C	C	C	C	C
	1961	T	C	C	C	C	C
	1962	T	T	C	C	C	C
	1963	T	T	T	C	C	C
	1964	T	T	T	T	C	C
	1965	T	T	T	T	T	C

Notes: Birth cohorts by region under the old regime (controls, "C") or in the new system (treatment cells, "T"). Region 1972 refers to the region that underwent the reform in 1972, region 1973 refers to the region that underwent the reform in 1973, and so on.

most fierce in the densely populated Helsinki region, and figure 5.2 shows that Helsinki was the last region to implement the reform.

In the year of the reform, pupils who were between 12 and 15 years old continued in the old system, while students starting grades five and below were transferred to the new system. To illustrate this, we look at the first implementation region, where the reform was implemented in 1972. The 1961 birth cohort that started the 5th grade in the year of the reform went through the old system for 4 years and then through the new system for 5 years. Pupils from this region who were born in 1965 were the first to only attend comprehensive school. This implementation scheme is illustrated in table 5.1.

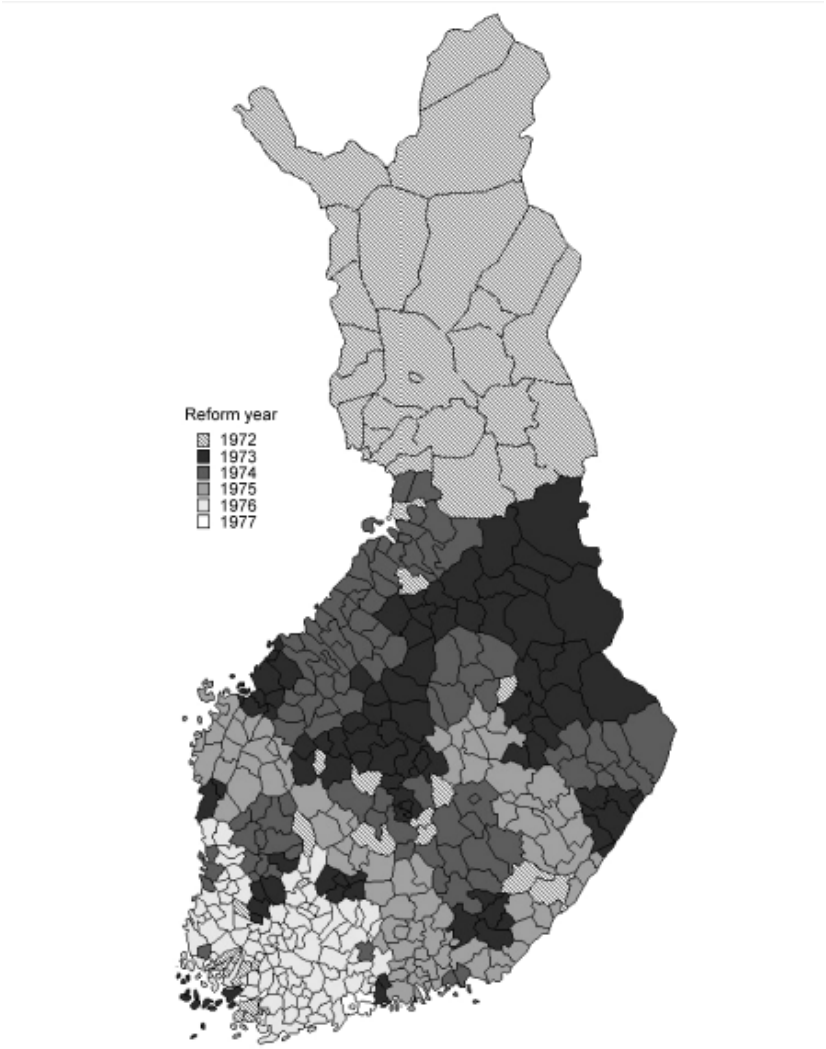


Figure 5.2: The year of the implementation of the reform varied between municipalities.

Source: Pekkarinen (2009)

5.4 Data on socioeconomic status, education, and health in Finland

We use an 11 percent random sample of Finnish residents who were born between 1960 and 1965 and who were residing in Finland between 1987 and 2007 obtained from the Labour Market Data File and Census Records maintained by the Statistics Finland. These data were further linked with information on health outcomes as available in the death records (Statistics Finland), hospital discharge records (National Institute for Health and Welfare) and prescription medication records (Finnish Social Insurance Institution).

The total sample size for our main analysis is 56,502 individuals. We will analyze males and females separately to allow for a differential response to the schooling reform by gender, since the gradient for some of the outcome measures differs substantially by gender. In this section, we will discuss our variables relating to family socioeconomic status, personal educational attainment, and personal health outcomes, respectively.

5.4.1 Family socioeconomic status

Through a household identification number, we can link children to their parents and observe household income during childhood. Information on income is obtained from the tax register, and our main measure of family socioeconomic status is the inflation-adjusted average of household taxable incomes observed in 1970 and 1975 (i.e. during the childhood of the individual). Table 5.2 shows sample characteristics for males and females in the 1960 to 1965 birth cohorts. Average household income did not differ between boys and girls, and the high inflation level led to a dramatic increase in nominal household income. Household income was top-coded at 17,705 in 1970 and at 17,800 in 1975 – both in 1975 prices. To emphasize that household income was measured during childhood, we will use the term parental income interchangeably with household income.

5.4.2 Educational attainment

Retrospective information on educational attainment predating 1970, the first census year, is available from the 1980 census. Education is measured as the highest degree that was obtained, while precise information on years of schooling is not available. Table 5.2 shows that close to sixteen percent of men and eighteen percent of women obtained at least a college degree.

Table 5.2: Descriptive statistics

	Total	Men	Women
Average household income in 1970 and 1975	14,229 (7,421)	14,225 (7,412)	14,232 (1,430)
College degree	.168	.156	.181
Total accidents	1.139 (2.965)	1.576 (3.645)	.640 (1.787)
Hypertension medication	.276	.254	.300
Antidepressants	.236	.204	.274
Deceased	.037	.049	.023

Notes: Sample means; standard deviations in parentheses. Income in 1975 prices. Mortality for the period between 1987 and 2012, medication and accidents for the period between 1995 and 2007. Source: Statistics Finland.

5.4.3 Health outcomes

We focus on selected health outcomes that show sufficient health variability and a socioeconomic gradient by middle age, as the cohorts affected by the schooling reform were only in their mid-forties by 2007, our last observation year for all health outcomes except mortality, which we observe until 2012. Health outcome data were obtained from linking the sample to three other data sources: the mortality registry, the hospitalization registry, and a dataset with drug prescription information by diagnosis for outpatient treatment. The latter information is fairly unique as in most countries, data on chronic diseases can only be derived from the admission diagnosis of hospitalized patients, which misses out on all non-hospitalized patients. All medically prescribed drug purchases in Finland are eligible for reimbursement, and therefore financial barriers for seeking pharmaceutical treatment of chronic diseases may be less than in other countries (Nihtilä et al., 2008).

We use four indicators for adult health, taken from the three linked data sources. (1) Survival until 2012, obtained from the cause-of-death registration, (2) the number of hospital admissions for an accident between 1987 and 2007 from the hospitalization registry, (3) a dummy variable which indicates the purchase of anti-hypertension medication between 1995 and 2007, and (4) a dummy variable which indicates the purchase of antidepressants between 1995 and 2007. Survival, accidents, hypertension and depression are all health indicators that display a socioeconomic gradient in Finland in middle-aged adults (Valkonen et al., 2000; Jakovljević et al., 2001; Talala et al., 2009; Martikainen et al., 2014). We chose a cause of significant morbidity (accidents) and two chronic conditions (hypertension and depression) that are sufficiently prevalent among young adults in order to have sufficient power to detect relevant health differences among this relatively healthy age group.

Table 5.2 shows that men were much more likely than women to be hospitalized for an accident, and less likely to purchase hypertension medication – 25 percent of men and 30 percent of women – and antidepressants – 20 and 27 percent, respectively – between 1995 and 2007. 3.7 percent of our sample passed away between 1987 and 2012, with mortality among men more than twice as high compared to women.

5.5 Estimation methods

We employ a variation on the differences-in-differences estimator to estimate the effect of the reform on the intergenerational income effect on human capital. In particular, we test whether the reform has changed the influence of household income during childhood on individuals' human capital outcomes later in life. Our identification strategy relies on the assumption that differences between treatment regions remain constant across the 1960 to 1965 birth cohorts: cohort effects are assumed to be the same for each of the six regions in the absence of the reform.

We do this by regressing our human capital outcomes—educational attainment and health indicators—on the natural logarithm of household income during childhood h—the inflation-adjusted average of 1970 and 1975 household income—, a treatment dummy t, and the interaction between the two. We control for five birth cohort dummies relating to the 1960 to 1965 birth cohorts in vector c, five reform region dummies in vector r, and the interaction between those two sets of dummies and our measure of household income to account for differential effects of household income across birth cohorts and regions. The regression equation becomes:

$$\begin{aligned} y_{icr} = \beta_0 &+ \beta_1 h_{icr} + \beta_2 t_{icr} \\ &+ \beta_3 h_{icr} t_{icr} + \beta_4 \boldsymbol{c}_c \\ &+ \beta_5 r_r + \beta_6 \boldsymbol{c}_c h_{icr} + \beta_7 \boldsymbol{r}_r h_{icr} + \varepsilon_{icr} \end{aligned} \tag{5.5.1}$$

Where y is the outcome variable for individual i in cohort c in region r, β_1 reflects the gradient between the natural logarithm of parental income and the relevant outcome for the baseline cohort 1960 in the first region that experienced the reform, β_2 reflects the treatment effect of the reform, and β_3 reflects how the gradient has changed due to the reform. Note that parental income was scaled by subtracting the median, such that the interpretation of the coefficient of the reform dummy is the effect on children from median-income parents. We exclude the 1966 birth cohort as all individuals in this cohort are treated and we do not have observations for an untreated region to construct the counterfactual based on the common trend across regions. We cluster standard errors at the regional level, which is the level of variation in the treatment

variable. Following Cameron et al. (2008) and Donald and Lang (2007) we use a t-distribution with $g - 1 = 5$ degrees of freedom. This leads to higher threshold values for the t-statistic for each given significance level (Angrist and Pischke, 2008).

We use OLS estimation because of the difficulties associated with interaction terms in nonlinear models (Lechner, 2011; Norton et al., 2004; Puhani, 2012). Nonetheless, we checked robustness to using logit models for the binary outcomes, and in section 6.3 we estimate a Cox proportional hazard model for mortality to exploit the duration dimension of our data. The use of non-linear models requires a different set of assumptions compared to the linear model described before. While the index function is linear, the link function is not. We can therefore no longer assume parallel trends in the outcomes, but only parallel trends in the index function. We present these results as a robustness analysis to check whether the results are confirmed under these alternative assumptions. The survival data is left truncated because we observe survival from 1987 onwards, and right-censored at January 1, 2013.

5.6 Results

5.6.1 Educational attainment

Table 5.3 presents the results for our first outcome measure: the probability of completing at least a college degree. Panel A of table 5.3 shows the association between the natural logarithm of household income during childhood and educational attainment for the cohort-region combinations in our sample that did not experience the reform (the cells in table 5.1 that are denoted with a C). On average, a one percent increase in household income during childhood is associated with a .095 percentage point predicted increase of the probability of obtaining at least a college degree.

Panel B presents the estimated average treatment effects, obtained from the estimation of equation 5.5.1. It shows that while there is no significant effect of the reform on higher education for individuals at the median of household income, the reform did reduce the gradient between household income and educational attainment by an estimated eighteen percent, from .095 to .078. For men and women separately the point estimates suggest a similar pattern, although these effects do not reach statistical significance at conventional levels. The fact that the reform reduced disparities in educational attainment by household income suggests a pathway for the effect of the reform on the reduction of the intergenerational income elasticity, as reported by Pekkarinen et al. (2009).

While children from poor families benefited from the reform, the equalization of opportunities seems to have come at the expense of children from the richest families.

Table 5.3: Results for college attainment

	Total	Men	Women
A. Association with parental income before reform			
Log parental income	.095***	.093***	.098***
	(.003)	(.004)	(.005)
Observations	34,990	18,637	16,353
B. Linear interactions			
Reform × log parental	-.017**	-.017	-.020
income	(.007)	(.014)	(.010)
Reform at median	-.003	-.012	.008
parental income	(.005)	(.011)	(.004)
Observations	56,502	30,268	26,234
C. Dummy interactions			
Reform for high income	-.015	-.009	-.021*
	(.010)	(.012)	(.009)
Reform for low income	.007	-.005	.021
	(.004)	(.005)	(.009)
Observations	56,502	30,268	26,234

Notes: The sample is limited to the 1960 to 1965 birth cohorts. The coefficients of the cohort and region dummies and their interactions with household income are omitted from this table. Standard errors are clustered at the region level. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level. Source: Statistics Finland.

Figure 5.3 plots the change in the gradient due to the reform and the probability density function of parental income. The dashed line lies above the solid line at incomes below the median parental income, indicating a positive effect on educational attainment. For children from families with income above the median level, however, the estimated effect is negative.

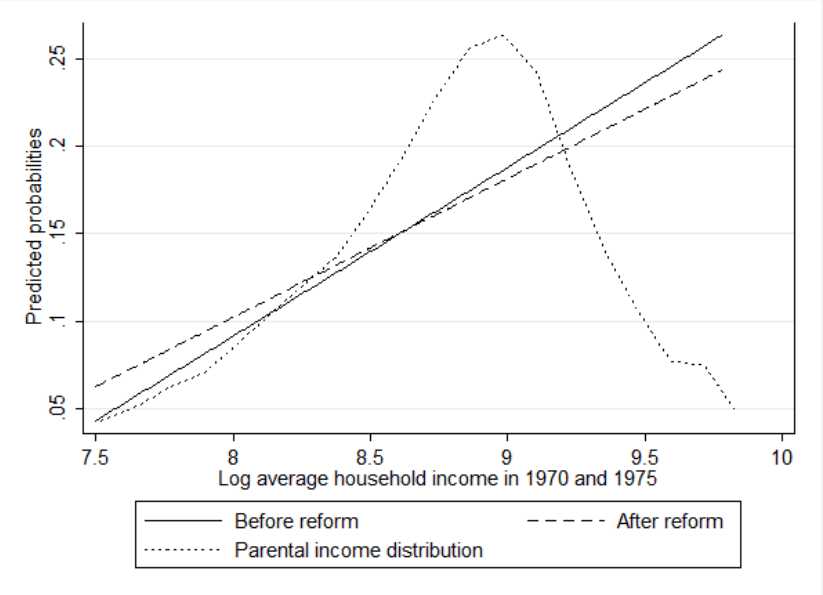


Figure 5.3: The change in the gradient. Source: Statistics Finland

Panel C of table 5.3 shows an alternative specification which is identical to equation 5.5.1, but with parental income now dichotomized, which allows us to estimate the average treatment effects for the bottom half and top half of the parental income distribution separately. While the point estimates tell a similar story that disparities in education were reduced by the reform, only the negative point estimate for girls from high-income families is significant at the ten percent significance level.

5.6.2 Mortality

Apart from educational attainment, another critical component of human capital is health, and mortality is the most absolute measure of health. Panel A of table 5.4 shows that a 10 percent increase in parental income is associated with a .14 percentage point reduction in the likelihood of death, indicating socioeconomic disparities in mortality among men. Panel B shows a specification with a linear interaction between the reform dummy and the natural logarithm of parental income, which suggests that the reform reduced inequalities in mortality with respect to parental income, but this effect is not significant. Our point estimate of the effect for children from median-income families is close to zero. However, panel C indicates that this result may be due to the linear

Table 5.4: Mortality

	Total	Men	Women
A. Association with parental income before reform			
Log parental income	-.014***	-.022***	-.006***
	(.002)	(.003)	(.002)
Observations	34,990	18,637	16,353
B. Linear interactions			
Reform × log parental income	.010	.016	.003
	(.011)	(.016)	(.008)
Reform median parental income	.003	.005	.000
	(.003)	(.005)	(.004)
Observations	56,502	30,268	26,234
C. Dummy interactions			
Reform for high income	.022**	.035**	.005
	(.005)	(.012)	(.008)
Reform for low income	-.010**	-.016	-.003
	(.004)	(.010)	(.006)
Observations	56,502	30,268	26,234

Notes: The sample is limited to the 1960 to 1965 birth cohorts. The coefficients of the cohort and region dummies and their interactions with household income are omitted from this table. Standard errors are clustered at the region level. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level. Source: Statistics Finland.

specification: estimation of separate effects for low- and high-income groups shows a significant negative effect for children with a high-income background, and a positive effect for children from low-income families. The negative effect for children from high-income families seems to be larger among men. This confirms our earlier conclusion that children from low-income families benefited from the reform, but that it harmed the survival prospects of children from high-income families, while the aggregate effect is zero.

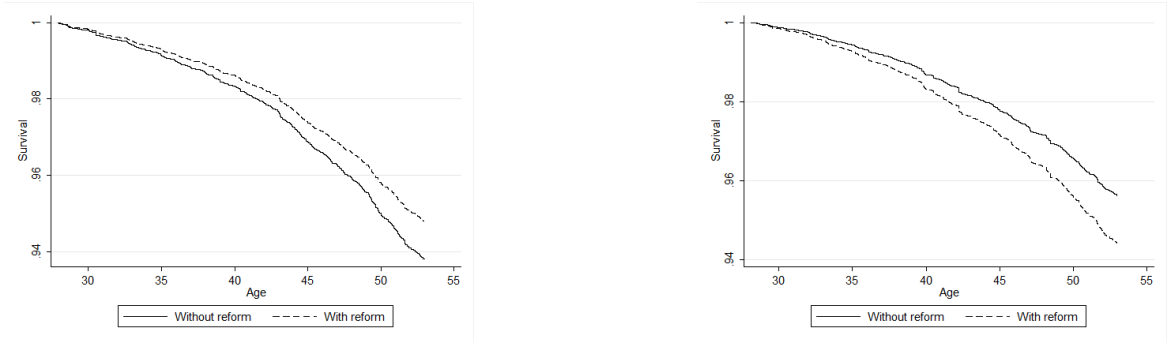
Panel A of table 5.5 shows that throughout the life cycle, the estimated probability of dying for individuals who were raised in above-median-income households is 19 percent lower compared to those from low-income families. The point estimates of the difference for men and women are similar. Panel B shows that the reform has substantially improved longevity for individuals who were raised in below-median income families, while panel C shows that it increased mortality for individuals who were raised in above-median families. This confirms our earlier finding that the effect of the reform varies by the parental income and that the gains of children from poor families came at the expense of (primarily male) children from rich families.

Figures 5.4a and 5.4b show that the survival curve of individuals from low-income families before the reform (the solid line in figure 5.4a) lies well below the survival curve

Table 5.5: Cox proportional hazard models for mortality

	Total	Men	Women
A. High versus low parental income pre-reform			
High income	.81***	.80***	.81*
	(.06)	(.05)	(.10)
Observations	34,990	18,637	16,353
B. For low parental income			
Reform	.83***	.81	.91
	(.06)	(.11)	(.18)
Observations	27,649	14,845	12,804
C. For high parental income			
Reform	1.29***	1.38***	1.06
	(.04)	(.06)	(.17)
Observations	30,311	16,221	14,090

Notes: Hazard ratios for mortality from a Cox proportional hazard model with left truncation in 1987 and right censoring in 2013. The first panel refers to the association between the sum of household income (the sum of 1970 and 1975 in 1975 prices) during childhood and mortality for the 1959 and 1960 birth cohorts in the regular sample. The coefficient of the 1959 birth cohort dummy was omitted from this table. The second panel refers to the estimated effect of the reform on mortality across the distribution of household income (the sum of 1970 and 1975 in 1975 prices) during childhood. The sample is limited to the 1960 to 1965 birth cohorts in the regular sample. The coefficients of the cohort and region dummies and their interactions with household income are omitted from this table. Standard errors are clustered at the region level.* indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level. Source: Statistics Finland.



(a) Survival after age 28 for children from below-median-income households

(b) Survival after age 28 for children from above-median-income households

Figure 5.4: Effect on survival, split by parental income.

Source: Statistics Finland

of those from high-income parents before the reform (the solid line in figure 5.4b). The dashed lines in both figures plot the (predicted) survival curves after the reform. After the reform, the difference between both groups is no longer significant (results not shown).

5.6.3 Health behavior outcomes

Unlike mortality, hospitalizations and medication purchases are behavioral outcomes: they require individuals to decide to seek treatment. As such, they do not only reflect health outcomes, but also an individual's propensity to seek care. We only use hospital admissions for accidents which may be argued to be an outcome measure that is less subject to discretion. Panel A of table 5.6 shows that pre-reform a one percent increase in parental income is associated with .23 fewer hospitalizations for accidents. Panel B shows that the reform eliminated much of the initial gradient between income and hospitalizations for accident reasons. A simple regression of the number of accidents on the natural logarithm of parental income for the individuals who were affected by the reform is no longer significant (output not shown). The reduction in the gradient is due to the effect of the reform on men, while the estimate of the effect on women is not significant. The point estimate of the effect of the reform is relatively large and positive, which explains why the negative point estimates for children from low-income families in panel C are not significant: among individuals from below-median income households, the predicted effect for the upper tier of this group is still positive.

Table 5.7 shows that a one percent increase in parental income was associated with a .025 percentage point reduction in the likelihood of purchasing hypertension medication before the reform. We find some weak evidence that the reform reduced the gradient for men (significant at the ten percent level only) in the linear specification

Table 5.6: Hospital admission for accidents

	Total	Men	Women
A. Association with parental income before reform			
Log parental income	-.233***	-.234***	-.230***
	(.023)	(.037)	(.029)
Observations	34,990	18,637	16,353
B. Linear interactions			
Reform × log parental income	.552**	.963**	.070
	(.158)	(.291)	(.098)
Reform median parental income	.494*	.859*	.068
	(.193)	(.291)	(.077)
Observations	56,502	30,268	26,234
C. Dummy interactions			
Reform for high income	1.15***	1.969***	.168
	(.275)	(.426)	(.172)
Reform for low income	-.144	-.253	-.015
	(.171)	(.385)	(.137)
Observations	56,502	30,268	26,234

Notes: The sample is limited to the 1960 to 1965 birth cohorts. The coefficients of the cohort and region dummies and their interactions with household income are omitted from this table. Standard errors are clustered at the region level. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level. Source: Statistics Finland

Table 5.7: Results for hypertension medication

	Total	Men	Women
A. Association with parental income before reform			
Log parental income	-.025***	-.022***	-.028***
	(.004)	(.005)	(.006)
Observations	34,990	18,637	16,353
B. Linear interactions			
Reform × parental income	.014	.040*	-.017
	(.013)	(.017)	(.024)
Reform median parental income	.009	.011	.007
	(.007)	(.009)	(.008)
Observations	56,502	30,268	26,234
C. Dummy interactions			
Reform for high income	.009	.027	-.014
	(.019)	(.016)	(.028)
Reform for low income	.003	-.007	.016
	(.010)	(.016)	(.014)
Observations	56,502	30,268	26,234

Notes: The sample is limited to the 1960 to 1965 birth cohorts. The coefficients of the cohort and region dummies and their interactions with household income are omitted from this table. Standard errors are clustered at the region level. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level. Source: Statistics Finland.

but not in panel C.

The evidence presented in Table 5.8 shows that before the reform, a one percent increase in parental income was associated with a .017 decrease in the likelihood of purchasing antidepressant medication, and that the gradient is steeper among women compared to men. Findings reported in Panel B suggest (at the 10 percent significance level only) that the reform increased the likelihood of purchasing antidepressants for men at the median of the parental income distribution. Panel C shows that even in the absence of an aggregate effect, the reform actually increased disparities for women: the gap in antidepressant use between high and low income women became bigger after the reform.

Table 5.8: Results for antidepressant use

	Total	Men	Women
A. Association with parental income before reform			
Log Parental income	-.017***	-.013**	-.022***
	(.004)	(.005)	(.006)
Observations	34,990	18,637	16,353
B. Linear interactions			
Reform × parental income	.004	.032	-.028
	(.012)	(.023)	(.014)
Reform median parental income	.022*	.031*	.011
	(.009)	(.023)	(.008)
Observations	56,502	30,268	26,234
C. Dummy interactions			
Reform for high income	-.004	.035	-.048*
	(.013)	(.020)	(.023)
Reform for low income	.025**	.012	.039**
	(.009)	(.016)	(.011)
Observations	56,502	30,268	26,234

Notes: The sample is limited to the 1960 to 1965 birth cohorts. The coefficients of the cohort and region dummies and their interactions with household income are omitted from this table. Standard errors are clustered at the region level. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level. Source: Statistics Finland.

5.7 Conclusion

Adjusting the tracking age is a popular educational policy to influence human capital formation. In particular, late tracking is often intended to reduce socioeconomic disparities in human capital in society. Despite some evidence that late tracking can reduce intergenerational income elasticities (Brunello and Checchi, 2007; Pekkarinen et al., 2009), strikingly little is known about the effect of late tracking on (socioeconomic disparities in) education and health outcomes.

Exploiting regional variation in the moment of tracking in Finland, our differences-in-differences framework adds to the literature by accounting for unobserved heterogeneity that remained constant across birth cohorts but varied over the six reform regions, and for unobserved heterogeneity that affected birth cohorts differently but that was constant across regions. We find that delaying the tracking age generally reduces the gradient between socioeconomic circumstance during childhood and human capital outcomes later in life. This conclusion holds for college attainment, for mortality, and for accident-related hospitalizations. We merely find suggestive evidence that the reform reduced disparities for hypertension medication purchases, while the reform increased the gradient for antidepressant medication. This paper is the first to study the effect of late tracking on health outcomes. Health may have been affected by the reform directly, through a different set of peers and health knowledge obtained in school, but also indirectly through different career paths and income levels as a result of the reform. Apart from mortality, three of our health measures are related to health care use. Health care seeking behavior, such as medicine use, may simultaneously reflect both health and health knowledge. Our results for health care use therefore cannot separate possible effects on both health and health knowledge, and this may explain the more modest impact on those outcomes (see also Jones et al., 2014).

The reform did not affect between-school socioeconomic disparities, while there is ample interregional variation in socioeconomic conditions. Tracking will not have an impact on the intergenerational relationships between schools, and further research should make clear whether certain parents were more likely to move to school districts after the tracking age was raised.

Strikingly, the human capital gains for children from poorer families have partly come at the cost of children from affluent families. We have identified the winners and losers of late tracking. The reduction in human capital disparities came at a price: while the reform had a positive effect on college attainment and longevity for children from low-income families, it had the opposite effect on children from high-income families. Hence, while the reduction of inequalities with respect to parental socioeconomic status will be largely considered a success, and potentially policymakers are willing to accept

some reduction in income among the better-off, plausibly the penalty that is paid in terms of human capital outcomes including mortality was unintended. This paper therefore tells a cautionary tale that while late tracking may indeed be successful in reducing socioeconomic disparities, potentially this comes at the expense of worse human capital outcomes among the better-off.

Chapter 6

Conclusions and Discussion

In each of the five chapters of this dissertation an attempt is made at unraveling the effects of policy on socioeconomic health disparities. The policy implications depend on the validity of the results, as well as on a normative framework. I will discuss the policy implications of each paper in turn, then synthesize the general lessons, and conclude with a research agenda for future work.

6.1 The results from each chapter

In the first chapter, we show that blue-collar workers are in worse health than white-collar workers in the Netherlands. A survey of the literature discusses the methodological difficulties of estimating the causal effects of physical and psychosocial occupational characteristics on health.

In the second chapter, using data from a long-running German panel, we find that selection accounts for at least 60 percent of the association between health and both physical workload and control possibilities at work, while it explains away the association between psychosocial stress and health. Our causal effects estimates indicate that the negative effects of physical workload and low job control on health increase with age.

The major contribution of the third chapter is mostly methodological in nature: it formally shows how a widely-used method of estimating direct and indirect effect breaks down in most real-world applications. It shows how controlling for endogenous variables—that intervene in the causal pathway between the regressor of interest and the outcome variable—leads to bias. The chapter outlines a solution to this problem of endogenous intervening controls, which relies on having exogenous variation in both the regressor of interest and the intervening control variable under the condition that both coefficients refer to effects for the same underlying population. Applying these insights to the relationship between parental SES, education, and health in the UK,

I find that—contrary to the results of the conventional method—the indirect health effect of parental SES transmitted by years of schooling is close to zero.

The fourth chapter shows that parental background is an important determinant of later-life human capital outcomes in Finland, but that educational tracking policies can close this gap. Late tracking reduces the disparities in educational attainment and health by parental SES, improving the position of the worst off and setting off a chain of events over the life cycle. However, the gains of children from poor families seem to have come at the cost of losses for children from rich families, which means that the average treatment effect is zero. Whether the tracking age should be raised or lowered is therefore subject to political debate. These results show that heterogeneous effects can be obscured when only estimating average treatment effects.

6.2 Policy implications and research agenda

The policy lessons from this dissertation all relate to distributional implications of different policies. They depend on normative ethics and one's beliefs on whether disparities in health result from exogenous endowments. If one believes that (at least part of) the current socioeconomic disparities in health result from circumstance, and that these disparities should be reduced, there are five lessons to be learned from this dissertation.

First, occupational rank is strongly associated with morbidity, mortality and self-assessed health in the Netherlands and Germany. A uniform increase in the (statutory) retirement age places a greater burden on individuals on lower rungs of the occupational ladder. Their life expectancy is lower, which means that an increase in the retirement age leads to a loss of a much larger share of the expected years in retirement. If an increase in the retirement age is intended to be distributionally neutral, it should be progressively increased for higher occupational classes. Educational attainment or lifetime earnings can be used as indicators in the absence of information on occupational history.

Second, our estimators for Germany suggest that 40 percent of the association between low job control and high physical demands on the one hand and health on the other cannot be explained by selection, and that the effect seems stronger at an older age. A uniform increase of the retirement age now places and additional burden on individuals in occupations with high levels of these stressors at ages at which vulnerability to these stressors is highest. This is an additional reason to shield individuals in these occupations from the harmful effects of these stressors. This can be achieved by limiting the increase of the retirement age for individuals with a long history in these occupations.

Third, given that we found the vulnerability to low job control and high physical demands to be increasing with age, retraining programs can help people to shift away from exposure to these stressors at an older age, allowing them to have long and healthy careers.

Fourth, delayed educational tracking in Finland has benefited children from poor families, but at the expense of children from rich families. In the absence of an aggregate average effect, this makes very explicit how redistributional policies create winners and losers. In contrast, I confirm that the UK compulsory schooling reforms in the UK did not have an effect on health. Early-tracking countries such the Netherlands or Germany can delay the tracking age to reduce socioeconomic disparities in schooling and mortality.

A general methodological message from all chapters is in line with the move in the economics discipline away from estimation methods that rely on demanding assumptions which may drive results. The recent improvements in data quality and development of estimation methods based on exogenous variation in the regressor of interest offer exciting new opportunities of causal inference in the social sciences as a whole, in areas that have previously not been exposed to these new techniques. This is not to say that the internal validity of the estimates presented here is beyond doubt. But the so-called credibility revolution in economics does offer a yardstick along which the credibility of the methods in this dissertation can be assessed.

The key policy message of this dissertation is that uniform, one-size-fits-all policies often increase disparities. These are particularly blunt instruments where precision is required. New econometric methods, ever-improving data availability, and better administrative capabilities allow us to customize policies to ensure that policies (the "treatment") better meets the need of individuals (the "patients").

Appendices

Appendix A

Derivations

A.1 Setup

We first list some general properties of our OLS estimators. As depicted in figure 4.1, health is a linear function of schooling s, parental SES p, and error term ε. Schooling is a linear combination of parental SES and error term η. For a given sample, we would like to estimate

$$\boldsymbol{h} = \beta_s \boldsymbol{s} + \beta_p \boldsymbol{p} + \boldsymbol{\varepsilon} \tag{A.1.1}$$

Where β_s and β_p are scalars and $\boldsymbol{h}$, $\boldsymbol{s}$, $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ are $m \times 1$ dimensional vectors, m referring to the number of observations. β_p is the direct effect of parental SES on health that is not transmitted by schooling. Schooling is a linear function of parental SES and a random variable:

$$\boldsymbol{s} = \gamma \boldsymbol{p} + \boldsymbol{\eta} \tag{A.1.2}$$

The OLS estimator is:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{h} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \end{aligned} \tag{A.1.3}$$

If we run a regression of health on both regressors, $\hat{\boldsymbol{\beta}} = \begin{pmatrix} \beta_s \\ \beta_p \end{pmatrix}$ and $\boldsymbol{X}$ is the $m \times 2$-dimensional matrix $[\boldsymbol{s}\ \boldsymbol{p}$. We can write equation A.1.3 as two separate equations in summation notation:

$$\hat{\beta}_s = \beta_s + \frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i^2) - (\sum_{i=1}^n p_i s_i)^2} - \frac{(\sum_{i=1}^n p_i s_i)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i^2) - (\sum_{i=1}^n p_i s_i)^2} \quad \text{(A.1.4)}$$

$$\hat{\beta}_p = \beta_p + \frac{(\sum_{i=1}^n s_i^2)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i^2) - (\sum_{i=1}^n p_i s_i)^2} - \frac{(\sum_{i=1}^n p_i s_i)(\sum_{i=1}^n s_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i^2) - (\sum_{i=1}^n p_i s_i)^2} \quad \text{(A.1.5)}$$

A univariate OLS regression of health on either schooling or parental SES leads to:

$$\hat{\beta}_s = \beta_s + \frac{\sum_{i=1}^n s_i(\varepsilon_i + \beta_p p_i)}{\sum_{i=1}^n s_i^2} \quad \text{(A.1.6)}$$

$$\hat{\beta}_p = \beta_p + \frac{\sum_{i=1}^n p_i(\varepsilon_i + \beta_s s_i)}{\sum_{i=1}^n p_i^2} \quad \text{(A.1.7)}$$

A.2 Derivation of equations 4.2.3 and 4.2.4: the regression model with both regressors

If we take expectations of equations A.1.4 and A.1.5 we get

$$plim\hat{\beta}_s = \beta_s + plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i^2) - (\sum_{i=1}^n p_i s_i)^2} - plim\frac{(\sum_{i=1}^n p_i s_i)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i^2) - (\sum_{i=1}^n p_i s_i)^2} \quad \text{(A.2.1)}$$

$$plim\hat{\beta}_p = \beta_p + plim\frac{(\sum_{i=1}^n s_i^2)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i^2) - (\sum_{i=1}^n p_i s_i)^2} - plim\frac{(\sum_{i=1}^n p_i s_i)(\sum_{i=1}^n s_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i^2) - (\sum_{i=1}^n p_i s_i)^2} \quad \text{(A.2.2)}$$

And if ε has zero mean and is uncorrelated with parental SES and schooling, the final two terms of both equations A.2.1 and A.2.2 are zero and this completes the derivation of equations 4.2.3 and 4.2.4.

A.3 Derivation of equations 4.2.5 and 4.2.6: univariate regression models

We derive the estimators of the univariate regression models, with health as the dependent variable and either parental SES or schooling.

We can take expectations of equation A.1.6

$$plim\hat{\beta}_s = \beta_s + \beta_p\gamma plim\frac{\sum_{i=1}^n p_i^2}{\sum_{i=1}^n (\gamma p_i + \eta_i)^2} + \beta_p plim\frac{\sum_{i=1}^n p_i\eta_i}{\sum_{i=1}^n (\gamma p_i + \eta_i)^2} + \gamma plim\frac{\sum_{i=1}^n p_i\varepsilon_i}{\sum_{i=1}^n (\gamma p_i + \eta_i)^2} + plim\frac{\sum_{i=1}^n \eta_i\varepsilon_i}{\sum_{i=1}^n (\gamma p_i + \eta_i)^2} \tag{A.3.1}$$

If schooling and the zero mean error term of equation A.1.1 are uncorrelated, and the same holds for parental SES and the zero mean error term of equation A.1.2, then $\beta_p plim\frac{\sum_{i=1}^n p_i\eta_i}{\sum_{i=1}^n (\gamma p_i+\eta_i)^2} = 0$, $\gamma plim\frac{\sum_{i=1}^n p_i\varepsilon_i}{\sum_{i=1}^n (\gamma p_i+\eta_i)^2} = 0$, and $plim\frac{\sum_{i=1}^n \eta_i\varepsilon_i}{\sum_{i=1}^n (\gamma p_i+\eta_i)^2} = 0$. which completes the derivation of equation 4.2.5.

We take expectations of equation A.1.7 to get the estimator of the coefficient of parental SES:

$$\hat{\beta}_p = \beta_p + plim\frac{\sum_{i=1}^n p_i\varepsilon_i}{\sum_{i=1}^n p_i^2} + \beta_s plim\frac{\sum_{i=1}^n p_i s_i}{\sum_{i=1}^n p_i^2} \tag{A.3.2}$$

The last part of this equation can be seen as the OLS estimator of γ in equation A.1.2 where η is randomly distributed for now. We can rewrite equation A.3.2 as

$$\begin{aligned} plim\hat{\beta}_p &= \beta_p + plim\frac{\sum_{i=1}^n p_i\varepsilon_i}{\sum_{i=1}^n p_i^2} + \beta_s plim\frac{\sum_{i=1}^n p_i(\gamma p_i + \eta_i)}{\sum_{i=1}^n p_i^2} \\ &= \beta_p + \beta_s\gamma + plim\frac{\sum_{i=1}^n p_i\varepsilon_i}{\sum_{i=1}^n p_i^2} + \beta_s plim\frac{\sum_{i=1}^n p_i\eta_i}{\sum_{i=1}^n p_i^2} \end{aligned} \tag{A.3.3}$$

If parental SES and the zero mean error terms of equations A.1.1 and A.1.2 are uncorrelated, $plim\frac{\sum_{i=1}^n p_i\varepsilon_i}{\sum_{i=1}^n p_i^2} = 0$ and $\beta_s plim\frac{\sum_{i=1}^n p_i\eta_i}{\sum_{i=1}^n p_i^2} = 0$, which completes the derivation of equations 4.2.6.

A.4 Derivation of equations 4.2.7 and 4.2.8: regression model with both regressors and omitted variables affecting the intervening variable

With respect to the effect of schooling on health and with the use of equation A.1.2, we can rewrite equation A.2.1 as:

$$\begin{aligned}
plim\hat{\beta}_s &= \beta_s + plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \\
&\quad - plim\frac{(\sum_{i=1}^n p_i(\gamma p_i+\eta_i)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \\
&= \beta_s + plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n \eta_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \\
&\quad - plim\frac{(\sum_{i=1}^n p_i\eta_i)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2}
\end{aligned} \tag{A.4.1}$$

Since $u1$ in figure 4.1 affects both schooling (through η in equation A.1.2) and health, η and ε are correlated. We rewrite equation A.4.1 as

$$plim\hat{\beta}_s = \beta_s + plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n \eta_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \tag{A.4.2}$$

The other term in equation A.4.1 drops out because schooling and η are uncorrelated. This completes the derivation of equation 4.2.7.

Again using equation A.1.2, we rewrite equation A.2.2 as

$$\begin{aligned}
plim\hat{\beta}_p &= \beta_p + plim\frac{(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \\
&\quad - plim\frac{(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))(\sum_{i=1}^n(\gamma p_i+\eta_i)\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \\
&= \beta_p + 2\gamma plim\frac{(\sum_{i=1}^n p_i\eta_i)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \\
&\quad + plim\frac{(\sum_{i=1}^n \eta_i^2)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \\
&\quad - \gamma plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n \eta_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \\
&\quad - \gamma plim\frac{(\sum_{i=1}^n p_i\eta_i)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2} \\
&\quad - plim\frac{(\sum_{i=1}^n p_i\eta_i)(\sum_{i=1}^n \eta_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n(\gamma p_i+\eta_i)^2)-(\sum_{i=1}^n p_i(\gamma p_i+\eta_i))^2}
\end{aligned} \tag{A.4.3}$$

As before, η and ε are correlated. We can rewrite equation A.4.3 as

$$plim\hat{\beta}_p = \beta_p - \gamma plim \frac{(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} \eta_i \varepsilon_i)}{(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} (\gamma p_i + \eta_i)^2) - (\sum_{i=1}^{n} p_i(\gamma p_i + \eta_i))^2} \tag{A.4.4}$$

The other terms of equation A.4.3 are equal to zero since ability is not correlated with ε and η. This completes the derivation of equation 4.2.8.

A.5 Derivation of equations 4.2.9 and 4.2.10: regression model with both regressors and measurement error

Schooling is measured with error such that

$$\tilde{s} = s + w \tag{A.5.1}$$

And s is randomly distributed with mean zero. For clarity of exposition let us briefly define $\nu = \varepsilon - \beta_s w$. Instead of A.2.1, the estimator of the coefficient of schooling is:

$$\begin{aligned} \hat{\beta}_s = \beta_s &+ \frac{(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} \tilde{s}_i \nu_i)}{(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} \tilde{s}_i{}^2) - (\sum_{i=1}^{n} p_i \tilde{s}_i)^2} \\ &- \frac{(\sum_{i=1}^{n} p_i \tilde{s}_i)(\sum_{i=1}^{n} p_i \nu_i)}{(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} \tilde{s}_i{}^2) - (\sum_{i=1}^{n} p_i \tilde{s}_i)^2} \end{aligned} \tag{A.5.2}$$

Take expectations and rewrite this as:

$$\begin{aligned}
\hat{\beta}_s = \beta_s &+ plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i+w_i)^2) - (\sum_{i=1}^n p_i(s_i+w_i))^2} \\
&- \beta_s plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n s_i w_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i+w_i)^2) - (\sum_{i=1}^n p_i(s_i+w_i))^2} \\
&+ plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n w_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i+w_i)^2) - (\sum_{i=1}^n p_i(s_i+w_i))^2} \\
&- \beta_s plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n w_i^2)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i+w_i)^2) - (\sum_{i=1}^n p_i(s_i+w_i))^2} \\
&- plim\frac{(\sum_{i=1}^n p_i s_i)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i+w_i)^2) - (\sum_{i=1}^n p_i(s_i+w_i))^2} \\
&+ \beta_s plim\frac{(\sum_{i=1}^n p_i s_i)(\sum_{i=1}^n p_i w_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i+w_i)^2) - (\sum_{i=1}^n p_i(s_i+w_i))^2} \\
&- plim\frac{(\sum_{i=1}^n p_i w_i)(\sum_{i=1}^n p_i\varepsilon_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i+w_i)^2) - (\sum_{i=1}^n p_i(s_i+w_i))^2} \\
&+ \beta_s plim\frac{(\sum_{i=1}^n p_i w_i)(\sum_{i=1}^n p_i w_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i+w_i)^2) - (\sum_{i=1}^n p_i(s_i+w_i))^2} \qquad \text{(A.5.3)}
\end{aligned}$$

All but one bias terms drop out since schooling, ε, and w are uncorrelated and parental SES is uncorrelated with w and ε. We get:

$$\hat{\beta}_s = \beta_s - \beta_s plim\frac{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n w_i^2)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n (s_i+w_i)^2) - (\sum_{i=1}^n p_i(s_i+w_i))^2} \qquad \text{(A.5.4)}$$

This completes the derivation of equation 4.2.9.

Instead of equation A.2.2 the estimator of the direct effect of parental SES is:

$$\begin{aligned}
\hat{\beta}_p = \beta_p &+ \frac{(\sum_{i=1}^n \tilde{s_i}^2)(\sum_{i=1}^n p_i\nu_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n \tilde{s_i}^2) - (\sum_{i=1}^n p_i\tilde{s_i})^2} \\
&- \frac{(\sum_{i=1}^n p_i\tilde{s_i})(\sum_{i=1}^n \tilde{s_i}\nu_i)}{(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n \tilde{s_i}^2) - (\sum_{i=1}^n p_i\tilde{s_i})^2} \qquad \text{(A.5.5)}
\end{aligned}$$

Using equation A.5.1, and taking expectations:

$$
\begin{aligned}
plim\hat{\beta}_p = \beta_p &+ plim\frac{(\sum_{i=1}^{n}(s_i+w_i)^2)\,(\sum_{i=1}^{n}p_i\varepsilon_i)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \\
&- \beta_s plim\frac{(\sum_{i=1}^{n}(s_i+w_i)^2)\,(\sum_{i=1}^{n}p_iw_i)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \\
&- plim\frac{(\sum_{i=1}^{n}p_is_i)(\sum_{i=1}^{n}s_i\varepsilon_i)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \\
&+ \beta_s plim\frac{(\sum_{i=1}^{n}p_is_i)(\sum_{i=1}^{n}s_iw_i)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \\
&- plim\frac{(\sum_{i=1}^{n}p_iw_i)(\sum_{i=1}^{n}s_i\varepsilon_i)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \\
&+ \beta_s plim\frac{(\sum_{i=1}^{n}p_iw_i)(\sum_{i=1}^{n}s_iw_i)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \\
&- plim\frac{(\sum_{i=1}^{n}p_is_i)(\sum_{i=1}^{n}w_i\varepsilon_i)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \\
&+ \beta_s plim\frac{(\sum_{i=1}^{n}p_is_i)(\sum_{i=1}^{n}w_i^2)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \\
&- plim\frac{(\sum_{i=1}^{n}p_iw_i)(\sum_{i=1}^{n}w_i\varepsilon_i)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \\
&+ \beta_s plim\frac{(\sum_{i=1}^{n}p_iw_i)(\sum_{i=1}^{n}w_i^2)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \qquad \text{(A.5.6)}
\end{aligned}
$$

All but one bias terms drop out since true schooling s, ε, and w are uncorrelated and parental SES is not correlated with w. We get:

$$
plim\hat{\beta}_p = \beta_p + \beta_s plim\frac{(\sum_{i=1}^{n}p_is_i)(\sum_{i=1}^{n}w_i^2)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \qquad \text{(A.5.7)}
$$

And multiplying the numerator and denominator of the bias term by p_i^2 yields:

$$
plim\hat{\beta}_p = \beta_p + \beta_s\gamma plim\frac{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}w_i^2)}{(\sum_{i=1}^{n}p_i^2)(\sum_{i=1}^{n}(s_i+w_i)^2)-(\sum_{i=1}^{n}p_i(s_i+w_i))^2} \qquad \text{(A.5.8)}
$$

This completes the derivation of equation 4.2.10.

A.6 Both the regressor of interest and the intervening control are influenced by omitted variables

This subsection in the appendix discusses the setting where both regressors are correlated with the error term. Now both $\mathbb{E}[\varepsilon, p \neq 0$ and $\mathbb{E}[\varepsilon, \eta \neq 0$ in equations 4.2.1 and 4.2.2. The expected values of the estimators of the univariate regressions are similar to

equations 4.2.5 and 4.2.6 but now include additional bias terms due to the correlation between parental SES and the error term in equation 4.2.1:

$$plim\hat{\beta}_s = \beta_s + \beta_p\gamma plim\frac{\sum_{i=1}^{n} p_i^2}{\sum_{i=1}^{n}(\gamma p_i + \eta)^2} + plim\frac{\sum_{i=1}^{n}\gamma p_i\varepsilon_i}{\sum_{i=1}^{n}(\gamma p_i + \eta)^2} + plim\frac{\sum_{i=1}^{n}\eta_i\varepsilon_i}{\sum_{i=1}^{n}(\gamma p_i + \eta)^2} \quad \text{(A.6.1)}$$

$$plim\hat{\beta}_p = \beta_p + \beta_s\gamma + plim\frac{\sum_{i=1}^{n} p_i\varepsilon_i}{\sum_{i=1}^{n} p_i^2} \quad \text{(A.6.2)}$$

The estimator of the effect of schooling is biased because of the omission of parental SES, because of the correlation between parental SES and the error term, and because of the correlation between η and the error term. It now yields the total association between schooling and health.

The estimator of the total effect of parental SES estimates the total association between schooling and health, with the final term reflecting the omitted variable bias due to the correlation between parental SES and the error term in equation 4.2.1.

The expected values of the estimators of the regression model with both variables are:

$$plim\hat{\beta}_s = \beta_s + plim\frac{(\sum_{i=1}^{n} s_i^2)(\sum_{i=1}^{n}\eta_i\varepsilon_i)}{(\sum_{i=1}^{n} s_i^2)(\sum_{i=1}^{n}(\gamma s_i + \eta_i)^2) - (\sum_{i=1}^{n} s_i(\gamma s_i + \eta_i))^2} \quad \text{(A.6.3)}$$

$$plim\hat{\beta}_p = \beta_p + plim\frac{(\sum_{i=1}^{n} \eta_i^2)(\sum_{i=1}^{n} s_i\varepsilon_i)}{(\sum_{i=1}^{n} s_i^2)(\sum_{i=1}^{n}(\gamma s_i + \eta_i)^2) - (\sum_{i=1}^{n} s_i(\gamma s_i + \eta_i))^2} - \gamma plim\frac{(\sum_{i=1}^{n} s_i^2)(\sum_{i=1}^{n}\eta_i\varepsilon_i)}{(\sum_{i=1}^{n} s_i^2)(\sum_{i=1}^{n}(\gamma s_i + \eta_i)^2) - (\sum_{i=1}^{n} s_i(\gamma s_i + \eta_i))^2} \quad \text{(A.6.4)}$$

In addition to equation 4.2.9, the expected value of the estimator of the direct effect of parental SES includes the omitted variable bias due to the correlation with the error term. Equation A.6.4 is identical to equation 4.2.10, which shows that the endogeneity of the preceding variable parental SES does not have consequences for the estimator of the effect of schooling

Appendix B

Simulations

B.1 Setup

Consider a data generating process (DGP) as described table B.1. Similar to figure 4.1, health is a function of parental SES, schooling and two unobserved factors. Schooling is a function of parental SES and a randomly distributed term. Note that parental SES and schooling are uncorrelated with ε in equation 4.2.1.

Panel B of table B.2 shows regression results under DGP 1 using a simulated sample of 100,000 observations. A univariate regression of health on parental SES in column (1) yields valid estimation results of the total effect of parfental SES on health. However, the second line shows that a regression of health on schooling yields a biased estimator of the effect of schooling since parental SES is omitted from the regression model. The estimation results of the regression model in the third line are close to the effect of schooling and the direct effect of parental SES. The estimated ratio of the indirect effect of parental SES (that is transmitted by schooling) to the total effect is computed by $1 -$ Direct effect/Total effect. The fourth row of panel B shows that the estimate of this ratio is close to the true value. Hence, we may add a bad (intervening) control variable which is uncorrelated with the error term if we are interested in the direct effect of the regressor of interest that is not transmitted by this control variable. However, in practice intervening control variables are rarely exogenous.

Table B.1: —Data generating process

	DGP 1
h	$\beta_p p + \beta_s s + 5 * u1 + 5 * u2$
p	p^*
s	$\gamma p + s^*$
β_p	1
γ	.5
β_s	.5
s^*	$\sim N(0,1)$
p^*	$\sim U[0,10)$
$u1$	$\sim N(0,1)$
$u2$	$\sim N(0,1)$

Health is a linear function of parental SES, schooling and two unobserved factors. Schooling is a linear function of parental SES and an individual-specific effect, which is drawn from the standard normal distribution. Parental SES is drawn from a uniform distribution on the interval between 0 and 10.

B.2 Omitted variable bias and intervening controls

To illustrate the bad controls problem, suppose that DGP 2 and 3 are identical to DGP 1 in table B.1, with the exception that:

$$s = \gamma p + s^* + u1 \quad \text{(DGP 2)}$$

$$s = \gamma p + s^* - u1 \quad \text{(DGP 3)}$$

Schooling is now a linear function of parental SES, s^*, and $u1$ which also affects health. In terms of equations 4.2.1 and 4.2.2, ε and η are now correlated. Panel C of table B.2 shows estimation results under DGP 2. The first line shows that the estimate of the total effect of parental SES is still close to the true value. The second line shows that the estimate of the effect of schooling is larger than the true value and larger than the result in the second line of panel B due to the omission of both parental SES and $u1$ in the regression model. The third line shows the estimate of the direct effect of parental SES which—conform equation 4.2.10—differs from the true value by$-\gamma$ times the bias of the estimator of the effect of the bad control, or $-.5 \times 2.5 = -1.25$. Consequently, a careless interpretation of these results could lead to the erroneous conclusion that the direct effect of parental SES is negative, while in fact it is positive. The fourth line would lead us to conclude that the indirect effect of parental SES on health (estimated

True parameters	(1)	(2)	(3)	(4)
Panel A	$\widehat{\beta_p + \gamma\beta_s}$	$\hat{\beta}_s$	$\hat{\beta}_p$	1- $\frac{\hat{\beta}_p}{\widehat{\beta_p+\gamma\beta_s}}$
True parameters	1.25	.50	1.00	.20
Panel B				
Univariate regression under **DGP 1**	**1.25 (.01)**			
Univariate regression under **DGP 1**		1.85 (.01)		
Regression with both regressors under **DGP 1**		**.51 (.02)**	**.99 (.01)**	
Indirect effect as a proportion of total effect				**.20**
Panel C				
Univariate regression under **DGP 2**	**1.26 (.01)**			
Univariate regression under **DGP 2**		2.76 (.01)		
Regression with both regressors under **DGP 2**		3.02 (.01)	-.25 (.01)	
Indirect effect as a proportion of total effect				1.20
Panel D				
Univariate regression under **DGP 3**	**1.26 (.01)**			
Univariate regression under **DGP 3**		.30 (.01)		
Regression with both regressors under **DGP 3**		- 2.00 (.01)	2.26 (.01)	
Indirect effect as a proportion of total effect				-.79
Panel E				
Univariate regression under **DGP 1 and measurement error**	1.26 (.01)			
Univariate regression under **DGP 1 and measurement error**		1.41 (.01)		
Regression with both regressors under **DGP 1 and measurement error**		.27 (.02)	1.12 (.01)	
Indirect effect as a proportion of total effect				.11

Table B.2: —Correlation between the intervening control variable and the error term

Panel A shows the true parameters of the data generating process. Panel B shows estimation results if the regressors are not correlated with $u1$ and $u2$. In panel C, schooling is positively correlated with $z1$. In panel D, schooling is negatively correlated with $z1$. Unbiased estimators in bold. Each analysis is on the basis of 100,000 simulated observations. Panel E shows estimation results if the regressors are not correlated with $u1$ and $u2$ but if schooling is measured with standard normally distributed measurement error.

as $1 - \hat{\beta}_p/(\widehat{\beta_p + \gamma\beta_s})$) is larger than the total effect, while the opposite is true.

Panel D of table B.2 shows the results under DGP 3. Due to the omission of $u1$ from the model, the estimate in the second line is now smaller than the true causal effect in panel A and much smaller than the unconditional association in the second line of panel B. The estimate of the direct effect of parental SES in the third line of panel C is much larger than the true effect—the difference is approximately 1.25—and the estimate of the effect of schooling is strongly negative, while the true value is .50. Both panels C and D illustrate the sensitivity of the estimator of the direct effect of schooling to correlation between the intervening control variable and the error term. While we would expect the indirect effect as a proportion of the total effect to be bounded between 0 and 1, the estimate in the fourth line is now negative and we could mistakenly conclude that the indirect effect is negative.

B.3 Measurement error and intervening controls

Consider DGP 1 in table B.1. To illustrate the problem of intervening control variables that are measured with error w, schooling is now $\tilde{s} = s + w$ with $w \sim N(0, 1)$. The first row of panel E in table B.2 shows a correct estimate of the total effect of parental SES. The univariate regression model in the second row is $h = \beta_s \tilde{s} + \mu$ where $\mu = \varepsilon + \beta_p p - \beta_s w$.

The estimate of the total association between schooling and health is biased towards 1.35, which is the unconditional association between schooling and health (the total association in the second row of panel B) minus the true causal effect of schooling in panel A, for which estimation is attenuated towards zero by the measurement error.

The estimate of the direct effect of parental SES in the third row is now biased towards the total effect, and the estimate of the effect of schooling is biased towards zero. The bias of the estimator of the direct effect of parental SES is equal to $-\gamma$ (-.50) times the bias of the estimator of the direct effect of schooling. The estimate of the indirect effect as a proportion of the total effect in the fourth line is biased towards 1.

B.4 Omitted variables affecting both regressors

DGP 4 and 5 are the same as DGP 1 in table B.1, with the exception that:

$$\begin{aligned} p &= p^* + u2 \\ s &= \gamma p + s^* + u1 \end{aligned} \qquad \text{(DGP 4)}$$

$$\begin{aligned} p &= p^* + u2 \\ s &= \gamma p + s^* - u1 \end{aligned} \qquad \text{(DGP 5)}$$

$u1$ has a positive effect on schooling in DGP 4 and a negative effect on schooling in DGP 5, similar to DGP 2 and DGP 3, respectively. In addition, $u2$ has a positive effect on parental SES in both DGP's. The estimator of the total effect in the first row of panel B in table B.3 is biased by .55 due to the omission of preceding $u2$ from the regression model. The bias of the second row of panel B has three components: (i) bias due to the omission of parental SES (which is a preceding control if we are interested in the coefficient of schooling) , (ii) bias due to the omission of $u2$ while parental SES is missing, and (iii) bias due to the omission of $u1$.

In the third row, the bias due to (i) and (ii) is eliminated by the inclusion of parental SES and we get the same point estimate as in panel C of table B.2. This illustrates that it harmless to include a preceding control variable even when it is correlated with the error term. In fact, the preceding variable proxies for other omitted variables, eliminating more than its "own" share of omitted variable bias.

Remember that the indirect effect through schooling is .25 while the omitted variable bias due to $u2$ is .55. This means that the direct effect plus the $u2$-related omitted variable bias is 1.55. However, as in panel C of table B.2, the bad controls problem leads to a bias of -1.25. The same is true in panel D of table B.3 although consistent with DGP 5 the bias term is of the opposite sign.

We have seen that a change in the estimate of the coefficient of the regression of interest after the inclusion of a preceding control can be interpreted as the part of the association that was explained by selection based on and proxied by the preceding control variable. In contrast, inclusion of bad control variables does not explain anything about the underlying relationships, since any change in the coefficient of the regressor of interest may be due to the spurious nature of the bad control.

True parameters	(1)	(2)	(3)	(4)
Panel A	$\widehat{\beta_p + \gamma\beta_s}$	$\hat{\beta}_s$	$\hat{\beta}_p$	1- $\frac{\hat{\beta_p}}{\widehat{\beta_p+\gamma\beta_s}}$
True parameters	1.25	.50	1.00	.20
Panel B				
Univariate regression under **DGP 4**	1.80			
	(.01)			
Univariate regression under **DGP 4**		3.32		
		(.01)		
Regression with both regressors under		3.02	.29	
DGP 4		(.01)	(.01)	
Indirect effect as a proportion of total effect				.84
Panel C				
Univariate regression under **DGP 5**	1.80			
	(.01)			
Univariate regression under **DGP 5**		1.02		
		(.01)		
Regression with both regressors under		-	2.79	
DGP 5		2.00	(.01)	
		(.01)		
Indirect effect as a proportion of total effect				-.55

Table B.3: —Correlation between the regressors and the error term

Panel A shows the true parameters of the data generating process. Unbiased estimators in bold. Each analysis is on the basis of 100,000 simulated observations. In panel B, parental SES is positively correlated with $u2$ and schooling is positively correlated with $u1$. In panel C, parental SES is positively correlated with $u2$ and schooling is negatively correlated with $u1$.

Bibliography

Abraído-Lanza, A., M. Chao, and K. Florez (2005). Do healthy behaviors decline with greater acculturation?: Implications for the latino mortality paradox. *Social science & medicine*.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy 113*(1), 151–184.

Alwin, D. and R. Hauser (1975). The decomposition of effects in path analysis. *American Sociological Review*, 37–47.

Anderson, M. and M. Marmot (2012). The effects of promotions on heart disease: Evidence from whitehall*. *The Economic Journal 122*(561), 555–589.

Angrist, J. D. (1990). Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records. *The American Economic Review*, 313–336.

Angrist, J. D. and A. B. Krueger (1992). Estimating the payoff to schooling using the vietnam-era draft lottery. Technical report, National Bureau of Economic Research.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Angrist, J. D. and J.-S. Pischke (2009, December). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29 – 51.

Bakija, J. (2013). A non-technical introduction to regression. `http://web.williams.edu/Economics/wp/Bakija-Non-Technical-Introduction-to-Regression.pdf`. Accessed: December 6, 2015.

Bakker, A. B. and E. Demerouti (2007). The job demands-resources model: State of the art. *Journal of managerial psychology 22*(3), 309–328.

Baron, R. and D. Kenny (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology 51*(6), 1173.

Becker, G. S. and K. M. Murphy (1988). A theory of rational addiction. *The journal of political economy*, 675–700.

Bernard, B. P., V. Putz-Anderson, et al. (1997). *Musculoskeletal disorders and workplace factors: a critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back*. Number 97-141. US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health Cincinnati, OH, USA.

Bertrand, M., E. Duflo, S. Mullainathan, et al. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115 – 143.

Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal 1*(2), 141–162.

Bongers, P. M., C. T. J. Hulshof, L. Dijkstra, H. C. Boshuizen, H. J. M. Groenhout, and E. Valken (1990). Back pain and exposure to whole body vibration in helicopter pilots. *Ergonomics 33*(8), 1007–1026.

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 & medicine 51*(7), 1019–1030.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

Bosma, H., R. Peter, J. Siegrist, and M. Marmot (1998). Two alternative job stress models and the risk of coronary heart disease. *American journal of public health 88*(1), 68–74.

Brunello, G. and D. Checchi (2007). Does school tracking affect equality of opportunity? new international evidence. *Economic Policy 22*(52), 782–861.

Burke, M. (2012). Bad control. `http://www.g-feed.com/2012/10/bad-control.html`. Accessed: December 6, 2015.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan (2014). Early childhood investments substantially boost adult health. *Science 343*(6178), 1478–1485.

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

Case, A. and A. S. Deaton (2005b, July). Broken down by work and sex: How our health declines. In *Analyses in the Economics of Aging*, NBER Chapters, pp. 185–212. National Bureau of Economic Research, Inc.

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., D. Lubotsky, and C. Paxson (2002). Economic status and health in childhood: The origins of the gradient. *The American Economic Review 92*(5), 1308–1334.

Case, A. and C. Paxson (2011). The long reach of childhood health and circumstance: Evidence from the whitehall ii study*. *The Economic Journal 121*(554), F183–F204.

Cavelaars, A. E., A. E. Kunst, J. J. Geurts, U. Helmert, O. Lundberg, A. Mielck, J. Matheson, A. Mizrahi, A. Mizrahi, N. Rasmussen, et al. (1998). Morbidity differences by occupational class among men in seven european countries: an application of the erikson-goldthorpe social class scheme. *International Journal of Epidemiology 27*(2), 222–230.

Cawley, J. and C. J. Ruhm (2011). The economics of risky health behaviors1. *Handbook of health economics 2*, 95.

Choo, E. and M. Denny (2006a, October). Wearing out – the decline in health. Working Paper tecipa-258, University of Toronto, Department of Economics.

Choo, E. and M. Denny (2006b). Wearing out–the decline in health. Technical report, University of Toronto, Department of Economics.

Clark, D. and H. Royer (2013). The effect of education on adult mortality and health: Evidence from britain. *The American Economic Review 103*(6), 2087–2120.

Conti, G., J. Heckman, and S. Urzua (2010). The education-health gradient. *The American economic review 100*(2), 234.

Cottini, E. and P. Ghinetti (2012). Working conditions, lifestyles and health. Technical report, School of Economics and Management, University of Aarhus.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education 1*, 697–812.

Currie, A., M. Shields, and S. Price (2007). The child health/family income gradient: Evidence from england. *Journal of health economics 26*(2), 213–232.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature 47*(1), 87–122.

Currie, J. (2011). Inequality at birth: Some causes and consequences. *The American Economic Review 101*(3), pp. 1–22.

Currie, J. and B. C. Madrian (1999). Health, health insurance and the labor market. *Handbook of labor economics 3*, 3309–3416.

Cutler, D., A. Deaton, and A. Lleras-Muney (2006). The determinants of mortality. *The Journal of Economic Perspectives 20*(3), 97–120.

Cutler, D. and E. Glaeser (2005). What explains differences in smoking, drinking, and other health-related behaviors? *The American Economic Review 95*(2), 238–242.

Cutler, D. and A. Lleras-Muney (2010). Understanding differences in health behaviors by education. *Journal of health economics 29*(1), 1–28.

Cutler, D. M., A. Lleras-Muney, and T. Vogl. Socioeconomic status and health: Dimensions and mechanisms.

Cutler, D. M., A. Lleras-Muney, and T. Vogl (2008, September). Socioeconomic status and health: Dimensions and mechanisms. Working Paper 14333, National Bureau of Economic Research.

Deaton, A. and C. Paxson (2001). Mortality, education, income, and inequality among american cohorts. In *Themes in the Economics of Aging*, pp. 129–170. University of Chicago Press.

Del Valle, A. (2013). The economics of happiness: An applied approach. `https://www.ocf.berkeley.edu/~adv/wp-content/uploads/2014/11/Happinessintro.pdf`. Accessed: December 6, 2015.

Devereux, P. J. and R. A. Hart (2010). Forced to be rich? returns to compulsory schooling in britain*. *The Economic Journal 120*(549), 1345–1364.

Dolan, P., D. Fujiwara, and R. Metcalfe (2013). The impact of family composition on human capital formation: Methodology and evidence. `http://www3.grips.ac.jp/~econseminar/20130130CCLL2013Jan25.pdf`. Accessed: December 6, 2015.

Donald, S. G. and K. Lang (2007). Inference with difference-in-differences and other panel data. *The review of Economics and Statistics 89*(2), 221–233.

Duflo, E., P. Dupas, and M. Kremera (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *The American Economic Review 101*(5), 1739–1774.

Duncan, O. (1966). Path analysis: Sociological examples. *American journal of Sociology*, 1–16.

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers 56*(2), 209–230.

Ferrer-i Carbonell, A. and P. Frijters (2004). How important is methodology for the estimates of the determinants of happiness?*. *The Economic Journal 114*(497), 641–659.

Fletcher, J. M. (2012). The effects of first occupation on long term health status: Evidence from the wisconsin longitudinal study. *Journal of Labor Research 33*(1), 49–75.

Fletcher, J. M. and J. L. Sindelar (2009). Estimating causal effects of early occupational choice on later health: Evidence using the psid. Working Paper 15256, National Bureau of Economic Research.

Fletcher, J. M., J. L. Sindelar, and S. Yamaguchi (2011a). Cumulative effects of job characteristics on health. *Health economics 20*(5), 553–570.

Fletcher, J. M., J. L. Sindelar, and S. Yamaguchi (2011b). Cumulative effects of job characteristics on health. *Health Economics 20*(5), 553–570.

Fleurbaey, M. and E. Schokkaert (2009). Unfair inequalities in health and health care. *Journal of health economics 28*(1), 73–90.

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.

Frölich, M. and M. Huber (2015). Direct and indirect treatment effects: causal chains and mediation analysis with instrumental variables. Technical report, IFAU-Institute for Evaluation of Labour Market and Education Policy.

Galama, T. and H. van Kippersluis (2010). A theory of socioeconomic disparities in health over the life cycle. Working Paper 2010-079/3, Tinbergen Institute.

Galama, T. and H. van Kippersluis (2013). Health inequalities through the lens of health capital theory: Issues, solutions, and future directions. *Research on economic inequality 21*, 263–284.

Galama, T. J. and H. Van Kippersluis (2015). A theory of socioeconomic disparities in health over the life cycle. *CESR-Schaeffer Working Paper* (2015-016).

Goldman, D. and J. Smith (2002). Can patient self-management help explain the ses health gradient? *Proceedings of the National Academy of Sciences 99*(16), 10929–10934.

Goodman, E. (1999). The role of socioeconomic status gradients in explaining differences in us adolescents' health. *American Journal of Public Health 89*(10), 1522–1528.

Grossman, M. (1972a). On the concept of health capital and the demand for health. *The journal of political economy*, 223–255.

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

Gupta, N. D. and N. Kristensen (2008). Work environment satisfaction and employee health: panel evidence from denmark, france and spain, 1994–2001. *The European Journal of Health Economics 9*(1), 51–61.

Hall, A., U. Braun, H. Herget, E. M. Krekel, I. Leppelmeier, H.-J. Schade, K. Troltsch, and J.-G. Ulrich (2010). Die bibb/baua-erwerbstätigenbefragung 2006. *Arbeit und Beruf im Wandel, Erwerb und Verwertung beruflicher Qualifikationen–Abschlussbericht (Bundesinstitut für Berufsbildung) Bonn*.

Hanushek, E. A. et al. (2006). Does educational tracking affect performance and inequality? differences-in-differences evidence across countries*. *The Economic Journal 116*(510), C63–C76.

Harmon, C. and I. Walker (1995). Estimates of the economic return to schooling for the united kingdom. *The American Economic Review*, 1278–1286.

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*(1), pp. 21–37.

Imbens, G. and J. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica: Journal of the Econometric Society*, 467–475.

Jakovljević, D., C. Sarti, J. Sivenius, J. Torppa, M. Mähönen, P. Immonen-Räihä, E. Kaarsalo, K. Alhainen, K. Kuulasmaa, J. Tuomilehto, et al. (2001). Socioeconomic status and ischemic stroke the finmonica stroke register. *Stroke 32*(7), 1492–1498.

Jones, A. M., J. E. Roemer, and P. R. Dias (2014). Equalising opportunities in health through educational policy. *Social Choice and Welfare 43*(3), 521–545.

Karasek, Robert A., J. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly 24*(2), pp. 285–308.

Karasek Jr, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative science quarterly*, 285–308.

Kelly, I. R., D. M. Dave, J. L. Sindelar, and W. T. Gallo (2012). The impact of early occupational choice on health behaviors. *Review of Economics of the Household 10*, 1–34.

Kelly, I. R., D. M. Dave, J. L. Sindelar, and W. T. Gallo (2014). The impact of early occupational choice on health behaviors. *Review of Economics of the Household 12*(4), 737–770.

Kenkel, D. (1991). Health behavior, health knowledge, and schooling. *Journal of Political Economy*, 287–305.

Kerr, S. P., T. Pekkarinen, and R. Uusitalo (2013). School tracking and development of cognitive skills. *Journal of Labor Economics 31*(3), 577–602.

Kunst, A. E., F. Groenhof, O. Andersen, J. K. Borgan, G. Costa, G. Desplanques, H. Filakti, M. d. R. Giraldes, F. Faggiano, S. Harding, et al. (1999). Occupational class and ischemic heart disease mortality in the united states and 11 european countries. *American Journal of Public Health 89*(1), 47–53.

Kunst, A. E., C. Looman, and J. P. Mackenbach (1990). Socio-economic mortality differences in the netherlands in 1950–1984: a regional study of cause-specific mortality. *Social science & medicine 31*(2), 141–152.

Lantz, P., J. House, J. Lepkowski, D. Williams, R. Mero, and J. Chen (1998). Socioeconomic factors, health behaviors, and mortality. *JAMA: the journal of the American Medical Association 279*(21), 1703–1708.

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends (R) in Econometrics 4*(3), 165–224.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics 30*(1), 67–80.

Lindeboom, M. and E. van Doorslaer (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics 23*(6), 1083 – 1099.

Lucifora, C. and E. Cottini (2013). Mental health and working conditions in european countries. *Industrial & Labor Relations Review* (4), 958–988.

Mackenbach, J., J. Simon, C. Looman, and I. Joung (2002). Self-assessed health and mortality: could psychosocial factors explain the association? *International Journal of Epidemiology 31*(6), 1162–1168.

Mackenbach, J. P., V. Bos, O. Andersen, M. Cardano, G. Costa, S. Harding, A. Reid, Ö. Hemström, T. Valkonen, and A. E. Kunst (2003). Widening socioeconomic inequalities in mortality in six western european countries. *International Journal of Epidemiology 32*(5), 830–837.

Mackenbach, J. P., I. Stirbu, A.-J. R. 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.

MacKinnon, D., A. Fairchild, and M. Fritz (2007). Mediation analysis. *Annual review of psychology 58*, 593.

MacKinnon, D., C. Lockwood, J. Hoffman, S. West, and V. Sheets (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological methods 7*(1), 83.

Malamud, O. and C. Pop-Eleches (2011). School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics 95*(11), 1538–1549.

Marmot, M. (1994). Social differentials in health within and between populations. *Daedalus*, 197–216.

Marmot, M. G., H. Bosma, H. Hemingway, E. Brunner, and S. Stansfeld (1997). Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *The Lancet 350*(9073), 235–239.

Marmot, M. G., M. J. Shipley, and G. Rose (1984). Inequalities in death—specific explanations of a general pattern? *The Lancet 323*(8384), 1003–1006.

Marmot, M. G., S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner, A. Feeney, and G. D. Smith (1991a). Health inequalities among british civil servants: the whitehall ii study. *The Lancet 337*(8754), 1387–1393.

Marmot, M. G., S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner, A. Feeney, and G. D. Smith (1991b). Health inequalities among british civil servants: the whitehall ii study. *The Lancet 337*(8754), 1387 – 1393.

Martikainen, P., P. Mäkelä, R. Peltonen, and M. Myrskylä (2014). Income differences in life expectancy: the changing contribution of harmful consumption of alcohol and smoking. *Epidemiology 25*(2), 182–190.

Meara, E., S. Richards, and D. Cutler (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs 27*(2), 350–360.

Meghir, C., M. Palme, and E. Simeonova (2013). Education, cognition and health: Evidence from a social experiment. Technical report, National Bureau of Economic Research.

Michaud, P.-C. and A. Van Soest (2008). Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of Health Economics 27*(5), 1312–1325.

Morefield, B., D. C. Ribar, and C. J. Ruhm (2012a). Occupational status and health transitions. *The BE Journal of Economic Analysis & Policy 11*(3).

Morefield, B., D. C. Ribar, and C. J. Ruhm (2012b). Occupational status and health transitions. *The BE Journal of Economic Analysis & Policy 11*(3), 1–29.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society 49*(6), 1417–1426.

Nihtilä, E. K., P. T. Martikainen, S. V. Koskinen, A. R. Reunanen, A. M. Noro, and U. T. Häkkinen (2008). Chronic conditions and the risk of long-term institutionalization among older people. *The European Journal of Public Health 18*(1), 77–84.

Norton, E. C., H. Wang, and C. Ai (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal 4*, 154–167.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician 46*(1), pp. 27–29.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *The American Economic Review*, 152–175.

Pekkarinen, T. (2008). Gender differences in educational attainment: Evidence on the role of tracking from a finnish quasi-experiment. *The Scandinavian Journal of Economics 110*(4), 807–825.

Pekkarinen, T., R. Uusitalo, and S. Kerr (2009). School tracking and intergenerational income mobility: Evidence from the finnish comprehensive school reform. *Journal of Public Economics 93*(7), 965–973.

Puhani, P. A. (2012). The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models. *Economics Letters 115*(1), 85–87.

Ravesteijn, B. (2015). When a bad control variable turns good: Is the effect of parental socioeconomic status on health transmitted by schooling? Mimeo.

Ravesteijn, B., H. van Kippersluis, and E. van Doorslaer (2013). The contribution of occupation to health inequality. In *Research on economic inequality*, Volume 21, pp. 311–332. Emerald Group Publishing Limited.

Ravesteijn, B., H. Van Kippersluis, and E. Van Doorslaer (2013). Long and healthy careers? the relationship between occupation and health and its implications for the statutory retirement age. *Netspar Panel Paper*.

Ravesteijn, B., H. Van Kippersluis, and E. Van Doorslaer (2014). The wear and tear on health: What is the role of occupation? SOEPpaper 15256.

Roemer, J. E. (2009). *Equality of opportunity*. Harvard University Press.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? *Handbook of the Economics of Education 3*, 249–277.

Scholte, R. S., G. J. van den Berg, and M. Lindeboom (2015). Long-run effects of gestation during the dutch hunger winter famine on labor market and hospitalization outcomes. *Journal of health economics 39*, 17–30.

Schuster, M., M. Elliott, D. Kanouse, J. Wallander, S. Tortolero, J. Ratner, D. Klein, P. Cuccaro, S. Davies, and S. Banspach (2012). Racial and ethnic health disparities among fifth-graders in three cities. *New England Journal of Medicine 367*(8), 735–745.

Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal of occupational health psychology 1*(1), 27.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations.*

Smith, G. D., C. Hart, D. Hole, P. MacKinnon, C. Gillis, G. Watt, D. Blane, and V. Hawthorne (1998). Education and occupational social class: which is the more important indicator of mortality risk? *Journal of epidemiology and community health 52*(3), 153–160.

Smith, R. S. (1974). The feasibility of an" injury tax" approach to occupational safety. *Law and Contemporary Problems 38*(4), 730–744.

Stock, J. H. and M. W. Watson (2010). *Introduction to econometrics* (Third ed.). Prentice Hall.

Talala, K., T. Huurre, H. Aro, T. Martelin, and R. Prättälä (2009). Trends in socio-economic differences in self-reported depression during the years 1979–2002 in finland. *Social psychiatry and psychiatric epidemiology 44*(10), 871–879.

Valkonen, T., P. Martikainen, M. Jalovaara, S. Koskinen, T. Martelin, and P. Mäkelä (2000). Changes in socioeconomic inequalities in mortality during an economic boom and recession among middle-aged men and women in finland. *The European Journal of Public Health 10*(4), 274–280.

Van den Berg, G. J., M. Lindeboom, and F. Portrait (2006). Economic conditions early in life and individual mortality. *The American Economic Review*, 290–302.

Viscusi, W. K. (1978a). Labor market valuations of life and limb: Empirical evidence and policy implications. *Public Policy 26*(3), 359–386.

Viscusi, W. K. (1978b). Wealth effects and earnings premiums for job hazards. *The review of economics and statistics*, 408–416.

Viscusi, W. K. (1979). *Employment hazards: An investigation of market performance.* Number 148. Harvard University Press.

Wooldridge, J. (2012). *Introductory econometrics: A modern approach* (Fifth ed.). Cengage Learning.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*.

Acknowledgment

I want to thank Eddy van Doorslaer, who has provided inspiring and encouraging supervision, offered insight, oversight and structure, and who is responsible for creating the collaborative and stimulating health economics group at Erasmus University Rotterdam.

I also want to thank Hans van Kippersluis, who was always willing to discuss the best way to approach an apparent obstacle, the deeper implications of certain econometric techniques, how to be an effective teacher, and how to communicate ideas clearly. My development as a researcher has benefited greatly from working with Eddy, Hans, and with my dear colleagues at Erasmus University and the Tinbergen Institute. For that I am grateful.

In my private life, my parents have always been supportive of my choices and taught me to withstand conformity now and then. My grandmother on my mother's side relentlessly encouraged me to learn and grow intellectually, and my grandparents on my father's side taught me about the importance of equality.

I have always felt the support of my brother Jorik and his girlfriend Milou, of my fellow PhD students Paul, Taylan, Martin, and Kim, and of my other friends, even in times when graduate school absorbed a large part of my attention.

And of course I want to thank Rosa for her love, support, and companionship.

CV

Bastian Ravesteijn

ravesteijn@hcp.med.harvard.edu
sites.google.com/site/bastianravesteijn
Harvard Medical School
Department of Health Care Policy
180 Longwood Avenue, Room 330-D
Boston, MA 02115
(617) 432 3138

EDUCATION

Ph.D. in Economics, Erasmus University Rotterdam (the Netherlands), 2011-2016 (expected award date May 19, 2016).

M.Phil. in Economics, Tinbergen Institute (the Netherlands), 2009-2011.

B.A. in Political Science, University of Amsterdam (the Netherlands), 2001-2005.

RESEARCH AND TEACHING FIELDS

Primary fields: Health Economics, Labor Economics.

Secondary field: Microeconometrics.

TEACHING EXPERIENCE

2012-15	Research Project, Instructor.
2011-15	Impact Evaluation, Teaching Assistant for H. Van Kippersluis.
2011-12	Research Project, Teaching Assistant for A. Baillon and H. Van Kippersluis.
2008-09	Academic Skills, Teaching Assistant.

RELEVANT POSITIONS

2015-	Commonwealth Fund Harkness Research Fellow at Harvard University.
2011-15	PhD Candidate at Erasmus University Rotterdam, defense expected 2015.
2010-11	Research Assistant at Hugo Sinzheimer Institute for Labor Law, University of Amsterdam.
2005-07	Policy Advisor at the Municipality of Amsterdam.
2004-05	Public Relations and Information Officer at the Municipality of Amsterdam.

GRANTS

2015	Ministry of Social Affairs research grant.
2011	Netspar Panel Paper Grant (with E. Van Doorslaer and H. Van Kippersluis).
2011	Netspar Research Paper Grant (with E. Van Doorslaer and H. Van Kippersluis).

RESEARCH

The Contribution of Occupation to Health Inequality (with E. Van Doorslaer and H. Van Kippersluis). Research on economic inequality 21 (2013): 311-332.

The Wear and Tear on Health: What Is the Role of Occupation? (with E. Van Doorslaer and H. Van Kippersluis). Submitted.

When a Bad Control Variable Turns Good: Is the Effect of Parental Socioeconomic Status on Health Transmitted by Education?

The Effect of Delayed Educational Tracking on Mortality: Evidence from the Finnish Comprehensive Schooling Reform (with M. Avendano, P. Martikainen, E. Van Doorslaer, and H. Van Kippersluis).

The Social Returns to Mental Health Care (with E. Schachar).

The Medical and Social Returns to Early Intervention for First Episode Psychosis (with T. McGuire and R. Douven).

SEMINARS AND CONFERENCE PRESENTATIONS

2016	University of Chicago.
2015	Netspar Breukelen, SOLE-EALE Montreal, Ministry of Health The Hague, Harvard University, Massachusetts General Hospital.
2014	Ministry of Social Affairs The Hague, RES Manchester, SOLE Washington, ASHE Los Angeles, USC Los Angeles, Applied Micro Workshop Odense.
2013	PhD Seminar on Health Economics Grindelwald, International Pension Workshop Frankfurt, IHEA Sydney, ES Australasia Meeting Sydney, EEA Gothenborg, Tinbergen Institute Amsterdam.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University 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:

606 D. KARSTANJE, *Unraveling Dimensions: Commodity Futures Curves and Equity Liquidity*

607 T.C.A.P. GOSENS, *The Value of Recreational Areas in Urban Regions*

608 Ł.M. MARĆ, *The Impact of Aid on Total Government Expenditures*

609 C. LI, *Hitchhiking on the Road of Decision Making under Uncertainty*

610 L. ROSENDAHL HUBER, *Entrepreneurship, Teams and Sustainability: a Series of Field Experiments*

611 X. YANG, *Essays on High Frequency Financial Econometrics*

612 A.H. VAN DER WEIJDE, *The Industrial Organization of Transport Markets: Modeling pricing, Investment and Regulation in Rail and Road Networks*

613 H.E. SILVA MONTALVA, *Airport Pricing Policies: Airline Conduct, Price Discrimination, Dynamic Congestion and Network Effects.*

614 C. DIETZ, *Hierarchies, Communication and Restricted Cooperation in Cooperative Games*

615 M.A. ZOICAN, *Financial System Architecture and Intermediation Quality*

616 G. ZHU, *Three Essays in Empirical Corporate Finance*

617 M. PLEUS, *Implementations of Tests on the Exogeneity of Selected Variables and their Performance in Practice*

618 B. VAN LEEUWEN, *Cooperation, Networks and Emotions: Three Essays in Behavioral Economics*

619 A.G. KOPÁNYI-PEUKER, *Endogeneity Matters: Essays on Cooperation and Coordination*

620 X. WANG, *Time Varying Risk Premium and Limited Participation in Financial Markets*

621 L.A. GORNICKA, *Regulating Financial Markets: Costs and Trade-offs*

622 A. KAMM, *Political Actors playing games: Theory and Experiments*

623 S. VAN DEN HAUWE, *Topics in Applied Macroeconometrics*

624 F.U. BRÄUNING, *Interbank Lending Relationships, Financial Crises and Monetary Policy*

625 J.J. DE VRIES, *Estimation of Alonso's Theory of Movements for Commuting*

626 M. POPŁAWSKA, *Essays on Insurance and Health Economics*

627 X. CAI, *Essays in Labor and Product Market Search*

628 L. ZHAO, *Making Real Options Credible: Incomplete Markets, Dynamics, and Model Ambiguity*

629 K. BEL, *Multivariate Extensions to Discrete Choice Modeling*

630 Y. ZENG, *Topics in Trans-boundary River sharing Problems and Economic Theory*

631 M.G. WEBER, *Behavioral Economics and the Public Sector*

632 E. CZIBOR, *Heterogeneity in Response to Incentives: Evidence from Field Data*

633 A. JUODIS, *Essays in Panel Data Modelling*

634 F. ZHOU, *Essays on Mismeasurement and Misallocation on Transition Economies*

635 P. MULLER, *Labor Market Policies and Job Search*

636 N. KETEL, *Empirical Studies in Labor and Education Economics*

637 T.E. YENILMEZ, *Three Essays in International Trade and Development*

638 L.P. DE BRUIJN, *Essays on Forecasting and Latent Values*

639 S. VRIEND, *Profiling, Auditing and Public Policy: Applications in Labor and Health Economics*

640 M.L. ERGUN, *Fat Tails in Financial Markets*

641 T. HOMAR, *Intervention in Systemic Banking Crises*

642 R. LIT, *Time Varying Parameter Models for Discrete Valued Time Series*

643 R.H. KLEIJN, *Essays on Bayesian Model Averaging using Economic Time Series*

644 S. MUNS, *Essays on Systemic Risk*

645 B.M. SADABA, *Essays on the Empirics of International Financial Markets*

646 H. KOC, *Essays on Preventive Care and Health Behaviors*

647 V.V.M. MISHEVA, *The Long Run Effects of a Bad Start*

648 W. LI, *Essays on Empirical Monetary Policy*

649 J.P. HUANG, *Topics on Social and Economic Networks*

650 K.A. RYSZKA, *Resource Extraction and the Green Paradox: Accounting for Political Economy Issues and Climate Policies in a Heterogeneous World*

651 J.R. ZWEERINK, *Retirement Decisions, Job Loss and Mortality*

652 M. K. KAGAN, *Issues in Climate Change Economics: Uncertainty, Renewable Energy Innovation and Fossil Fuel Scarcity*

653 T.V. WANG, *The Rich Domain of Decision Making Explored: The Non-Triviality of the Choosing Process*

654 D.A.R. BONAM, *The Curse of Sovereign Debt and Implications for Fiscal Policy*

655 Z. SHARIF, *Essays on Strategic Communication*