Universität Paderborn
Fakultät für Wirtschaftswissenschaften

Essays in Empirical Health and Education Economics

Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
– Doctor rerum politicarum –

vorgelegt von

Valentin Schiele, Master of Science

geboren am 22.07.1987 in Filderstadt

Paderborn, 2022

Acknowledgments

Many people have supported me in one way or another during the preparation of this dissertation and on the way to it. At this point I want to thank them all. I am especially grateful to my supervisor and co-author Hendrik Schmitz. Thank you, Hendrik, for always supporting me on the way to my doctorate and for being available to me at all times with very valuable advice and guidance. Moreover, I thank my co-supervisor Wendelin Schnedler for the time he spends on this dissertation and the report, as well as for his helpful comments and suggestions so far. I want to thank Verena for introducing and welcoming me so warmly in Paderborn. Without you, Verena, I would have been lost more than once. I would like to thank Christian for not being scared off by our coffee and giving us a second chance. You have been a great colleague, co-author, and friend. I am also very grateful to all my other colleagues who have worked with me over the years and on whom I could always rely. My greatest gratitude goes to my family and friends for their unconditional support and motivation. Thank you, Judith and Jakob, for always being there for me and always believing in me. You are the greatest treasures in my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and Scope

This dissertation consists of four individual contributions to the health economics and economics of education literature. Each of them is self-contained and can therefore stand well on its own. Nevertheless, they are linked by common features. The greatest commonality is of methodological nature. The research question is addressed empirically, using large microdata and modern microeconometric methods in all four contributions to this dissertation. Moreover, two of the four independent contributions are also closely related thematically. They examine the role of important life events, namely job loss and serious illness, on later mental and physical well-being. In doing so, they contribute to our understanding of determinants of human capital depreciation, a topic that - compared to human capital accumulation - has received rather little attention in the economic literature. The two remaining contributions study consequences of more global determinants, namely institutional regulations, for individuals' health and labor market outcomes. They focus on unintended side-effects of policies affecting a large number of individuals. By doing so, they contribute to a more comprehensive understanding of the economic effects of such regulations.

The first of these studies, presented in Chapter 2 ("Quantile Treatment Effects of Job Loss on Health"), deals with a question that has a long-standing tradition in social science research (see e.g. Ezzy, 1993): Does job loss affect health? It builds on previous results indicating that the effect of job loss depends on individual and contextual characteristics (e.g. Kuhn et al., 2009; Green, 2011; Gathergood, 2013) and asks whether job loss impacts differently along the health distribution. Because people who have all of their physical and mental capacities at the time of job loss may be more able to find new employment or alternative sources of income, meaningful and rewarding tasks and activities, they might be less affected by job loss. Individuals with already compromised health, on the other hand, might be especially vulnerable and might suffer large from job loss with respect to their health. Consequently, looking only at average effects, which is common

in the related literature, could lead to a substantial underestimation of the individual as well as the social costs of job loss. To overcome this problem, the consequences of job loss throughout the entire health distribution are examined in this study.

The second contribution to this dissertation, presented in Chapter 3 ("Understanding Cognitive Decline in Older Ages: The Role of Health Shocks"), centers on health, too. Here, health is considered not only as an outcome or output of individual behavior and social, economic, and other conditions, but also as an important resource or input that determines agency later in life. In this chapter, we examine how a sudden deterioration in physical health affects another important aspect of human capital, namely cognitive abilities. Cognitive abilities are crucial factors for rational decision-making, social participation and - above all - autonomy. Dementia, as a (drastic) form of cognitive decline, is the main cause of need for long-term care in old age (Pick et al., 2004). With increasing population aging, the absolute numbers of individuals in need of care will continue to rise (see e.g. Patterson, 2018) and with them the strain on professional and informal caregivers and social security systems. However, we still know rather little about the determinants and potential moderators of age-related cognitive decline and thus about the reasons why healthy people become care-dependent ones. This is where Chapter 3 aims to contribute.

While the first two chapters focus on the consequences of individual experiences and events for health, the following two chapters examine effects of institutional settings or regulations that are beyond the responsibility and control of the individual. Chapter 4 ("Labor Market Spillover Effects of a Compulsory Schooling Reform in Germany") deals with the consequences of a large-scale educational reform for labor market outcomes of individuals who were not targeted by this reform. This reform increased length of compulsory schooling in Germany and has often been used in the empirical economics of education literature to estimate returns to education (e.g. Pischke and von Wachter, 2008; Cygan-Rehm, 2021). These studies typically rely on the assumption that such reforms do not exhibit externalities on individuals who are not targeted by the reform. Yet, given that these reforms affect a large share of the population, they might have spillover or general equilibrium effects and thus might impact individuals who are assumed not to be affected by the reform, too. Indeed, the data shows that the compulsory schooling reform in Germany led to a shock to the number of graduates in the reform year. I study whether this labor supply shock affected later life labor market outcomes of individuals who graduated in close distance to the reform year. This not only contributes to our understanding of externalities of education policies and the role of labor market conditions at the time of school completion for subsequent labor market outcomes, but has also implications for the interpretation of earlier findings in the literature on returns to education.

2

Chapter 5 ("Spring Forward, Don't Fall Back: The Effect of Daylight Saving Time on Road Safety"), the final contribution to this dissertation, deals with unintended consequences of a policy regulation, too. This regulation has been the subject of public attention and debate around the world twice a year since the beginning of time. It is the question of the sense or nonsense of setting the clocks forward one hour in spring and back one hour in fall that keeps stirring people's minds. In recent years, it has become increasingly evident that the time change fails to achieve its intended effect of saving energy (Havranek et al., 2018) and, even worse, that the practice of time change has harmful side effects, especially on health (e.g. Toro et al., 2015). As a result, the issue has recently made its way back into cabinets and plenary halls. The European Parliament voted to abolish the biannual time change from 2021 (European Parliament, 2019), but member states still cannot agree on which time regime (permanent summer or winter time) should govern in the future.[1] From an economist's point of view, the decision for one and against the other regime should take into account the permanent costs and benefits of the time regimes. While the costs of the time *change* under the current time regime are well documented, very little is known about the potential permanent costs of the alternative time regimes. One area where shifting daylight permanently from the morning to the afternoon/evening hours or vice versa could make a big difference is road safety, since traffic density and composition and thus accident risk differs by time of day. In this study, we investigate the influence of light conditions on the frequency of road accidents and, based on this, estimate how the introduction of different time regimes affects the frequency of accidents and the associated societal costs.

1.2 Data and Methods

Despite different thematic foci, all four studies follow a common principle: They address the respective research question with the help of data and modern microeconometric methods. The data used for this purpose come from different sources, can be assigned to different data types and are structured differently. They include information collected in surveys or obtained administratively. In some of the datasets, the units of observation are observed at one (cross-sectional data) or more (panel data) predefined points in time; in others, the units of observation are tracked continuously over time, by recording start and end times for each state (spell data) or enter the dataset only when a specific event occurs (event data). In all cases, however, the data represent individual (rather than aggregate) measurements that can be analyzed using microeconometric methods. The underlying statistical methods range from standard linear regression to modern semi-/non-parametric methods. They are all applied in a particular setting, e.g. institutional

---

[1]A similar legislative initiative, the Sunshine Protection Act, has been introduced (several times) in the U.S. Congress (United States Congress, 2019) and has now been referred to the relevant committee there.

or environmental, which, in combination with certain characteristics of the data, helps to solve the fundamental problem of causal inference: For each unit of observation, we can observe the outcome in only one (potential) state, the realized state. Alternative states, the counterfactual states, can never be observed for that unit of observation at the same time. Consequently, and by definition, neither individual treatment effects – the difference in (potential) outcomes between two or more alternative states – can be observed, but must be inferred based on reasonable assumptions (Holland, 1986). Suppose, for example, we are interested in the effects of unemployment on health. In this case, the two mutually exclusive states are being unemployed and not being unemployed, respectively. Obviously, we can observe only one of these states and thus only one of the possible outcomes for each individual at a time. To estimate the impact of unemployment on health, we need thus need to impute the outcome in the counterfactual state based on well-reasoned assumptions. In this section, I will briefly describe the data sources and methods used in this dissertation to overcome the fundamental problem of causal inference. In doing so, I will highlight certain features of the data that make them particularly attractive in the context of the research question at hand, and describe how these features and the chosen microeconometric approach are interrelated.

Chapter 2 ("Quantile Treatment Effects of Job Loss on Health") is based on an analysis of representative data from the German Socio-Economic Panel ("Sozio-oekonomisches Panel", SOEP). The SOEP is a well-known multidisciplinary household survey that has been conducted annually since 1984, has covered about 30,000 individuals in 15,000 German households in recent years, and provides information on a wide range of topics. The following features of the SOEP make it particularly valuable for this study. First, the SOEP combines detailed information on labor market participation with detailed information on health. Specifically, it includes background information on job loss that makes it possible to identify individuals who lost their jobs for reasons that are arguably unrelated to individual characteristics and thus independent of their health. This feature of the data is important for identification of the effects of job loss. Furthermore, the SOEP comes with continuous measures of physical and mental health making it possible to employ quantile regressions and thus to estimate heterogeneous effects along the range of possible health states. Second, because it is a panel study, individuals can be tracked over time. This is an asset especially when looking at the effects on physical health, since it presumably takes a while for changes in behavior or mental constitution due to job loss to manifest in physical health.

The following Chapter 3 ("Understanding Cognitive Decline in Older Ages: The Role of Health Shocks") deals with another determinant of human capital depreciation and is based on survey data, too. This study combines data from three sister surveys all targeting at individuals aged 50 and older: The Health and Retirement Study (HRS), the English Longitudinal Study of Ageing (ELSA) and the Survey of Health, Ageing

and Retirement in Europe (SHARE). These surveys cover a broad range of topics but provide especially valuable information on health. Particularly important for this study is experimentally collected information on grip strength and on cognitive abilities. These pieces of information heal common disadvantages of most survey data: that the information comes from self-assessments and is often subjective. Using this information, it is possible to assess respondents' physical and mental constitution objectively. Another important characteristic of the data, especially for identification and estimation, is the panel structure. As all three surveys follow respondents over time, it is possible to employ an event study approach. This approach is closely related to a classical difference-in-differences design, but offers some advantages over the latter. In this very application it utilizes differences in the timing of health shocks to estimate its effect on cognition. Thereby it also reveals potential dynamics of the effects, which is an important feature in the context of research on determinants of the gradual process of age-related cognitive decline. At the same time it also allows to directly assess the credibility of its main identifying assumption, i.e. that the timing of (but not necessarily the probability of experiencing) a health shock is as good as random. Plausibly, it is this transparency which has led to a boom of studies employing event study methods in the field of applied microeconometrics in recent years. The analysis, however, does not stop with the estimation of the effects of health shocks on cognition, but contributes also by shedding light on the role of two potentially important moderators of the effect. To do so, the analysis takes advantage of another feature of the data: it covers individuals from 20 different countries with quite different institutional settings. These differences in institutions are used to define instrumental variables that are then used in the moderation analysis to identify whether retirement and more education can dampen or amplify the consequences of a health shock for cognition.

Chapter 4 ("Labor Market Spillover Effects of a Compulsory Schooling Reform in Germany") investigates in potential spillover effects of a large compulsory schooling reform on wages and employment of individuals who were not directly subject to the reform. In this chapter, I make use of large-scale administrative data from the Sample of Integrated Labor Market Biographies (SIAB) and data from the Qualification and Course of Employment (QaC) surveys. The SIAB is particularly suitable in the context of the research question because of its size: it covers 2% of German employees subject to social security contributions on a daily basis and thus also makes it possible to estimate effects that may seem small at the individual level with high precision. Since it is based on employers' (mandatory) social security declarations, the information on wages and employment can be considered highly reliable and largely free of measurement errors. A disadvantage of this data set is, however, its rather broad measure of education. I therefore complement the SIAB data with the QaC data, which is way smaller than the SIAB but entails more detailed information on education and thus allows for a more precise assignment of treatment status. Both data sources come with information on the

home state of individuals. This allows to make use of the staggered introduction of the compulsory schooling reform in a difference-in-differences approach to separate the spillover effects from other spatial and temporal shocks on wages and employment.

Chapter 5 ("Spring Forward, Don't Fall Back: The Effect of Daylight Saving Time on Road Safety") studies how time regime choice affects road accident counts and severity using large administrative data from the Stats19 data base. Three noteworthy characteristics of the data make it especially valuable for this study. First, it provides very detailed information on date and time of each accident. Second, it also entails information on the exact geographical position of accidents. Third, it is exceptionally large: the data covers almost every road accident on public roads in England, Scotland or Wales between 1996 and 2016. The first characteristic allows to aggregate accidents by date and thus, as it is common in the related literature, to employ a regression discontinuity design in order to identify the (short-run) effects of the biannual time change. The second characteristic allows in combination with the third to add the geographic position as another dimension when aggregating the data. This unique combination makes it possible to go a step beyond existing research on effects of the current time regime and to look also on the long-run consequences of time regime choice. It allows to study whether it is better for road safety to have more light during evening hours and less during morning hours than vice versa, i.e. it allows to compare road safety under a year-round summer time regime with road safety under a year-round winter time regime. This is done in a simulation based on estimates for the effect of darkness derived from fixed-effects regressions. The underlying approach exploits deterministic variation in darkness for identification: day-by-day variation in darkness for a given region and hour; east-west variation for a given day of the year and hour; and north-south variation for a given day of the year and hour.

1.3 Summary of the Studies

This final section provides a brief overview of each of the four contributions to this dissertation. To this end, I briefly describe the respective research question and the methods and data used, I state the main findings, while also discussing the contribution to the literature and possible policy implications. This is done in the order in which the studies appear in the main part of the dissertation.

Chapter 2: Quantile Treatment Effects of Job Loss on Health

The first study in this thesis examines the consequences of job loss on mental and physical health, thereby focusing on possible heterogeneity of effects with respect to

the health status itself. The analysis builds on representative survey data and methods for the estimation of quantile treatment effects. To address endogeneity concerns, we follow the related literature and focus on job loss due to plant/establishment closure as an arguably exogenous reason for job loss. The results show that job loss impacts differently along the mental and physical health distribution. We find rather large and (partly) significant effects of job loss at the bottom parts of the health distributions, while both the effects at the upper parts of the health distributions but also the average effects are always small and never statistically significant.

These results can be interpreted such that the negative consequences of job loss for mental and physical health are concentrated among those who are most at risk: Individuals whose health is already compromised prior to job loss. This finding, which is overlooked when only the average impact is considered, has implications for the interpretation of previous research findings. It shows that considering only average impacts, which is common in the relevant literature, can lead to a substantial underestimation of both the individual and social costs of job loss. Moreover, it provides indications as to which group of people should be given preferential support by policy makers, e.g. with regard to a possibly necessary reintegration into the labor market or preventive measures or for the restoration of physical and mental health.

Chapter 3: Understanding Cognitive Decline in Older Ages: The Role of Health Shocks

In this chapter, we study how physical health shocks affect cognitive decline and ask whether more education or later retirement, both factors that can be influenced by policy and that are known to increase health capital and the cognitive reserve, can dampen adverse consequences of health shocks and slow down cognitive decline. To do so, we combine large panel data from continental Europe, the UK and the US, i.e. regions with different institutional settings including different health insurance systems. Using event study methods, we show that physical health shocks lead to a strong and persistent reduction of cognitive abilities. The estimates imply that a health shock in older ages has the potential to bring a care episode forward by up to 4 years. This result is largely independent of individual characteristics and the underlying health insurance system.

To circumvent endogeneity problems with respect to education and retirement in the moderation part of the analysis, we look at the interaction between health shocks and indicators for the potential to retire and years of education, respectively. We do not find evidence that more education or later retirement can slow down cognitive decline due to a health shock. Taken together, the study contributes to our understanding of the gradual process of cognitive decline and human capital depreciation in older ages and highlights the importance of health prevention for healthy cognitive aging.

Chapter 4: Labor Market Spillover Effects of a Compulsory Schooling Reform in Germany

In this chapter, I study whether the introduction of the compulsory ninth grade in West German states during the 1940ies-1960ies affected employment and wages of individuals who were not directly subject to the reform. Specifically, I show that the reform led to a negative shock to the supply of basic track graduates in the reform year. Building on previous results on the importance of labor market conditions at graduation, I argue that individuals who were not directly affected by the introduction of the compulsory ninth grade but graduated in close temporal distance to the reform may have benefited in the long-run from the negative shock to labor supply induced by the reform. To investigate these kinds of spillover effects, I exploit the staggered introduction of the compulsory ninth grade across German federal states in a difference-in-differences approach. Based on register and survey data, I find no evidence for persistent spillover effects for men. For women, however, my results suggest that the labor supply shock resulting from the reform may have led to a persistent increase in employment and wages.

This study contributes to the understanding of the role of labor market conditions at the time of graduation from school on subsequent labor market outcomes, focusing on the role of supply-side shocks. It also adds to the growing literature examining the externalities of educational interventions. Finally, the results provide insights into the plausibility of the assumption of no spillover effects in the literature on returns to education, which is commonly made there (at least implicitly) but rarely discussed.

Chapter 5: Spring Forward, Don't Fall Back: The Effect of Daylight Saving Time on Road Safety

In this final contribution to the thesis, we examine how the choice of time zone affects road safety. In particular, we are interested in the costs or benefits of establishing one time zone (e.g. year-round summer time) instead of another (year-round winter time). To assess the costs associated to road safety under the alternative time regimes, we first examine how darkness affects accident counts. The resulting estimates are then used in a simulation where we, depending on the underlying time regimes, shift light either from morning to evening hours or vice versa to evaluate the long-term costs of the alternative time regimes.

The analysis draws on extensive administrative data from England, Scotland, and Wales. Using a regression discontinuity design, we are able to replicate the results of previous studies showing in a first step that the biannual time change reduces road safety. We then add to the existing literature by examining the permanent consequences of time

regime choice. In particular, we show that darkness is a significant risk factor for road traffic and that the costs of the current time regime are significantly underestimated when focusing on the time change itself. The results suggest that the introduction of permanent summer time can prevent fatal accidents and thus reduce social costs compared to year-round winter time. These results may also provide guidance to policy makers who have recently decided to abolish daylight saving time but do not yet know what should come after.

Chapter 2

Quantile Treatment Effects of Job Loss on Health*

2.1 Introduction

The aim of this paper is to investigate if job loss causally deteriorates mental and physical health, thereby focusing on heterogeneous effects with regard to the health status itself. Despite a large body of convincing research on the effects of job loss on health, where effects on the conditional mean are analyzed, the findings are somewhat inconclusive. We employ quantile regression techniques to yield a broader picture of the effects on different parts of the health distribution. Theories as well as some previous empirical results deliver arguments showing that adverse effects of job loss on health are not the same for everybody and that initial health might matter. The knowledge about such differences in the effect of job loss on health is important, as they potentially influence individual and social costs of job loss.

There are good reasons to assume that job loss may deteriorate health, both through a pecuniary and a non-pecuniary channel.[1] Several theories and models – mostly originating in fields like social medicine, psychology, and sociology[2] – emphasize the beneficial role of work for well-being. These theories suggest that job loss negatively affects mental health through a reduction in self-esteem, a loss of daily routine and social contacts and a rise in stress. Several of the theoretical models dealing with job

---

[1]However, is not fully clear whether income causally affects health, see e.g. Frijters et al. (2005), Lindahl (2005) and Ahammer et al. (2015).

[2]See, e.g., Jahoda et al. (1933), Jahoda (1981) and Jahoda (1983), Warr (1987), Fryer (1986), Nordenmark and Strandh (1999). Readers who are interested in a wider overview of models and theories dealing with work, job loss, unemployment and mental health are referred to Hollederer (2011), Ezzy (1993) and Fryer (1992).

*This chapter is joint work with Hendrik Schmitz and is published as: Schiele, V. and Schmitz, H. (2016). Quantile treatment effects of job loss on health. *Journal of Health Economics*, 49: 59-69.

loss, unemployment and health suggest that there are heterogeneous effects on health and some of them predict that initial health matters (e.g., Warr's vitamin model, Warr, 1987).

The main challenge in the empirical literature on job loss and health is the potential endogeneity of job loss. Indeed, there is a well documented selection of unhealthy individuals into unemployment (e.g. Arrow, 1996; van de Mheen et al., 1999, Jusot et al., 2008) and of healthier individuals out of unemployment (Stewart, 2001). To deal with this problem, plant closures or mass layoffs as exogenous source of variation in job loss are often used in the literature.[3] Examples are Salm (2009), Mandal et al. (2011), Marcus (2014), Browning and Heinesen (2012), Kuhn et al. (2009) or Eliason and Storrie (2009a,b).[4] The results of these studies are mixed which may have several reasons including different outcome measures (also subjective vs. objective), different data sources (surveys and administrative data) and different social and institutional contexts.

A full synthesis of the literature is beyond the scope of this article. Yet, as these studies look into effects on the conditional mean they might not be able to detect effects which apply only to certain parts of the distribution. As discussed before, one would not expect that the effect of job loss on health is the same for everybody. Some of the results mentioned above indicate that the effect of unemployment on health varies by certain individual and contextual factors like sex, age, employability or local unemployment (e.g. Kuhn et al., 2009; Green, 2011; Gathergood, 2013). The findings of Huber et al. (2011) support our presumption of heterogeneous effects of job loss on health with respect to initial health, as they indicate that people who were in poor health before they left welfare are those gaining most from a transition into employment or a welfare-to-work program. Therefore, it seems to be likely that the causal effect of job loss on health, if there is such a relationship, varies with the health distribution. Analyzing this is the contribution of this paper.

We use representative panel data from the German Socio-Economic Panel (SOEP) for the years 2000-2014 and methods for the estimation of unconditional quantile treatment effects. Although there is a growing number of studies using quantile regression approaches, these methods have only recently been introduced to the literature on job loss and unemployment. Korkeamäki and Kyyrä (2014) use quantile regression to study the effect of job loss on the conditional earnings distribution. Binder and Coad (2015a,b) employ quantile regression to study differences in the effects of unemployment on measures of (mental) well-being and life satisfaction throughout the distributions of

---

[3]Other approaches are employed, too, however. In a recent study Gathergood (2013) uses industry-specific unemployment rates as instruments for unemployment.

[4]Other studies analyze the effect of unemployment on health e.g. Green (2011), Böckermann and Ilmakunnas (2009), Schmitz (2011), Marcus (2013). This, however, is more challenging because even if job loss due to plant closure is exogenous, certain individual quickly find a new job while others stay unemployed.

these variables. They indeed find different effects of unemployment at different parts of well-being and life satisfaction distributions. However, they focus on unemployment (and all reasons of unemployment) and fairly strong assumptions need to be imposed to give their results a causal interpretation.

We employ quantile regression techniques proposed by Firpo (2007) and mainly focus on job loss due to plant closure as an arguably exogenous reason for job loss. Our outcome variables are the Body-Mass-Index (BMI) as well as the Physical Component Summary Scale (PCS) and the Mental Component Summary Scale (MCS) gathered from the SF12v2-questionnaire in the SOEP. From a statistical point of view these outcome measures are preferred over other possible health measures, such as, e.g., specific diagnoses, as they are continuous variables and allow to employ quantile regression methods.

The results suggest that the effect of job loss significantly varies across the mental and physical health distribution. Job loss affects physical health adversely for individuals in the middle and lower part of the health distribution while those in best physical condition do not seem to be affected. The results for mental health, though less distinct, point in the same direction, indicating that people in the lowest part of the mental health distribution suffer most from job loss, while those in the upper parts do not seem to be affected. If rank preservation holds, this can be interpreted as a finding that those who were worse off already before job loss suffer more than those who were in better health before. Mean estimations would have missed these findings as standard matching methods show no average effect of job loss on health in our application. We find no effects on BMI.

The remainder of this paper is structured as follows: Section 2.2 describes the empirical strategy to identify causal effects of job loss on health. Section 2.3 provides an overview of the data while the estimation results are presented in Section 2.4. Section 2.5 concludes and discusses policy implications.

2.2 Empirical Approach

2.2.1 General Idea and Framework

The estimation of quantile treatment effects (QTE) enables to evaluate the effect of a variable on different points of the outcome distribution and therefore allows for the identification of effects even in situations where the mean of the outcome variable remains unchanged (see Firpo, 2007 and Frölich and Melly, 2013 for technical details in deriving and estimating QTE). As the main point of interest of this work is to examine if the effect of job loss varies along the overall health distribution (or the marginal

distribution), methods proposed by Firpo (2007) for the estimation of unconditional quantile treatment effects – in contrast to conditional treatment effects that inform about effects conditional on covariates – are used here.

Let Y denote the outcome variable, which is a continuous measure of health, D the treatment variable taking on the value 1 in case of job loss and 0 in case of no job loss, and X a set of covariates. Then Y_i^0 denotes the health variable in a situation where individual i is continuously employed with the same employer whereas Y_i^1 tells us about the value of the health variable in the situation where i experiences a job loss. The observed health state of individual i is therefore given by $Y_i = Y_i^1 D_i + Y_i^0(1 - D_i)$. Taking the difference between the τ quantile of the potential outcome distribution in the (hypothetical) situation where all individuals experience a job loss ($Q_{Y^1}^\tau$) and the τ quantile of the potential outcome distribution in the (hypothetical) situation where all individuals are continuously employed ($Q_{Y^0}^\tau$) gives the effect of job loss on the potential outcome distribution at quantile τ (Δ^τ):

$$\Delta^\tau = Q_{Y^1}^\tau - Q_{Y^0}^\tau \qquad (2.1)$$

These quantile treatment effects (QTE) are defined for all $\tau \in (0,1)$ and, thus, in principle allow for the identification of effects over the entire health distribution (Frölich and Melly, 2013).

2.2.2 Identification

If the treatment status was completely randomly assigned to individuals, one could simply use the outcome distributions for the treated and untreated sub-populations to obtain the QTE. Obviously, this is not the case with respect to job loss, as there are many factors related to job loss like education, occupation or unobservable factors like willingness to take risks or other preferences that might also determine health.

Firpo's estimator is based on a selection-on-observables, or *unconfoundedness* assumption, which requires potential health states to be independently distributed from job loss, conditional on X. While this assumption is hard to justify with respect to job loss for all reasons, we argue that it is much more likely to hold for job loss due to plant closure. Indeed, plant closures have often been regarded as an exogenous reason for job loss in the previous literature. This is mainly because the reason of job loss does not lie at the responsibility of the individual or is linked to the individual's characteristics and abilities, but is associated with cyclical developments, (wrong) management decisions or other conditions and developments that are considered not to be self-inflicted. Nevertheless, it seems possible that employees with better outside options, for example higher

educated or better skilled persons, quit their job before the plant closure takes place. As a result, less educated and low skilled persons would be the only employees who stay at the plant until it is closed. However, using information about predetermined individual, job and firm-characteristics, we try to control for a selection into job loss due to plant closure that might be related to the health status and assume job loss due to plant closures to be randomly assigned conditional on X. Note, however, that the definition of the unconditional QTE does not depend on the set of covariates X included. This is because the unconditional distributions of Y^0 and Y^1 do not depend on X. In principle, one could estimate the unconditional QTE without any covariates.

In order to infer from distributional effects of job loss on health to individual effects one has to assume *rank preservation*, which requires the relative rank of an individual in the outcome distribution to be the same in the (hypothetical) situations with and without treatment. Recently, formal tests for rank preservation and the weaker rank similarity assumption have been proposed by Frandsen and Lefgren (2018) and Dong and Shen (2018). Frandsen and Lefgren (2018) apply their test to data from the STAR class size reduction experiment and the National Job Training Partnership Act (JTPA) Study and find that the rank similarity assumption does not hold for the effect of class size on math scores, but find no evidence for a violation for the effect of JTPA participation on earnings. In contrast, the results of Dong and Shen (2018), also using JTPA data, suggest that rank similarity is violated at least for male trainees.

These results indicate, that the rank preservation assumption might be violated in many applications. We argue, however, that there is a good reason to assume that the rank preservation assumption is not systematically violated with respect to job loss and health. There are, for example, other determinants of health like hereditary tendencies and the provision of medical care which are not affected by job loss. The most plausible explanation for large adverse effects at the lower part of the outcome distribution and no effects at the upper part of the distribution is that people in poor health suffer most from job loss, while the health of people in (initially) good health is not impaired. Therefore, we assume that people rank similarly in the health distribution of the (hypothetical) situation where all people experience job losses and the health distribution of the (hypothetical) situation where all people are continuously employed. But even if rank preservation does not hold, it is not useless to estimate unconditional quantile treatment effects. In general, one could, for example, calculate bounds for the individual effects or simply focus on distributional aspects, which are sometimes even more relevant than individual effects (Firpo, 2007). In the context of job loss and health, for example, large negative effects at the lower part of the distribution and no or positive effects at the upper part would indicate that higher rates of job termination increase health inequality. Furthermore, as this implies that the fraction of people who are in bad health increases with increasing rates of job termination, one would also

expect health expenditures to rise with job termination. [5] The estimation is carried out with the Stata command `ivqte` by Frölich and Melly (2010). The Appendix provides details of the estimation procedure.

2.3 Data

Our data comes from the German Socio-Economic Panel (SOEP). Starting in 1984 as a representative cross-section of the adult population living in private households, the survey was repeated in each year, resulting in a panel of 31 waves (1984-2014) consisting of 24,000 individuals (Wagner et al., 2007). In total, we use data from the waves 2000-2014. As the detailed health measures from the SOEP health module are only available in every second year starting in 2002, observations from these seven waves build the main database of the analysis. The data set is restricted to people aged 16 to 63 and who are either working full- or part-time at wave $t - 2$ as job loss is defined to be between the wave $t - 2$ and t, see below. Furthermore, observations with missing values on either all outcome variables, the job loss variable or one of the main covariates are dropped from the data set. Depending on the outcome and treatment evaluated and the number of covariates accounted for, the sample size varies for the different regression results. The largest sample contains 47,446 observations stemming from 17,901 different individuals.

2.3.1 Treatment Variable

The baseline treatment variable is *job loss (plant closure)*. It takes on the value one in year t if the respondent reports a job loss since the interview at $t - 2$ because their place of work or office has closed and zero if the respondent reports no job loss, that is, continuous employment between $t - 2$ and t. Thus, when using this treatment variable, we exclude all observations with job loss for other reasons than plant closures and estimate the effects of losing a job due to a plant closure as opposed to being continuously employed with the same employer. As a benchmark, we also look at job loss for any reason, *job loss (all reasons)*. It takes on the value one if the individual has lost their job since the interview at $t - 2$ for any reason and the value zero otherwise (i.e. continuous employment between $t - 2$ and t). Around 9,800 person-year observations in the final estimation sample suffered from a job loss, almost 900 thereof due to plant closure. Certainly, as all variables in the SOEP are self-reported, measurement error in the

---

[5] A common support assumption and uniqueness of quantiles completes the set of assumptions. We restrict the sample to the region of common support. Uniqueness of quantiles means that potential health states are continuously distributed with positive density. As our outcome measures are continuous variables that are defined on a broad range of values we are confident that the quantiles are unique and well defined.

treatment status cannot be ruled out. In principal this could lead to an underestimation of the true effects ("downward bias") but it is unclear to what extent.

2.3.2 Outcome Measures

Three variables from the health module of the SOEP are used as outcome measures: the Mental Health Component Summary Score (MCS), the Physical Health Component Summary Score (PCS) and the Body Mass Index (BMI). Although the theoretical models cited in the introduction focus more on mental health, we argue that there might also be effects on physical health. First, psychological stress is discussed to be a determinant of cardiovascular diseases in the medical literature (Cohen et al., 2007). Second, coping strategies related to an increased level of stress and a lack of duties and responsibilities, like smoking or drinking, are known to have adverse health effects (e.g. Boden et al., 2010; Cornfield et al., 2009; Fergusson et al., 2009). Furthermore, mental health problems might be accompanied by psychosomatic disorders that at least affect physical well-being and the ability to cope with challenges of everyday life. Finally, weight gains due to dietary change or changes in physical activity level can be considered as health problems, at least if they apply to people at the upper part of the BMI distribution. Although one cannot directly infer from BMI on health or even subjective well-being, very high and very low values of the BMI are known to be associated with higher mortality (Campos et al., 2006) and very high values are linked to a higher prevalence of diseases like cardiovascular disease, hypertension and diabetes (Wyatt et al., 2006).

While job loss might affect mental health quickly, adverse effects of job loss on physical health can rather be expected in the medium run. To take this into account, we estimate the effects of job loss between $t-2$ and t on mental health in t (i.e., at most two years after job loss) while we measure PCS and BMI in $t+2$ (i.e., two to four years after job loss).

The summary scores for mental and physical health (MCS and PCS) are based on the SF12v2 questionnaire, which is an internationally standardized set of 12 items regarding eight dimensions of the individual's health status.[6] These eight dimensions comprise physical functioning, physical role functioning, bodily pain, general health perceptions, vitality, social role functioning, emotional role functioning and mental health. Using one or two items, a scale ranging from 0 to 100 (with mean=50 and standard deviation=10 in the SOEP 2004 population) is calculated for each of these eight dimensions. The eight dimensions or subscales are then aggregated to the two main dimensions mental and physical health, using weights derived from an explorative factor analysis. Like in the case of the sub-scales, the aggregated scales (MCS and PCS) are standardized

---

[6]See the questionnaire in Table A2.1 in the Appendix.

and transformed to have a mean of 50 and a standard deviation of 10 in the SOEP 2004 population with lower values indicating lower health states (Andersen et al., 2007).

In epidemiological research the SF-12 is commonly used to measure general health and functioning (Ware et al., 1996). It includes information on subjective health but the component summary scales are correlated actual with health diagnoses. With respect to mental health, Gill et al. (2007) find that MCS "is a useful screening instrument for depression and anxiety disorders in the general community, and thus, a valid measure of mental health". This view is supported by Vilagut et al. (2013). Salyers et al. (2000) regard it as a valid and reliable instrument to measure health-related quality of life. Recently, MCS has also been used in the health economic literature (e.g., Schmitz and Westphal, 2015; Reichert and Tauchmann, 2011).

The BMI, a standard measure of the body shape, is the weight of a person adjusted to the person's height. It is calculated as the ratio of weight in kilograms to the square of height in meters. As MCS and PCS are based on the same set of sub-scales, they might overlap to a certain degree. However, due to the different weighting schemes, they are expected to absorb different components of overall health. A negative and small correlation coefficient between MCS and PCS (around -0.04 in our sample) confirms this. Table A2.2 in the Appendix reports deciles of the outcome variables in the sample and shows, for instance, that the median individual is overweight (BMI > 25) and more than 10 per cent are obese (BMI > 30) according to standard classifications (WHO, 2000).

Figure 2.1 shows empirical density functions of the three outcome measures for people with job loss due to plant closure, people with job loss (all reasons), and those without job loss. The distributions of all outcome measures differ for all three groups where, all in all, individuals without a job loss are on average in slightly better health than those who suffered from a job loss due to plant closure. Interestingly, however, for those with a job loss due to plant closure, the distribution of health outcomes is not simply shifted to the left (indicating worse health) but rather flattened, retaining more probability mass at the tails of the distribution. Assuming unconditional exogeneity of job loss due to plant closure, it would imply that while some people's mental and physical health suffers from involuntary job loss, some other people's mental and physical health benefits from it. With respect to BMI it rather points to adverse effects of involuntary job loss at both tails of the distribution.

2.3.3 Covariates

The effects of job loss on the unconditional health distribution, in principle, do not depend on the covariates accounted for. Nevertheless, it seems useful to include some covariates when estimating unconditional quantile treatment effects, since the inclusion of these variables can reduce the variance of the estimator and makes the

Figure 2.1: Distribution of Outcome Variables by Treatment Status

(a) MCS (at t)

(b) PCS (at $t+2$)

(c) BMI (at $t+2$)
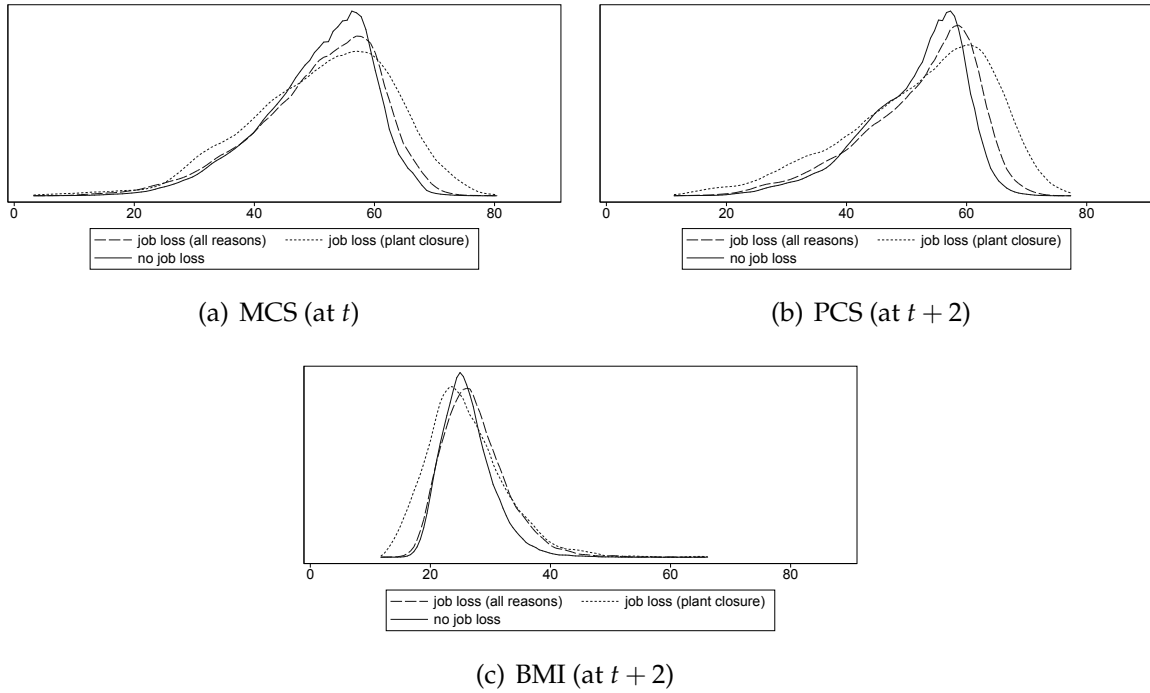
Note: Authors' calculations based on the SOEP.

unconfoudedness assumption more credible. Even for job loss due to plant closure the unconfoundedness assumption might hold only conditional on some covarariates X.

Plant closures probably do not happen like a bolt from the blue, but are consequences of (wrong) decisions of the firms' management or macroeconomic developments and thus might be preceded by a period of downsizing.[7] This could be problematic for two reasons. First, the management might try to get rid of certain employees, e.g. the least productive. To take this into account we control for *education, working experience, occupational position, gross labor income, unemployment experience*. Second, the period of downsizing is probably recognized also by employees, who might try to leave the sinking ship and get a position at another firm. However, it is likely that not all employees are able or willing to leave the plant before it closes down. Whether they (try to) find a new job or stay at the firm depends on their outside options. We argue that the outside options largely depend on sex and ethnic background (in case of racial discrimination in the labor market), education, working experience, age, occupation, spatial flexibility, the dependence on income from (own) employment and local labor market conditions. Thus, we additionally control for *sex, German nationality, migration background, age, private health insurance, state of residence, living with partner, household size, net household income*. As there are marked differences in insolvencies between sectors and firms of different sizes as well as across time (Statistisches Bundesamt, 2016) we also control for the *year* of interview and the following firm characteristics: *size* and

---

[7]However, Browning and Heinesen (2012) report that in Denmark 70% of displaced employees having worked in a plant that finally closes down are displaced in the year of plant closure.

sector. Furthermore, Reichert and Tauchmann (2011) provide evidence that fear of job loss adversely effects mental health. To separate this effect from effects of actual job loss, we control for *job worries*. Some of the variables mentioned above might themselves be affected by job loss. To avoid the problem of "bad controls", we measure all controls before job loss, i.e. at $t - 2$.

Finally, we check whether the results are robust against the inclusion of lagged health variables by capturing pre-treatment outcomes MCS, PCS and BMI. This should also capture unobservable but time-invariant factors that affect the outcomes (like genetic factors, general frailty) and might be correlated with job loss. Table A2.3 in the Appendix reports descriptive statistics of the controls by treatment status. One possible explanation for this finding is that in some cases not the entire plant but only a certain office is closed and that employees in better health could be more likely to get a job at another office within the same firm.

2.4 Estimation Results

Obviously, the raw differences in outcome distributions for the different treatment states (as illustrated in Figure 2.1) do not necessarily reflect the effect of job loss on the health distribution. Therefore it is reasonable to account for selection into job loss in a regression framework.

Table 2.1 reports the estimated effects of job loss on our first outcome measure, the Mental Component Summary Scale (MCS). As a benchmark, the first row shows the results for the average treatment effects on the treated (ATT) for *job loss (plant closure)* (columns I, II and III) and *job loss (all reasons)* (column IV). The ATT is estimated using standard propensity-score-kernel-matching.[8] In column I, we report the ATT when we do not control for any potential confounder. In all other specifications we control for differences in pre-treatment individual and household as well as firm characteristics (see Section 2.3.3). In columns III and IV, we additionally allow for a non-random selection into job loss based on pre-treatment health status, i.e., we also take potential differences in MCS, PCS and BMI between the treated and control sub-populations at $t - 2$ into account. This, however, comes at the cost of a reduced sample size, as the lagged health variables are not available for all individuals.[9]

Without any controls (I), we find average differences in MCS by *job loss (plant closure)* of 1.169, suggesting that, on average, those who experienced a job loss are less healthy. As soon as we control for differences in pre-treatment individual, household and firm

---

[8]This uses an Epanechnikov kernel and a bandwidth of 0.02 which ensures that the normalized difference between treated and matched controls is smaller than 5% for all variables.

[9]Since we have no information about health in the year 2000, we cannot use observations from the year 2002 and thus lose around 17% of observations.

Table 2.1: Quantile Treatment Effects of Job Loss on MCS (at *t*)

| | Job Loss Plant Closure | | | | | | Job Loss all Reasons | |
| | I | | II | | III | | IV | |
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
ATT	-1.169***	0.345	-0.492	0.348	-0.582	0.372	0.415***	0.134
QTE								
0.1	-2.322***	0.704	-1.686**	0.859	-1.459	1.069	-0.657	0.469
0.2	-2.152***	0.486	-1.328	0.960	-1.036	1.007	0.006	0.282
0.3	-2.229***	0.554	-0.959	0.694	-1.212	0.944	0.524*	0.292
0.4	-1.682***	0.542	-0.577	0.842	-0.944	0.871	0.674**	0.268
0.5	-1.262***	0.450	0.000	0.741	-0.588	0.915	0.990***	0.187
0.6	-1.491***	0.419	-0.175	0.638	-0.516	0.765	0.721***	0.217
0.7	-0.739	0.464	0.097	0.756	-0.583	0.837	1.169***	0.146
0.8	-0.213	0.335	0.751	0.678	0.391	0.675	1.536***	0.155
0.9	0.018	0.347	0.018	0.367	0.000	0.579	1.232***	0.229
Standard Controls	no		yes		yes		yes	
Pre-Treatment Outcomes	no		no		yes		yes	
Obs	47,446		44,071		36,058		43,008	

Note: Authors' calculations based on the SOEP. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Bootstrapped standard errors based on 500 replications, clustered at individual level. The ATT is estimated using standard propensity-score-kernel-matching. Column I gives the effects without any controls. In column II we control for: sex, age, education, German nationality, migration background, private health insurance, working experience, unemployment experience, job worries, living with partner, household size, household income, year, state, gross labor income, occupation, sector, firm size (measured at $t-2$); in columns III and IV additionally for MCS, PCS, and BMI (measured at $t-2$).

characteristics (II), the effect decreases in size. Additionally controlling for pre-treatment outcomes (III), does not affect the estimates much. For this specification the average effect of *job loss (plant closure)* is -0.582, a rather small effect (roughly one twentieth of a standard deviation of which is around 10), suggesting that, on average, there is no adverse effect of job loss due to plant closure on mental health. This is in line with Schmitz (2011) who studies the average effect of unemployment on the same mental health score.

Turning to the QTE results, however, we find some evidence for negative effects of job loss due to plant closure on the distribution of mental health states. However, they seem to apply only to the lowest part of the distribution.[10] For all specifications, the largest estimates can be found for the first three deciles, while the estimates for the middle and upper part of the mental health distribution are small in size and for some deciles even positive. In general, controlling for individual, household and firm characteristics

---

[10]As a reading example, the second row of Table 2.1 reports the quantile treatment effects (QTEs) on the 0.1 quantiles (the first deciles) of the MCS distribution. The third row reports the QTEs on the 0.2 quantile (second deciles) and so on.

decrease the estimated QTEs considerably, indicating that job loss due to plant closure is not purely randomly assigned. Additionally controlling for health before job loss, however, attenuates the estimated QTEs only slightly. We thus focus in the following on the specification with standard controls (column II) and the specification additionally controlling for health before job loss (column III). Although we control for health in $t-2$ in column III (and the differences to column II are small), we cannot rule out that job loss due to plant closure might be somewhat endogenous, at least if our measures for pre-treatment health do not capture all differences in health before job loss.

The effect of *job loss (plant closure)* on the first decile of MCS is estimated to be between -1.69 (standard controls) and -1.46 (additionally controlling for health before job loss), a considerable though not dramatic impact – approximately one sixth of a standard deviation. This impact is statistically significant only if we do not control for pre-treatment health. However, as the point estimates are of similar size for both specifications, the differences with respect to statistical significance seem to be driven mainly by a loss in statistical power due to reduced sample size for the latter one. The effects on the second and third decile, though somewhat smaller and not statistically significant, point in the same direction and provide additional evidence that job loss stretches the lower part of the distribution of mental health states. We do not find any evidence for effects of job loss on the upper part of the mental health distribution. All in all, the results suggest a small effect of job loss on mental health in the lower part of the distribution which is as the margin of statistical significance. This finding is in contrast to the much smaller average treatment effect on the treated.

The results for *job loss (all reasons)* (column IV) differ considerably as they indicate positive effects on mental health in large parts of the distribution. One may speculate that this is because *job loss (all reasons)* comprises all possible changes in employment status that involve job loss, including voluntary job change, (early) retirement, sabbaticals et cetera, for which even positive effects are imaginable. However, as this variable is probably endogenous we focus, in the following, entirely on *job loss (plant closure)* and note that the results for *job loss (plant closure)* possibly do not carry over to other (especially voluntary) reasons for job loss.

Table 2.2 reports the (medium term) effects of job loss on physical health as measured by the Physical Component Summary Scale (PCS) at $t+2$. Again, the estimated ATT does not provide any evidence for economically relevant average effects of job loss on physical well-being. The estimates are rather small in size (less than one twentieth of a standard deviation) as soon as we allow for a non-random selection into job loss due to plant closure.

In contrast, the estimates for the QTEs do suggest that there are effects of job loss on the distribution of physical health and that these effects apply to the lower two thirds of the outcome distribution. Again, the largest estimates can be found for the first decile. This

Table 2.2: Quantile Treatment Effects of Job Loss on PCS (at $t + 2$)

| | Job Loss Plant Closure | | | | | | Job Loss all Reasons | |
| | I | | II | | III | | IV | |
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
ATT	-0.869**	0.344	-0.168	0.349	-0.457	0.376	0.076	0.120
QTE								
0.1	-2.138***	0.802	-2.464**	1.234	-2.531*	1.364	-1.771***	0.372
0.2	-1.304**	0.591	-1.521	1.253	-1.694	1.398	-1.126***	0.290
0.3	-1.246**	0.522	-0.938	0.982	-0.898	1.232	-0.931***	0.294
0.4	-0.911	0.591	-1.487*	0.826	-1.181	0.983	-0.694**	0.273
0.5	-1.048*	0.531	-1.593**	0.689	-1.453*	0.761	-0.527*	0.279
0.6	-0.684**	0.368	-1.528**	0.608	-1.320*	0.700	-0.225	0.161
0.7	-0.069	0.419	-1.559*	0.671	-1.487	0.969	0.000	0.070
0.8	0.000	0.254	-0.320	0.472	0.000	0.707	0.172	0.175
0.9	0.000	0.253	0.000	0.441	0.185	0.558	0.296	0.197
Standard Controls	no		yes		yes		yes	
Pre-Treatment Outcomes	no		no		yes		yes	
Obs	33,885		31,765		26,260		31,038	

Note: Authors' calculations based on the SOEP. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Bootstrapped standard errors based on 500 replications, clustered at individual level. The ATT is estimated using standard propensity-score-kernel-matching. Column I gives the effects without any controls. In column II we control for: sex, age, education, German nationality, migration background, private health insurance, working experience, unemployment experience, job worries, living with partner, household size, household income, year, state, gross labor income, occupation, sector, firm size (measured at $t − 2$); in columns III and IV additionally for MCS, PCS, and BMI (measured at $t − 2$).

impact is relatively large (about one fourth of a standard deviation). The QTEs for the following six deciles (quantiles 0.2 to 0.7) are negative, somewhat smaller than the effect on the first decile, but of considerable size and statistically significant for the fifth and sixth decile. This indicates that job loss due to plant closure stretches the distribution of physical health states to lower values.

The results for the BMI, again measured at $t + 2$, are reported in Table 2.3. The estimate for the ATT in our preferred specification is small and, thus, comparable to the results of Marcus (2014). With respect to the quantile treatment effects, we find some evidence for positive and significant effects on the upper deciles (column II). However, the estimated effects decrease substantially in size as soon as we control for differences in our outcome variables between treated and controls at $t − 2$. When controlling for these differences, there are only minor and unsystematic differences between the estimates for the different deciles. The largest (0.629) but insignificant effect corresponds to a change in body weight by roughly 2 kilogram for a person of average height (1.75m). This sounds quite much at first, but given that this effect applies to the 0.8 quantile, which is 29 (see Table A2.2) in our sample and corresponds to almost 90 kilogram for a person of

Table 2.3: Quantile Treatment Effects of Job Loss on BMI (at $t + 2$)

| | Job Loss Plant Closure | | | | | | Job Loss all Reasons | |
| | I | | II | | III | | IV | |
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
ATT	0.401**	0.186	0.227	0.192	0.197	0.123	0.051	0.038
QTE								
0.1	0.114	0.280	0.114	0.392	-0.194	0.377	-0.032	0.109
0.2	0.307	0.222	0.498	0.305	0.114	0.453	0.141	0.089
0.3	0.309	0.208	0.386	0.282	0.339	0.341	0.207**	0.093
0.4	0.047	0.163	0.193	0.283	0.163	0.336	0.222**	0.090
0.5	0.262	0.246	0.485	0.395	0.258	0.404	0.298***	0.105
0.6	0.469*	0.261	0.711*	0.374	0.395	0.428	0.317**	0.123
0.7	0.516*	0.265	0.935*	0.501	0.192	0.608	0.505***	0.166
0.8	0.703**	0.315	1.020**	0.515	0.629	0.675	0.358**	0.139
0.9	0.954**	0.414	0.708	0.678	0.004	0.948	0.306	0.216
Standard Controls	no		yes		yes		yes	
Pre-Treatment Outcomes	no		no		yes		yes	
Obs	33,695		31,588		26,183		30,947	

Note: Authors' calculations based on the SOEP. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Bootstrapped standard errors based on 500 replications, clustered at individual level. The ATT is estimated using standard propensity-score-kernel-matching. Column I gives the effects without any controls. In column II we control for: sex, age, education, German nationality, migration background, private health insurance, working experience, unemployment experience, job worries, living with partner, household size, household income, year, state, gross labor income, occupation, sector, firm size (measured at $t − 2$); in columns III and IV additionally for MCS, PCS, and BMI (measured at $t − 2$).

average height, a gain in body weight by 2 kilogram would not be dramatic. All in all, we conclude that there is no evidence for an effect of job loss on the BMI, neither on the mean nor on different quantiles.

Taken together, we find some evidence that job loss impacts differently at different parts of the distribution of mental and physical health. While the adverse effects of job loss due to plant closure for mental health seem to concentrate at the bottom part of the distribution, the estimated effects for the PCS suggest that job loss impacts also in the middle of the distribution, stretching the whole distribution of physical health states considerably to lower values. We do not find evidence for economically relevant effects of job loss on the BMI distribution, however. Assuming rank preservation, our results suggest that job loss (*i*.) affects mental health for those who were in poor mental health already before they lost their job; (*ii*.) deteriorates physical health for those who initially rank in the lowest or middle part of the distribution of physical well-being; and (*iii*.) does not affect mental and physical health for those in better health.

2.5 Conclusion

The aim of this study is to evaluate empirically if job loss causally deteriorates health, thereby focusing on heterogeneous effects of job loss on health with respect to the health distribution. Some previous empirical results as well as theories suggest that job loss and unemployment deteriorate health the more, the lower initial health is. Using quantile regression methods that allow the effect of an explanatory variable to vary along the outcome distribution, we are able to detect effects of job loss on health even if they apply only to certain parts of the health distribution. In order to estimate causal effects of job loss we focus on job loss due to plant closure.

We find evidence that job loss stretches the distribution of mental and physical health states to lower values. The estimated effects of job loss on the lower 7 deciles of the distribution of physical health states are negative and considerable in size, while the effects on the upper two deciles are virtually zero. This result can be interpreted such that adverse effects of job loss on physical health are more pronounced for people who were in poor physical health already before they lost their job and, that job loss does not deteriorate physical health of people who are in best physical condition. Although the results for the mental health are less distinct from a statistical point of view, they point in the same direction. In contrast, we find no or only very small medium term effects of job loss on BMI. Our findings highlight the benefit of using quantile regression to study health effects of job loss. In our application, all average treatment effects on the treated are fairly small which may lead one to conclude the absence of a causal effects. This picture changes when quantile treatment effects are estimated.

The main implication of our results is that policy makers should especially focus on people in bad health, if they are interested in reducing individual and social costs of job loss and unemployment. The results of Huber et al. (2011) show that reemployment improves health and that adverse effects of job loss and unemployment on health are in principle reversible. In addition target-group-specific efforts should be undertaken to dampen adverse health consequences of job loss and following periods of unemployment. This could be done, for example, within the case management of the federal employment agency. A pilot project for the promotion of employment with health-related orientation (”*Arbeitsförderung mit gesundheitsbezogener Ausrichtung*”), conducted in Brandenburg, might serve as a model (MASGF, 2008). The results of this project suggest that the health status of unemployed can be improved by incorporating brief psychological interventions and programmes targeted to promote social skills into the case management. Such programmes might prevent people who are in poor health and suddenly lose their jobs from experiencing a further impairment of health, an approach which, by the way, most likely increases their employment prospects.

As our results are based on job losses that happened because a plant closed down they cannot be extended directly to all reasons for job loss. For example, it is possible that people who quit their job do not suffer from job loss with respect to their mental and physical health at all, even if their health was already impaired before they lost their job. The health of people who lost their job because they were fired, on the other hand, might suffer even more as these people might reproach themselves for being fired.

Appendix

Estimation

The main estimator used in this work, a root-N consistent and asymptotically normal estimator QTE-estimator originally proposed by Firpo (2007), is given by (Frölich and Melly, 2013) $(\widehat{Q}^{\tau}_{Y^0|c}, \widehat{\Delta}^{\tau}_c) = \underset{a,b}{argmin} \frac{1}{n} \sum_{i=1}^{n} \rho_{\tau}(Y_i - a - bD_i)\widehat{W}_i$, where ρ_{τ}, a check function evaluated at the real number u, is: $\rho_{\tau}(u) = u(\tau - (u < 0))$. In this representation a gives us the τ quantile of the distribution of Y in a situation where no one looses their job and b gives the quantile treatment effect of job loss. A first step estimator of the weights $\widehat{W}_i$ is given by $\widehat{W}_i = \frac{D_i}{\widehat{\pi}(X_i)} + \frac{1 - D_i}{1 - \widehat{\pi}(X_i)}$. The propensity score $\pi(X) = Pr(D = 1|X = x)$ is estimated here by local logit regression, but one could also use other nonparametric estimators like local linear regression or series regression without affecting the asymptotic distribution of the estimator (Frölich and Melly, 2013).

To estimate the propensity score by local logit regression one has to select the optimal values of the smoothing parameters (window width λ and bandwidth h) of the kernel function. We follow Frölich and Melly (2010) and make use of a cross-validation procedure for the selection of the parameters. As we bootstrap the estimation results and since the mere cross validation procedure is computationally very demanding, we determine the optimal smoothing parameters only once for a 20% sample of our data set and use the same parameters for each bootstrap replication. We do not expect this approach to influence our results much, as the results are fairly robust with respect to the choice of λ and h. Table A2.4 reports estimated QTEs for the preferred specification when we use global instead of local smoothing to estimate the propensity score. They are very similar to the main resultsreported in column III of tables 2.1, 2.2 and 2.3.

Since we use several waves of the SOEP it is necessary to adjust the standard errors to the clustered structure of the data. To do so, all parameter estimates are bootstrapped and the bootstrap resampling is carried out by replacement over clusters (individuals).

Additional Tables

Table A2.1: SF-12v2 questionnaire in the SOEP

	Very Good	Good	Satis-factory	Poor	Bad
How would you describe your current health?					
	Greatly	Slightly	Not at all	–	–
When you ascend stairs, i.e. go up several floors on foot: Does your state of health affect you greatly, slightly or not at all?					
And what about having to cope with other tiring everyday tasks, i.e. where one has to lift something heavy or where one requires agility: Does your state of health affect you greatly, slightly or not at all?					
Please think about the last four weeks. How often did it occur within this period of time, . . .	Always	Often	Some-times	Almost never	Never
⋄ that you felt rushed or pressed for time? ⋄ that you felt run-down and melancholy? ⋄ that you felt relaxed and well-balanced? ⋄ that you used up a lot of energy? ⋄ that you had strong physical pains? ⋄ that due to physical health problems . . . you achieved less than you wanted to at work or in everyday tasks? . . . you were limited in some form at work or in everyday tasks? ⋄ that due to mental health or emotional problems . . . you achieved less than you wanted to at work or in everyday tasks? . . . you carried out your work or everyday tasks less thoroughly than usual? ⋄ that due to physical or mental health problems you were limited socially, i.e. in contact with friends, acquaintances or relatives?					

Note. Source: SOEP Individual question form. Available at http://panel.gsoep.de/soepinfo2008/.

Table A2.2: Deciles of Outcome Distributions

	MCS (at t)	PCS (at $t+2$)	BMI (at $t+2$)
0.1	37.299	39.644	21.2240
0.2	42.824	44.351	22.6420
0.3	46.493	47.690	23.7650
0.4	49.353	50.604	24.7550
0.5	51.710	53.039	25.6900
0.6	54.315	55.024	26.7070
0.7	56.232	56.594	27.7780
0.8	57.898	58.084	29.3880
0.9	60.417	60.023	31.8770

Note: Authors' calculations based on the SOEP. Numbers are based on the estimation sample.

Table A2.3: Descriptive Statistics by Treatment Status

	No Job Loss	Job Loss		Overall
		all reasons	plant closure	
Binary variables (values in %):				
No Job loss	100.00	0.00	0.00	82.90
Job Loss (All Reasons)	0.00	100.00	100.00	17.10
Job Loss (Plant Closure)	0.00	10.08	100.00	1.85
Male	55.46	48.96	57.53	54.35
German	94.30	92.48	89.63	93.99
Migration Background	16.93	19.91	22.19	17.44
Education (according to ISCED):				
In School	0.04	0.14	0.00	0.05
Inadequately	0.71	1.12	1.34	0.78
General Elementary	8.11	9.08	12.15	8.27
Middle Vocational	46.62	48.87	54.96	47.01
Vocational + Abitur	7.01	7.59	5.57	7.11
Higher Vocational	9.56	8.28	7.47	9.34
Higher Education	27.95	24.92	18.51	27.43
Having Been Unemployed before	29.54	44.17	41.58	32.05
Private Health Insurance	18.20	10.73	6.69	16.93
Living with Partner	79.29	73.26	78.37	78.26
Household Size:				
Household Size: 1	11.51	12.95	10.70	11.76
Household Size: 2	30.22	35.50	30.32	31.12
Household Size: 3	23.81	23.86	25.31	23.82
Household Size: 4 to 5	32.27	25.52	32.33	31.11
Household Size above 5	2.19	2.18	1.34	2.19
Year of Interview:				
2002	15.09	19.52	21.63	15.85
2004	16.21	17.08	22.85	16.36
2006	14.83	13.59	14.16	14.62

2008	13.63	11.54	10.81	13.28
2010	12.37	11.72	11.82	12.26
2012	10.42	8.78	6.13	10.14
2014	17.44	17.77	12.60	17.50
Federal State:				
Schleswig-Holstein	2.82	2.50	2.68	2.76
Hamburg	1.38	1.60	2.12	1.42
Lower Saxony	8.75	8.75	7.80	8.75
Bremen	0.68	0.99	1.56	0.73
North Rhine-Westfalia	20.65	19.40	23.08	20.44
Hesse	7.22	7.39	8.36	7.25
Rhineland-Palatinate	4.82	4.27	4.12	4.72
Baden-Wuerttemberg	12.43	12.01	10.03	12.36
Bavaria	15.23	13.83	11.15	14.99
Saarland	1.13	0.88	0.89	1.08
Berlin	3.55	4.12	4.35	3.65
Brandenburg	3.80	4.33	4.24	3.89
Mecklenbrg.-West Pomerania	2.24	2.87	3.46	2.35
Saxony	7.38	7.42	7.02	7.39
Saxony-Anhalt	3.92	4.81	4.12	4.08
Thuringia	4.00	4.82	5.02	4.14
Occupational Position:				
Training, Internship etc.	0.15	0.71	0.11	0.24
Unskilled/Semi-Skilled Worker	10.54	14.78	16.95	11.26
Skilled Worker	14.04	15.51	20.29	14.29
Master Craftsman	1.50	1.33	2.01	1.47
Self Employed	9.69	5.75	3.90	9.02
Clerk Low Skilled	10.30	14.19	15.83	10.96
Clerk Middle Skilled	26.31	25.56	23.97	26.18
Clerk High Skilled	18.19	17.04	15.05	17.99
Civil Servant	9.29	5.13	1.90	8.58
Firm Size:				
<20 Employees	23.31	30.75	35.34	24.58
20 - 200 Employees	27.39	29.45	33.11	27.74
200 - 2000 Employees	21.88	18.52	14.83	21.31
>2000 Employees	24.37	18.78	15.50	23.42
Self Employed - No Employees	3.05	2.50	1.23	2.96
Sector (according to NACE Rev1.1):				
Sectors A and B	1.51	1.61	0.56	1.53
Sectors C and E	1.44	1.16	0.67	1.39
Sector D	23.95	20.67	29.10	23.39
Sector F	5.68	8.55	13.27	6.17
Sector G	11.22	13.68	20.62	11.64
Sector H	1.68	3.59	4.12	2.01
Sector I	5.13	5.65	6.80	5.22
Sector J	4.83	3.49	3.12	4.60
Sector K	8.24	9.93	7.58	8.53
Sector L	9.81	5.48	1.90	9.07
Sector M	8.45	7.68	2.56	8.32
Sector N	12.04	12.31	4.24	12.08
Sector O	3.92	3.92	3.01	3.92
Sectors P and Q	2.10	2.27	2.45	2.13
Job Worries:				
Job Security: Very Concerned	12.25	21.53	33.00	13.84
Job Security: Somewhat Concerned	39.44	38.41	43.26	39.27

Job Security: Not Concerned At All	48.31	40.06	23.75	46.90
Continuous variables:				
Age	43.64	40.39	42.26	43.09
Working Experience	20.31	16.79	19.12	19.71
Net Household Income	3222.72	2831.04	2762.28	3155.76
Gross Labor Income	2849.21	2238.03	2225.93	2744.71
MCS	50.53	49.15	49.62	50.30
PCS	52.39	52.14	51.65	52.35
BMI	25.79	25.44	26.13	25.74

Note: Authors' calculations based on the SOEP. All variables are measured at $t-2$ and all but age, working experience, net household income, gross labor income, MCS, PCS, and BMI are binary.

Table A2.4: QTE for Job Loss due to Plant Closure (Propensity Score Estimated by Global Smoothing)

	MCS (at t)		PCS (at $t+2$)		BMI (at $t+2$)	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
QTE						
0.1	-1.459	0.953	-2.613*	1.546	-0.426	0.420
0.2	-1.069	0.908	-2.243	1.481	0.067	0.538
0.3	-1.213	0.934	-1.253	1.306	0.207	0.424
0.4	-0.944	0.827	-1.448	0.982	0.132	0.329
0.5	-0.588	0.874	-1.523**	0.739	0.249	0.419
0.6	-0.665	0.720	-1.320**	0.682	0.388	0.427
0.7	-0.583	0.790	-1.487	0.955	0.427	0.554
0.8	0.391	0.612	0.000	0.682	0.629	0.690
0.9	0.000	0.601	0.185	0.565	-0.070	0.965
Standard Controls	yes		yes		yes	
Pre-Treatment Outcomes	yes		yes		yes	
Obs	36,058		26,260		26,183	

Note: Authors' calculations based on the SOEP. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Bootstrapped standard errors based on 500 replications, clustered at individual level. Controls include: sex, age, education, German nationality, migration background, private health insurance, working experience, unemployment experience, job worries, living with partner, household size, household income, year, state, gross labor income, occupation, sector, firm size, MCS, PCS, and BMI (all measured at $t-2$).

Chapter 3

Understanding Cognitive Decline in Older Ages: The Role of Health Shocks*

3.1 Introduction

Age-related cognitive decline is among the key reasons for the transition of healthy individuals into care dependent ones. As an example, its most drastic form, dementia, is responsible for around 50% of all nursing home stays in Germany (Sütterlin et al., 2011).[1] And this is becoming an increasing challenge for societies. The number of people suffering from dementia is projected to triple to 152 million in the developed countries between 2018 and 2050 (Patterson, 2018). Estimated dementia costs are expected to increase from $1 trillion today to an estimated $2 trillion by 2030 (Patterson, 2018). Yet, also milder forms of cognitive impairment are among the risk factors of becoming care dependent.

Apart from care dependence, cognitive abilities are in itself of growing importance in an ever more complicated world, even more so in combination with ongoing population aging. They are an important determinant of social participation. Complex decisions involve those on medical treatments, insurance coverage, or financial markets for example (Mazzonna and Peracchi, 2018). Studies find that lower cognitive abilities lead to lower investments in stocks and other risky assets (Christelis et al., 2010) and lower wealth (Banks et al., 2010; Smith et al., 2010). Cognitive performance is important for labor productivity and, more general, for wellbeing (Engelhardt et al., 2010; Maurer,

---

[1]Strong cognitive impairments as predictors for dementia and care dependence on the individual level are also reported in (Celidoni et al., 2017; American Psychiatric Association, 2000).

*This chapter is joint work with Hendrik Schmitz and is published as a working paper: Schiele, V. and Schmitz, H. (2021). Understanding cognitive decline in older ages: The role of health shocks. Ruhr Economic Papers, 919, RWI.

2011). Moreover, individuals are often not aware of their cognitive decline, which in particular leads to bad economic and financial decisions (Mazzonna and Peracchi, 2018).

While economists have devoted a great deal of effort to understand the process of human capital accumulation (e.g. determinants of education and the effects of education on productivity), not so much is known about its depreciation. Repeating the notion of McFadden (2008): in the past decades "...economists have given less attention to the process of human capital depreciation, and technologies for human capital maintenance. Natural questions to ask are (...) the degree to which the depreciation of human capital components is an exogenous consequence of aging or can be controlled through work, study, and behavioral choices; and the degree to which depreciation is predictable or random." It is our goal to contribute here and to learn about triggers of strong cognitive decline and what policy makers can or cannot do about that.

Cognitive decline is a gradual process but there is a small economic literature suggesting that it may be accelerated by adverse life events such as the loss of beloved persons, economic shocks, or bad economic circumstances at birth (see e.g. Lindeboom et al., 2002, van den Berg et al., 2011 and Doblhammer et al., 2013). In this paper, we seek to understand how a specific adverse life event, a health shock, affects cognitive decline among individuals aged 50 and older. Anecdotal evidence tells about the hale and hearty old person that accidentally falls and where the resulting hip fracture marks the beginning of a care episode that goes along with a strongly accelerated reduction in cognitive functioning, ending in a nursing home. In a second step of the paper, we study how potential policy measures can moderate effects of these shocks. These are labor force participation (affected by retirement regimes) and education. That is, we ask: can education or labor force participation increase the cognitive reserve such that adverse health effects have a less drastic effect on cognitive decline? In a previous study, van den Berg et al. (2010) assess how economic conditions early in life moderate the effect of life-events on cognitive functioning. They find that individuals born under adverse early-life conditions (that is, in recessions) suffer from a stronger cognitive decline after a stroke than those born under beneficial early-life conditions. No such moderation effects are found for several other conditions, however, such as peripheral arterial disease, heart disease, diabetes, cancer, respiratory disease, or arthritis. Surgery or illness of a partner are only harmful for individuals born under adverse economic conditions.

We use data from the Survey of Health, Ageing, and Retirement (SHARE), the Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA). Measures of cognitive abilities are based on experimentally retrieved test scores for *recall* and *verbal fluency* while health shocks are experimentally and/or objectively measured by strong reductions in handgrip strength and onset of severe conditions such as heart attack, stroke, or hip fracture. We carry out event study estimations to see whether

health shocks are anticipated (in terms of cognitive decline) and to learn about the longer run effects up to eight years after the shock. It turns out that there are no significant differences in cognitive decline trends before the shock for those who suffer from a health shock but that there is a strong immediate and persistent drop in cognitive abilities upon the health shock. Comparing the effect size to the general age-related decline in cognitive functioning, a health shock, on average, induces a similar cognitive decline as growing up to four years older. Thus, health shocks also have the potential to bring a long-term care episode for mental health reasons forward by several years.

We then analyze whether results differ by demographic characteristics or between Europe and the USA and whether retirement or education moderate the effect of a health shock. We use reduced form regressions and early retirement ages as set by the governments as well as compulsory schooling regulations to circumvent problems of potential endogeneity of retirement and education. It turns out that, while there is a considerable effect of health shocks on cognitive abilities in all specifications, this effect does not seem to be mitigated by retirement behavior or education. The effect sizes also do not vary much between Continental Europe, England and the US, representing regions with different institutional settings including different health insurance systems.

We mainly contribute to the literature in the following ways. Apart from presenting the first study on this topic that combines micro data from several countries and two continents with a large sample size that potentially also allows to identify smaller effects, we expand the previous study by van den Berg et al. (2010)[2] and focus on potential moderating variables that, arguably, are more susceptible to policy action than early-life circumstances. Moreover, in contrast to previous studies on effects of health shocks, we show transparent event-study results.

The results of our analysis suggest that physical health shocks have a significant impact on cognitive ageing and thus on human capital depreciation. While they are consistent with previous studies that concluded that higher education and active ageing can slow cognitive decline, they do not suggest that a higher cognitive reserve also helps to dampen cognitive decline after a health shock. Thus, they underline the importance of health prevention and point out that investing in physical health could pay off twice: through its direct return on physical health, but also through its indirect return in terms of cognitive functioning in old age.

The structure of the paper is as follows. We present the data and descriptive statistics in Section 3.2. In Section 3.3 we lay out the empirical approach and present baseline results, while we show the moderation analysis in Section 3.4. Section 3.5 concludes.

---

[2]While there is some literature on the direct effect of retirement/education on cognitive abilities (e.g. Bonsang et al., 2012; Rohwedder and Willis, 2010; Celidoni et al., 2017; Mazzonna and Peracchi, 2012, 2017; Coe et al., 2012; Schneeweis et al., 2014), we are aware of only one study that analyses the effect of a health shock.

3.2 Data

3.2.1 Sample Selection

Our main data sources are the Survey of Health Ageing, and Retirement (SHARE), the Health and Retirement Study (HRS) and the English Longitudinal Study of Aging (ELSA), three large representative micro data sets providing information on health and a great deal of other socioeconomic characteristics for individuals aged 50 and older. HRS was initiated in 1992. By now, 14 interview waves are available, each covering information about 20,000 Americans. Its sister study ELSA is fielded biennially since 2002 and is now containing data from 8 interview waves. SHARE started in 2004 as a cross-national survey. Since then data of 8 interview waves have been released covering information about 140,000 individuals living in 28 European countries plus Israel. HRS, ELSA and SHARE are highly harmonized and can be used for pooled analyses.[3]

We use all waves containing at least one measure of cognitive ability, one health shock measure and all covariates from the baseline specification, i.e. from SHARE waves 1, 2, 4, 5, 6, 7 and 8[4] from the HRS waves 4 to 13[5] and from ELSA waves 1 to 8.[6] In addition to the biennial data sets from HRS and ELSA, we include information from the RAND HRS Data file[7] and Harmonized ELSA. We exclude individuals from the sample who were interviewed only once, as our empirical strategy requires measurements from two consecutive waves. We drop individuals younger than 50 and older than 90. Our final sample consists of 421,656 observations from 124,167 individuals living in 20 countries.[8]

Given that we have a sample of older individuals with a focus on health shocks, non-random panel attrition is an obvious issue. We refer to Celidoni et al. (2017) who use the SHARE in an analysis of the effect of retirement on cognitive abilities. They both test for potential problems of panel attrition and, later, also account for it by including an inverse Mills ratio. Yet, they find that non-random panel attrition does not seem to be a relevant issue and, not surprising then, the selection model with the inverse Mills ratio does not yield different findings. We also find that our long-run results do not seem to be affected by potential attrition problems (see below).

---

[3]For comprehensive information on the sampling procedure, questionnaire contents, and fieldwork methodology of HRS, ELSA, and SHARE see Sonnega et al. (2014), Steptoe et al. (2003), and Börsch-Supan and Jürges (2005).

[4]See Börsch-Supan (2019a,b,c,d,e,f, 2020, 2021).

[5]Health and Retirement Study (2016a,b,c,d, 2014a,b, 2017a,b,c, 2019)

[6]See Banks et al. (2019)

[7]The RAND HRS Data file is an easy to use longitudinal data set based on the HRS data. It was developed at RAND with funding from the National Institute on Aging and the Social Security Administration.

[8]Countries covered in our sample: Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Greece, Switzerland, Belgium, Israel, Czech Republic, Poland, Luxembourg, Slovenia, Estonia, Croatia, USA, England.

3.2.2 Measures of Cognitive Ability

Cognitive abilities summarize the "ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought" (Neisser et al., 1996), where the sum of these abilities is referred to as intelligence. SHARE, HRS and ELSA offer a number of potential measures for cognitive abilities: orientation in time, numeracy, verbal fluency and word recall tests. Some of the measures are not available in all waves and, thus, not suitable for our analysis.[9]

In the *word recall test*, the interviewer reads ten words and the interviewed is asked which of these words they can remember. The number of words they can recall is counted. This word recall test is done twice: directly after the words are read (immediate recall test) and about 5 minutes later (delayed recall test). The total number of words recalled in these two occasions is added up to yield the word recall test score. This score can range between 0 and 20. The average in our final sample is 9.616 with a standard deviation of 3.658. Word recall is a measure of episodic memory, which is found to react most strongly to ageing (Rohwedder and Willis, 2010). It is considered a measure of "fluid intelligence". Broadly speaking, fluid intelligence is the innate cognitive ability while crystallized intelligence is what people learn in their lifetime (using their fluid intelligence).

In the *verbal fluency test* respondents are asked to name as many animals as they can in one minute, where the number of animals they can tell is their test score. Here the lower limit is 0, but there is no upper limit (the maximum number in the sample is 100). The sample mean is 20.003, the standard deviation is 7.595. Verbal fluency is a measure of both fluid and crystallized intelligence as it is both important to know many animals (crystallized knowledge) and to remember them quickly (fluid intelligence). Obviously, both *recall* and *verbal fluency* only capture specific parts of the multidimensional concept "cognitive ability".

The general bivariate relationship between the two measures of cognitive abilities and age can be seen in Figure 3.1. Obviously, getting older goes along with a steady decline in cognitive abilities.

3.2.3 Measures of a Health Shock

In defining a health shock, we follow two approaches. Approach 1 is to use the onset of a serious illness between two waves. Survey respondents state whether they suffered from any of the following illnesses since the previous interview: heart attack, stroke,

---

[9]The description of measures of cognitive ability in this section closely follows Schmitz and Westphal (2021).

Figure 3.1: Measures of Cognitive Abilities by Age
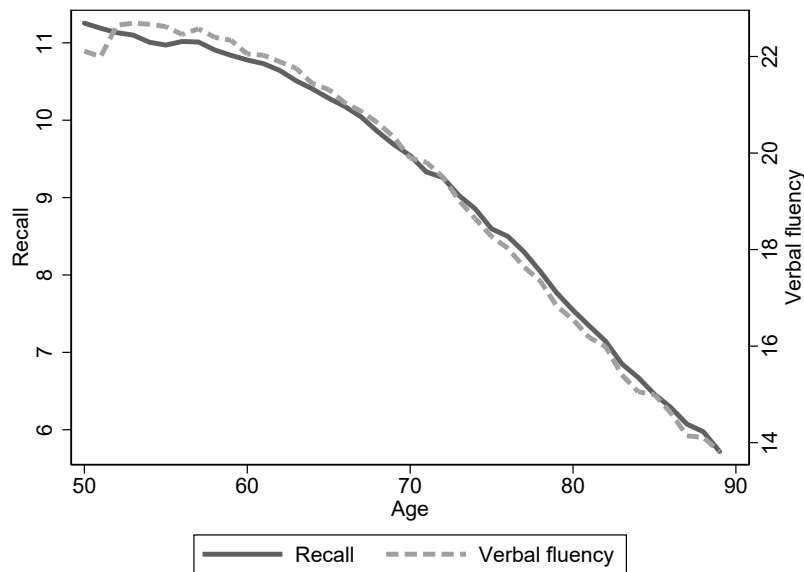
Notes: Based on data from SHARE, HRS and ELSA. The graph plots unconditional averages of the two scores by age in full years.

cancer (any kind), hip fracture.[10] We consider all of these as health shocks. While these illnesses are highly relevant and objective health measures, they probably are not free from measurement error because they are self-stated by the respondents. Onsets of these conditions are serious negative life events. Moreover, the trained interviewers compare answers to those in previous waves. Nevertheless, in approach 2, we rely on changes in hand grip strength as another objective measure.

Grip strength in the SHARE, HRS, and ELSA is measured by the regular interviewers. The interviewers are equipped with so-called dynamometers, receive instruction on the usage of these small, non-invasive devices and are then able to assess the grip strength of the survey respondents.[11] The actual measurement procedure is as follows: The interviewer illustrates the use of the dynamometer first and then asks the respondents to press it twice with each hand as hard as they can, starting with the right hand and alternating afterwards. Test trials are not allowed.

There exist several alternatives in the medical literature on how to summarize this information (Roberts et al., 2011). Ambrasat and Schupp (2011) and Ambrasat et al. (2011) have analysed the case of the German Socio-Economic Panel (SOEP) rigorously. They suggest using the maximal value from all available measurements as a measure for the grip strength of an individual as, due to the absence of test trials, some individuals might not exert their full grip strength in the first measurements. We follow this suggestion. We are mostly interested in *relative* changes in individual grip strength (*GS*)

---

[10]Respondents are asked about hip fractures in HRS and ELSA only if they are older than 65 years (60 years in ELSA). Since hip fractures are quite seldom according to the SHARE data below these ages, we assume that individuals below these age cut-offs did not suffer from hip fractures.

[11]The following text on grip strength is largely taken from Decker and Schmitz (2016).

over time. Specifically, we calculate $\Delta GS_t = (GS_t - GS_{t-1})/GS_{t-1}$, where t indicates the wave.

Grip strength contains more information than just the muscle strength of the hands. Based on broad empirical evidence from the medical literature, grip strength is known to be a valid indicator of the overall health status of an individual. Several medical studies document the association between a low level of grip strength and certain negative health outcomes, such as decreased overall muscular strength, the onset of chronic diseases, nutritional depletion, physical inactivity and mortality (see Rantanen et al. (2003), Bohannon (2008) or Ambrasat et al. (2011) and the references therein). The underlying mechanisms are not yet completely understood but it is suggested that "poor muscle strength could be a marker of disease severity, which in turn is associated with mortality" (Rantanen et al., 2003, p. 637).

Analogous evidence exists for extreme losses of grip strength over time (Rantanen et al., 1998; Ling et al., 2010; Xue Q et al., 2011; Stenholm et al., 2012). For example, Stenholm et al. (2012) study a sample of adults aged 30-70 at baseline and their grip strength changes over 22 years. The evidence suggests that the onset of chronic conditions such as coronary heart diseases, other cardiovascular diseases, diabetes or chronic bronchitis is correlated with accelerated grip strength decline over time. Similarly, Ling et al. (2010) compare 89-year-old Dutch males with different developments of grip strength over four years. They find that those with a decline in grip strength of 25% or more have a significantly higher mortality risk than those with a lower decrease or even an increase. It is our aim to incorporate this medical evidence when defining our health shock measure. A general cut-off point that identifies those with extreme losses in grip strength would be ideal but despite intense literature research we are unaware of such information. We therefore take the aforementioned value of a loss of 25% or more from Ling et al. (2010) for our main analysis and define our health shock measure as a loss of 25% or more in individual grip strength over two years, but check whether using alternative cut-off points affect the results. This definition of health shocks is also in accordance with Decker and Schmitz (2016) who use strong grip strength changes to analyze the effect of health shocks on risk aversion. While SHARE provides grip strength for each individual in every wave, this is only the case in every other wave in the HRS and in ELSA, except for a couple of individuals. Thus, when using the grip strength measure, we need to restrict our analysis to the SHARE countries.

Table 3.1 reports descriptive statistics on the health shock measures and socioeconomic controls. A health shock is a rare event, in particular the single conditions. In the main regression analysis below, we do not use the conditions separately but use the indicator "(Any) New condition", taking on the value 1 if the individual exhibits at least one of the four conditions between two waves. In some additional regressions we show results for the four conditions separately. Figure 3.2 plots the incidence of health shocks by age

Table 3.1: Descriptive Statistics on Health Shocks and Socioeconomic Controls

Variable	Mean	Std. dev.	Min	Max
Heart attack	0.017	0.129	0.000	1.000
Stroke	0.015	0.122	0.000	1.000
Cancer	0.026	0.158	0.000	1.000
Hip fracture	0.007	0.082	0.000	1.000
New condition	0.061	0.240	0.000	1.000
Grip strength shock	0.068	0.252	0.000	1.000
Male	0.427	0.495	0.000	1.000
Age	67.510	9.305	50.000	89.000

Notes: Descriptive statistics on condition based on 384,692 observations from SHARE, HRS and ELSA, descriptive statistics for Grip strength shock based on 160,922 observations (SHARE countries only) and descriptive statistics for age and gender based on 421,656 observations.

in the sample. The average unconditional probability of a health shock monotonously increases from about 5 per cent at the age of 50 to about 10 per cent at the age of 90.

Figure 3.2: Health Shocks by Age

Notes: Based on data from SHARE, HRS and ELSA. The graph plots unconditional relative frequencies of the two health shocks by age in full years. The grip strength measure is only available in the SHARE data.

In Table 3.2 we take a look at the dynamics of health shocks. We show recovery rates by time distance to the health shock for those who experienced one. The results indicate that among those experiencing a health shock, defined as a 25% decline in grip strength between two waves $t-1$ and t, around 40% reached their initial grip strength by $t+1$, that is, two years later (Δ grip strength $\geq$ 25%). Another 20% seem to have recovered at least partly (25%> Δ grip strength $\geq$ 10%) with respect to their physical health in the same period. Only around 40% experienced a stagnation (10%> Δ grip strength $\geq$ -10%) or worsening (-10%> Δ grip strength). Even though the proportion of those who

(partly) recover decreases somewhat the more time has passed since the health shock, the overall impression that a large share recovers physically from the shock does not change when looking at time lags larger than two years from the onset of the shock.

Table 3.2: Recovery of a Health Shock

	$t+1$		$t+2$		$t+3$		$t+4$	
	N	%	N	%	N	%	N	%
Δ grip strength since t								
≥ 25%	1142	38.74	504	0.34	180	0.34	117	0.27
+10 to +25%	626	21.23	309	0.21	109	0.21	72	0.17
-10 to +10%	766	25.98	407	0.28	125	0.24	125	0.29
<-10%	414	14.04	260	0.18	108	0.21	122	0.28
Total	2948		1480		522		436	

Notes: Based on SHARE data. The sample for this table only includes individuals who experienced a health shock between two waves. These individuals are followed over time. $t+1$ indicates, for instance, one wave after the health shock (which occurred between $t-1$ and t). Percentage shares in each columns add up to 100.

3.3 Effects of Health Shocks on Cognitive Abilities

3.3.1 Baseline Results

Before looking at potential moderators, we are interested in the baseline effect of a (physical) health shock on cognitive abilities. Obviously, individuals who experience a health shock at some point in time might differ in many aspects including their level or development of cognitive abilities from individuals who do not suffer from such a shock. One way to deal with this potential endogeneity of health shocks, has been to condition on a large set of controls including pre-treatment outcomes (e.g. van den Berg et al., 2010). We estimate the effects of health shocks on cognitive abilities in a event study design instead.[12] This approach has the advantage that it does not only allow for fixed effects, but can also be used to assess whether individuals with health shocks follow different trajectories with respect to their cognitive abilities already before the shock. Furthermore, compared to the classical Diff-in-Diff model, the event study design does not require the assumption of static effects of a health shocks but visualizes potential dynamics of the effects directly.[13]

---

[12]In the supplementary materials we report results of specifications in the spirit of (van den Berg et al., 2010).

[13]This flexible approach already account for most of the problems outlined in, e.g., de Chaisemartin and D'Haultfoeuille (2020) and Goodman-Bacon (2021) of "Diff-in-Diff with staggered entry". However, we also use the approach of Sun and Abraham (2021) as a robustness check below.

As a first step we define the event time r which is survey wave relative to the wave the health shock occurred

$$r_{it} = t - h_i$$

where t is the wave, i is the individual and h_i denotes the wave, a health shock is observed first. That is $r = -1$ is the last wave before the health shock, while $r = 0$ is the first wave after the health shock. We do not observe the exact date of the health shock. Due to the biannual nature of the data, on average, the health shock should have occurred one year before $r = 0$. We run the following fixed effects regression with relative time indicators:

$$Y_{it} = \sum_{r=-3}^{r=-2} \mu_r + \sum_{0}^{3} \mu_r + \mu_a + \mu_b + \alpha_i + \lambda_t + \tau_{it} + \varepsilon_{it} \tag{3.1}$$

Here, Y_{it} is a measure of cognitive abilities, either recall or verbal fluency, and μ_r is the coefficient of the indicator variable of event time r. We leave observations with event times smaller than -3 or larger than 3 in the sample but account for this by μ_a (the coefficient of the indicator for $r < -3$) and μ_b (the coefficient of the indicator for $r > 3$. Furthermore, α_i denotes individual, λ_t time and τ_{it} a full set of age fixed effects and ε_{it} is the error term. In this setting, the dynamic effects of a health shock on cognitive abilities given by μ_r (for $r \geq 0$) are identified solely by different timings of the health shock.

Figure 3.3 shows the results for both health shocks and both outcome measures. The upper two use recall, the lower two verbal fluency as outcomes. All figures are remarkably similar and show a significant and immediate decline in cognitive ability after a physical health shock. This decline in cognitive ability amounts to approximately 0.25 fewer words in the memory test and approximately 0.6-1 fewer animals/fruits, etc. named in the verbal fluency test, corresponding to a reduction in test scores of around 2.5-5 percent, as measured against the respective mean, or around 10 per cent of a standard deviation.

All specifications indicate that the effect of a health shock is also persistent and thus that people do not recover, at least in the mid-run, from its adverse consequences for cognition. While the estimates for the *any condition* health shock measure and both outcomes are precisely estimated and unambiguously point to such persistent effects even more than 6 years after the shock, the estimates for the *grip strength* measure are somewhat less conclusive. The point estimate for grip strength and the mid-run effect ($r = 3$) on recall is, in absolute terms, smaller than the estimates for the short-run. At

Figure 3.3: Impact of Health Shocks on Cognitive Abilities
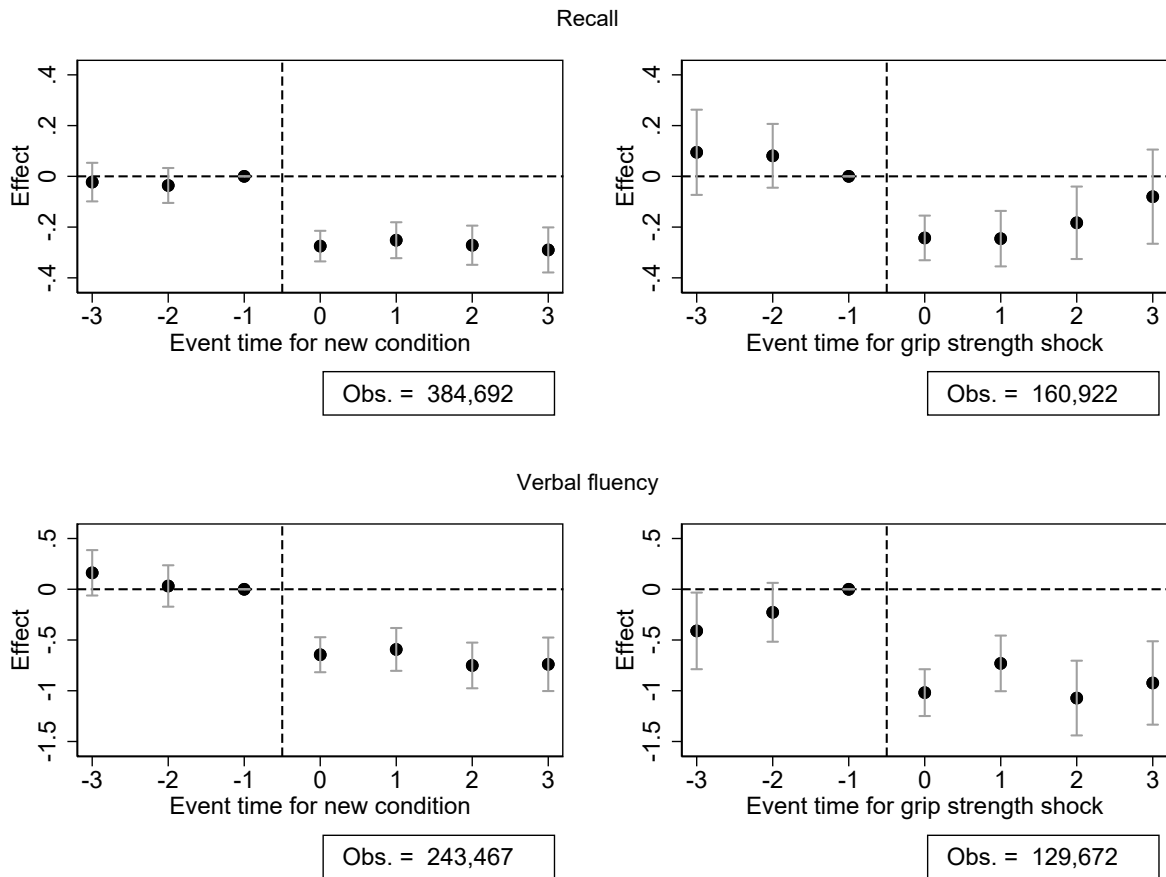
Notes: Coefficients μ_r from estimations of Equation 3.1 based on data from SHARE (Grip strength shock) and SHARE, HRS and ELSA (New condition), respectively. μ_{-1} is restricted to zero. 95% confidence intervals reported. Standard errors clustered on individual level.

the same time, the confidence bands are large[14] and still include the estimates for the short-run effects. Thus, we do not interpret this single result as evidence against the overall picture, which is a persistent effect of health shocks on cognitive abilities.

The results also contradict the concern that individuals who experience a health shock are on a path of declining cognitive abilities already before the health shock. When looking at the estimates for the pre-treatment periods, no structural pre-health-shock pattern can be identified, neither in terms of statistical nor economic significance.

To get a better impression of effect sizes, Figure 3.4 compares the short- and medium-run effects of a health shock with the general age decline in cognitive abilities. The figure compares two individuals with average characteristics in the sample, where only age and the event of a health shock are varied. Technically, these are predicted values from the baseline regression with all control variables except for age and health shock set to average values. We then add a hypothetical health shock at the age of 68 by using the coefficients from the event study estimates. We find that a health shock compares

---

[14]This is not surprising as the sample size is clearly reduced for the *grip strength* measure due to the rather short panel length in SHARE.

to a general age decline over around 1 to 4 years, depending on the health shock and outcome measure used. For instance, the average individual who experiences a health shock (new condition, left panel) at the age of 68 has the cognitive abilities of an average individual at the age of 72-73 without a health shock when he turns 70. To the amount that strong cognitive declines lead to care dependence, a health shock brings a care episode forward by up to 4 years.[15]

Figure 3.4: Stylized Medium-run Effect of a Health Shock
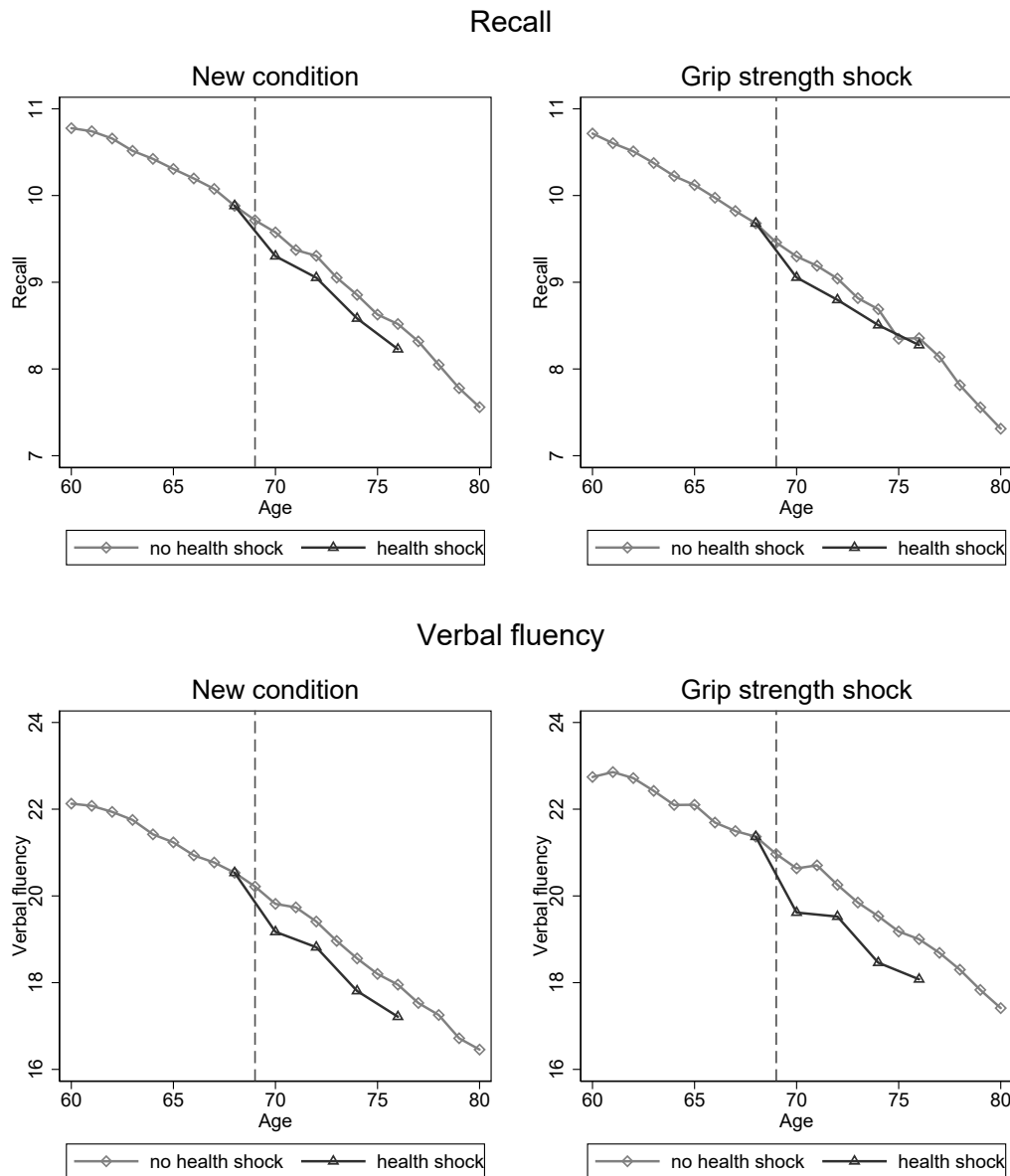
Notes: Based on SHARE, HRS and ELSA data. The graph plots the predicted cognitive decline for persons with average characteristics.

To test whether there are differences by region, we repeat the analysis separately for Continental Europe, the USA and the UK. We do this for the health shock measure

---

[15]See Figure 3.6 below that shows that the average effect of a health shock between 60 and 70, does not differ strongly from a health shock between 70 and 90. Thus, it does not make a difference at what age we start our hypothetical health shock in Figure 3.4.

"New condition" and the outcome "recall" as the figure for grip strength already reports findings for Continental Europe only and verbal fluency is available only in a few recent waves in the HRS. Figure 3.5 reports the results. There are only slight differences across regions. In all regions we observe an immediate and persistent drop in cognitive abilities after the health shock. This drop seems to be a little larger in the USA then in Europe, however, the difference is not statistically significant for most event time indicators. This might shed light on broader effects of institutional differences. For instance, most European countries have universal health insurance systems that might buffer the effects of health shocks while health insurance is more restricted in the US. While this obviously cannot be interpreted as evidence that health insurance systems do not matter at all, it at least indicates that existing institutional differences between the USA, UK and other European countries are no decisive factors.

Figure 3.5: Impact of Health Shocks on Cognitive Abilities (Recall) - by Region
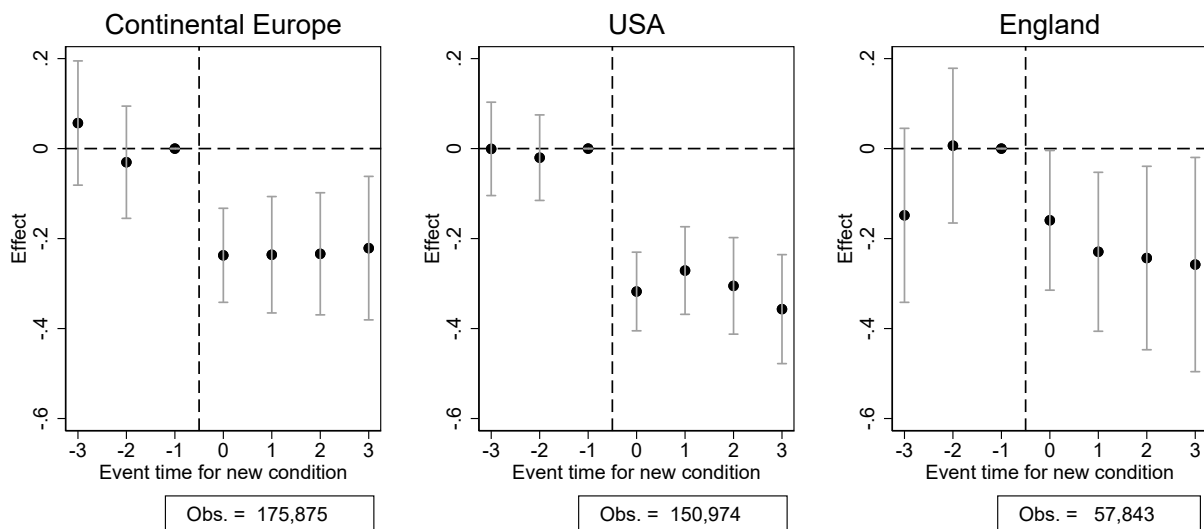
Notes: Coefficients μ_r from estimations of Equation 3.1 based on data from SHARE, HRS and ELSA. μ_{-1} is restricted to zero. 95% confidence intervals reported.

We so far yield three conclusions from this exercise. First, the clear pattern of no pre-treatment trends combined with an immediate drop in cognitive abilities after the health shock makes us confident that potential endogeneity might not be the driver of the observed relationship. Of course, this cannot be proven as we do not clearly observe what happens between $r = -1$ and $r = 0$. Nevertheless, it does not seem to be the case that individuals are on a general path of declining cognitive abilities before experiencing a health shock. Second, the effects are still visible in the longer run, several years after the health shock. Specifically, physical health shocks seem to induce a persistent downward shift in cognitive abilities immediately after the shock. That is, the effects of health shocks seem to be rather static than dynamic. Third, there is no heterogeneity across regions (USA, England, Continental Europe).

3.3.2 Further Results

Before proceeding with some robustness checks and the moderation analysis, we check for more general differences in the effect of health shocks on cognitive abilities. To overcome the above-mentioned problems with reduced power, which are partly due to the specific data structure (grip strength only available in SHARE, verbal fluency only available in some waves in ELSA and HRS) and which come into effect especially in subgroup analyses, we collapse the event-time indicators into a single (post)treatment indicator in the following, i.e., we run classical difference-in-differences regressions of the form:

$$Y_{it} = \alpha_i + \lambda_t + \tau_{it} + \beta post_{it} + \varepsilon_{it} \tag{3.2}$$

where $post_{it}$ is 1 for individuals with a health shock in periods after the health shock, i.e. for $r \geq 0$, and 0 otherwise. This approach does not only reduce noise in the estimates, but also allows for a more comprehensive presentation of the results of heterogeneity analysis and robustness checks. Note that using a classical difference-in-differences approach instead of the more flexible event study, should not alter the validity of the analysis, given that there is little evidence for dynamic effects in the pooled sample.

Figure 3.6 shows the results when estimating Equation 3.2 by subgroups defined by region, gender, age groups and quartiles of cognitive abilities. To start with, we do not find evidence for a gender difference, irrespective of the health shock and outcome measure used.

We do not find an age gradient or gradient in initial cognitive abilities (cognitive abilities at first appearance in the sample) either. If any, the last segment of Figure 3.6 suggests, that the loss in cognitive abilities in the aftermath of a health shock is larger for individuals with a higher cognitive reserve. Yet, as this result seems to hold only for some health shock/outcome measures, it might be that the role of initial cognitive abilities depends on the type of health shock.

3.3.3 Robustness Checks

Recent literature on the estimation of dynamic treatment effects shows that the event-time coefficients derived from two-way fixed effects estimation in settings comparable to ours, i.e. settings with variation in treatment timing and thus potentially effect heterogeneity, can be contaminated by the effects of other periods (Sun and Abraham, 2021). This might cause misleading interpretations not only of the dynamics of the effects, but also with respect to the identifying assumptions. To deal with this problem

Figure 3.6: Effect Heterogeneity
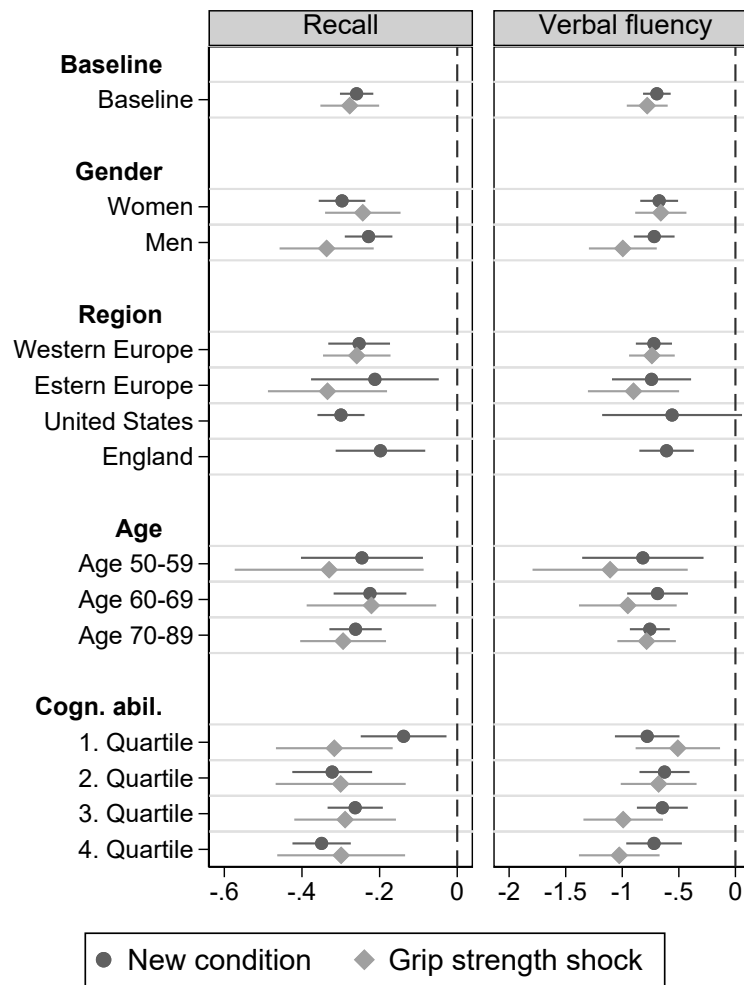
Notes: The figure shows the coefficient β along with 95% confidence intervals from estimations of Equation 3.2 based on data from SHARE, HRS and ELSA. Each estimate comes from a separate regression.

Sun and Abraham (2021) propose an alternative weighting estimator which is robust also in the presence of effect heterogeneity.

In order to address this concern, Figure 3.7 shows the estimated effects of a health shock corresponding to μ_r in Equation 3.1 when we apply their estimator to our study using the Stata package *eventstudyweights* (Sun, 2020). The estimated effects for "New condition" and recall almost exactly match our baseline results. The results for the alternative outcome "Verbal fluency" are smaller in size than the corresponding baseline results and estimated less precisely, but nevertheless point to adverse persistent effects of health shocks on cognition. Note, that due to the specific data structure in SHARE (wave 3 and wave 7 were no regular interviews but focused on initial living conditions) and the necessity to observe information on grip strength in two consecutive waves it was not feasible to apply the estimator to study the effects of a grip strength shock.[16]

---

[16]The lack of information on grip strength in wave 3 and (partially) wave 7, combined with the requirement to observe grip strength in at least two consecutive waves to define a health shock, means

Taken together, the results do not provide evidence against our interpretation of the baseline results.

Figure 3.7: Event Study Results using the Sun and Abraham (2020) Estimator
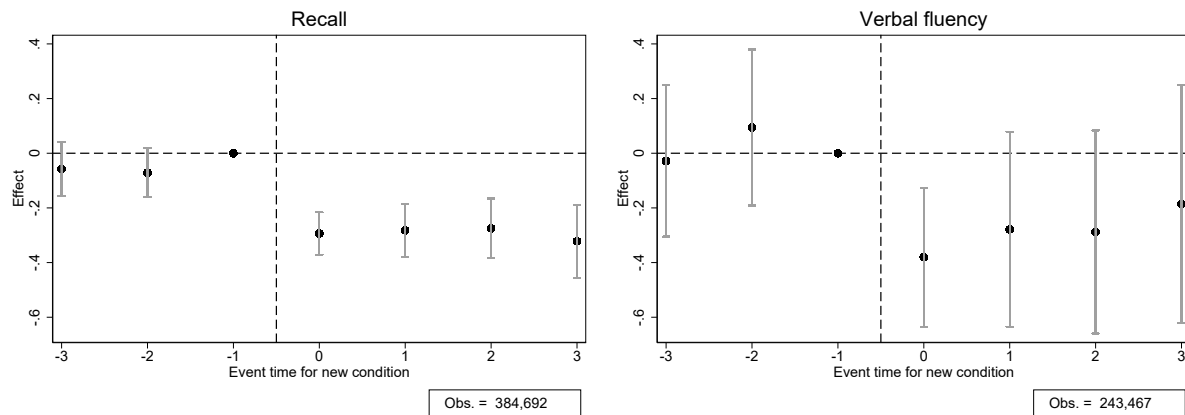
Notes: Coefficients corresponding to μ_r in Equation 3.1 based on data from SHARE, HRS and ELSA. μ_{-1} is restricted to zero. 95% confidence intervals reported. Standard errors clustered on individual level.

A second concern relates to the definitions of health shocks in our analysis. So far, we have combined different health conditions into one health shock measure and set the threshold for grip strength shock rather arbitrarily. In what follows, we examine whether the exact definition of health shocks is crucial for our results and whether panel attrition might bias our results. As each condition is a rather rare event we again pool over all post- (and pre-) treatment periods in the following and estimate a single effect of health shocks according to Equation 3.2.

Figure 3.8 reports the result of the difference-in-differences estimation for the entire sample and both health shock and outcome measures in the first segment (first line). These estimates mirror the baseline event study estimates presented in Figure 3.3. In the following segments we address potential concerns with respect to the definition of health shocks and panel attrition.

First, we report the effects of the onset of single conditions in the second segment of Figure 3.8. As mentioned earlier, the onset of one of these conditions, either a heart attack, a stroke, cancer or a hip fracture, is a rather rare event. Thus, we have used an aggregate measure (any of these four conditions new) in the event study, to make sure that we have enough power to detect effects of health shocks, especially when estimating medium term effects. Now, we try to see how each single condition contributes to the overall effect. One potential concern, for example, might be that the overall effect is solely driven by the onset of strokes, which, as might be argued, have a rather mechanic adverse effect on cognitive abilities. Indeed, the largest estimates can be found for strokes. The estimated effects for hip fractures are somewhat smaller but still relatively

---

that the event time coefficients cannot be estimated for all cohorts that have a weight greater than zero because of perfect collinearity.

Figure 3.8: Alternative Health Shock Definitions
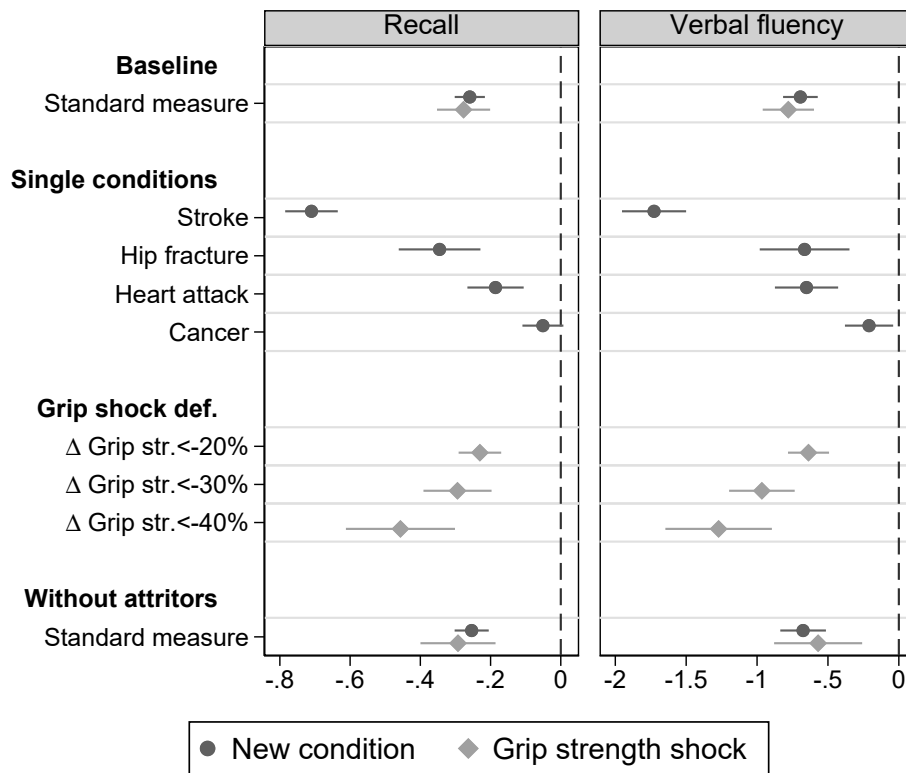
large and clearly significant. Also the estimates for heart attacks are in the range of the baseline estimates. Only for new cancer diagnosis we find no clear evidence for adverse effects on cognitive decline. An explanation for the latter finding might be that cancer is a disease progressing rather gradually while other conditions have a clear onset and require immediate treatment. While cancer is doubtlessly a serious health issue, it does not necessarily imply longer hospital stays at the onset of the condition but rather at the end. Thus, one might not expect to see large effects of newly detected cancer on cognitive abilities. Irrespective of the exact reason for the finding of zero effects of cancer, the results presented in here indicate that health shocks requiring immediate medical treatment affect cognitive abilities, even if the shock does not directly affect the functionality of the brain.

Another potential problem concerns the threshold that defines a health shock based on grip strength loss. Since there is no natural cutoff point, we followed the literature (see section 3.2.3) and defined a health shock as a reduction in grip strength of at least 25 percent. Segment 3 of the figure shows that the exact choice of cutoff point is not critical to the overall picture. Although we observe an increase in effect sizes as we decrease the cutoff point, it seems to make little difference whether we use -20 or -30 percent as the cutoff point instead of -25. If a cutoff of -40 percent is used, the effect size increases more notably, but the precision of the estimate also decreases.

Finally, one might worry that panel attrition affects the results. Although related previous research using SHARE data did not find evidence for influential non-random panel attrition (see Section 3.2), one can argue that this result cannot be transferred easily to our setting, because sample composition, variable definitions as well as the empirical approach differ. To address this concern, we repeat the analysis only looking at individuals who are at least 4 waves in the sample. Thus, we exclude those who drop out of the panel, possibly after a health shock. The results shown in the last row of the figure are very similar to the results for the full sample and indicate that attrition has no relevant effect on the estimates.

3.4 Moderation Analysis

The results presented so far provide evidence for adverse effects of health shocks on cognitive abilities in older ages. They imply that health shocks accelerate cognitive decline and thus might increase the risk of early care dependence. This raises the question whether there are factors that are susceptible to policy action and are able to improve health capital or increase the cognitive reserve and, thus, make people less prone to cognitive impairments following health shocks.

To shed light on this question we interact the event time dummies in our final specification with potential moderators (mod_{it}) which are either education or retirement:

$$
\begin{aligned}
Y_{it} =\ & \sum_{r=-3}^{r=-2} \mu_r + \sum_{0}^{3} \mu_r + \mu_a + \mu_b \\
& + \sum_{r=-3}^{r=-2} \mu_r \times mod_{it} + \sum_{0}^{3} \mu_r \times mod_{it} + \mu_a \times mod_{it} + \mu_b \times mod_{it} \\
& + mod_{it} + \alpha_i + \lambda_t + \tau_{it} + \varepsilon_{it}
\end{aligned}
\tag{3.3}
$$

Both retirement and education have been shown to affect cognitive abilities (see e.g. Bonsang et al., 2012; Rohwedder and Willis, 2010, for retirement and Kamhöfer et al., 2019; Schneeweis et al., 2014, for education). According to the "use-it-or-lose-it hypothesis" cognitive abilities decline faster if individuals do not use their cognitive capacities. We hypothesize that retirement as well as education might not only directly affect cognitive decline but that people who are cognitively more stimulated, are less prone to cognitive impairments following health shocks.

Equation (3.4) shows that a straightforward approach to test this presumption within an event study design is to regress cognitive abilities on the event time dummies, the moderator and all interactions of both along with controls. This approach obviously

raises endogeneity concerns of both the retirement decision as well as educational attainment. We thus focus on *retirement eligibility* instead of actual retirement status and on *compulsory schooling* instead of educational attainment, i.e. we present the reduced form relationships with variables that may be considered instrumental variables. Specifically, *mod* in Equation (3.4) is either a dummy variable for being above the (early) retirement age set by the retirement system or a dummy for being affected by a compulsory schooling reform that increased years of compulsory schooling. We follow the related literature and include – aside from individual, year and age fixed effects – linear country-specific trends to account for correlated changes in cognitive abilities and retirement regulations/compulsory schooling reforms across age groups/birth cohorts.

Obviously, the resulting parameter estimates then represent intention to treat (ITT) parameters rather than structural estimates. By focusing on the reduced form, we circumvent the problem that the interpretation of estimates in an IV model with multiple endogenous variables is not completely clear in the case of heterogeneous effects. Furthermore, the reduced form estimates are probably sufficient to assess whether there is an effect also in the structural model. In this sense, Angrist and Krueger (1991) note that if one cannot see an effect in the reduced form, then, most likely, it does not exist.

To estimate the moderating effects of retirement, we make use of arguably exogenous variation in early retirement regulations. This is a frequently employed instrument in the literature, see, e.g. Mazzonna and Peracchi (2012), Celidoni et al. (2017), Mazzonna and Peracchi (2017). We follow Schmitz and Westphal (2021) and focus on *early retirement* ages only instead of also using *official retirement* ages. The reason is that the jump in retirement probability at the early retirement threshold is much larger, leading to a sufficiently strong instrument only for early retirement. Early retirement appears to be the more important institutional feature as it allows individuals to retire (at the cost of penalties on retirement benefits) while official retirement age only is the age threshold that abandons penalties on retirement benefits. Our indicator for early retirement age takes on the value one if the individual has reached the early retirement age and zero otherwise. As can be seen in the first column of Table 3.3 (*ERA*), there is considerable within- and across-country variation in early retirement ages. Within variation is due to reforms in the observation period.

When looking at education as a second potential moderator of health shocks, we make use of a binary variable that captures whether an individual went to school before or after a reform that raised years of compulsory education went into effect. Note that although we can estimate the parameter of the interaction between the reform and the event time dummies, we can not estimate the direct effect of the reforms on cognitive abilities, as the value of the reform dummy is fixed for each individual and we include individual fixed effects in all specifications. The last column of Table 3.3 (*compulsory schooling*) gives an overview of the changes in years of compulsory

Table 3.3: Retirement Ages and Compulsory Schooling

	ERA		Compulsory schooling	
	men	women	change in years	pivotal cohort
Austria	60-65	55-60	8-9	1951
Belgium	58-60	58-60		
Flanders			8-9	1939
Czech Republic	-	-	8-9	1947
Denmark	60	60	4-7	1947
England	65-66	60-66	10-11	1957
France	60	60	8-10	1953
Germany	63	62-63		
BW			8-9	1953
BY			8-9	1955
HB			8-9	1943
HH			8-9	1934
HE			8-9	1953
NI			8-9	1947
NRW			8-9	1953
RLP			8-9	1953
SL			8-9	1949
SH			8-9	1941
Greece	-	-	6-9	1963
Italy	57-58	57-58	5-8	1949
Netherlands	62	62	7-9	1936
Spain	61	61	6-8	1957
Sweden	61	61	7-9	1950
Switzerland	63	62		
USA	62	62		

Notes: The table shows for each country and gender the Early Retirement Age (ERA) and for each compulsory schooling reform the change in years of compulsory schooling as well as the first cohort affected by the reform. As ERA depends on e.g. the birth cohort in some countries, we provide the ERA range in our sample for these countries. Information about the compulsory schooling reforms in most countries is taken from Brunello et al. (2016). Additional information about the reforms in Spain, Greece and England is taken from Brunello et al. (2013). Detailed information on retirement rules for each country are in the appendix.

education for each reform and states the first birth cohort (*pivotal cohort*) affected by the reform.

For our approach to yield meaningful estimates, retirement age regulations have to actually affect the retirement decisions and compulsory schooling reforms must have an effect on educational attainment. Panel a) of Figure 3.9 shows that there is a clear jump in retirement rates as soon as people reach the early retirement age. With respect

to education, panel b) of Figure 3.9 shows an increase in average years of education[17] for the first cohorts affected by the compulsory schooling reform. When estimating the first stage regressions for retirement and education, the resulting estimates suggests that crossing the early retirement age increases the likelihood to retire by around 8.55 percentage points ($\widehat{\beta} = 0.0855$ with $s.e. = 0.0055$) and that the average compulsory schooling reform increased years of education by 0.48 years ($\widehat{\beta} = 0.4757$ with $s.e. = 0.0866$) on average.

Figure 3.9: First Stage Relationships
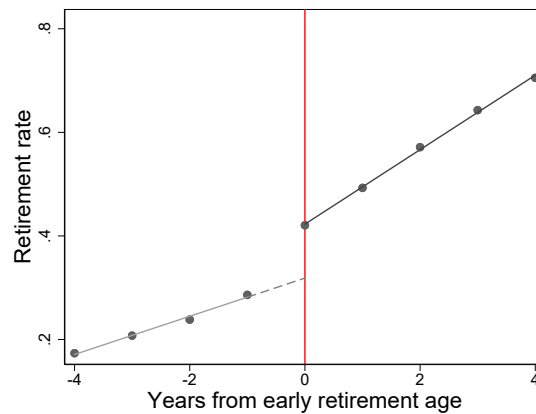
(a) Retirement rates by distance to early retirement ages
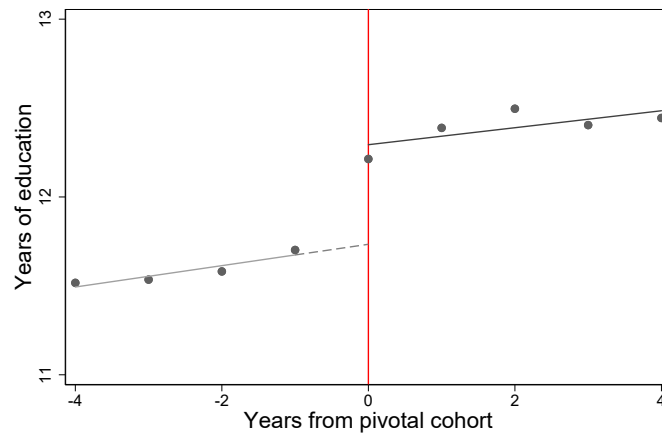
(b) Education by distance to compulsory schooling reforms

Notes: Panel a) based on data from SHARE, HRS and ELSA, panel b) based on SHARE data. Panel a) plots retirement rates by years from legislated early retirement ages. Panel b) shows average years of education by distance to the compulsory schooling reform.

With respect to exogeneity of the instruments, we borrow from the related literature and argue that they are based on legislated rules and are unrelated to individual characteristics (except for gender, country, age, cohort and year of interview which are flexibly controlled for in all regressions). A potential concern, however, is that

---

[17]We rely on self reported years of full time education here, which is available only for SHARE countries. For individuals with less years of education than years of compulsory education, we set years of education to years of compulsory education.

retirement or education affects the likelihood of having a health shock. Table 3.4 shows that auxiliary regressions[18] of the health shock variables on the instruments do not provide much evidence for this concern. All estimates are rather small and far away from being statistically significantly different from zero.

Table 3.4: The Effects of Retirement and Education on Health Shocks

	New condition		Grip strength shock	
Above early retirement age (in $t-1$)	0.004	(0.003)	0.005	(0.006)
Observations	306,108		126,410	
Affected by comp. schooling reform	-0.000	(0.003)	0.002	(0.003)
Observations	114,532		100,694	

Notes: The results for *New condition* are based on data from SHARE, ELSA and HRS (only for retirement), the results for *Grip strength shock* are based on SHARE data. Standard errors (clustered at individual level) in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Controls include: year and age fixed effects as well as country-specific linear age/cohort trends; Additional controls in retirement equation: individual fixed effects; Additional controls in education equation: gender and cohort fixed effects.

Figures 3.10 and 3.11 present the results of the moderation analysis. Both graphs show the estimated effects of a health shock on cognitive abilities differentiated by moderator/instrument status. In both graphs the grey markers show the effects when the instrument is switched on (i.e. for individuals above the early retirement age or individuals who were affected by a compulsory schooling reform that increased minimum years of schooling) and the black markers show the corresponding effects when the instrument is switched off.

To start with, Figure 3.10 shows the dynamic effects of health shocks on recall (panel (a)) and verbal fluency (panel (b)) by retirement eligibility. Some of the estimates, mainly those for the grip strength measure and the long run effects but also those for individuals below the early retirement age, are somewaht noisy and thus have to be interpreted with caution. Nevertheless, the overall impression is rather robust: Health shocks negatively and persistently affect cognitive abilities among both individuals who are eligible due to their age as well as individuals (of the same age) who are not yet allowed to retire. Thus, other than expected, retirement eligibility neither seems to amplify nor to dampen the adverse consequences of health shocks.

Figure 3.11 repeats the exercise for education and interacts the event time with an indicator for being affected by a compulsory schooling reform which raised years of compulsory education. It suggests that education likewise does not seem to moderate the effects of a health shock on cognitive abilities.

---

[18]All regressions include age and year fixed effects. The regressions for retirement additionally include individual fixed effects and country-specific linear age trends, the regression for education country fixed effects, a gender dummy and country-specific linear cohort trends, instead.

Figure 3.10: Moderation Analysis – Retirement
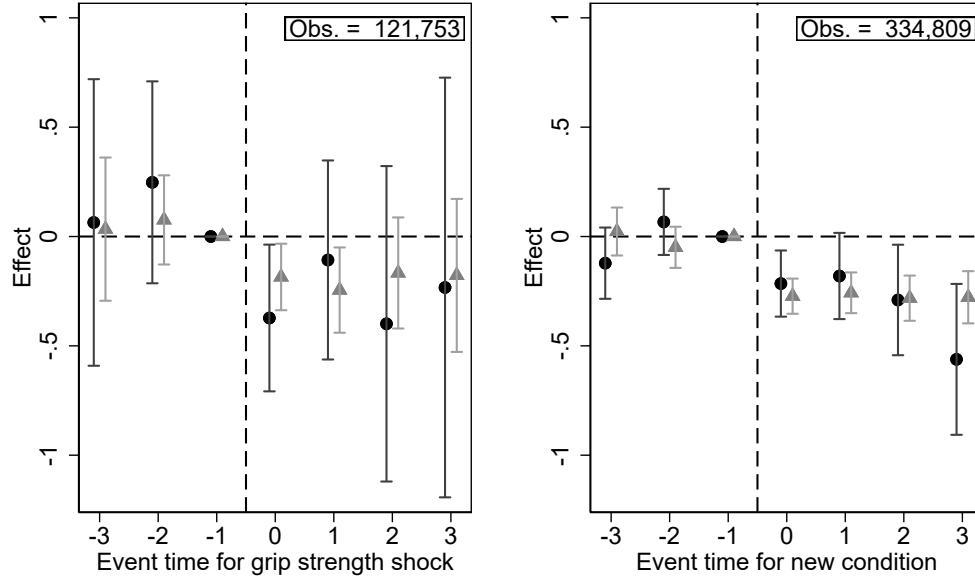
(a) Recall
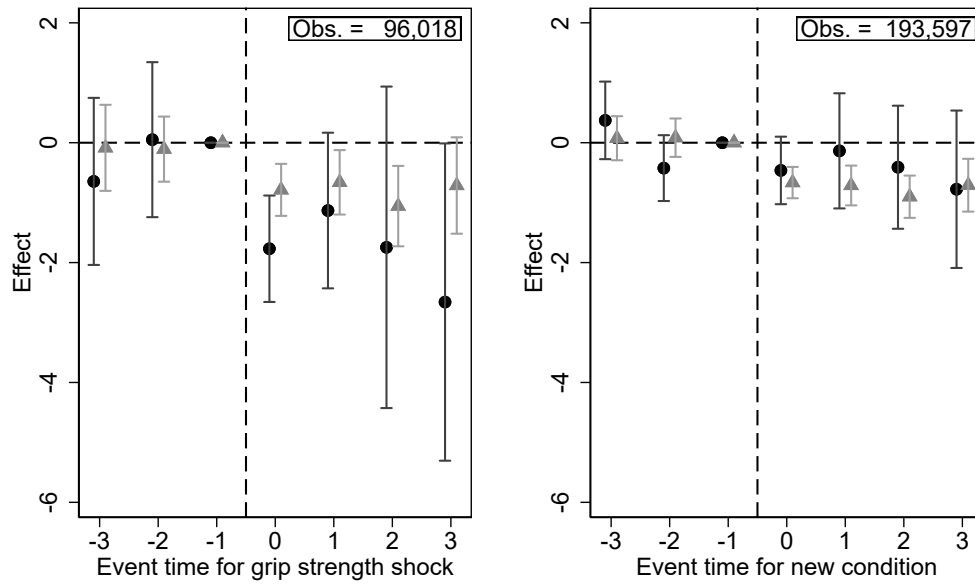
(b) Verbal fluency

● below early retirement age ▲ above early retirement age

Notes: Coefficients $\mu_r + \mu_r \times mod_{it}$ from estimations of Equation 3.4 based on data from SHARE, HRS and ELSA. $\mu_{-1} + \mu_{-1} \times mod_{it}$ is restricted to zero. 95% confidence intervals reported.

Figure 3.11: Moderation Analysis – Education
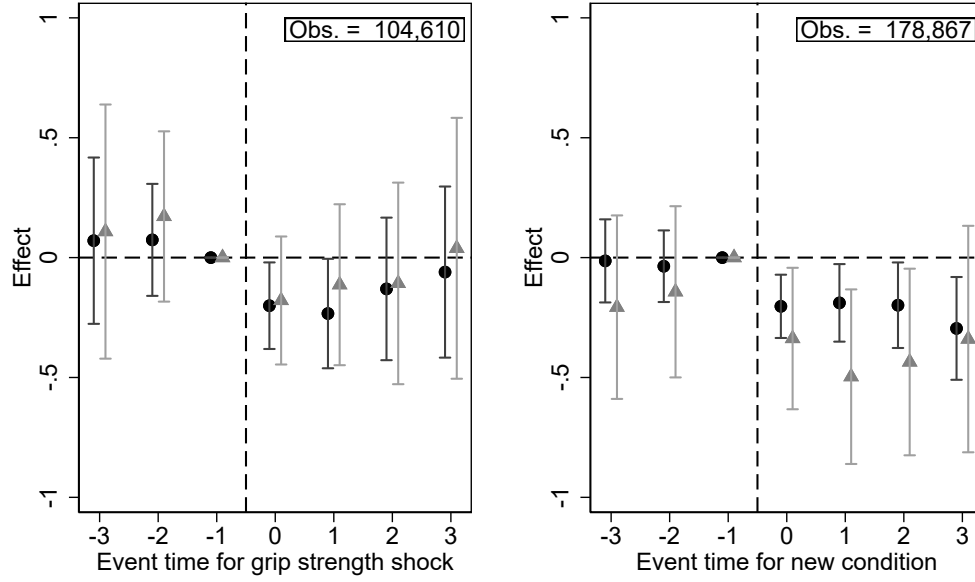
(a) Recall
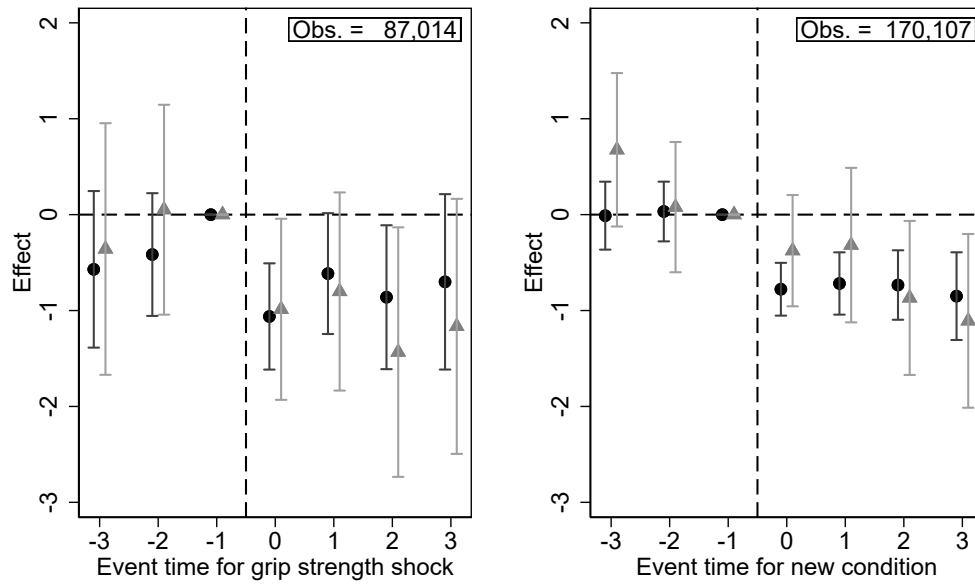
(b) Verbal fluency

● before comp. school. reform   ▲ after comp. school. reform

Notes: Coefficients $\mu_r + \mu_r \times mod_{it}$ from estimations of Equation 3.4 based on data from SHARE, HRS and ELSA. $\mu_{-1} + \mu_{-1} \times mod_{it}$ is restricted to zero. 95% confidence intervals reported.

3.5 Conclusion

We analyze the short- and longer-run effects of a health shock on cognitive decline in older individuals from Continental Europe, UK, and the US. Health shocks are measured by strong declines in grip strength and the onset of health conditions, while cognitive abilities are determined experimentally by the *word recall* and *verbal fluency*. We also ask whether the potential effect is moderated by variables that can be comparably easily changed by policy makers such as the retirement or education system.

In an event study, we find robust evidence that health shocks negatively affect cognitive functioning. The effects are persistent over a longer time even though most individuals have recovered from their health shock after some years. Comparing the effect size to the general age-related decline in cognitive functioning, a health shock, on average, induces a similar cognitive decline as growing up to four years older. Thus, physical health shocks also have the potential to bring a long-term care episode for mental health reasons forward by some years.

The effects of health shocks on cognitive decline do not seem to be moderated by retirement and education. Thus, we find that higher cognitive capacities (possibly due to labor force participation or education) do not prevent negative effects of health shocks. Of course, this does not mean that labor force participation and education do not pay off in terms of cognitive abilities as the direct effect is usually found to be positive and significant in the literature.

Taken together, our analysis consistently suggests that physical health shocks significantly and persistently impair cognitive abilities in older ages. This finding seems to hold not only for different regions, representing different health insurance systems, but also independently of socio-economic characteristics that are, at least some of them, susceptible to policy action. Therefore, this analysis cannot directly point to concrete policy measures, such as further promotions of work in older ages, that could help to curb cognitive decline after a health shock. Nevertheless, we believe that it provides valuable insights, as it highlights the role of physical health for human capital maintenance and suggests that investments in physical health pay double: with a healthy body and a healthy mind.

Appendix

Alternative Approach

Here we specify regression models in the spirit of van den Berg et al. (2010) as follows

$$\text{Cogn. abilities}_t = \beta_0 + \beta_1 \text{health shock}_t + X'_{t-1}\gamma + \varepsilon \qquad (3.4)$$

where a health shock is defined by a strong change in health between the waves $t-1$ and t, that is, over a period of about two years. The vector X includes gender, age fixed effects, measures of education, marital status, labor force status, income and wealth, baseline health, health behavior, country-specific fixed effects and year fixed effects. All these variables are measured in $t-1$ to make sure they are not affected by a health shock. Table A3.1 lists all variables in detail and reports their descriptive statistics.

Table A3.1: Descriptive Statistics

Variable	Mean	Std. dev.	Min	Max
Socioeconomic controls:				
Male	0.413	0.492	0.000	1.000
Married	0.669	0.471	0.000	1.000
Separated	0.014	0.118	0.000	1.000
Divorced	0.100	0.300	0.000	1.000
Widowed	0.154	0.361	0.000	1.000
Total household income/1000	44.215	145.355	0.000	60014.375
Household net worth/1000	334.773	798.879	-2245.500	68156.547
0-10 years of education	0.301	0.459	0.000	1.000
11-13 years of education	0.376	0.484	0.000	1.000
Employed	0.309	0.462	0.000	1.000
Unemployed	0.022	0.148	0.000	1.000
Disabled	0.030	0.172	0.000	1.000
Retired	0.548	0.498	0.000	1.000
Household size	2.151	1.039	1.000	19.000
Number of children	2.598	1.814	0.000	21.000
Never drinking	0.333	0.471	0.000	1.000
Ever smoked	0.536	0.499	0.000	1.000
Currently smoking	0.157	0.364	0.000	1.000
Self-assessed health	2.859	1.087	1.000	5.000
# difficulties with ADL	0.233	0.764	0.000	6.000
# difficulties with IADL	0.133	0.525	0.000	5.000

Notes: Descriptives statistics based on 240,810 observations from SHARE, HRS and ELSA.

A way to account for the potential issue that individuals with low cognitive abilities might be more likely to experience a health shock suggested by van den Berg et al. (2010) is to also condition on pre-treatment outcomes (Cogn. abilities$_{t-1}$).

$$\text{Cogn. abilities}_t = \beta_0 + \beta_1 \text{health shock}_t + \beta_2 \text{Cogn. abilities}_{t-1} + X'_{t-1}\gamma + \varepsilon \quad (3.5)$$

A third specification allows for fixed effects and takes first differences in cognitive abilities, thus, implicitly assuming that changes in cognitive abilities do not pre-date or even cause physical health shocks:

$$\Delta\text{Cogn. abilities}_t = \beta_1\text{health shock}_t + X'_{t-1}\gamma + \varepsilon \qquad (3.6)$$

where $\Delta\text{Cogn. abilities}_t$ is defined as a change in cognitive abilities between the waves $t-1$ and t.

If changes in cognitive abilities pre-date or cause physical health shocks, the estimates for β_1 derived from these regressions will be biased. This seems to be less of a problem for severe health shocks such as myocardial infarctions or cancer diseases, but might be more relevant for injuries due to falls like hip fractures. Also, time-varying unobservables that affect both, a health shock and cognitive abilities lead to biased results. We discuss these issues in more detail in the next subsection but start here with benchmark results.

The results from six separate regressions (two health shock measures times three specifications) are reported in Table A3.2. Each cell reports a coefficient of a health shock measure from a single regression. Apparently, both kinds of health shocks strongly affect cognitive abilities. Depending on the type of health shock and the regression model, a health shock goes along with around 0.24–0.46 recalled words less. The effect size is around 7–13 percent of a standard deviation which is typically regarded as a considerable amount.

Table A3.2: Baseline Regression Results

Dep. var.:	Recall (1)	Recall (2)	Δ Recall (3)
New condition	-0.291***	-0.269***	-0.240***
	(0.029)	(0.026)	(0.029)
Grip strength shock	-0.464***	-0.406***	-0.337***
	(0.042)	(0.038)	(0.045)
Further control variables	yes	yes	yes
Pre-treatment outcome	no	yes	no

Standard errors clustered at individual level in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Each of the six cells is the result of a different regression with either New condition or Grip strength shock as explanatory variable and several control variables. Column numbers (1), (2), (3) match the regression equations (1), (2), (3) in the text. Regressions on New condition based on 240,810 observations from SHARE, HRS and ELSA, regressions on Grip strength shock based on 88,416 observation from SHARE. Further control variables as indicated in the text.

Early Retirement Eligibility Criteria

Early retirement eligibility criteria are mainly based on Celidoni et al., 2017. If there are deviations, sources are reported with country specific rules below.

Austria

For men: Before 2001, early retirement age (ERA) is 60. From 2001 onwards, ERA is still 60 for those with at least 45 contribution years. Otherwise, ERA depends on the year of birth from 2001 on as follows. From 2001 to 2004, ERA is 61 for those born until 1942 and 62 for those born 1943 and later. From 2005 onwards, ERA is still 61 for those born until 1942, 62 between 1943 and 1944, 63 between 1945 and 1947, 64 between 1948 and 1950, and 65 for those born in 1951 and later.

For women: Before 2001, ERA is 55. From 2001 onwards, ERA is still 55 for those with at least 40 contribution years. Otherwise, ERA depends on the year of birth from 2001 on as follows. From 2001 to 2004, ERA is 56 for those born until 1947, 57 for those born between 1948 and 1951, and 58 for those born in 1952 and later. From 2005 onwards, ERA is still 56 for those born until 1947, 57 between 1948 and 1949, 58 between 1950 and 1952, 59 between 1953 and 1955, and 60 for those born in 1956 and later.

Belgium

For men: From 1967 to 1997, ERA is 60.

For women: From 1967 to 1986, ERA is 55 and from 1987 to 1997, ERA is 60.

For both: From 1998 on, ERA is 60 for both men and women, depending on contribution years: In 1998, at least 20 contribution years are needed, 24 in 1999, 26 in 2000, 28 in 2001, 30 in 2002, 32 in 2003, 34 in 2004 and 35 from 2005 on. For individuals employed in the public sector ERA is 58 from 1986 to 2008.

Czech Republic (see CSSZ, 2019b Ministry of Labour and Social Policy (Poland), 2019, Rabušic, 2004, CSSZ, 2019a)

For men: Until 2009, ERA is 57. From 2010 onwards, ERA is 60.

For women: ERA depends on the number of children. For women without children until 2009 ERA is 54. From 2010 to 2014 ERA is 59. From 2015 onwards ERA is 60. For women with one child until 2009, ERA is 53. From 2010 to 2014 ERA is 58. From 2015 to 2017 ERA is 59. From 2018 onwards ERA is 60. For women with two children until 2009 ERA is 52. From 2010 to 2014 ERA is 57. From 2015 to 2016 ERA is 58. From 2017 to 2018 ERA is 59. From 2019 onwards, ERA is 60. For women with 3 to 4 children until 2009 ERA is 51. From 2010 to 2014 ERA is 56. From 2015 to 2017 ERA is 57. From 2018 to 2020 ERA is 58. From 2021 to 2023 ERA is 59. From 2024 onwards ERA is 60. For women with 5 or more children until 2009 ERA is 50. From 2010 to 2017 ERA is 56. From 2018 to 2020 ERA is 57. From 2021 to 2023 ERA is 58. From 2024 to 2026 ERA is 59. From 2027 onwards, ERA is 60.

For both: Contribution years depend on the year, where ERA is reached. Until 2009 CY=25, in 2010 CY=26, in 2011 CY=27, in 2012 CY=28, in 2013 CY=29 in 2014 CY=30, in 2015 CY=31, in 2016 CY=32, in 2017 CY=33, in 2018 CY=34 and from 2019 onwards CY=35.

Denmark (see Angelini et al., 2009)
For both: From 1976 to 1978, ERA is 60. From 1979 onwards, ERA is 60 for those people with at least 30 contribution years.

Estonia (see Puur et al., 2015, Sotsiaalkindlustusamet, 2019)

For men: Before 2001: ERA is 45 if the man is visually impaired or a lilliputian with at least 20 contribution years. ERA is 55 for a widower with a disabled child and with 20 contribution years. ERA is 60 for those with 5 contribution years. From 2001 to 2020 ERA is reached 3 years before statutory retirement age, resulting in: ERA is 60 for those born from 1941 to 1956, ERA is 61 for those born from 1957 to 1960 and 62 for those born since 1961 with 15 contribution years, respectively.

For women: Before 2001: ERA is 40 if the woman is visually impaired or a lilliputian with at least 15 contribution years. ERA is 50 for those with a disabled child and 20 contribution years. ERA is 55 for those with at least 5 children and 15 contribution years. ERA is 55 for those with 5 contribution years. From 2001 to 2020 ERA is reached 3 years before statutory retirement age, resulting in: ERA is 56 for those born in 1946, ERA is 57 for those born from 1947 to 1948, ERA is 58 for those born from 1949 to 1950, ERA is 59 for those born form 1951 to 1952, ERA is 60 for those born from 1953 to 1956, ERA is 61 for those born from 1957 to 1960 and ERA is 62 for those born since 1961 with 15 contribution years, respectively.

For both: From 2021 onwards, ERA is 60 with at least 40 contribution years, ERA is 61 with at least 35 contribution years, ERA is 62 with at least 30 contribution years, ERA is 63 with at least 25 contribution years and 64 with at least 20 contribution years. From 2027 onwards, ERA will be bounded on life expectation. Having three children reduces the statutory retirement age by 1 year, four children reduces it by 3 years and five or more children (or a disabled child) reduces it by 5 years for one parent, respectively. For civil servants, retirement is possible at every age for those with at least 25 contribution years.

France (see Godard, 2016)
For both: From 1963 onwards, ERA is 60.

Germany
For men: From 1973 to 2003, ERA is 60 for those with at least 15 contribution years and 63 from 2004 onwards with at least 15 contribution years.

For women: From 1962 to 2003, ERA is 60 for those with at least 15 contribution years, 62 from 2004 to 2005 with at least 15 contribution years, and 63 from 2006 onwards with at least 15 contribution years.

Greece (see European Commission, 2019, Hauser and Strengmann-Kuhn, 2004)

For men: For men who started working before 1993: ERA is 58 with 35 contribution years. For all men: ERA is 60 with 15 contribution years. ERA is 50 for a widower with a disabled child and 18 contribution years.

For women: For women who started working before 1993: ERA is 55 with 15 contribution years. ERA is 50 for women with underage children and 18 contribution years. For women who started working since 1993: ERA is 60 with 15 contribution years. ERA is 50 for women with underage children and 20 contribution years.

For both: ERA is 62 with 15 contribution years.

Israel (see Kol-Zchut, 2019, Shai, 2018, Ministry of Justice (Israel), 2019)

For men: ERA is 60 for men.

For women: Until 2004, ERA is 55. From 2005 onwards, ERA is 58 for those born between May 1951 and April 1953, 59 for those born between May 1953 and April 1955 and 60 for those who were born after April 1955.

For both: (Kindergarten-)Teacher can retire at every age with at least 20 contribution years. ERA is 57 for kindergarten teachers born between March 1947 and April 1948, 58 for those born between May 1948 and April 1950 and 59 for those born after April 1950 with at least 10 contribution years, respectively. For other civil servants ERA is 55 for those born between March 1949 and April 1950, 56 for those born between May 1950 and April 1952, 57 for those born after April 1952 with 25 contribution years, respectively. For other civil servants ERA is 60 with at least 10 contribution years.

Italy (see Angelini et al., 2009)

For both: From 1965 to 1995, ERA is at any age possible for those with at least 35 contribution years (25 in the public sector).From 1996 to 1997 ERA is 52 in the private and public sector with at least 35 contribution years (or 36 contribution years independently of age), for self-employed, ERA is 56 with at least 35 contribution years. In 1998, ERA is 53 for the public sector, 54 for the private sector and 57 for self-employed. In 1999 ERA is 53 for the public sector, 55 for the private sector and 57 for self-employed. In 2000, ERA is 54 for the public sector, 55 for the private sector, 57 for self-employed. In 2001, ERA is 55 for the public sector, 56 for the private sector, 58 for self-employed. In 2002,

ERA is 55 for the public sector, 57 for the private sector, 58 for self-employed. In 2003, ERA is 56 for the public sector, 57 for the private sector, 58 for self-employed. From 2004 onwards, ERA is 57 for both the private and public sector, 58 for self-employed. The requirements in terms of years of contributions remain the same in the period from 1996 onwards.

Netherlands
For both: From 1975 to 1994, ERA is 60 for those with at least 10 contribution years. From 1995 onwards, ERA is 62 with at least 35 contribution years.

Slovenia (see ZPIZ, 2019, Government Office for Legislation (Slovenia), 2013)

For men: ERA is 59 for a father of one child and 58 for a father of two or more children with at least 40 contribution years.

For women: ERA is 56 for a mother of five or more children, 57 for a mother of three to four children, 58 for a mother of 2 children and 59 for a mother of 1 child with 40 contribution years, respectively.

For both: From 2013 onwards ERA is 60.

Spain
For both: Until 1982, ERA is 64. From 1983 to 1993, ERA is 60. From 1994 to 2001, ERA is 61, and from 2002 onwards, ERA is 61 for those with at least 30 contributions years.

Sweden
For both: From 1963 to 1997, ERA is 60. From 1998 onwards, ERA is 61.

Switzerland
For men: From 1997 to 2000, ERA is 64. From 2001 onwards, ERA is 63.

For women: From 2001 onwards, ERA is 62. Note, that before 2001, the official retirement age for women was at most 63. Thus, women are allowed to retire earlier than men at any point in time.

Chapter 4

Labor Market Spillover Effects of a Compulsory Schooling Reform in Germany[*]

4.1 Introduction

Compulsory schooling reforms are often used to estimate returns to education as they are unrelated to individual characteristics and preferences and thus arguably eliminate selection bias.[1] However, since these reforms affect a large number of individuals and thus have a high share of compliers, they might have spillover effects. Such spillover effects constitute a problem for identification and consistent estimation of returns to schooling as they imply that individuals who are used to estimate the counterfactual outcome distribution for treated units can be affected by the policy, too. Because this problem is difficult to address, it is often ignored or assumed away in the returns to education literature. This can also be problematic, because the total (general equilibrium) effect of a policy intervention can differ largely from partial equilibrium estimates based on standard policy evaluation techniques that assume away any kind of policy externalities on individuals or the economy as a whole (Abbring and Heckman, 2007).[2]

In this paper I study potential long-run spillover effects of a large scale educational reform that raised length of compulsory schooling from eight to nine years in all West

---

[1]For Germany see e.g. Pischke and von Wachter (2008); Kemptner et al. (2011); Kamhöfer et al. (2019); Cygan-Rehm (2021); Margaryan et al. (2021).

[2]For example, Heckman and Lochner (1998) simulate the effects of a tuition policy on college attendance and earning in a framework allowing for general equilibrium effects and compare the estimates to their partial equilibrium counterparts. Their results indicate that the partial equilibrium effects for such a policy differ largely from the general equilibrium effects of the reform.

[*]This chapter is published as a working paper: Schiele, V. (2022). Labor market spillover effects of a compulsory schooling reform in Germany. Working Paper Dissertations 84, Paderborn University, Faculty of Business Administration and Economics.

German federal states. I argue that the reform likely exhibited spillover effects on individuals not directly subject to the reform, since the reform led to a negative shock in the supply of basic track graduates in the reform year. I suspect that this negative labor supply shock led to a persistent increase in employment and wages of individuals who graduated in close temporal distance to the reform. In a recent study Morin (2015) estimates the wage effects of a positive shock in cohort size that was due to the abolition of the 13th grade in Ontario (Canada). He indeed finds substantial wage losses (between 5 and 9 percent) in the short-run for graduates who left school in the double-cohort year. Previous literature has shown that the role of initial labor market conditions are decisive factors also for later life labor market outcomes. Unfavorable labor market conditions such as high unemployment at labor market entry of graduates are known to have large and persistent effects on wages and employment (see e.g. Schwandt and Von Wachter, 2019; Oreopoulos et al., 2012; Kahn, 2010, or Von Wachter, 2020 for an overview). These negative effects can be explained to a large extend by an increase in mismatch between skills provided by graduates and skills demanded by employers during recessions (Liu et al., 2016).

To estimate the long-run spillover effects of the reform on labor market outcomes of individuals who were not directly affected by the increase in years of compulsory education, I exploit the staggered introduction of the compulsory ninth grade during the 1940ies, 50ies and 60ies across states in a difference-in-differences design. By doing so, I essentially compare wage and employment differentials of basic track graduates who left school close to the reform year and individuals who left school earlier to the wage/employment differentials of basic track graduates of the same cohorts who lived in other states and thus were not affected by the reform at that time. In order to rule out that the estimates are driven by wage or employment effects of schooling or wage and employment effects due to a changes in the skill composition of the workforce, I exclude individuals who were directly affected by the reform from the estimation sample. The data come from the *Sample of Integrated Labor Market Biographies (SIAB)* and the *Qualification and Course of Employment (QaC)* surveys. While the SIAB provides social security records on earnings on a daily basis for nearly 1.5 million workers and therefore even allows for identification of small effects, its measure of schooling is quite broad. I therefore check the robustness of the results using data from the QaC survey, which provides more detailed information on schooling but has a smaller sample size.

The contribution of the study is threefold. First, it contributes to our understanding of the role of labor market conditions at the time of graduation on subsequent labor market outcomes. While the existing literature mainly examines the effects of (unfavorable) labor market conditions due to a reduction in labor demand, I focus on potentially favorable conditions due to a reduction in labor supply. Second, the results may provide important insights into whether previous studies of the monetary returns to additional years of compulsory education also report effects close to zero because they use as

control units individuals who are affected by the reform, contrary to the assumptions made there. Finally, the paper adds to a small but growing strand of the economics of education literature that uses microdata and methods from the program evaluation literature to examine the externalities of educational interventions (Bianchi, 2020; Morin, 2015; Moretti, 2004; Duflo, 2004).

My results show that, although there was a substantial drop in the number of basic track graduates in the reform year, there were no long-term spillover effects of the reform on wages and employment for male graduates who left school shortly before the reform. For women, I find some evidence that graduating just before the reform increased employment and hourly wages in the long run, suggesting that the relative shortage of (male) entrants may have been compensated by female high school graduates who otherwise would not have entered the labor market at that time.

The remainder of the paper is structured as follows. In Section 4.2 I provide a short description of the educational system and the compulsory schooling reform. Section 4.3 describes the data and explains the empirical strategy. The results are presented in Section 4.4 and Section 4.5 concludes.

4.2   Institutional Background

In Germany students are tracked after they have finished primary school, i.e. after finishing fourth grade. Basically, there are three secondary school tracks available:[3] *Hauptschule*, the basic track that leads to a school leaving certificate after grade 8 or 9; *Realschule*, the intermediate track that leads to a school leaving certificate after grade 10; and *Gymnasium* the academic track that leads to a higher education entrance qualification after grade 13. While students from the basic and intermediate track typically start an apprenticeship after school completion, students from the academic track typically attend a university.

After Word War II all West German states stuck to a three-tiered school system with the aforementioned tracks. However, the length of schooling in the basic track varied over time and states. This is because education policies lay at the responsibility of individual states and not of the federal government. Some, mainly northern states introduced a compulsory ninth grade in the 1940ies, 50ies and early 60ies while others, foremost southern states decided to maintain eight years of compulsory schooling at that time. However, in 1964 all West German states agreed in the *Hamburg Accord* to harmonize length of compulsory schooling and the start of the school year across states. Until 1966

---

[3]Comprehensive schools that do not track students before the tenth grade were introduced only in the early 1970ies and played a negligible role (as measured by enrollment) in the education system at that time.

the school year started in April and ended in March of the following year in all states but Bayern, where students finished their grades in July and were promoted to the next grade in August. From 1967 onwards the school year started in August and ended in July all over West Germany. Thus, all states but Bayern had to move the start of the school year from April to August. This was done either by extending the length of the 1966/1967 school year by four months (i.e. having one long school year) or by splitting the transition period April 1966 - July 1967 into two short school years, the first lasting from April 1966 until November 1966 and the second lasting from December 1966 until July 1967 (see e.g. Pischke, 2007; Pischke and von Wachter, 2008; Helbig and Nikolai, 2015).

When the accord was signed, only five states had laws requiring students to stay in school for only eight years (see Table 4.1). Four of these states, namely Baden-Württemberg, Hessen, Nordrhein-Westfalen, and Rheinland-Pfalz, decided to use the school year transition period to introduce the compulsory grade 9. Thus, the compulsory schooling reform coincided with the short school years in these states. Although this coincidence might affect treatment intensity[4], it should not bias my estimates, as I exclude individuals from cohorts that were directly affected by the reforms.

Table 4.1: Years of Schooling for Basic Track Students by State

	BW	BY	HB	HE	HH	NI	NW	RP	SH	SL
First year when all students left school after grade 9	'67	'70	'59	'67	'50	'63	'67	'67	'57	'65
'66/'67 school year type	ssy	–	ssy	ssy	lsy	lsy	ssy	ssy	ssy	ssy
Last year when all students left school after grade 8	'66	'68	'57	'66	'48	'61	'66	'66	'55	'63
Last birth cohort with 8 years of comp. schooling	'51	'53	'42	'51	'33	'46	'51	'51	'40	'48
Months between graduation of last cohort with 8 and first cohort with 9 years of schooling	16	24	24	16	24	24	16	16	24	24

Notes: Largely based on Pischke and von Wachter (2005), Pischke and von Wachter (2008) and Pischke (2007), for details see Footnote 5; ssy: short school year, lsy: long school year; BW: Baden-Württemberg; BY: Bayern; HB: Bremen; HE: Hessen; HH: Hamburg; NI: Niedersachsen; NW: Nordrhein-Westfalen; RP: Rheinland-Pfalz; SH: Schleswig-Holstein; SL: Saarland.

If all basic track students who finished eighth grade in the reform year had to stay one year longer in school due to the introduction of the compulsory ninth grade, then there should be a drop in the number of basic track graduates in the reform year. Figure

---

[4]While the first cohort with 9 years of compulsory schooling left school 24 months after the last cohort with 8 years of compulsory schooling in all other states, this gap was only 16 months in the four states where the short school years coincided with the compulsory schooling reform (Table 4.1).

shows the number of basic track graduates sampled in the *Qualification and Course of Employment (QaC)* surveys by year of graduation (centered around the year of the reform in the respective home state). There is indeed a sizable drop in the number of graduates from basic track schools in the reform year (almost -20 percent compared to the pre-reform year) which is, however, smaller than expected.

Figure 4.1: Basic Track Graduates by Year of Graduation
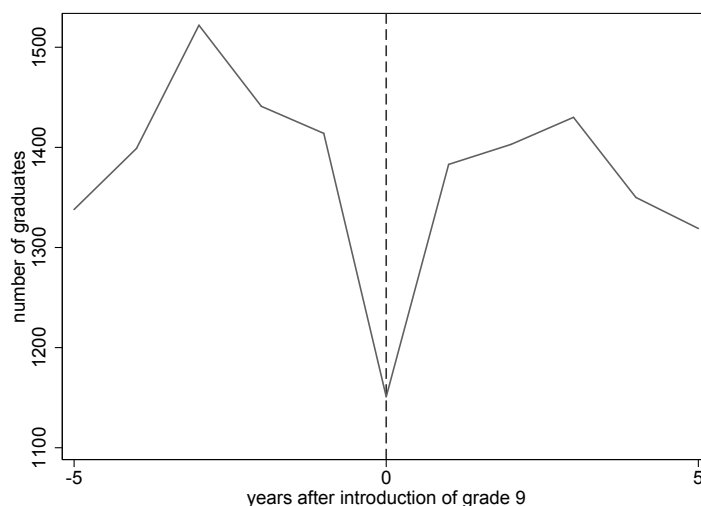
Notes: Own calculations based on QaC data. The graph shows the number of basic track graduates by year of graduation, where the year of graduation is centered around the year of introduction of the compulsory ninth grade.

An obvious explanation for the weaker drop in basic track graduated is the coincidence of the short school years and the compulsory schooling reform in Nordrhein-Westfalen, Hessen, Rheinland-Pfalz and Baden-Württemberg. Students from the last cohort with eight years of compulsory schooling left school in March 1966 in these states, while the following cohort – the first cohort with nine years of compulsory schooling – left school in July of the following year after having finished grade 7 in March 1966, grade 8 in November 1966, and grade 9 in July 1967.[5] Other explanations include early introductions of the compulsory ninth grade in some counties (see Pischke and von Wachter (2008) for details) and grade repetitions.

---

[5]The coding in Pischke and von Wachter (2008) seems to differ slightly for these states. According to the working paper version of their article Pischke and von Wachter (2005), the first cohort with nine years of compulsory schooling in theses states was the 1953 birth cohort. However, using the same rule to calculate year of school enrollment and graduation as Pischke and von Wachter (2005), I end up with the 1952 birth cohort being the first cohort with nine years of compulsory schooling. This cohort, was enrolled into school in April 1959, finished grade 7 in March 1966 and then had to stay in school for two additional (school) years. As these two school years were the short school years, students from this birth cohort should have graduated in July 1967, the first year when all basic track students were obliged to graduate after nine years of schooling.

4.3 Empirical Approach

4.3.1 Identification and Estimation

If basic school graduates achieve better job matches when there are less potential labor market entrants and thus less competitors for a job or apprenticeship, then raising compulsory schooling might increase wages in the long run for those basic track graduates who left school close to the reform. I define treatment as being part of the last cohort of basic track students with eight years of compulsory schooling. Since the timing of the introduction of the compulsory ninth grade varies across states, I use a generalized difference-in-differences (Diff-in-Diff) model with multiple groups as to investigate whether the compulsory schooling reform might have led to persistent wage gains for basic track students who left school in the pre-reform year:

$$y_{isct} = \alpha_s + \lambda_c + \delta_t + \alpha_s \times c_i + \beta last\_cohort_{sc} + \epsilon_{isct} \tag{4.1}$$

where y_{isct} is either the log wage of or an employment dummy for individual i as measured in year t, who lives in state s and is part of cohort c; α_s are state fixed effects, λ_c denote birth cohort and δ_t year fixed effects, $\alpha_s \times c_i$ are state specific linear birth cohort trends. Finally, $last\_cohort$ is a dummy for being part of the last cohort with 8 years of compulsory schooling, and ϵ_{isct} is an error term.

In this setting, β is the Diff-in-Diff parameter which gives the effect of graduating as part of the last cohort with 8 years of schooling. Note that other than in a classical Diff-in-Diff setting, the variable $last\_cohort$ is not determined by a combination of group membership and time in which the outcome is measured (i.e. before or after treatment), but by a combination of group membership (state) and cohort membership (year of birth). Still identification is based on the staggered introduction of the compulsory ninth grade across states. However, because this study focuses on the long-term effects of the reform, it takes advantage of the fact that the staggered implementation of the reform has led to differences in how birth cohorts are affected across states. Thus, by focusing on β_1, I compare the difference in wages between the last cohort of basic track graduates with 8 years of compulsory schooling and previous cohorts to the difference in wages between graduates from the same cohorts but different states, where a compulsory ninth grade was not yet introduced at that time. As it is possible that the labor market shock in the reform year did not only affect the last but also the second to last, third to last etc. cohort with 8 years of education I estimate alternative versions of equation 4.1. In these versions $last\_cohorts$ is either a dummy for being part of the last 2 or 3 cohorts with 8 years of compulsory schooling.

For several reasons individuals who left school after the introduction of the ninth grade are excluded from the analysis. First, these graduates might achieve higher wages due

to the additional year of schooling. Second, if I would use these cohorts as control units, I had to assume that there were no changes in relative skill prices due to the compulsory schooling reform. Third, since students were enrolled into school in August from 1967 onwards, the cut-off date for enrollment was in the mid of the year for most students with 9 years of compulsory schooling, which makes it more difficult to impute cohort membership and year of graduation from year of birth. Thus, by excluding these cohorts, I can rule out that the estimates capture direct wage effects of the compulsory schooling reform and avoid that the results are driven by changes in skill prices due to the reform.

The main identifying assumption in this setting, the Common Trends (CT) assumption, implies that in absence of the reform, the wage differential between the last cohort with 8 years of compulsory schooling and earlier cohorts in treatment states would have been the same as the wage differential between these cohorts in states without a compulsory schooling reform (at that time). To assess whether this is a reasonable assumption, I estimate a more flexible version of Equation 4.1 that has the following form:

$$
y_{isct} = \alpha_s + \lambda_c + \delta_t + \alpha_s \times c_i + \mu_a + \sum_{r=-9}^{r=-4} \mu_r + \epsilon_{isct} \tag{4.2}
$$

where α_s, λ_c and δ_t are again state, birth cohort and year fixed effects and $\alpha_s \times c_i$ are state-specific trends in birth cohorts. Here, however I do not include one indicator for the last cohort with eight years of education, but instead include several so called event-time indicator variables μ_r for graduation cohorts based on their year of graduation measured relative to the reform year. Let k_s denote the reform year in state s and l_i the year of graduation of individual i, then r_{is} is defined as $r_{is} = l_i - k_s$. Thus, μ_{-9} is a dummy for having graduated nine years before the reform, μ_{-8} is a dummy for having graduated eight years before the reform, etc. Finally, μ_a is a dummy variable that takes on the value 1 if an individual has graduated more than 10 years before the reform, i.e. if $r_{is} < -10$, and μ_{-10} is normalized to zero.

Note that when estimating 4.2, I exclude cohorts that graduated in the 3 years prior to the introduction of the compulsory ninth grade, i.e. all cohorts that graduated in close temporal proximity to the reform and thus are arguably affected by the potential labor market shock of the reform. In doing so, I follow arguments of de Chaisemartin and D'Haultfoeuille (2020) who suggests to exclude treated individuals from the estimation when assessing the credibility of the common trends assumption in settings with potential effect heterogeneity, since otherwise the estimates for μ_r that are used to assess

the common trends assumption can be contaminated by (the dynamics of) treatment effects.[6]

Another potential thread to the empirical strategy is that students might manipulate treatment status. This, however, requires them either to move to another state, to change the school track, or to delay graduation/push graduation forward. I argue, that it is rather unlikely that students or their parents move to another state due to a reform that does not affect them directly, especially since moving across state borders is quite seldom.[7] Moreover, there seems to be little incentive to delay graduation by repeating a grade, as repeating a grade could be seen as a bad signal on the labor market. Skipping classes, on the other hand, is the absolute exception in the German education system and only possible under exceptional circumstances. Since students or their parents have to decide on a school track already after the fourth grade and switching from one school track to another is rare, I expect selection into treatment not to be a serious problem here.

4.3.2 Data and Measurement

The main data source is the *Sample of Integrated Labour Market Biographies (SIAB)*,[8] a data set that provides social security records on a daily basis for about 2 percent of all individuals who i. have been in employment subject to social security, or ii. received unemployment benefits in the period 1975-2010.[9] For people sampled in the SIAB, the data provides information on all employment spells between 1975 and 2010, including the start and end date of each spell, daily wages up to the social security contribution ceiling (*Beitragsbemessungsgrenze*), occupation, a very broad measure of working time, the district region of the work place, and some basic demographic information such as year of birth, gender, nationality as well as schooling and vocational training (Vom Berge et al., 2013; Dorner et al., 2010). I use the spell data to generate a panel data set with yearly observations that includes the relevant information for the period 1975-1990 as present at the 1st of July of each year.

While the great strength of SIAB is its high quality information about daily wages for a large number of employees (in total about 1.5 million), it also has two disadvantages. First, and more important, there is no distinct measure of school track and no information about year of graduation available in the SIAB. Instead, highest educational

---

[6]As shown by Goodman-Bacon (2021) the "contamination" results from using already treated units as control units and can occur in the presence of effect heterogeneity.

[7]Pischke (2003) reports that according to the ALLBUS survey more than 80 percent of respondents live in the same state as they have lived at the age of 6 (the usual school entrance age in Germany)

[8]The data basis is the scientific use file of the Sample of Integrated Labor Market Biographies (SIAB-R 7510) provided by the Federal Institute for Employment Research (IAB).

[9]SIAB also covers information on individuals who have been employed in marginal part-time employment, participated in programs of active labor market policies, or have been registered as job-seekers, but only for some periods and not the entire period covered by SIAB.

attainment and vocational training is combined in one variable, making it impossible to identify individuals' school tracks unambiguously. Second, SIAB does not contain the number of working hours. The only information related to working time included in the data set is a variable that allows to distinguish between people who work more than a certain number of hours a day and people working less than this amount (depending on the year the threshold varies between 15 and 20 hours). I thus cannot calculate hourly wages, but have to stick to daily wages when using SIAB.

To deal with these issues I also use the *Qualification and Course of Employment (QaC)*[10] surveys as a second data source. The QaC surveys have been conducted by the Federal Institute for Vocational Education and Training (BIBB) in cooperation with the Institute for Employment Research (IAB) in the years 1979, 1985/86, 1991/92 und 1998/99. They provide representative information about current and past employment (wages, occupations, vocational requirements, working conditions) and education among employees and self-employed persons. Although the sample size of these cross-sectional surveys are much smaller than the sample size of SIAB, they have some attractive features. First, they provide more detailed information about educational attainment than SIAB. The QaC data does not only include information about highest educational attainment but also about year of graduation. Second, the QaC data contains the number of hours worked per week which can be used – in combination with weekly wages – to calculate hourly wages.

For both data sets applies that they do not include direct information about school track choice and state of residence at the age of school entry. I thus follow the standard procedure (see e.g. Pischke and von Wachter, 2008; Kemptner et al., 2011; Margaryan et al., 2021) and use highest educational attainment to impute school track and approximate state of residence during time at school by current state of workplace (SIAB) and state of residence (QaC), respectively. I do not expect this approach to have a sizable impact on the estimates, as mobility across state borders is quite low (Pischke, 2003) and changing the school track is quite complicated in Germany. Based on (imputed) school track, I then calculate the year of graduation for basic track graduates according to the following rule: year of birth plus 7 (the year after birth in which students enter school) plus 8 (number of years in school). Finally, comparing expected year of graduation and the year of the introduction of the compulsory ninth grade, I define the treatment status. Table A4.1 in the appendix gives an overview about the treatment status (*last_cohort*) by year of birth and state and cohorts included in the sample (see also Section 4.3.1).

When using the SIAB, I need an additional assumption to define treatment status. In the SIAB one can only distinguish between individuals who attended basic or intermediate track schools and individuals who attended academic track schools. I therefore assume

---

[10]The data used in this paper were made available by the gesis – Leibniz-Institute für Sozialwissenschaften.

that all students who do not hold a higher education entrance qualification, to have attended basic track schools when using the SIAB data. This, of course, means that some individuals are misclassified with respect to school track attendance.[11] Contrasting the results based on the SIAB with the results based on the QaC data helps to evaluate whether this assumption is crucial for my findings. It does not seem to be.

As a first outcome measure I use the log daily wages in the SIAB data and log hourly wages in the QaC data (both in 1995 Euro). Due to the relatively high level of collective bargaining coverage in Germany, which might prevent employers to pay higher wages to attract employees in times of scarce labor supply, especially in the apprenticeship occupations, I also consider possible effects of the reform on the employment probability of those not directly affected by the reform. I do this using a pseudo-panel that is constructed based on the SIAB panel. This pseudo-panel contains the same information as the panel for the wage regression plus a dummy variable for employment. This dummy variable indicates whether a person is in employment (subject to social security) or not. It takes the value 1 for persons who are in employment (subject to social security) according to the SIAB data on the reference date. Persons who do not appear in the raw data on the reference date of the respective year (i.e. who were not employed subject to social security on the reference date), but were employed on any other reference date during the period 1975-1990, are assigned the value 0.

The data are restricted in all analysis to basic track graduates (as defined above) who are between 20 and 58 years old and who were born after 1922 and were not subject to the introduction of the compulsory ninth grade in all analyses. Table 4.2 shows sample statistics for both data sets. The average log wage is 4.1 in the SIAB and 2.2 in the QaC data, corresponding to a daily wage of 61.07 Euro (SIAB) and a hourly wage of 9.22 Euro (QaC). Individuals who are part of the sample are employed (subject to social security) on average on 74 percent of all 1st of July during the period 1975 to 1990. About 4 percent of the observations belong to individuals who are part of the last cohort with 9 years of compulsory schooling. The average age is 43-45 years. In general, the descriptive statistics in SIAB and QaC are similar. The largest differences can be found for the proportion of females. This is about 30 percent in the SIAB data and 36 percent in the QaC data. The difference is not surprising, since women are relatively often in atypical employment and thus less often sampled in the SIAB compared to the QaC data.

---

[11]Note, however, that the intermediate track was not very popular until the 1970ies. In the period 1952-1969 only between one in ten and less than one in five students in the eighth grade attended a intermediate track school (Köhler and Lundgreen, 2014).

Table 4.2: Descriptive Statistics

	SIAB		QaC	
	mean	sd	mean	sd
Outcome variables				
Log wage[+]	4.112	0.462	2.221	0.439
Employed (d)[−]	0.739	0.439		
Treatment variables				
Last cohort (d)	0.036	0.185	0.043	0.202
Last 2 cohorts (d)	0.070	0.255	0.088	0.284
Last 3 cohorts (d)	0.113	0.317	0.143	0.350
Control variables				
Female (d)	0.366	0.482	0.307	0.461
Age	43.281	7.669	44.540	7.536
Year of birth	1938.677	7.057	1940.163	6.911
Year	1981.958	4.498	1984.703	6.392
BW (d)	0.177	0.382	0.156	0.363
BY (d)	0.201	0.401	0.202	0.401
HB (d)	0.011	0.105	0.009	0.095
HH (d)	0.009	0.096	0.004	0.061
HE (d)	0.099	0.299	0.102	0.302
NI (d)	0.100	0.299	0.106	0.308
NW (d)	0.306	0.461	0.306	0.461
RP (d)	0.057	0.232	0.076	0.266
SH (d)	0.023	0.150	0.022	0.146
SL (d)	0.016	0.127	0.017	0.129
Observations	2,592,588		24,055	

Notes: (d) indicates dummy variables; [+] Wage refers to the daily wage in the SIAB and to the hourly wage in the QaC data; [−] The number of observations when using the pseudo-panel and employment as dependent variable is 3,512,494; BW: Baden-Württemberg; BY: Bayern; HB: Bremen; HE: Hessen; HH: Hamburg; NI: Niedersachsen; NW: Nordrhein-Westfalen; RP: Rheinland-Pfalz; SH: Schleswig-Holstein; SL: Saarland.

4.4 Results

Table 4.3 gives the estimation results for the Diff-in-Diff model based on the SIAB for all basic track graduates and separately for men and women. Each row represents a separate regression. The first block shows the results for the wage regressions, the second block those for the regressions with the employment dummy as dependent variable. Within each block, the estimates reported in the first row are for the baseline treatment where the introduction of the compulsory ninth grade is assumed to affect only the last cohort with 8 years of compulsory schooling. The estimates in the second and third row account for the possibility that the negative shock in basic track graduates in the reform year might not only affect the employment prospects of the last cohort of graduates with 8 years of schooling but also of the second to last and third to last,

respectively.

Table 4.3: Baseline Results

dep. var.	All		Men		Women	
	Coeff.	(S.e.)	Coeff.	(S.e.)	Coeff.	(S.e.)
log daily wages						
Last cohort	-0.001	(0.006)	0.002	(0.003)	-0.004	(0.017)
Last 2 cohorts	0.000	(0.007)	0.004	(0.003)	-0.000	(0.020)
Last 3 cohorts	0.009*	(0.004)	0.007**	(0.003)	0.018	(0.013)
Obs.	*2,617,600*		*1,644,659*		*972,941*	
employment						
Last cohort	0.003	(0.002)	0.000	(0.004)	0.008*	(0.004)
Last 2 cohorts	0.002	(0.002)	0.002	(0.005)	0.005	(0.006)
Last 3 cohorts	0.003	(0.003)	-0.003	(0.004)	0.013*	(0.007)
Obs.	*3,512,494*		*2,029,057*		*1,483,437*	

Notes: Coefficient β from estimations of Equation 4.2 based on data from the SIAB; Standard errors clustered at state level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Each estimate comes from a separate regression; Controls include state, year of birth and year fixed effects, a gender dummy and state-specific year of birth trends.

The estimates suggest that there was no persistent effect of graduating shortly before the reform for basic track graduates. According to the results, graduating in the pre-reform year (row 1) is associated with a wage loss by 0.1 percent. The estimated coefficients for graduating in one of the last two (row 2) are positive, but economically and statistically not significant. Only for those who graduated 3 years prior to the introduction of the compulsory ninth grade (row 3) I observe a slightly larger effect that is statistically also marginally significant. The effect size is, however, still rather small (less than 1 percent). Furthermore, if the shock in potential labor market entrants due to the introduction of the compulsory ninth grade led to better job (or apprenticeship) matches and thus persistent wage gains for labor market entrants one would expect a positive effect first and foremost for those graduating in the last pre-reform year and, to a smaller degree, for those graduating one or two years earlier. However, my estimates do neither suggest positive wage effects for either of these cohorts nor do they decrease with increasing distance to the reform year.

The results stratified by gender in the following columns are similar with the pooled results. The estimates for graduating in the last pre-reform year are small and not significant. Not surprisingly, considering the results for both genders, I find the largest estimates for the broadest treatment definition. As these estimates are nevertheless rather small and since I cannot provide a reasonable explanation for these results, I

prefer to interpret them as a statistical artifact at this stage of the analysis instead of evidence for spillover effects of the compulsory schooling reform.

The second part of Table 4.3 shows the results for the employment dummy as dependent variable. In line with the results for wages, I find no evidence that graduating prior but close to the reform year had an effect on employment prospects for men. All estimates for men are almost indistinguishable from zero. For women the results are somewhat less clear cut. All of them are positive and two out of three also statistically significant and thus seem to be compatible with the aforementioned hypothesis, at least to some extend. The largest estimate for women indicates that graduating in the 3 years prior to the introduction of the ninth grade increased the likelihood of being employed by 1.3 percentage points, corresponding to an increase of 2 percent as measured at the mean of the dependent variable.

The validity of the estimation approach used here hinges on the assumption of common trends. Figure 4.2 shows how wages (left panels) and employment (right panels) of graduates developed by year of graduation (centered around the reform year), conditional on fixed effects and relative to wages/employment of the cohort that graduated 10 years before the reform (reference group). For neither outcomes there is any evidence for a violation of the CT assumption. This impression holds when looking on men and women separately. In total, two out of the 36 (i.e. 5.56 percent of) estimates differ statistically from zero at the 5 percent significance level. This is what one would expect simply by chance. Even if ignoring statistical significance, I do not find any indication for a systematic deviation from the common trend assumption, instead the estimates fluctuate randomly around zero.

As the SIAB data does not include information on the year of graduation, the treatment assignment is based on year of birth in the results presented so far. Using calculated instead of actual year of graduation to assign treatment status might introduce measurement error and thus bias. To prove robustness of results with respect to the assignment to treatment and control group, Table 4.4 repeats the analysis but now using QaC data. The first three columns of the table repeat the baseline analysis and report the results for the treatment assignment based on year of birth (column header *calculated year of graduation*). Columns 4-6 give the results for the treatment assignment based on *Stated year of graduation*. Note that the dependent variable is now the log hourly wage instead of the log daily wage.

When looking at the results for the QaC data, there are two notable things. First, whether I assign treatment status on the basis of home state and year of birth or whether I use self-stated year of graduation does not seem to matter much, suggesting that bias due to measurement error or self-selection into treatment is negligible. Second, there are marked differences between the results for men and women when using the QaC data instead of SIAB. While the results for men are in line with the baseline results based on

Figure 4.2: Assessment of CT Assumption
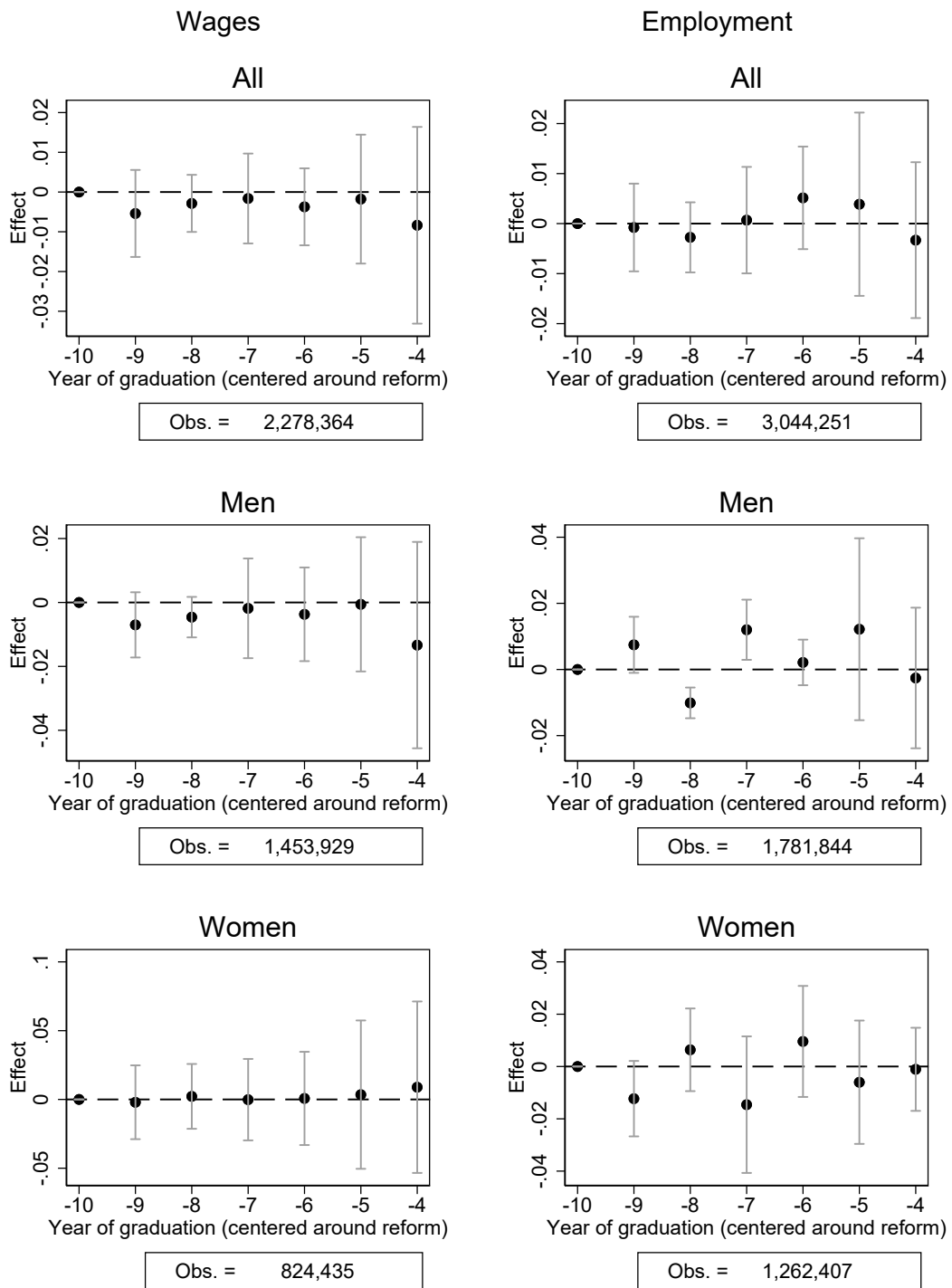
Notes: Coefficients μ_r from estimations of Equation 4.2 based on data from the SIAB. μ_{-10} is restricted to zero. 95% confidence intervals reported. Standard errors clustered on state level.

the SIAB data, pointing to zero effects of graduating shortly before the reform, the results for women strongly suggest positive and statistically significant effects. Specifically, I find that graduating in the last year before the reform took effect, increased hourly wages by 6.4-7.4 percent. For the broader treatment definitions which include also

graduates who graduated two and three years, respectively, before the reform, the estimates are somewhat smaller but still sizable. Due the reduced sample size and thus low precision they are not statistically significant. The overall impression, however, fits to the hypothesis that the advantage of graduating close to the reform decreases with increasing distance between year of graduation and year of the reform.

Table 4.4: Results QaC

	Calculated year of graduation			Stated year of graduation		
	All	Men	Women	All	Men	Women
dep. var.	Coeff. (S.e.)	Coeff. (S.e.)	Coeff. (S.e.)	Coeff. (S.e.)	Coeff. (S.e.)	Coeff. (S.e.)
log hourly wages						
Last cohort	0.027 (0.024)	-0.000 (0.021)	0.074** (0.028)	0.020 (0.019)	0.001 (0.020)	0.064** (0.021)
Last 2 cohorts	0.016 (0.022)	-0.007 (0.014)	0.056** (0.036)	0.016 (0.023)	-0.010 (0.014)	0.061 (0.040)
Last 3 cohorts	0.024* (0.015)	0.012 (0.010)	0.045* (0.029)	0.024 (0.016)	0.009 (0.010)	0.050 (0.032)
Obs.	24,055	16,682	7,373	23,915	16,584	7,331

Notes: Coefficient β from estimations of Equation 4.2 based on data from the QaC; Standard errors clustered at state level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Each estimate comes from a separate regression; Controls include state, year of birth and year fixed effects, a gender dummy and state-specific year of birth trends

Assuming that the results for women in the QaC data really capture the effects of the reform, there are at least two potential explanations for the differences compared to the results for women in the SIAB.[12] The first explanation relates to the fact that SIAB and QaC do not sample the same population. While the SIAB sample includes only individuals subject to social security (e.g. no self-employed or marginally employed workers and no civil servants), the QaC sample includes individuals in any type of work. Thus, if graduating shortly before the reform has increased wages only for women and only in sectors/occupations not sampled in the SIAB, e.g. for women who are in marginal employment, this might explain the differences. The second, also speculative explanation relates to differences in the definition of the dependent variable, i.e. that I use daily wages in the SIAB but hourly wages in the QaC data: It could be that the shortage of (male) workers and potential (male) trainees led to better job offers and higher demand for female workers. This may have led to more women entering the labor/apprenticeship market immediately after graduation who otherwise would not have entered (at that time). In the long run, this might then reflect in higher hourly

---

[12]A violation of the CT assumption for the QaC sample does not appear to be one of them, as shown in Figure A4.1 in the appendix.

wages. If, however, these women respond to an increase in higher hourly wages with a reduction of working hours (e.g. because women "choose" their total income), this could explain why effects can be found for women in the QaC for hourly wages but not for daily wages in the SIAB data. Considering that the women in the estimation sample were born before 1953 and thus on average had a rather low labor market attachment compared to men and remembering that in the SIAB I indeed find weak evidence for positive employment effects of the labor supply shock, this seems to be a reasonable (though still somewhat speculative) explanation.

4.5 Conclusion

In this paper I investigate in potential long-run labor market spillover effects of the introduction of a compulsory ninth grade in West German basic track schools during the 1940ies, 1950ies and 1960ies. The reform led to a negative shock in basic track graduates in the reform year which might have improved job match quality, career opportunities and thus employment prospects and wages of individuals who graduated close to the reform year. To estimate the effects of the labor supply shock on later wages and employment, I make use of large administrative and survey data and exploit the staggered introduction of the compulsory ninth grade across West German federal states in a Diff-in-Diff design.

I do not find evidence that the shock in the supply of labor market entrants due to the compulsory schooling reform had a long-run effect on labor market outcomes of men who graduated shortly before the reform. However, I find some evidence that the reform has increased employment and hourly wages (but not earnings from employment) of women who graduated in close temporal distance. A potential explanation for these results is that employers tended to fill vacant (apprenticeship) positions with female graduates who otherwise would not have entered the labor market at that time and that the observed effect for women is a result of increased work experience among these women. Note that the finding of zero long-run effects for men does not necessarily imply zero effects in the short or medium run. It might also be, that positive wage or employment effects of favorable labor market conditions vanish after some years, just as the negative wage effects of disadvantageous labor market conditions (e.g. high unemployment) seems to fade after 10-15 years (Von Wachter, 2020), and that these dynamics are not symmetric for men and women. While I cannot provide estimates for short- and medium run effects due to data limitations and thus cannot provide evidence for one explanation or the other, I can at least show that such effects – if they exist – are not very persistent for men.

Appendix

Figure A4.1: Assessment of CT Assumption in the QaC Data
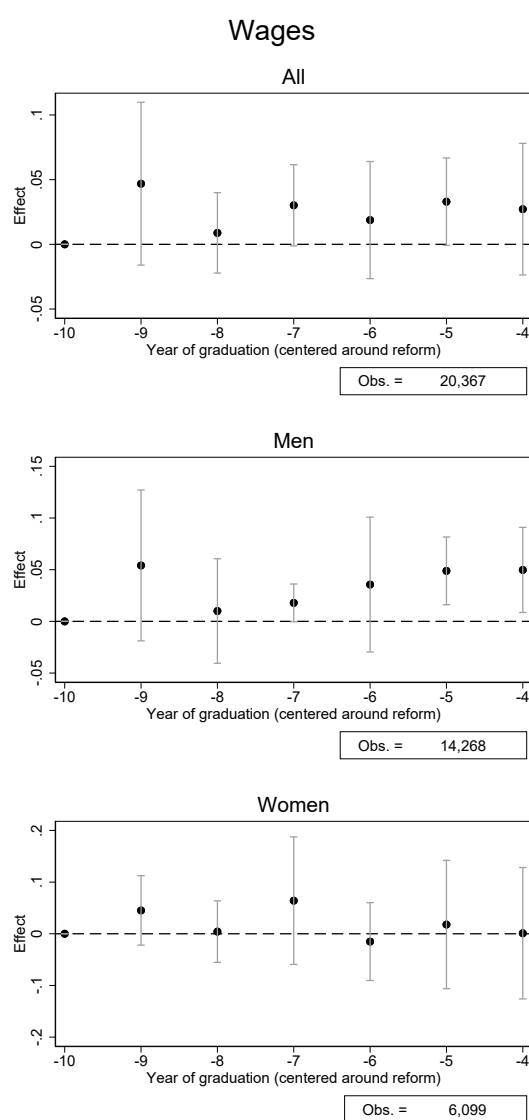
Wages

Table A4.1: Sample and Baseline Treatment by Year of Birth (Basic Track)

	'21	'22	…	'32	'33	'34	…	'39	'40	'41	'42	'43	'44	'45	'46	'47	'48	'49	'50	'51	'52	'53	'54	'55
HH	0	0	0	0	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
SH	0	0	0	0	0	0	0	0	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
HB	0	0	0	0	0	0	0	0	0	0	1	–	–	–	–	–	–	–	–	–	–	–	–	–
NI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	–	–	–	–	–	–	–	–	–
SL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	–	–	–	–	–	–	–
NW	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	–	–	–
HE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	–	–	–
RP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	–	–	–
BW	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	–
BY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	–

Notes: Cohorts highlighted in gray are included in the sample; 0 indicates graduation before reform year; 1 indicates graduation in reform year.

Chapter 5

Spring Forward, Don't Fall Back: The Effect of Daylight Saving Time on Road Safety*

5.1 Introduction

Daylight Saving Time (DST), also summer time, refers to the practice of moving clocks forward by one hour from standard winter time in spring and backward by one hour to winter time in fall. It was first introduced in Germany followed by Great Britain, France and the US during World War I in an attempt to conserve energy by shifting daylight from morning to evening hours. Currently, around 70 countries (e.g. Europe and large parts of the US and Canada) covering about one quarter of the world's population adopt DST regimes, yet, concerns about the usefulness of DST are growing. In March 2019, the "Sunshine Protection Act", a bill that would introduce year-round DST in most areas of the US, was brought to US Congress (United States Congress, 2019). Virtually at the same time, the European Parliament voted for a resolution to scrap the biannual clock change from 2021 onwards (European Parliament, 2019). Although both bills have not been approved yet, they mirror the rising concerns about whether the harms of changing the clocks twice a year potentially outweigh the intended benefits of this policy regulation.

Recent literature provides empirical evidence that DST does not have the originally intended effect on energy conservation. To the contrary, DST may even increase total energy consumption, since energy savings in lighting are at least offset by increased energy use in other areas such as heating or air conditioning (Kellogg and Wolff, 2008;

Momani et al., 2009; Krarti and Hajiah, 2011; Kotchen and Grant, 2011; Sexton and Beatty, 2014).[1] In addition, several studies indicate that the transition in and out of DST causes disruptions in the circadian rhythm and adversely affects the duration and quality of sleep (Lahti et al., 2006; Kantermann et al., 2007), which in turn may have unintended negative side effects in various (economic) dimensions other than energy saving. These negative short term effects range from lower general well-being (Kountouris and Remoundou, 2014) and life satisfaction (Kühnle and Wunder, 2016), decreases in stock market returns (Kamstra et al., 2000) and students' performance (Gaski and Sagarin, 2011) to higher risk of work injuries (Barnes and Wagner, 2009; Lahti et al., 2011), acute myocardial infarction (Janszky and Ljung, 2008; Jiddou et al., 2013; Toro et al., 2015), suicides (Berk et al., 2008) and fatal road accidents (Varughese and Allen, 2001; Sullivan and Flannagan, 2002; Sood and Ghosh, 2007; Smith, 2016).[2]

While this literature provides evidence in favor of abolishing the yearly ritual of changing the clocks twice, at least from a short term perspective, it is less clear for which time regime – DST or standard winter time – a society eventually should opt. From an economists point of view, this decision should be based on the long term costs and benefits of establishing one of these time regimes instead of the other. One area in which permanent costs and benefits might arise is road safety. This is mainly because choosing one time regime instead of the other affects the distribution of natural light across hours of the day. Since traffic density also differs by time of the day, shifting light from the morning hours to the evening hours or the other way round might have consequences for annual road accident counts, if light levels affect road safety. It is the aim of this paper to empirically assess to which extent this is indeed the case and how different time regimes affect road safety.

To estimate the effect of darkness on accident counts we make use of large, administrative data from England, Scotland and Wales covering all accidents on public roads that resulted in a personal injury and were reported to the police. Our identification strategy exploits arguably exogenous variation in darkness stemming from three sources of variation in sunrise and sunset times: day-by-day variation for a given region and hour; east-west variation for a given day of the year and hour; and north-south variation for a given day of the year and hour. The resulting estimates are used to simulate the number of fatal, serious and slight accidents under two different time regimes: setting the clocks permanently to DST vs. permanent standard winter time.

Our contribution to the literature is twofold. First, we extend the literature on the effects of DST on road safety by looking at the long-run consequences for accident counts

---

[1]See Aries and Newsham (2008) and Havranek et al. (2018) for a literature review covering earlier studies on the relationship between DST and lighting energy usage.

[2]There are also studies finding no effect of the transition into DST on stock market returns (Gregory-Allen et al., 2010), students' performance (Herber et al., 2017), myocardial infarction (Sandhu et al., 2014), hospital admissions (Jin and Ziebarth, 2020) and fatal accidents Lahti et al. (2010).

of different time regimes. This allows to assess whether year-round DST or all-year standard winter time should be implemented, at least from a road safety perspective. Most of the related literature on the effect of DST on traffic accidents exploits the abrupt transition from standard winter time to DST in spring and vice versa in fall (e.g. Smith, 2016). This approach allows for credible identification of the effects of the transition into and out of DST under plausible assumptions, however, the resulting estimates cannot easily be used to assess the effects of DST over the entire year, as they are valid only locally (i.e. around the dates of transition into/ out of DST). There are only a few early papers that investigate the long-run consequences of time regime choice for road safety (e.g. Ferguson et al., 1995; Coate and Markowitz, 2004).[3] Yet, as these paper essentially compare accident counts during the morning / evening hours across weeks, strong assumptions have to be imposed to give their estimation results a causal interpretation.[4] Second, exploiting information about all individuals involved in each accident, we shed some light on potential mechanisms through which darkness might affect road safety. Given that light levels do not only affect vision but might also influence concentration and fatigue due to adjustments of the circadian rhythm, it is unclear whether measures that aim at increasing visibility such as more road lighting can be effective in reducing road casualties.

We find that darkness increases accident counts by around 7 percent per hour, which translates into annual costs of more than £500 million caused by darkness. As the estimated effects are larger for fatal and serious accidents than for slight accidents, our results indicate that light conditions do not only increase accident risk but also accident severity. Simulations based on these estimates suggest that shifting light from the morning to the evening by setting the clocks permanently to DST could save lives and costs. Compared to an all-year standard winter time regime, setting the clocks to DST during the entire year could avoid at least £40 million per year across England, Scotland and Wales. This indicates that existing literature on the effects of the transition itself underestimates the total costs of the current time regime. Our heterogeneity analyses indicate that the positive effects of darkness on accident counts are likely due to a reduction in vision during darkness.

The remainder of the paper is structured as follows. Section 5.2 describes the data and the empirical approach. In Section 5.3 we present our main results as well as the results from heterogeneity and sensitivity analyses. Section 5.4 concludes and discusses policy implications.

---

[3]For a broader overview about the previous literature see Carey and Sarma (2017).

[4]These studies assume that gradual changes in accident counts in the course of the year are only the result of gradual changes in the number of light hours and not influenced by other factors, such as road conditions, traffic composition or the number of road users.

5.2 Empirical Framework

5.2.1 Data and Measurement

Our accident data comes from the STATS19 data base (Department for Transport, 2018b). It covers all accidents on public roads in England, Scotland and Wales that resulted in a personal injury and were reported to the police between 1996 and 2017. The data is collected by the police using a nationally standardized form for the collection of data on each road accident, on each casualty, on the vehicles involved in the accident as well as on drivers and passengers.[5]

There are three properties making the STATS19 data especially suitable to study the effects of ambient light condition on road safety. First, it provides detailed information on the level of individual accidents. For each accident we do not only have information on the location and the exact date and time of the accident, but also on the type of accident, the number of slightly, seriously and fatally injured casualties and the age of all drivers. Thus, we can assign each accident to a geographic region and are able to distinguish between certain types of accidents, e.g. we can distinguish car accidents from accidents involving pedestrians. Second, the coverage of the STATS19 data is exceptionally large. As our identification strategy is based on variation in sunrise and sunset time across latitude, longitude and day of the year, we are reliant on data covering a sufficiently large area and time span. Our sample comprises information on more than 4.1 million accidents from all across England, Scotland and Wales between 1996 and 2017. Finally, the data can be considered to be highly reliable. According to our sample, a police officer attended the scene of accident to obtain the details for the report in almost 80 percent of all accidents. The information for another 20 percent of accidents were gathered in questionings at the police station while the details about the accidents were obtained using a self-completion form only in less than 0.2 percent of all accidents.

The individual level data is aggregated in order to yield hourly accident counts for a panel of 265 regions. We use a grid references system based on longitude and latitude to define regions. This system divides Great Britain into 265 regions, each of them spanning an area of the size $0.5°$ longitude $\times$ $0.5°$ latitude.[6] Using information about the type and severity of the accident as well as the age of the drivers,[7] we distinguish between the

---

[5]Although the local police forces are not obliged to use the standardized form to collect the relevant data, a great majority of forces use the precise form or a minor variation (Department for Transport, 2013).

[6]$0.5°$ longitude corresponds to approximately 36 kilometers in the very south of GB and 27 kilometers in the very north; $0.5°$ latitude corresponds to approximately 55.5 kilometers. With respect to the grid size there is a trade-off between spatial precision and computational feasibility. Increasing spatial precision, for instance, by a factor of four ($0.25°$ longitude $\times$ $0.25°$ latitude), quadruples the number of observations and increases the computational burden by a multiple of four.

[7]The term driver includes all active road users, i.e. car drivers, pedestrians, cyclists etc. but no passengers.

following accident categories: all accidents, accidents involving pedestrians, accidents involving drivers younger than 25, between 25 and 45, between 45 and 65 or above the age of 65, as well as accidents with only slightly injured casualties, accidents with seriously injured casualties, and fatal accidents. The left panel of Figure 5.1 illustrates the grid reference approach and shows how accidents are distributed across regions.

Figure 5.1: Distribution of Accidents and Weather Stations across GB
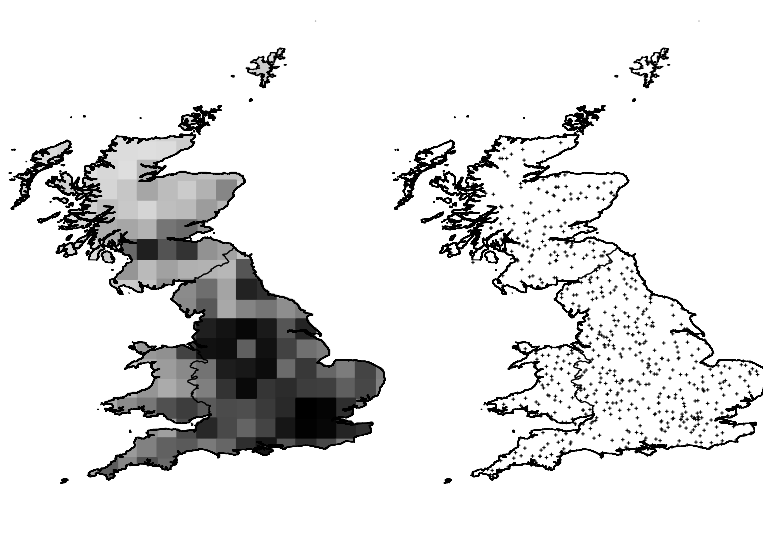
Notes: Own calculations and illustration based on STATS19 and MIDAS data. The left part of the figure shows grid based regions and the distribution of accidents during the year 2017 across these regions. Darker shading indicates more accidents. The right part of the figure shows the distribution of weather stations across GB.

As one might worry that changes in light intensity are systematically related to changes in weather conditions, we amend the accident data with information about weather conditions at site. Hourly weather data comes from the Met Office's MIDAS database (Met Office, 2006b,a) and includes information on wind speed, precipitation and temperature. In order to make the weather data compatible with the accident data, we first calculate simple averages of all stations within each region and hour for all weather variables. These regional averages are then used to generate three variables with four (wind speed), two (precipitation) and six (temperature) categories.[8] Furthermore, we include a dummy variable to capture icy road conditions (i.e. positive precipitation and temperature below 2°C). As there are regions without operating weather stations during all or some time periods, the sample clearly decreases in size when we include weather information (see the right panel of Figure 5.1 for an overview about the distribution of relevant weather stations across Great Britain). In the following, we thus present results for both samples, the sample covering all observations from all regions during the whole observation period as well as the sample covering only those observations for which weather data is available.

---

[8]As shown in Table A5.3 in the appendix, using alternative cutoffs (quintiles and deciles) does not alter the results.

Although the STATS19 data does provide some information about lighting conditions at accident site, we cannot use it to derive a measure for darkness, since the raw data naturally provides such information only for regions (and hours) with at least one accident. Thus, if we would decide to make use of the information about lighting conditions, we had to discard a large share of potentially informative observations (i.e. observations with a zero accident count) and had to run our estimations on the truncated sample. Instead, we treat a region-hour as a dark/light hour based on information about the position of the sun relative to the horizon in this region and at this date and time. For each region and date, we first calculate the times when the sun's position is six degree below the horizon as seen from an observer who is located in the center of the region. This happens twice a day, once in the morning before sunrise and once in the evening after sunset and marks the beginning or ending, respectively, of civil twilight. For a detailed description of the procedure used here to calculate the position of the sun based on longitude, latitude, date and time see Cornwall et al. (2015). During civil twilight, no ray of sunlight touches the ground. Yet, higher air layers are still directly illuminated by the sun and diffract a relevant part of the sunlight to the ground. This indirect lighting during civil twilight is sufficient for the human eye to clearly distinguish objects. We thus make use of a definition of darkness that is based on the onset and offset of civil twilight rather than on sunrise and sunset. Similar definition of darkness have been used in the related literature (Sullivan and Flannagan, 2002) but also became part of UK law. The Road Vehicles Lighting Regulations 1984, for example, also refers to the beginning and ending of civil twilight when it defines the *daytime hours* as the time between half an hour before sunrise and half an hour after sunset and requires drivers to keep lamps lit during the *hours of darkness* (Department for Transport, 1984).

Based on the exact information about the beginning and ending of twilight at the center of each region, day and hour, we are able to define our main treatment variable for darkness: *dark* gives the fraction of the hour that is dark. Thus, it takes on the value 1 if the hour of observation is before the hour of transition from nautical twilight (dark) into civil twilight (light) in the morning or after the hour of transition from civil twilight into nautical twilight in the evening and the value 0 if the hour of observation is after the hour of transition into civil twilight in the morning and before the hour of transition out of civil twilight in the evening. If the hour of observation is instead the same as the hour of transition from or into civil twilight, *dark* can take on some value between 0 and 1, which depends on the minute of transition out of / into civil twilight.[9]

---

[9]Descriptive statistics are shown in Table A5.4 in the appendix.

5.2.2 Identification and Estimation

The basic intuition of our main empirical strategy is to make use of variation in darkness induced by variation in sunrise and sunset times across time and space to estimate the effect of darkness on road safety and then to simulate road safety under alternative time regimes. In order to be able to put the results of our main analysis into perspective, we follow the related literature and also present results based on a regression discontinuity (RD) design. The RD exploits the discrete change from standard winter time to DST in spring and vice versa in autumn and has been used (e.g. Smith, 2016) to directly estimate the effect of DST on road safety. The intuition behind this approach is that one might expect to see a sharp increase (decrease) in the number of accidents around the date of DST transition if DST affects road safety. Specifically, we estimate regressions based on Equation (5.1):

$$\ln accidents_{dt} = f(days_{dt}) + \pi post_{dt} + f(post_{dt} \times days_{dt}) + \eta_{dt} \tag{5.1}$$

where $accidents_{dt}$ is the total number of accidents (of a certain type and net of day of the week and year fixed effects)[10] at day d in year t, $days_{dt}$ is the running variable and denotes the number of days from DST transition (either in spring or in autumn), $post_{dt}$ is a dummy equal to one for days after DST transition and η_{dt} is an error term. The interaction between *post* and *days* is included to allow for different slopes at both sides of the cut-off. The parameter of interest in this setting is π and gives the effect of the transition into or out of DST on accident counts. To account for the fact that the length of the day is 23 instead of 24 hours on the day of DST transition in spring, we follow Smith (2016) and count accidents during the hour 3 in the morning twice. Correspondingly, we drop half of the accidents during the hour 2 on the day of DST transition in autumn. To further increase comparability of daily accident counts, we drop holidays[11] from the estimation sample.

The assumption required for π being consistently estimated is that in the absence of DST transition, *ln accidents* would change continuously in *days* around the transition date. If this assumption holds, comparing accident counts just before with accident counts just after the cut-off date yields a reasonable estimate of the short-run effect of DST on road safety. Note that this assumption is likely to hold, if people cannot manipulate treatment – as in this setting where the whole country is treated – and if there are no other rules that might affect the outcome differently at both sides of the cut-off. To our knowledge there are no other discontinuities around the DST transition

---

[10] To get rid of differences in accident counts by day of the week and year, we follow Smith (2016) and use the residuals from a regression of (logged) daily accident counts on day of the week and year fixed effects as dependent variable.

[11] January 1, Easter holidays (Good Friday until Easter Monday), Early May and Spring Bank Holidays as well as Christmas Holidays (December 25 and December 26).

dates that might have an effect on road safety. To determine the number of days to be used at both sides of the cut-off, we make use of mean squared error optimal bandwidth selectors (see Calonico et al., 2017).

While the RD provides causal estimates of the effect of DST *transition* on road safety under plausible assumptions and thus can be used to assess the costs of DST *transition*, its use is rather limited when it comes to the main goal of this paper, i.e. assessing the costs of alternative time regimes throughout the entire year. This is because the RD estimates are valid only very locally, namely just a few days around the date of transition into and out of DST and cannot be used easily to extrapolate the number of accidents under DST far away from the transition date, e.g. in January. Our main approach, which aims at assessing the long-run consequences of different time regimes, is therefore to estimate the effect of darkness on accident counts and then – using these estimates – to simulate the number of accidents under the different time regimes. This approach uses variation in darkness and accident counts throughout the entire year and thus is more suitable to assess the relative costs of establishing one time regime instead of the other.

To identify the effect of darkness on accident counts we estimate regressions based on the following equation:

$$accidents_{ihdt} = f(\beta dark_{ihdt} + X'_{ihdt}\gamma + \alpha_i + \rho_h + \delta_d + \lambda_t + \varepsilon_{ihdt}) \tag{5.2}$$

where $accidents_{ihdt}$ denotes the number of accidents (of a certain type) in region i during hour h on day d of year t, $dark_{ihdt}$ is the dark share in hour h in region i on day d of year t (see above), α_i are region, ρ_h hour of the day, δ_d day of the year, and λ_t year fixed effects. Finally, X_{ihdt} includes weather variables and indicator variables for the day of week and ε_{ihdt} is an error term.

The parameter β is the parameter of interest and gives the causal effect of darkness on hourly accident counts. To estimate β consistently the necessary assumptions have to hold, most importantly, *dark* has to be conditionally exogenous. We argue that there are good reasons to assume that *dark* is exogenous conditional on the fixed effects and the weather controls included in *X*. First, by including hour of the day fixed effects (ρ_h) we account for the fact that both darkness as well as accidents are distributed unevenly across hours. Darkness concentrates on some hours during night when people are at home and road traffic and thus accident counts are quite low (see Figure 5.2), while daylight concentrates on hours when people go about their daily tasks and streets are rather crowded. Second, by including day of the year fixed effects (δ_d) we avoid bias due to correlated changes in road and light conditions in the course of the year. While most hours are dark and road conditions can be hazardous due to snow, ice and fog during winter, the opposite is true for the summer months. Third, to capture that differences

Figure 5.2: Distribution of Accidents by Year, Week, Hour and Day of Week
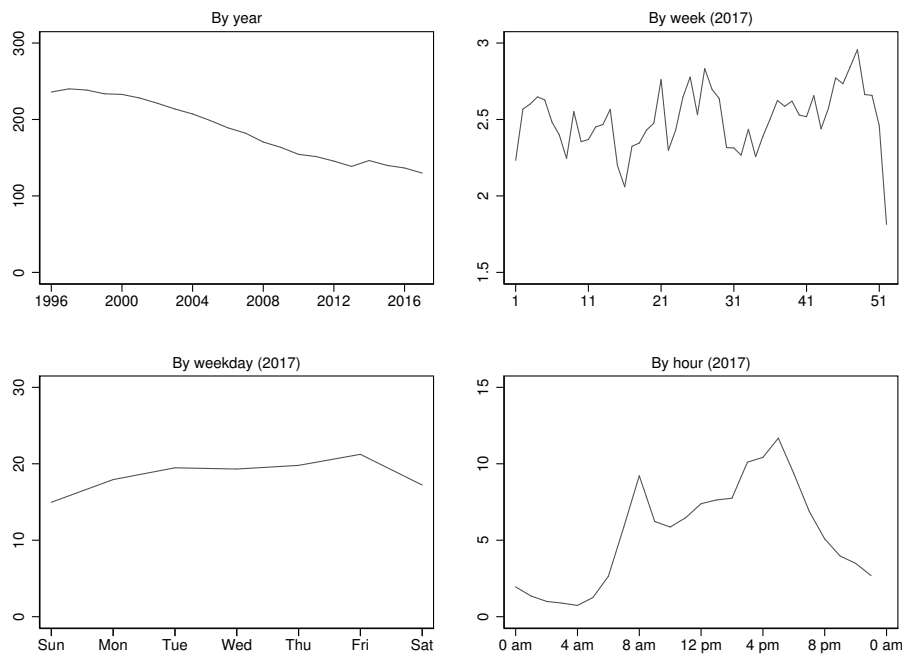
Notes: Own calculations and illustration based on STATS19 data. The figure shows the total number of accidents in thousands by year (upper left), the total number of accidents in thousands during the year 2017 by week (upper right), day of the week (lower left) and hour (lower right).

in the number of accidents might be related to differences in the share of dark hours between regions (i.e. northern vs. southern regions), we include region fixed effects (α_i). Finally, as changes in light conditions might be directly related to changes in weather conditions, we also estimate variations of Equation 5.2 where we control for weather conditions at accident site.

The variation in *dark* that is not captured by one of the fixed effects is used to identify β. It comes, as will be shown in the following figures, from variation in the offset and onset of darkness across time and space. The first figure, Figure 5.3, shows the timing of transition from darkness to daylight in the morning (upper two panels) and from daylight to darkness in the evening (lower two panels) over the course of the year. In the left two panels we distinguish between two places on the same line of latitude: one in the very east of Great Britain (dashed line) and one in the very west (solid line). Similarly, in the right panels we differentiate between two places on the same line of longitude: one in the very north (dashed line) and one in the very south (solid line). What can be seen immediately are three sources of variation in darkness: changes in the timing of onset and offset of darkness throughout the year and by longitude (east-west comparison) as well as latitude (north-south comparison).

Figure 5.4 shows for selected hours of the day, days of the year and regions how this variation in the timing of onset and offset of darkness translates into variation in darkness. The shaded areas in each panel give the proportion of the respective hour

## Figure 5.3: Variation in Offset and Onset of Darkness
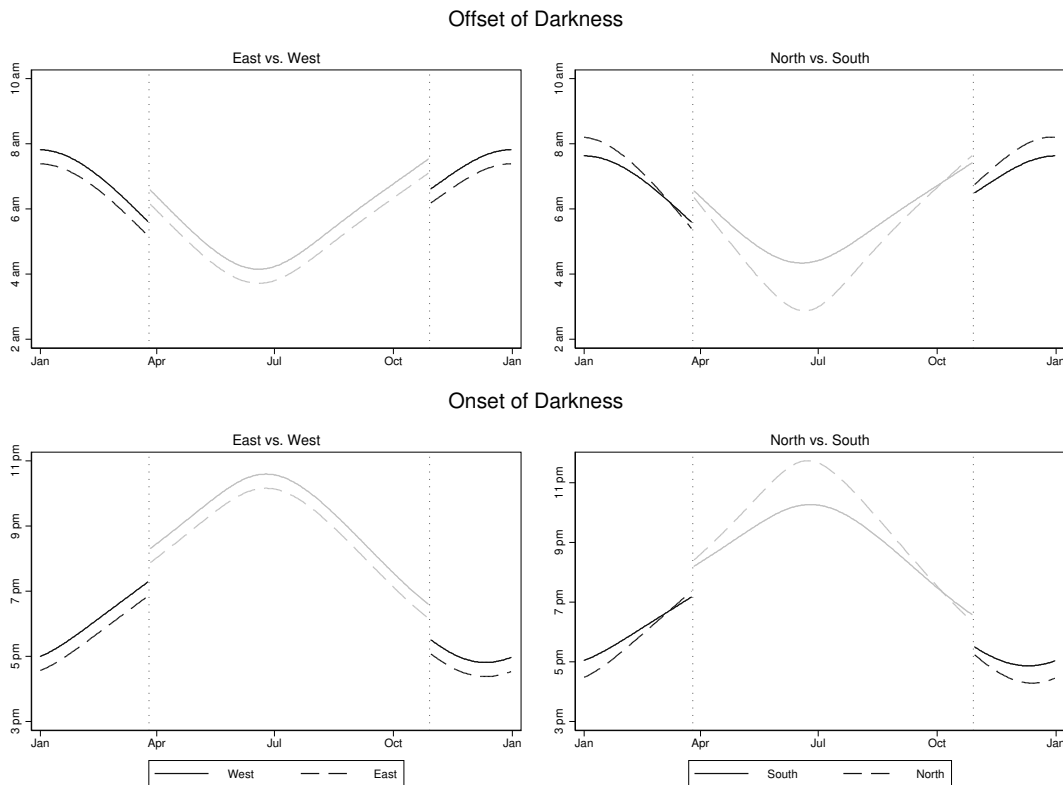
Notes: Own calculations and illustration. The figure shows the timing of offset (upper panels) and onset (lower panels) of darkness at four different places in GB and over the course of the year (2017). The panels on the left side differentiate between two places on the same line of latitude (52.25° latitude), a place in the very west (-5.25° longitude, solid line) and a place in the very east (1.25° longitude, dashed line) of the country. Correspondingly, the panels on the right side differentiate between two places on the same line of longitude (-4.25° longitude), a place in the very south (50.25° latitude, solid line) and a place in the very north (58.25° latitude, dashed line).

that is dark. The hour five in the morning, for example, is completely dark until the end of February. Then, however, the dark proportion of this hours starts to decrease such that by March, 25 at least 30 percent of this hour is light all across Great Britain. As all clocks are then set one hour ahead due to daylight saving time transition,[12] the dark proportion of hour five jumps back to 100 percent but then decreases again. Similar variation in darkness across time can be found for other hours of the day but at different days of the year, e.g. during October, November and December for the hour four in the afternoon.

Aside from variation in darkness for a given hour and region across day of the year there is also variation in darkness for a given hour and day of the year across regions. As can be seen in the left part of Figure 5.4, the dark proportion of an hour is 0-50 percentage points higher in the morning in regions of the very west of Great Britain compared to regions (on the same line of latitude) in the very east. The opposite is true for hours in the evening. The left panel of the graph shows that the dark proportion of

---

[12]Note that our estimates are not sensitive to the exclusion of observations from two weeks after daylight saving time transition and thus do not reflect effects of daylight saving time *transition*.

Figure 5.4: Variation in Darkness used for Identification
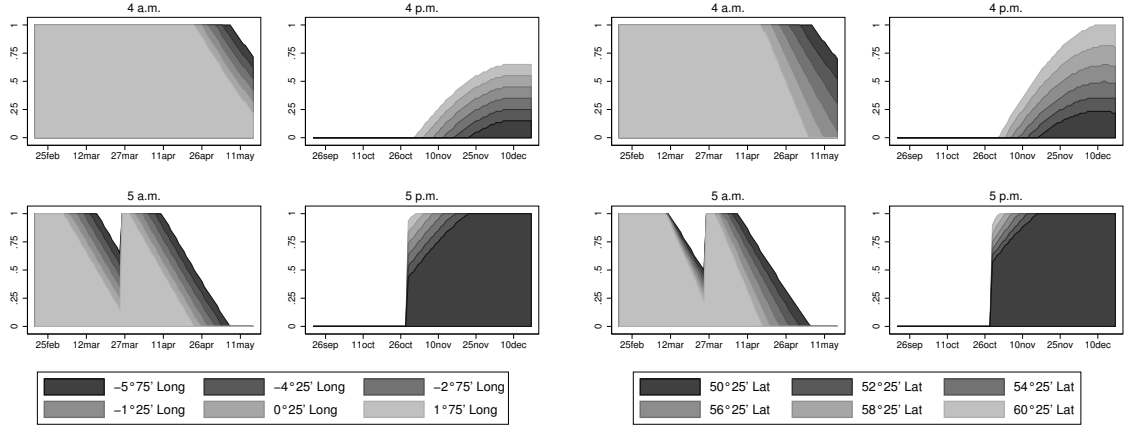
(a) East vs. West   (b) North vs. South

Notes: Own calculations and illustration. The figure shows the dark proportion of an hour by day of the year and for selected hours and places. The four panels on the left side differentiate between six places on the same line of latitude but on different lines of longitude (the darker the colour the more in the west), the four panels on the right differentiate between six places on the same line of longitude but different lines of latitude (the darker the colour the more in the south).

an hour varies also by latitude. By isolating these three parts of variation in darkness to estimate β and additionally controlling for weather conditions, we are confident to give our estimation results a causal interpretation.

Equation 5.2 is estimated using a negative binomial (NB) model, as our outcome variables are count variables with a large proportion of zeros and show signs of overdispersion. In order to avoid the well known incidental parameter bias problem, we estimate the NB fixed effects models by including full sets of dummies for regions, hours, days of the years and years. Although there is no formal proof available that this approach rules out inconsistencies due to the incidental parameter bias problem, simulations by Allison and Waterman (2002) and Greene (2004) suggest that the bias should be small even if the number of time periods is small (Cameron and Trivedi, 2015). Given that the number of time periods is rather high in our application, the incidental parameter bias problem should not pose a threat here, anyway. Nevertheless, we also estimated OLS regressions. The OLS results are mostly quite similar to the NB results, but less precisely estimated. We thus focus in the following on the NB results and present OLS estimates in the appendix.

5.3   Results

5.3.1   Main Results

Before turning to the results of our main analysis, we show the results from the replication part of our analysis in Table 5.1. The corresponding graphical representation to

Table 5.1: RD Estimates of DST Transition on Accidents

Bandwidth selector	Spring Transition		Fall Transition	
	one	two	one	two
ln # all accidents	-0.028 (0.022)	-0.012 (0.017)	0.037* (0.021)	0.013 (0.020)
Observations	466	990	638	638
ln # slight accidents	-0.043* (0.022)	-0.004 (0.018)	0.031 (0.022)	0.011 (0.021)
Observations	544	1,034	682	638
ln # serious accidents	0.025 (0.030)	0.016 (0.025)	0.021 (0.027)	0.034 (0.025)
Observations	754	1,100	770	792
ln # all accidents	0.083 (0.055)	0.099* (0.056)	0.013 (0.055)	0.051 (0.052)
Observations	1,830	1,659	1,071	1,441

Notes: Own calculations based on STATS19 data. Each estimate gives the effect (semi-elasticity) of DST transition on the number of accidents in the respective category and comes from a separate local linear regression based on Equation 5.1. Bandwidth selectors are either a common mean squared error (MSE)-optimal bandwidth selector (denoted by one) or two different MSE-optimal bandwidth selectors below and above the cut-off (denoted by two). Robust standard errors are in parentheses. Significance levels: * p<0.1; ** p<0.05; *** p<0.01.

the RD estimates can be found in Figure A5.2 in the appendix. The first two columns of Table 5.1 report the estimates for the spring transition, while column three and four show the results for the fall transition. We do not find clear evidence that the transition into or out of DST increases accident counts. Only the RD estimates and graphs for fatal accidents consistently suggest that DST transition adversely affects road safety. According to the point estimates entering DST increases the number of fatal accidents by 8-10 percent. While both estimates are of considerable size and in line with the results of Smith (2016), who finds that entering DST increases the number of fatal crashes in the US by around 6 percent, only one of them is marginally significantly different from zero ($p < 0.1$). The results for the other accident types as well as the results for the total number of accidents rather point to small or zero effects of DST transition on road safety.

Table 5.2 shows the baseline results of our main analysis. The first column gives the estimated coefficients of darkness on the total number of accidents and the number of accidents by accident severity for all regions. Columns 2 and 3 present the estimates for the same outcomes but the sample including only observations for which weather information is available. While column 2 gives the estimates from the regression including the standard controls, column 3 shows the results for our preferred specification in which we additionally control for weather conditions at accident site.

Table 5.2: NB Estimates of the Effects of Darkness on Accidents

	All regions		Regions with weather data			
Weather controls	no		no		yes	
# all accidents	0.071***	(0.006)	0.072***	(0.006)	0.072***	(0.007)
# slight accidents	0.062***	(0.006)	0.063***	(0.007)	0.062***	(0.007)
# serious accidents	0.108***	(0.007)	0.111***	(0.008)	0.113***	(0.008)
# fatal accidents	0.339***	(0.018)	0.315***	(0.021)	0.318***	(0.022)
Observations	51,103,130		21,352,015		21,352,015	

Notes: Own calculations based on STATS19 and MIDAS data. Each estimate gives the effect (semi-elasticity) of darkness on the number of accidents of the respective category and comes from a separate negative binomial regression. All regressions include region, year, day-of-the-year, day-of-the-week, and hour fixed effects. Standard errors clustered at region level are in parentheses. Significance levels: * p<0.1; ** p<0.05; *** p<0.01.

The results from our preferred specification, shown in column 3, suggest that darkness increases the total number of accidents for a given hour by 7.2 percent. As the effect of darkness is larger for the number of fatal (+31.8 percent) and serious (+11.3 percent) than for the number of slight accidents (+6.2 percent), darkness does not only seem to affect accident frequency considerably but also to increase accident severity. The corresponding estimates derived from OLS regressions, though less precisely estimated and partly smaller, point in the same direction and are shown in Table A5.5 in the appendix.

Our results are robust to the inclusion of weather controls. There are no relevant differences, neither with respect to size nor to statistical significance, between the estimates presented in columns 2 and 3. For both specifications and all outcomes we find relatively large positive and significant effects of darkness on accident counts. This finding also holds when we use the entire sample (column 1), i.e. also observations for which weather information is not available.

Using simulations based on the results for all regions, we can now address two questions. First, how many accidents per year have been caused by darkness in England, Scotland and Wales under the current regime with Greenwich Mean Time (GMT) during winter and DST during summer? This number shows how many accidents could potentially be avoided if darkness-related accidents could be brought to zero. The second, related question is, whether having more light in the evening than in the morning can help to bring down accident counts, i.e. whether setting the clocks permanently to DST can prevent at least some accidents that would happen under an all-year GMT regime. Our approach to answer these questions is to predict the number of accidents in the hypothetical situations without darkness as well as under an all-year GMT and an all-year DST regime by adjusting the darkness variable accordingly.[13] The resulting accident counts are then multiplied with the following average accident costs for 2017 as provided by the Department for Transport (2018a) to broadly assess the social costs: £25,451 for a slight accident; £243,635 for a serious accident; £2,130,922 for a fatal accident. As these costs are quite low compared to estimates used in the related literature,[14] our results for the social costs rather represent lower bounds of the true costs.

The credibility of this simulation approach mainly rests on the assumption that darkness impacts uniformly over the course of the day. This assumption, however, might be violated, e.g. if darkness interacts with fatigue or traffic density. We thus do not only present simulation results for the baseline models allowing only for one uniform effect of darkness, but also present results where we relax this assumption. Specifically, we also show simulated accident counts based on regressions which allow for different effects during early morning (hours 0-4 a.m.), late morning (hours 5-8 a.m.), early evening (hours 3-6 p.m.) and late evening (hours 7-11 p.m.).[15] The underlying regression results

---

[13]E.g. to get the predicted accident count under an all-year DST regime, we determine sunrise and sunset times in this hypothetical situation, adjust the value of dark accordingly and then predict accident counts based on the regression results presented in column 1 of Table 5.2 and Table 5.3

[14]Smith (2016), for example, assumes average social costs of $4-$10 million per fatality.

[15]We abstain from presenting simulation results from even more flexible models (e.g. where darkness is interacted with individual hours), as it is not clear what some of the resulting coefficients – namely those for the very early/late morning/evening hours – identify. This is because the whole variation in darkness during these hours comes from regions in the very north, which is problematic here for several reasons. First, as these regions are located on rather small islands there is neither east-west nor north-south variation in darkness for these hours. Second, given that these hours turn light/dark only during a few days in June/December, there is also very little day-by-day variation. Finally, these islands are located far away from the mainland, are very sparsely populated and thus might exhibit very

Table 5.3: NB Estimates of the Effects of Darkness on Accidents by Time of Day

	Morning hours		Evening hours	
	Hours 0-4 a.m.	Hours 5-8 a.m.	Hours 3-6 p.m.	Hours 7-11 p.m.
# slight accidents	0.078*** (0.028)	0.274*** (0.017)	0.064*** (0.005)	0.013 (0.013)
# serious accidents	0.136*** (0.040)	0.351*** (0.021)	0.138*** (0.010)	0.022 (0.014)
# fatal accidents	-0.053 (0.097)	0.428*** (0.044)	0.472*** (0.028)	0.214*** (0.027)

Notes: Own calculations based on STATS19 data. Each estimate gives the effect (semi-elasticity) of darkness on the number of accidents of the respective category at the particular hours. All three regressions are based on 51,103,130 observations and include region, year, day-of-the-year, day-of-the-week, and hour fixed effects. To allow the effect of darkness to vary over the course of the day, we interact darkness with time of the day. Standard errors clustered at region level are in parentheses. Significance levels: * p<0.1; ** p<0.05; *** p<0.01.

for this more flexible specifications are shown in Table 5.3 and indicate that the effect of darkness is larger during busy hours in late morning and early evening. For slight and serious but not for fatal accidents, the effect also seems to be larger on average during morning hours than evening hours.

The first two columns of Table 5.4 show the observed and predicted number of accidents by accident class for the status quo with GMT during winter and DST during summer. In columns 3-6 we distinguish between two cases: results from the baseline model are marked by *M1* in the column header and results from the more flexible model allowing for different effects of darkness over the course of the day are marked by *M2*. Columns 3 and 4 report how many accidents could be prevented in the hypothetical situation where none of the hours is dark. The remaining results of the simulation shown in columns 5 and 6 address the question whether setting the clocks permanently to DST instead of setting them to GMT for the whole year can save lives and prevent injuries.

Table 5.4: Annual Accident Counts and Simulated Preventable Accidents

| | Current GMT/DST | | Preventable accidents | | | |
| | | | All hours light | | DST vs. GMT | |
	Observed	Predicted	M1	M2	M1	M2
Slight	157,542	157,641	2,202	1,887	358	-487
Serious	26,246	26,247	768	646	109	-46
Fatal	2,490	2,490	275	149	26	30

Notes: Own calculations based on STATS19 data. The table shows the average annual (preventable) number of accidents by time regime. The first column (*Observed*) gives the observed number of accidents by accident severity and the following column (*Predicted*) gives the predicted number of accidents under the Status Quo with GMT during winter and DST during summer. The remaining columns show the estimated number of accidents that could be prevented if all hours were light (columns *All hours light*) and by setting the clocks to DST during the entire year instead of setting them to GMT year-round (columns *DST vs. GMT*), where the column header *M1* marks the results based on the baseline model and *M2* those based on the model allowing the effect of darkness to vary by time of the day. Positive/ negative numbers in the last two columns indicate more/ less preventable accidents under an all-year DST then under an all-year GMT.

Depending on the underlying model, we estimate that darkness causes around 150-275 fatal, 650-750 serious and 1,900-2,200 slight accidents per year or annual total costs of £520-830 million. Thus, measures that can reduce accidents due to darkness (almost) to zero would have a positive net value only if the costs to implement these measures would not exceed around £500-800 million per year. Our results provide evidence supporting the view that setting the clocks permanently to DST instead of setting them to GMT for the whole year can save lives and reduce costs. While there is some ambiguity with respect to the predicted number of slight and serious accidents, the

different traffic conditions. Thus, using these estimates in the simulation and applying them to other regions (with quite different traffic situations) might yield very misleading results. On the other hand, differentiating only between morning and evening hours seems to be too restrictive (compare Figure A5.3 in the appendix).

results for fatal accidents, which cause the bulk of accident related costs, are clear cut. We find that by setting DST instead of GMT year-round, at least 25 fatal accidents could be prevented. Multiplying the estimated number of preventable accidents of each severity type with the average costs of the respective type, we arrive at an estimated annual cost savings potential of £40-90 million under the all-year DST regime compared to the GMT regime, representing 7.7-11 percent of all accident costs caused by darkness.

Given that the simulation results based on the model including the interaction term require less restrictive assumptions, we refer to these results and thus the more conservative cost saving potential of £40 million in the following. Although this seems to be not too much, one has to bear in mind that these £40 million are only due to the long term effects of DST on road safety, i.e. are the consequence of shifting light from the quiet morning hours to the rather busy afternoon hours. Abolishing the yearly transition into and out of DST has the potential to prevent additional road casualties that are due to low concentration and fatigue following the clock change.

5.3.2 Mechanisms

While the simulation reveals that establishing DST as the standard time throughout the year could prevent a considerable number of fatal accidents, the general mechanisms of the effect of darkness on accident counts are unknown. Although we cannot disentangle these mechanisms unambiguously with the data at hand, we exploit some features of the data to provide suggestive evidence in favor of one or the other channel. Theoretically, there are at least three potential channels through which ambient light conditions might affect road safety. First, darkness might influence vision and thus the ability to recognize other road users early enough to prevent a collision. Second, natural light conditions influence the circadian rhythm. Thus, darkness might increase accident risk by increasing fatigue and reducing concentration. Third, people might expect driving to be more dangerous during darkness than during daylight. Consequently, they avoid driving during darkness, especially if they consider themselves as insecure drivers, which in turn might reduce accident counts during darkness.[16] As we find *positive* effects of darkness on accident counts, we can rule out that such behavioral responses are the main channel through which the effect operates.

Depending on the underlying mechanism, politicians might want to establish different measures to avoid at least some of the accidents that are due to poor light conditions. If, for example, darkness decreases road safety due to reduced vision of each and every driver during night, an option to reduce accidents might be to increase road lighting. However, it might also be that the risk of poor night vision is not the same for everybody

---

[16]Note, that this does not imply any kind of selection that renders our results inconsistent, but would be a mechanism and thus a part of the effect.

but starts to increase at the age of 45, as the medical literature suggests (Darius et al., 2018). In case that this is the main driver of the effect, a more cost-efficient way to reduce road casualties might then be to identify drivers with impaired night vision, e.g. by establishing compulsory night vision screenings for older drivers. If, instead, the main driver of the effect is an increase in fatigue and poor concentration during darkness, neither an expensive expansion of road lighting nor visions screenings would help to reduce road casualties.

We cannot provide direct evidence for any of the mechanisms proposed above, as we neither have information about the vision of road users nor about their physical and mental constitution. However, we can use the information on the timing and about the persons involved in the accident to see whether one of the mechanism is likely to be more important than others. Our first approach is the following: We again estimate the effect of DST transition on accident counts using a RD, but now look only at hours that were light/dark before DST transition and turned dark/light due to the transition. For the spring transition into DST these are the hours 5 and 6 in the morning and the hours 7 and 8 in the evening. For the autumn transition back to standard time these are the hours 6 and 7 in the morning and 6 and 7 in the evening. If vision is a key mechanism, we would expect a positive effect of DST transition in spring for the morning hours that turn dark (again) due to the transition and a negative effect for the evening hours that turn light. For the autumn transition back to standard time we would expect exactly the opposite, i.e. a decrease in accidents for morning hours and an increase for evening hours. Table 5.5 gives the results of this analysis. While all significant estimates show the expected signs, those pointing in the wrong direction are statistically not different from zero. Thus, the general impression is pretty much in line with our expectations and indicates that vision is an important mechanism.

This impression is also supported by the results of a series of regressions (Table 5.6), where we estimate the effect of darkness on the number of accidents involving pedestrians and the number of accidents differentiated by age of the driver. The intuition behind this second approach is the following: If vision is an important mechanism, one would expect that darkness has a larger effect i. on the number of pedestrian accidents, as especially small and unlit objects are difficult to spot during night, and ii. on the number of accidents involving older drivers, as the risk of poor night vision increases in age. Indeed, compared to the baseline effect (7 percent), we find quite large effects for pedestrian accidents (12 percent) and accidents involving drivers above the age of 45 (between 14 and 21 percent). The estimates for accidents involving only younger drivers on the other hand are either rather small and negative or not significant.[17] The results of interaction-like regressions show that the effect on pedestrian accidents is

---

[17]We cannot provide an unambiguous explanation for the negative and significant estimate for drivers younger than 25. Note that this is a very specific subgroup since every actively involved person must be 25 or younger. One speculative explanation could be that this group of drivers is less frequent on the

Table 5.5: RD Estimates of DST Transition on Accidents during Affected Hours

	Spring Transition		Fall Transition	
	Morning	Evening	Morning	Evening
ln # all accidents	0.091* (0.054)	-0.115*** (0.032)	0.038 (0.043)	0.136*** (0.028)
Observations	1,042	832	814	682
ln # slight accidents	0.129** (0.065)	-0.104*** (0.037)	0.070 (0.044)	0.212*** (0.019)
Observations	1,084	793	858	1,386
ln # serious accidents	-0.057 (0.064)	-0.082 (0.052)	-0.145** (0.070)	0.254*** (0.044)
Observations	1,508	1,670	1,029	1,209
ln # fatal accidents	-0.033 (0.058)	0.037 (0.055)	0.006 (0.056)	0.076 (0.060)
Observations	614	834	794	1184

Notes: Own calculations based on STATS19 data. Each estimate gives the effect (semi-elasticity) of DST transition on the number of accidents of the respective category during those hours in the morning/evening that turn light/dark due to the DST transition and comes from a separate local linear regression based on Equation 5.1. Bandwidth selected using common mean squared error (MSE)-optimal bandwidth selector. Robust standard errors are in parentheses. Significance levels: * p<0.1; ** p<0.05; *** p<0.01.

Table 5.6: NB Estimates of the Effects of Darkness on Additional Accident Categories

# accidents involving pedestrians	0.118***	(0.015)
# accidents by age of oldest driver		
$0 < \text{age} \leq 25$	-0.040***	(0.011)
$25 < \text{age} \leq 45$	0.001	(0.010)
$45 < \text{age} \leq 65$	0.143***	(0.007)
$65 < \text{age}$	0.210***	(0.017)
# accidents involving pedestrians & age of oldest driver < 45	0.028*	(0.016)
# accidents involving pedestrians & age of oldest driver > 45	0.251***	(0.016)

Notes: Own calculations based on STATS19 and MIDAS data. Each estimate gives the effect (semi-elasticity) of darkness on the number of accidents of the respective category and comes from a separate negative binomial regression. All regressions use 21,352,015 observations and include region, year, day-of-the-year, day-of-the-week, and hour fixed effects as well as weather controls. Standard errors clustered at region level are in parentheses. Significance levels: * p<0.1; ** p<0.05; *** p<0.01.

mainly driven by accidents that feature both "risk types" older drivers and pedestrians (25 percent), while the effect for accidents also involving pedestrian but only younger drivers is quite small (3 percent).[18]

5.3.3 Sensitivity Analyses

We provide sensitivity analyses in the appendix. By employing different aggregation schemes, we show that our results are not sensitive to the sources of variation in darkness (north-south, east-west and over the year) used for identification. We also present estimates from specifications where we interact darkness with weather conditions. The results reveal that the effect of darkness is not driven by the coincidence of darkness and hazardous road conditions.

5.4 Conclusion

In a time when the political debate about the usefulness of DST flares up again, a thorough knowledge of all potential costs and benefits associated to the alternative time regimes is essential for policy makers who consider to abolish the biannual clock change. Since the choice of the time regime affects the distribution of light and darkness throughout the day, it has been hypothesized that establishing an all-year DST regime could prevent road casualties as this would shift light from the morning hours to the afternoon hours when streets are crowded and accident risk is high. To quantify the

---

streets during darkness. Although this might not be the case on weekend days, it is conceivable that young drivers are less frequent on the streets during darkness on the remaining weekdays.

[18]Figure A5.4 in the appendix shows the estimated effects of darkness on accident counts differentiated by year. Although there have been advances in vehicle lighting and vehicle safety technology, we do not find evidence that the effect of darkness on accident risk has diminished over time.

number of accidents and the associated costs under the alternative time regimes, we estimated how darkness affects accident counts and used the resulting estimates to simulate the number of accidents under the two time regimes.

We find that darkness considerably affects road safety. Our results show that darkness increases accident counts for a given hour on average by around 7 percent. This implies that darkness causes around 150 fatal, 600 serious and 1,800 slight accidents per year or total annual costs of at least £500 million in England, Scotland and Wales. Comparing the simulated accident counts under the different time regimes, our results suggest that road safety is indeed somewhat higher under an all-year DST than under an all-year standard time. According to our estimates, establishing DST throughout the year could reduce social costs related to darkness by around 8 percent per year compared to a situation with all-year standard time. These numbers result only from shifting daylight from the morning hours to the evening hours and do not include the number of prevented accidents that are due to the abolishment of the transitions, especially the spring transition. The latter has been reported to be quite substantial, at least for fatal accidents, with an increase of 6 percent for the US (Smith, 2016) and 8 percent for GB.

To put the effect size into perspective, we apply the same back-of-the-envelope calculation as Smith (2016).[19] Assuming that the estimated effect of the transition into DST on fatal accidents persists for six days – until the circadian rhythm has adjusted after the clock change – one would expect that abolishing the clock change would prevent around 3 fatal accidents each year. Compared to the estimated annual saving potential due to changes in light conditions (around 25 fatal accidents), the absolute number of prevented accidents due to disruptions of the circadian rhythm seems rather low, indicating that – other than hypothesized by Smith (2016) – the distribution of natural light across hours of the day still seems to matter for road safety. This interpretation is also consistent with the results from the heterogeneity analysis, where we found that the effect of darkness on accident counts likely operates through a reduction in vision during darkness. Consequently, by solely focusing on the (spring) clock change, one runs the risk of significantly underestimating the total costs of the current time regime.

We acknowledge that our analysis covers only one, though, as we believe, important aspect of time regime choice and emphasize that an informed decision must take into account all potential benefits and side-effects of time regime choice. For example, it is often argued that setting DST year-round leads to an increase in mental illnesses and learning difficulties in school. Furthermore, we acknowledge that the simulation results do not necessarily hold for other countries, especially if these countries are not at the same latitudes and thus have different sunrise/sunset times or if e.g. peak traffic

---

[19]To calculate the number of fatal accidents that are due to the clock change in spring, we multiply the estimated coefficient of the spring transition (0.08) with the average number of fatal accidents per day on Sundays and weekdays in March and April (6.2) and the number of days (6). We look at fatal accidents only, since costs due to fatal accidents constitute the largest part of total costs.

times and road conditions differ fundamentally. Finally, people might adjust to the new sunrise and sunset times under permanent DST or permanent GMT. We cannot rule out such behavioral responses in the long run, however, this should if any alleviate but not reverse our main results.

Against the ongoing debate about which time regime should be implemented, our results challenge the notion of an all-year GMT standard time, at least from a road safety perspective. This is in line with results from previous empirical research that emphasizes the benefits of having more light during afternoon and evening hours, such as increased physical activity (Wolff and Makino, 2012) and reduced crime (Doleac and Sanders, 2015), and thus support the call for an all-year DST regime. Finally, adverse effects in various dimensions caused by the biannual transitions could be avoided.

Appendix

Sensitivity Analyses

The basic intuition behind our empirical approach is to capture the endogenous part of variation in darkness and accident counts by a set of fixed effects so that the remaining, arguably exogenous part of variation can be used for estimation of β. Specifically, as shown in Section 2.2, we have used three sources of variation in darkness to estimate the effect of darkness on accident counts so far: day-by-day variation for a given region and hour; east-west variation for a given day of the year and hour; and finally, north-south variation for a given day of the year and hour. While there are no obvious reasons why this approach should produce inconsistent results – at least after controlling for weather conditions at accident site – one might nevertheless worry that one or several of these parts of variation are endogenous. To address this concern, we show the results of a sensitivity analysis where we eliminate one (two) source(s) of the variation and use the remaining two (one) source(s) to estimate the effect of darkness. The idea behind this approach is to compare the resulting estimates to each other and the baseline results. If they do not differ too much from each other, we can be more confident that it is really darkness that drives our results.

In order to exclude one source of variation, we did not aggregate in cells as in the baseline specification, but either in vertical or in horizontal strips of width $0.5°$ longitude or latitude, respectively.[20] The resulting data provides information about the number of accidents (of the respective type), the mean proportion of the hour that is dark as well as about weather conditions for each hour and every strip (either along the lines of longitude or along the lines of latitude).

For the specification in which we eliminate day-by-day as well as east-west variation in darkness, we sum both the number of accidents and the treatment variable *dark* by region and day. By doing so we create a new treatment variable which gives the number of hours that are dark during the respective day in the respective region.[21] Thus, the interpretation of the estimates in this specification differs slightly from the interpretation of the other estimates, as it refers to daily instead of hourly accident counts.

Table A5.1 gives the results of this sensitivity analysis. Column (I) repeats the baseline results from Table 2 where we use all three sources of variation. The results for the

---

[20]An alternative way to separate the different sources of variation, would be to include interactions between day of the year and hour (capturing day-by-day variation), day of the year, hour and longitude (capturing east-west variation), and day of the year, hour and latitude (capturing north-south variation). This approach, however, requires estimation of a huge number of coefficients and thus is impracticable in our application.

[21]Note that for some regions weather information is not available for all hours of the day. As we only sum over hours of a region if weather information is available for all hours of a day, the number of observations in column 4 is smaller than 1/24 of the original sample size.

specification where we only use day-by-day and east-west variation but exclude north-south variation are shown in column (II). The results for the specification where we only use day-by-day and north-south variation are given in column (III). Finally, column (IV) shows the estimates for the specification where we exclude both east-west and day-by-day variation in darkness and accident counts and only rely on north-south variation to estimate the effect of darkness on road safety.

We find that the results are quite robust to the different aggregation approaches and do not substantively change if the sources of variation in darkness that are used to estimate β change. The point estimates are all positive, highly significant and vary between 3.6 and 7.6 percent.

Table A5.1: NB Estimates of the Effect of Darkness on Accidents using Different Sources of Variation

	(I)	(II)	(III)	(IV)
# all accident	0.072***	0.071***	0.076***	0.036***
	(0.007)	(0.009)	(0.010)	(0.011)
Variation				
Day-by-day	yes	yes	yes	no
East-west	yes	yes	no	no
North-south	yes	no	yes	yes
Observations	21,352,015	3,507,364	4,030,227	830,620

Notes: Own calculations based on STATS19 and MIDAS data. Each estimate gives the effect (semi-elasticity) of darkness on the total number of accidents and comes from a separate negative binomial regression. All regressions include region, year, day-of-the-year, day-of-the-week, and hour fixed effects as well as weather controls. Standard errors clustered at region level are in parentheses. Significance levels: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

A potential problem with the aggregations schemes along lines of latitude or longitude is that eastern regions (e.g. the region between 0.5° and 0.0° longitude) are on average located more in the south and northern regions (e.g. the region between 56.0° and 56.5° latitude) are on average located more in the east. Thus, if we aggregate all regions along lines of latitude or longitude we cannot perfectly separate north-south variation in darkness from east-west variation and vice versa. To check whether this is a problem we also estimate regressions where we follow the same aggregation approach but restrict the sample to regions between 50.75° and 52.75° latitude (when using east-west variation) and -3.75° and -2.25° longitude (when using north-south variation). The results of this auxiliary regressions are similar to the results for the entire sample and are shown in Table A5.2.

In our main specification we control for weather conditions at site by four categorical variables with four (wind speed), two (precipitation), six (temperature) and two (icy road) categories to avoid bias due to a correlation between weather conditions and

Table A5.2: NB Estimates of the Effect of Darkness on Accidents using Different Sources of Variation – Restricted Sample

	(I)	(II)
# all accident	0.093***	0.062***
	(0.005)	(0.014)
Variation		
Day-by-day	yes	yes
East-west	yes	no
North-south	no	yes
Observations	2,740,436	3,248,049

Notes: Own calculations based on STATS19 and MIDAS data. Each estimate gives the effect (semi-elasticity) of darkness on the total number of accidents and comes from a separate negative binomial regression. All regressions include region, year, day-of-the-year, day-of-the-week, and hour fixed effects as well as weather controls. Standard errors clustered at region level are in parentheses. Significance levels: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

darkness. This, however, does not exclude the possibility that the estimated effect of darkness is driven by the coincidence of darkness and specific weather conditions, in particular hazardous road conditions. To test whether this is the case, we re-estimate our main specification, but interact darkness with weather controls and calculate the effects of darkness for different weather conditions.

Figure A5.1 shows the estimated effects of darkness for different sub-samples resulting from stratifying the original sample by wind (rows), rain (columns) and temperature (x-axes). The estimates show very similar patterns across the four wind categories, which remain stable when further differentiating by rain/ no rain. Along the different categories of temperature, we observe positive and within each stratum quantitatively similar effects except for temperatures $< 0^\circ$C. For the latter we find negative point estimates, roughly half of which are statistically not significant. More importantly, the negative sign points into the unexpected direction as it suggests that darkness reduces the number of accidents when road conditions are dangerous. This might be explained by adjustments in driving behavior, e.g. more cautious and foresighted driving or even not driving at all, when road conditions are hazardous. In sum, we find no evidence that the estimated average effect of darkness on accident counts is the result of specific weather conditions.

Figure A5.1: The Effects of Darkness on Accidents by Weather Conditions
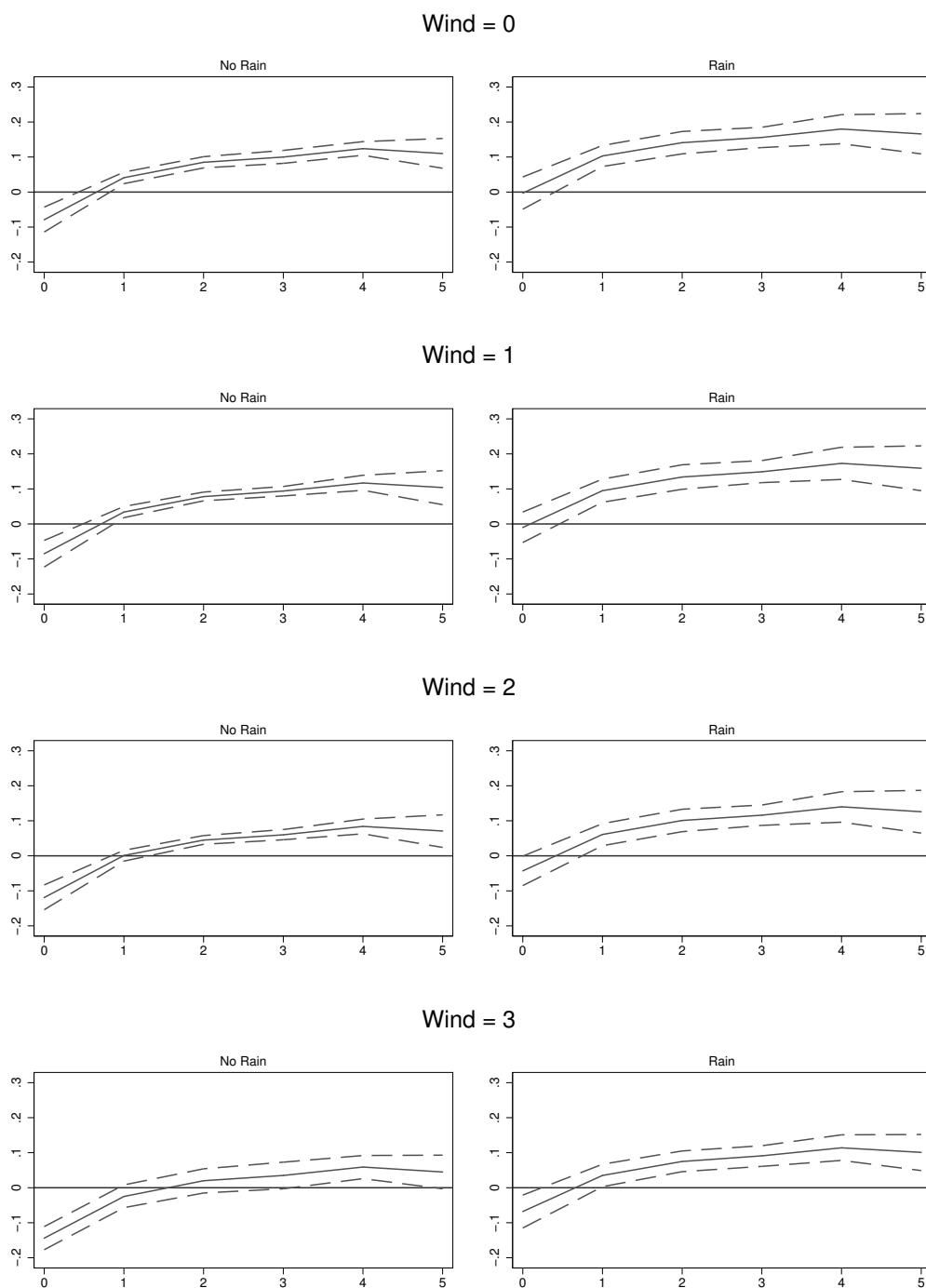
Notes: Own calculations and illustration based on STATS19 and MIDAS data. The results originate from a negative binomial regression of the total number of accidents on all control variables, where dark is interacted with all weather variables. The graphs show the effects (semi-elasticities) of darkness on the total number of accidents including the 95% confidence interval for different weather conditions. Rows (columns) represent the four (two) wind (rain) categories. The five temperature categories are assigned to the x-axes.

Additional Tables

Table A5.3: Using Alternative Cutoffs to Define Weather Variables

Categories (weather) based on	baseline		quintiles		deciles	
# all accidents	0.072***	(0.007)	0.074***	(0.007)	0.072***	(0.007)
# slight accidents	0.062***	(0.007)	0.064***	(0.007)	0.063***	(0.008)
# serious accidents	0.113***	(0.008)	0.115***	(0.008)	0.115***	(0.008)
# fatal accidents	0.318***	(0.022)	0.321***	(0.021)	0.321***	(0.021)
Observations	21,352,015		21,352,015		21,352,015	

Notes: Own calculations based on STATS19 and MIDAS data. Each estimate gives the effect (semi-elasticity) of darkness on the number of accidents of the respective category and comes from a separate negative binomial regression. All regressions include region, year, day-of-the-year, day-of-the-week, and hour fixed effects as well as weather controls. The first column repeats the baseline specification. In column two and three, the cutoffs used to define categories of the weather variables are based on quintiles or deciles, respectively. Standard errors clustered at region level are in parentheses. Significance levels: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

Table A5.4: Descriptive Statistics

Variable	Mean	Std. dev.	Min	Max
Dependent variables:[+]				
All accidents	0.080	0.386	0	20
Fatal accidents	0.001	0.033	0	3
Serious accidents	0.011	0.112	0	6
Slight accidents	0.068	0.346	0	19
Accidents with age of oldest driver ≤ 25	0.009	0.096	0	5
Accidents with $25 <$ age of oldest driver ≤ 45	0.028	0.197	0	11
Accidents with $45 <$ age of oldest driver ≤ 65	0.022	0.166	0	11
Accidents with $65 <$ age of oldest driver	0.007	0.085	0	4
Pedestrian accidents	0.014	0.129	0	9
Pedestrian accident & age of oldest driver ≤ 45	0.007	0.091	0	6
Pedestrian accident & age of oldest driver > 45	0.004	0.063	0	4
Darkness:[+]				
Dark proportion of the hour	0.427	0.480	0	1
Weather:[++]				
Temperature ($^{\circ}C$) < 0	0.033	0.179	0	1
$0 \leq$ temperature ($^{\circ}C$) < 5	0.166	0.372	0	1
$5 \leq$ temperature ($^{\circ}C$) < 10	0.321	0.467	0	1
$10 \leq$ temperature ($^{\circ}C$) < 15	0.315	0.465	0	1
$15 \leq$ temperature ($^{\circ}C$) < 20	0.138	0.345	0	1
$20 \leq$ Temperature	0.026	0.159	0	1
Wind speed (knots) < 5	0.225	0.418	0	1
$5 \leq$ wind speed (knots) < 10	0.362	0.481	0	1
$10 \leq$ wind speed (knots) < 20	0.343	0.475	0	1
$20 \leq$ wind speed (knots)	0.070	0.255	0	1
Positive precipitation	0.155	0.362	0	1
Snow or ice	0.008	0.089	0	1

Notes: Own calculations based on the STATS19 and MIDAS data. [+] refers to the full sample including 51,103,130 observations. [++] refers to the restricted sample comprising 21,352,015 observations.

Table A5.5: OLS Estimates of the Effects of Darkness on Accidents

	All regions		Regions with weather data			
	Coef. (S.E.)	Coef./$\overline{y}$	Coef. (S.E.)	Coef./$\overline{y}$	Coef. (S.E.)	Coef./$\overline{y}$
Weather controls	no		no		yes	
# all accidents	0.0007 (0.0011)	0.0082	.0080** (0.0032)	0.0504	0.0049* (0.0029)	0.0311
# slight accidents	-0.0005 (0.0010)	-0.0080	0.0050* (0.0027)	0.0369	0.0024 (0.0024)	0.0181
# serious accidents	0.0009*** (0.0002)	0.0757	0.0024*** (0.0005)	0.1088	0.0019*** (0.0005)	0.0859
# fatal accidents	0.0003*** (0.0000)	0.3195	0.0006*** (0.0001)	0.3204	0.0006*** (0.0001)	0.3090
Observations	51,103,130		21,352,015		21,352,015	

Notes: Own calculations based on STATS19 and MIDAS data. Each estimate gives the effect of darkness on the number of accidents of the respective category and comes from a separate OLS regression. All regressions include region, year, day-of-the-year, day-of-the-week, and hour fixed effects. To ease comparison with results from negative binomial regressions, we report also relative effect sizes (Coef./$\overline{y}$), where $\overline{y}$ refers to the sample mean of the respective outcome variable. Standard errors clustered at region level are in parentheses. Significance levels: * p<0.1; ** p<0.05; *** p<0.01.

Additional Figures

Figure A5.2: RD Residual Plots
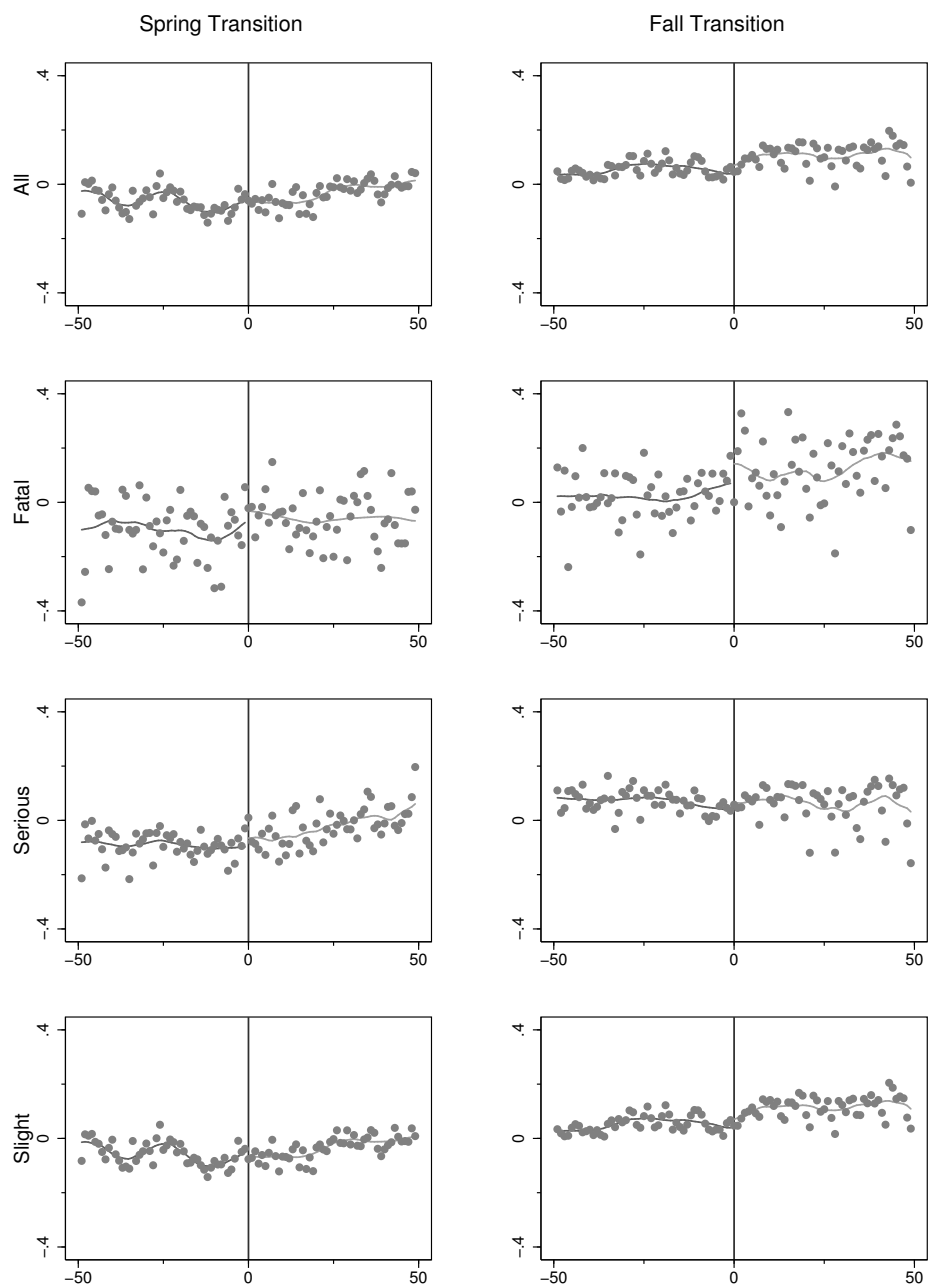
Notes: Own calculations and illustration based on STATS19 data. The residuals come from a OLS regression of $ln(accidentcounts)$ – either all (first row), slight (second row), serious (third row), or fatal (last row) – on year and day of week fixed effects. Each point gives the mean of all residuals at the respective date, where the date is defined relative to the spring (left part of the graph) or fall transition (right part). Fitted lines are based on locally weighted regressions.

Figure A5.3: The Effects of Darkness on Accidents by Time of Day

Slight accidents

Serious accidents

Fatal accidents
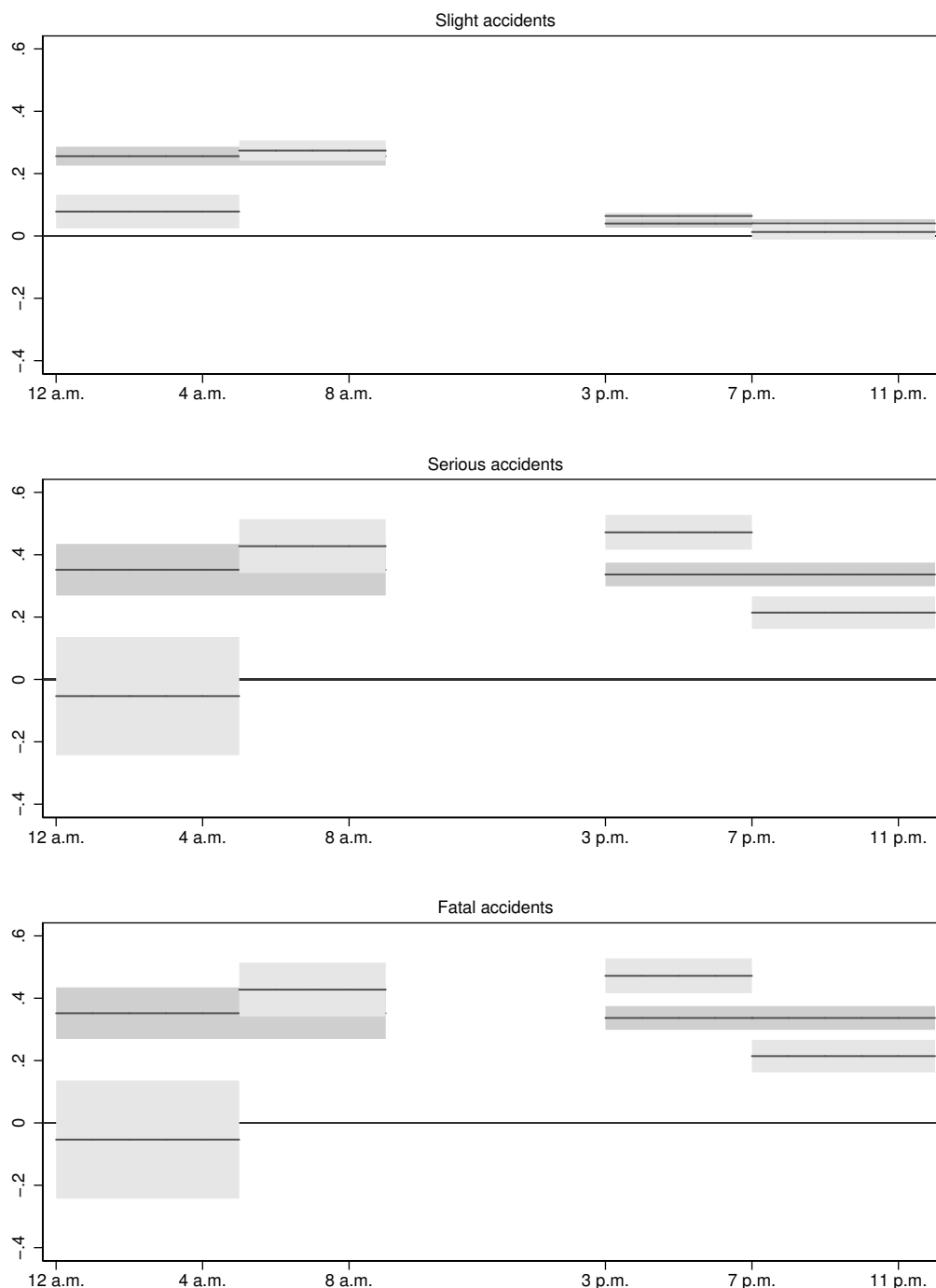
Notes: Own calculations and illustration based on STATS19 data. The graphs show the estimated effects of darkness including the 95% confidence intervals by hour of the day. The estimates result from two types of negative binomial regressions: the first allows the effect of darkness to differ between morning and evening hours (95% confidence intervals are in darker gray), the second to vary between early and late morning hours as well as early and late evening hours (95% confidence intervals are in light gray).

Figure A5.4: The Effects of Darkness on Accidents by Year
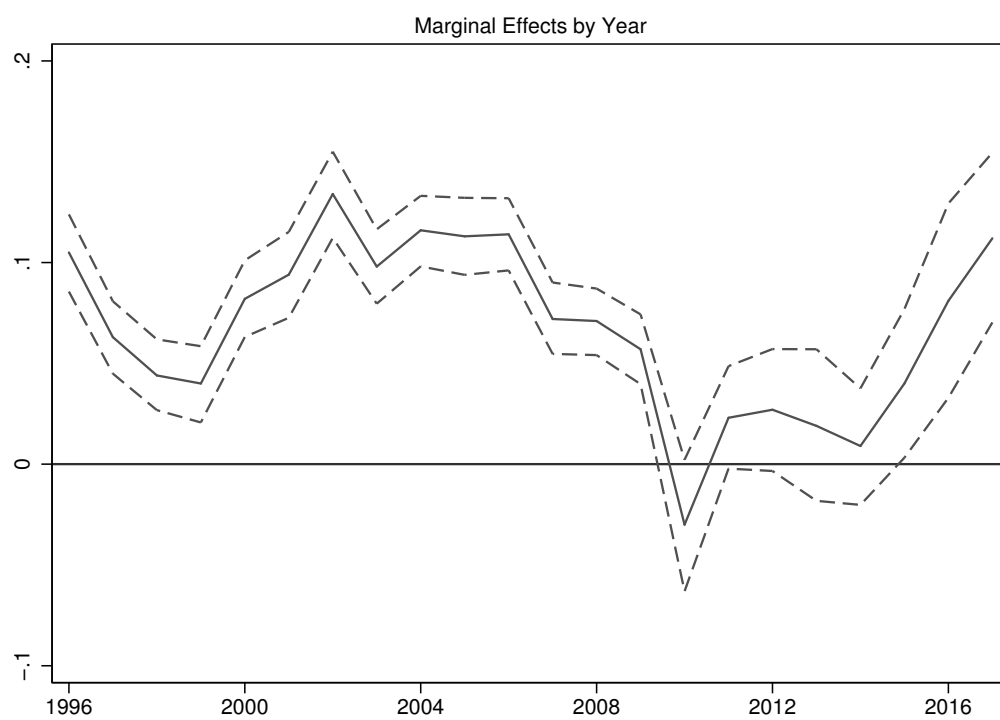
Notes: Own calculations and illustration based on STATS19 and MIDAS data. The results come from a negative binomial regression with all control variables, where darkness is interacted with year. The graphs show the effects (semi-elasticities) including the 95% confidence interval of darkness by year.

Bibliography

Abbring, J. H. and Heckman, J. J. (2007). Econometric evaluation of social programs, part iii: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In James J. Heckman and Edward E. Leamer, editor, *Handbook of Econometrics*, volume 6, part B, pages 5145–5303. Elsevier.

Ahammer, A., Horvath, G. T., and Winter-Ebmer, R. (2015). The effect of income on mortality: New evidence for the absence of a causal link. IZA Discussion Paper 9176, IZA Institute of Labor Economics, Bonn.

Allison, P. D. and Waterman, R. P. (2002). Fixed-effects negative binomial regression models. *Sociological Methodology*, 32(1):247–265.

Ambrasat, J. and Schupp, J. (2011). Handgreifkraftmessung im Sozio-oekonomischen Panel (SOEP) 2006 und 2008. Data Documentation 54, DIW Berlin, The German Socio-Economic Panel (SOEP), Berlin.

Ambrasat, J., Schupp, J., and Wagner, G. G. (2011). Comparing the predictive power of subjective and objective health indicators: Changes in hand grip strength and overall satisfaction with life as predictors of mortality. SOEPpapers on Multidisciplinary Panel Data Research 398, DIW Berlin, The German Socio-Economic Panel (SOEP), Berlin.

American Psychiatric Association (2000). Diagnostic and statistical manual of mental disorders. `http://www.psychiatry.org/practice/dsm`.

Andersen, H. H., Mühlbacher, A., Nübling, M., Schupp, J., and Wagner, G. G. (2007). Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2. *Journal of Applied Social Science Studies (Schmollers Jahrbuch)*, 127(1):171–182.

Angelini, V., Brugiavini, A., and Weber, G. (2009). Ageing and unused capacity in Europe: Is there an early retirement trap? *Economic Policy*, 24(59):463–508.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.

Aries, M. B. and Newsham, G. R. (2008). Effect of daylight saving time on lighting energy use: A literature review. *Energy Policy*, 36(6):1858–1866.

Arrow, J. O. (1996). Estimating the influence of health as a risk factor on unemployment: A survival analysis of employment durations for workers surveyed in the German Socio-Economic Panel (1984-1990). *Social Science & Medicine*, 42(12):1651–1659.

Banks, J., Batty, G., Coughlin, K., Dangerfield, P., Marmot, M., Nazroo, J., Oldfield, Z., Steel, N., Steptoe, A., Wood, M., and Zaninotto, P. (2019). English Longitudinal Study of Ageing: waves 0-9, 1998-2019. 33rd Edition. UK Data Service. SN: 5050, http://doi.org/10.5255/UKDA-SN-5050-20. Technical report.

Banks, J., O'Dea, C., and Oldfield, Z. (2010). Cognitive function, numeracy and retirement saving trajectories. *Economic Journal*, 120(548):381–410.

Barnes, C. M. and Wagner, D. T. (2009). Changing to daylight saving time cuts into sleep and increases workplace injuries. *Journal of Applied Psychology*, 94(5):1305–1317.

Berk, M., Dodd, S., Hallam, K., Berk, L., Gleeson, J., and Henry, M. (2008). Small shifts in diurnal rhythms are associated with an increase in suicide: The effect of daylight saving. *Sleep and Biological Rhythms*, 6(1):22–25.

Bianchi, N. (2020). The indirect effects of educational expansions: Evidence from a large enrollment increase in university majors. *Journal of Labor Economics*, 38(3):767–804.

Binder, M. and Coad, A. (2015a). Heterogeneity in the relationship between unemployment and subjective wellbeing: A quantile approach. *Economica*, 82(328):865–891.

Binder, M. and Coad, A. (2015b). Unemployment impacts differently on the distribution of a comprehensive well-being measure. *Applied Economics Letters*, 22(8):619–627.

Böckermann, P. and Ilmakunnas, P. (2009). Unemployment and self-assessed health: Evidence from panel data. *Health Economics*, 18(2):161–179.

Boden, J. M., Fergusson, D. M., and Horwood, L. J. (2010). Cigarette smoking and depression: Tests of causal linkages using a longitudinal birth cohort. *The British Journal of Psychiatry*, 196(6):440–446.

Bohannon, R. W. (2008). Hand-grip dynamometry predicts future outcomes in aging adults. *Journal of Geriatric Physical Therapy*, 31(1):3–10.

Bonsang, E., Adam, S., and Perelman, S. (2012). Does retirement affect cognitive functioning? *Journal of Health Economics*, 31(3):490–501.

Börsch-Supan, A. (2019a). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w1.710. Technical report.

Börsch-Supan, A. (2019b). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w2.710. Technical report.

Börsch-Supan, A. (2019c). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 3 - SHARELIFE. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w3.710. Technical report.

Börsch-Supan, A. (2019d). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.710. Technical report.

Börsch-Supan, A. (2019e). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w5.710. Technical report.

Börsch-Supan, A. (2019f). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w6.710. Technical report.

Börsch-Supan, A. (2020). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 7.1.1. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w7.711. Technical report.

Börsch-Supan, A. (2021). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. Release version: 8.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w8.100. Technical report.

Börsch-Supan, A. and Jürges, H. (2005). The Survey of Health, Aging and Retirement in Europe - Methodology. Technical report, Mannheim: Mannheim Research Institute for the Economics of Aging.

Browning, M. and Heinesen, E. (2012). Effect of job loss due to plant closure on mortality and hospitalization. *Journal of Health Economics*, 31(4):599–616.

Brunello, G., Fabbri, D., and Fort, M. (2013). The causal effect of education on body mass: Evidence from Europe. *Journal of Labor Economics*, 31(1):195–223.

Brunello, G., Weber, G., and Weiss, C. T. (2016). Books are forever: Early life conditions, education and lifetime earnings in Europe. *The Economic Journal*, 127(600):271–296.

Bünnings, C. and Schiele, V. (2021). Spring forward, don't fall back: The effect of daylight saving time on road safety. *Review of Economics and Statistics*, 103(1):165–176.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). rdrobust: Software for regression discontinuity designs. *Stata Journal*, 17(2):372–404.

Cameron, C. A. and Trivedi, P. K. (2015). Count panel data. In Baltagi, B. H., editor, *The Oxford Handbook of Panel Data*. Oxford University Press, Oxford.

Campos, P., Saguy, A., Ernsberger, P., Oliver, E., and Gaesser, G. (2006). The epidemiology of overweight and obesity: Public health crisis or moral panic? *International Journal of Epidemiology*, 35(1):55–60.

Carey, R. N. and Sarma, K. M. (2017). Impact of daylight saving time on road traffic collision risk: A systematic review. *BMJ open*, 7(6):1–14.

Celidoni, M., Bianco, C. D., and Weber, G. (2017). Retirement and cognitive decline. A longitudinal analysis using SHARE data. *Journal of Health Economics*, 56:113–125.

Christelis, D., Jappelli, T., and Padula, M. (2010). Cognitive abilities and portfolio choice. *European Economic Review*, 54(1):18–38.

Coate, D. and Markowitz, S. (2004). The effects of daylight and daylight saving time on us pedestrian fatalities and motor vehicle occupant fatalities. *Accident Analysis & Prevention*, 36(3):351–357.

Coe, N. B., von Gaudecker, H., Lindeboom, M., and Maurer, J. (2012). The effect of retirement on cognitive functioning. *Health Economics*, 21(8):913–927.

Cohen, S., Janicki-Deverts, D., and Miller, G. E. (2007). Psychological stress and disease. *JAMA: The Journal of the American Medical Association*, 298(14):1685–1687.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (2009). Smoking and lung cancer: Recent evidence and a discussion of some questions. *International Journal of Epidemiology*, 38(5):1175–1191.

Cornwall, C., Horiuchi, A., and Lehman, C. (2015). NOAA solar calculator. National Oceanic and Atmospheric Administration, U.S. Department of Commerce. `https://www.esrl.noaa.gov/gmd/grad/solcalc/calcdetails.html`.

CSSZ (2019a). Čssz informuje: za jakých podmínek je možné přiznat předčasný starobní důchod. `https://www.cssz.cz/web/cz/-/cssz-informuje-za-jakych-podminek-je-mozne-priznat-predcasny-starobni-duchod`.

CSSZ (2019b). Starobní důchod podrobně. `https://www.cssz.cz/web/cz/starobni-duchod-podrobne`.

Cygan-Rehm, K. (2021). Are there no wage returns to compulsory schooling in Germany? A reassessment. *Journal of Applied Econometrics*, (forthcoming).

Darius, S., Bergmann, L., Blaschke, S., and Böckelmann, I. (2018). Influence of sex and age on contrast sensitivity subject to the applied method. *Klinische Monatsblätter für Augenheilkunde*, 235(2):2012–2018.

de Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

Decker, S. and Schmitz, H. (2016). Health shocks and risk aversion. *Journal of Health Economics*, 50(C):156–170.

Department for Transport (1984). The road vehicles lighting regulations act.

Department for Transport (2013). Reported road casualties in Great Britain: Guide to the statistics and data sources. `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/259012/rrcgb-quality-statement.pdf`.

Department for Transport (2018a). Average value of prevention per reported casualty and per reported road accident: Great Britain, latest available year. `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/665325/ras60001.ods`.

Department for Transport (2018b). Road safety data. `https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data`.

Doblhammer, G., Van den Berg, G. J., and Fritze, T. (2013). Economic conditions at the time of birth and cognitive abilities late in life: Evidence from ten European countries. *PloS one*, 8(9):e74915.

Doleac, J. L. and Sanders, N. J. (2015). Under the cover of darkness: How ambient light influences criminal activity. *Review of Economics and Statistics*, 97(5):1093–1103.

Dong, Y. and Shen, S. (2018). Testing for rank invariance or similarity in program evaluation. *Review of Economics and Statistics*, 100(1):78–85.

Dorner, M., Heining, J., Jacobebbinghaus, P., and Seth, S. (2010). The sample of integrated labour market biographies. *Journal of Contextual Economics: Schmollers Jahrbuch*, 130(4):599–608.

Duflo, E. (2004). The medium run effects of educational expansion: evidence from a large school construction program in indonesia. *New Research on Education in Developing Economies*, 74(1):163–197.

Eliason, M. and Storrie, D. (2009a). Does job loss shorten life? *Journal of Human Resources*, 44(2):277–302.

Eliason, M. and Storrie, D. (2009b). Job loss is bad for your health – Swedish evidence on cause-specific hospitalization following involuntary job loss. *Social Science & Medicine*, 68(8):1396–1406.

Engelhardt, H., Buber, I., Skirbekk, V., and Prskawetz, A. (2010). Social involvement, behavioural risks and cognitive functioning among older people. *Ageing and Society*, 30(5):779–809.

European Commission (2019). Griechenland – altersrenten. `https://ec.europa.eu/social/main.jsp?catId=1112&langId=de&intPageId=4567`.

European Parliament (2019). European parliament resolution of 26 march 2019 on discontinuing seasonal changes of time and repealing directive 2000/84/EC. `http://www.europarl.europa.eu/doceo/document/TA-8-2019-0225_EN.pdf`.

Ezzy, D. (1993). Unemployment and mental health: A critical review. *Social Science & Medicine*, 37(1):41–52.

Ferguson, S. A., Preusser, D. F., Lund, A. K., Zador, P. L., and Ulmer, R. G. (1995). Daylight saving time and motor vehicle crashes: The reduction in pedestrian and vehicle occupant fatalities. *American Journal of Public Health*, 85(1):92–95.

Fergusson, D. M., Boden, J. M., and Horwood, J. (2009). Tests of causal links between alcohol abuse or dependence and major depression. *Archives of General Psychiatry*, 66(3):260–266.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75(1):259–276.

Frandsen, B. R. and Lefgren, L. J. (2018). Testing rank similarity. *Review of Economics and Statistics*, 100(1):86–91.

Frijters, P., Haisken-DeNew, J. P., and Shields, M. A. (2005). The causal effect of income on health: Evidence from german reunification. *Journal of Health Economics*, 24(5):997–1017.

Frölich, M. and Melly, B. (2010). Estimation of quantile treatment effects with stata. *The Stata Journal*, 10(3):423–457.

Frölich, M. and Melly, B. (2013). Unconditional quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics*, 31(3):346–357.

Fryer, D. (1986). Employment deprivation and personal agency during unemployment. A critical discussion of Jahoda's explanation of the psychological effects of unemployment. *Social Behaviour*, 1(1):3–23.

Fryer, D. (1992). Psychological or Material Deprivation: Why Does Unemployment Have Mental Health Consequences? In McLaughlin, E., editor, *Understanding unemployment, New Perspectives on Active Labour Market Policies*, pages 103–125. Routledge, London, New York.

Gaski, J. F. and Sagarin, J. (2011). Detrimental effects of daylight-saving time on sat scores. *Journal of Neuroscience, Psychology, and Economics*, 4(1):44.

Gathergood, J. (2013). An instrumental variable approach to unemployment, psychological health and social norm effects. *Health Economics*, 22(6):643–654.

Gill, S. C., Butterworth, P., Rodgers, B., and Mackinnon, A. (2007). Validity of the mental health component scale of the 12-item short-form health survey (mcs-12) as measure of common mental disorders in the general population. *Psychiatry Research*, 152(1):63 – 71.

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

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, (forthcoming).

Government Office for Legislation (Slovenia) (2013). Pension and Disability Insurance Act (ZPIZ-2). `http://pisrs.si/Pis.web/pregledPredpisa?id=ZAKO6280`.

Green, F. (2011). Unpacking the misery multiplier: How employability modifies the impacts of unemployment and job insecurity on life satisfaction and mental health. *Journal of Health Economics*, 30(2):265–276.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, 7(1):98–119.

Gregory-Allen, R., Jacobsen, B., and Marquering, W. (2010). The daylight saving time anomaly in stock returns: Fact or fiction? *Journal of Financial Research*, 33(4):403–427.

Hauser, R. and Strengmann-Kuhn, W. (2004). Armut der älteren Bevölkerung in den Ländern der Europäischen Union. *DRV-Schriften des Verbandes der Rentenversicherungsträger (VDR)*, 54.

Havranek, T., Herman, D., and Irsova, Z. (2018). Does daylight saving save electricity? A meta-analysis. *Energy Journal*, 39(2):35–61.

Health and Retirement Study (2014a). HRS 2006 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Health and Retirement Study (2014b). HRS 2008 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Health and Retirement Study (2016a). HRS 1998 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Health and Retirement Study (2016b). HRS 2000 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Health and Retirement Study (2016c). HRS 2002 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Health and Retirement Study (2016d). HRS 2004 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Health and Retirement Study (2017a). HRS 2010 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Health and Retirement Study (2017b). HRS 2012 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Health and Retirement Study (2017c). HRS 2014 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Health and Retirement Study (2019). HRS 2016 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Heckman, J. J. and Lochner, L. (1998). General-equilibrium treatment effects: A study of tuition policy. *American Economic Review*, 88(2):381–386.

Helbig, M. and Nikolai, R. (2015). *Die Unvergleichbaren: der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Verlag Julius Klinkhardt, Bad Heilbrunn.

Herber, S. P., Quis, J. S., and Heineck, G. (2017). Does the transition into daylight saving time affect students' performance? *Economics of Education Review*, 61:130–139.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.

Hollederer, A. (2011). *Erwerbslosigkeit, Gesundheit und Präventionspotenziale. Ergebnisse des Mikrozensus 2005*. VS Verlag für Sozialwissenschaften, Wiesbaden.

Huber, M., Lechner, M., and Wunsch, C. (2011). Does leaving welfare improve health? Evidence for Germany. *Health Economics*, 20(4):484–504.

Jahoda, M. (1981). Work, employment and unemployment: Values, theories, and approaches in social research. *American Psychologist*, 36(2):184–191.

Jahoda, M. (1983). *Wieviel Arbeit braucht der Mensch? Arbeit und Arbeitslosigkeit im 20. Jahrhundert*. Beltz, Weinheim, Basel.

Jahoda, M., Lazarsfeld, P. F., and Zeisel, H. (1975 (1933)). *Die Arbeitslosen von Marienthal. Ein soziographischer Versuch über die Wirkungen langandauernder Arbeitslosigkeit, mit einem Anhang zur Geschichte der Soziographie*. Suhrkamp, Frankfurt am Main.

Janszky, I. and Ljung, R. (2008). Shifts to and from daylight saving time and incidence of myocardial infarction. *New England Journal of Medicine*, 359(18):1966–1968.

Jiddou, M. R., Pica, M., Boura, J., Qu, L., and Franklin, B. A. (2013). Incidence of myocardial infarction with shifts to and from daylight savings time. *American Journal of Cardiology*, 111(5):631–635.

Jin, L. and Ziebarth, N. R. (2020). Sleep, health, and human capital: Evidence from daylight saving time. *Journal of Economic Behavior & Organization*, 170:174–192.

Jusot, F., Khlat, M., Rochereau, T., and Serme, C. (2008). Job loss from poor health, smoking and obesity: a national prospective survey in france. *Journal of Epidemiology & Community Health*, 62(4):332–337.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2):303–316.

Kamhöfer, D. A., Schmitz, H., and Westphal, M. (2019). Heterogeneity in Marginal Non-Monetary Returns to Higher Education. *Journal of the European Economic Association*, 17(1):205–244.

Kamstra, M. J., Kramer, L. A., and Levi, M. D. (2000). Losing sleep at the market: The daylight saving anomaly. *American Economic Review*, 90(4):1005–1011.

Kantermann, T., Juda, M., Merrow, M., and Roenneberg, T. (2007). The human circadian clock's seasonal adjustment is disrupted by daylight saving time. *Current Biology*, 17(22):1996–2000.

Kellogg, R. and Wolff, H. (2008). Does extending daylight saving time save energy? Evidence from an australian experiment. *Journal of Environmental Economics and Management*, 56:207–220.

Kemptner, D., Jürges, H., and Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics*, 30(2):340–354.

Köhler, H. and Lundgreen, P. (2014). *Allgemein bildende Schulen in der Bundesrepublik Deutschland 1949 - 2010 (histat Studiennummer 8570)*. GESIS Datenarchiv, Köln.

Kol-Zchut (2019). Retirement age. `https://www.kolzchut.org.il/he`.

Korkeamäki, O. and Kyyrä, T. (2014). A distributional analysis of earnings losses of displaced workers in an economic depression and recovery. *Oxford Bulletin of Economics and Statistics*, 76(4):565–588.

Kotchen, M. J. and Grant, L. E. (2011). Does daylight saving time save energy? Evidence from a natural experiment in Indiana. *Review of Economics and Statistics*, 93(4):1172–1185.

Kountouris, Y. and Remoundou, K. (2014). About time: Daylight saving time transition and individual well-being. *Economics Letters*, 122(1):100–103.

Krarti, M. and Hajiah, A. (2011). Analysis of impact of daylight time savings on energy use of buildings in Kuwait. *Energy Policy*, 39(5):2319–2329.

Kuhn, A., Lalive, R., and Zweimüller, J. (2009). The public health costs of job loss. *Journal of Health Economics*, 28(6):1099–1115.

Kühnle, D. and Wunder, C. (2016). Using the life satisfaction approach to value daylight savings time transitions: Evidence from Britain and Germany. *Journal of Happiness Studies*, 17(6):2293–2323.

Lahti, T., Nysten, E., Haukka, J., Sulander, P., and Partonen, T. (2010). Daylight saving time transitions and road traffic accidents. *Journal of Environmental and Public Health*, 2010.

Lahti, T., Sysi-Aho, J., Haukka, J., and Partonen, T. (2011). Work-related accidents and daylight saving time in Finland. *Occupational Medicine*, 61(1):26–28.

Lahti, T. A., Leppämäki, S., Lönnqvist, J., and Partonen, T. (2006). Transition to daylight saving time reduces sleep duration plus sleep efficiency of the deprived sleep. *Neuroscience Letters*, 406(3):174–177.

Lindahl, M. (2005). Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. *Journal of Human Resources*, 40(1).

Lindeboom, M., Portrait, F., and van den Berg, G. J. (2002). An econometric analysis of the mental-health effects of major events in the life of older individuals. *Health Economics*, 11(6):505–520.

Ling, C. H. Y., Taekema, D., Craen, A. J. M. d., Gussekloo, J., Westendorp, R. G. J., and Maier, A. B. (2010). Handgrip strength and mortality in the oldest old population: the Leiden 85-plus study. *Canadian Medical Association Journal*, 182(5):429–435.

Liu, K., Salvanes, K. G., and Sørensen, E. Ø. (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Labor Market Issues*, 84:3–17.

Mandal, B., Ayvagari, P., and Gallo, W. T. (2011). Job loss and depression: The role of subjective expectations. *Social Science & Medicine*, 72(4):576–583.

Marcus, J. (2013). The effect of unemployment on the mental health spouses. Evidence from plant closures in Germany. *Journal of Health Economics*, 32(3):546–558.

Marcus, J. (2014). Does job loss make you smoke and gain weight? *Economica*, 81(324):626–648.

Margaryan, S., Paul, A., and Siedler, T. (2021). Does education affect attitudes towards immigration? Evidence from Germany. *Journal of Human Resources*, 56(2):446–479.

MASGF (2008). *Arbeitsförderung mit gesundheitsbezogener Ausrichtung. AmigA. Leitfaden für die praktische Umsetzung*. Ministerium für Arbeit, Soziales, Gesundheit und Familie des Landes Brandenburg, Öffentlichkeitsarbeit, Potsdam.

Maurer, J. (2011). Education and male-female differences in later-life cognition: International evidence from latin america and the caribbean. *Demography*, 48(3):915–930.

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

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

Mazzonna, F. and Peracchi, F. (2018). The economics of cognitive aging. In *Oxford Research Encyclopedia of Economics and Finance*.

McFadden, D. (2008). Human capital accumulation and depreciation. *Applied Economic Perspectives and Policy*, 30(3):379–385.

Met Office (2006a). MIDAS UK hourly rainfall data. NCAS British Atmospheric Data Centre. `http://catalogue.ceda.ac.uk/uuid/bbd6916225e7475514e17fdbf11141c1`.

Met Office (2006b). MIDAS: UK hourly weather observation data. NCAS British Atmospheric Data Centre. `http://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c`.

Ministry of Justice (Israel) (2019). Courts law (combined version). `https://www.nevo.co.il/law/74849`.

Ministry of Labour and Social Policy (Poland) (2019). Starobní důchody. `https://www.mpsv.cz/web/cz/starobni-duchody`.

Momani, M. A., Yatim, B., and Ali, M. A. M. (2009). The impact of the daylight saving time on electricity consumption – A case study from Jordan. *Energy Policy*, 37(5):2042–2051.

Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121(1-2):175–212.

Morin, L.-P. (2015). Cohort size and youth earnings: Evidence from a quasi-experiment. *Labour Economics*, 32:99–111.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., et al. (1996). Intelligence: knowns and unknowns. *American Psychologist*, 51(2):77.

Nordenmark, M. and Strandh, M. (1999). Towards a sociological understanding of mental well-being among the unemployed: The role of economic and psychosocial factors. *Sociology*, 33(3):577–597.

Oreopoulos, P., von Wachter, T., and Heisz, A. (2012). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1):1–29.

Patterson, C. (2018). World alzheimer report 2018 – The state of the art of dementia research: New frontiers. https://www.alz.co.uk/research/WorldAlzheimerReport2018.pdf?2.

Pick, P., Brüggemann, J., Grote, C., Grünhagen, E., and Lampert, T. (2004). Schwerpunktbericht: Pflege. Technical report, Robert Koch-Institut. https://edoc.rki.de/bitstream/handle/176904/3174/29qmd3FjnRtQ_48.pdf?sequence=1&isAllowed=y.

Pischke, J.-S. (2003). The impact of length of the school year on student performance and earnings: Evidence from the German short school year. NBER Working Paper 9964, National Bureau of Economic Research, Cambridge, Massachusetts.

Pischke, J.-S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. *The Economic Journal*, 117(523):1216–1242.

Pischke, J.-S. and von Wachter, T. (2005). Zero returns to compulsory schooling in Germany: Evidence and interpretation. NBER Working Paper 11414, National Bureau of Economic Research, Cambridge, Massachusetts.

Pischke, J.-S. and von Wachter, T. (2008). Zero returns to compulsory schooling in Germany: Evidence and interpretation. *Review of Economics and Statistics*, 90(3):592–598.

Puur, A., Leppik, L., and Klesment, M. (2015). Changes in pension take-up and retirement in the context of increasing the pension age: The case of Estonia in the 2000s. *Post-Communist Economies*, 27(4):497–516.

Rabušic, L. (2004). Why are they all so eager to retire? On the transition to retirement in the Czech Republic. *Sociologický časopis/Czech Sociological Review*, 40(03):319–342.

Rantanen, T., Masaki, K., Foley, D., Izmirlian, G., White, L., and Guralnik, J. M. (1998). Grip strength changes over 27 yr in Japanese-American men. *Journal of Applied Physiology*, 85(6):2047–2053.

Rantanen, T., Volpato, S., Luigi Ferrucci, M., Eino Heikkinen, M., Fried, L. P., and Guralnik, J. M. (2003). Handgrip strength and cause-specific and total mortality in older disabled women: Exploring the mechanism. *Journal of the American Geriatrics Society*, 51(5):636–641.

Reichert, A. and Tauchmann, H. (2011). The causal impact of fear of unemployment on psychological health. Ruhr Economic Papers 266, RWI.

Roberts, H. C., Denison, H. J., Martin, H. J., Patel, H. P., Syddall, H., Cooper, C., and Sayer, A. A. (2011). A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. *Age and Ageing*, 40(4):423–429.

Rohwedder, S. and Willis, R. J. (2010). Mental retirement. *Journal of Economic Perspectives*, 24(1):119–38.

Salm, M. (2009). Does job loss cause ill health? *Health Economics*, 18(9):1075–1089.

Salyers, M. P., Bosworth, H. B., Swanson, J. W., Lamb-Pagone, J., and Osher, F. C. (2000). Reliability and validity of the sf-12 health survey among people with severe mental illness. *Medical Care*, 38(11):1141–1150.

Sandhu, A., Seth, M., and Gurm, H. S. (2014). Daylight savings time and myocardial infarction. *Open Heart*, 1(1):e000019.

Schiele, V. (2022). Labor market spillover effects of a compulsory schooling reform in Germany. Working Papers Dissertations 84, Paderborn University, Faculty of Business Administration and Economics.

Schiele, V. and Schmitz, H. (2016). Quantile treatment effects of job loss on health. *Journal of Health Economics*, 49:59–69.

Schiele, V. and Schmitz, H. (2021). Understanding cognitive decline in older ages: The role of health shocks. Ruhr Economic Papers 919, RWI.

Schmitz, H. (2011). Why are the unemployed in worse health? The causal effects of unemployment on health. *Labour Economics*, 18(1):71–78.

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

Schmitz, H. and Westphal, M. (2021). The dynamic and heterogeneous effects of retirement on cognitive decline. Unpublished Working Paper.

Schneeweis, N., Skirbekk, V., and Winter-Ebmer, R. (2014). Does education improve cognitive performance four decades after school completion? *Demography*, 51(2):619–643.

Schwandt, H. and Von Wachter, T. (2019). Unlucky cohorts: Estimating the long-term effects of entering the labor market in a recession in large cross-sectional data sets. *Journal of Labor Economics*, 37(S1):S161–S198.

Sexton, A. L. and Beatty, T. K. (2014). Behavioral responses to daylight savings time. *Journal of Economic Behavior & Organization*, 107:290–307.

Shai, O. (2018). Is retirement good for men's health? Evidence using a change in the retirement age in Israel. *Journal of Health Economics*, 57:15–30.

Smith, A. C. (2016). Spring forward at your own risk: Daylight saving time and fatal vehicle crashes. *American Economic Journal: Applied Economics*, 8(2):65–91.

Smith, J. P., McArdle, J. J., and Willis, R. (2010). Financial decision making and cognition in a family context. *Economic Journal*, 120(548):363–380.

Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W. R., and Weir, D. R. (2014). Cohort profile: the Health and Retirement Study (HRS). *International Journal of Epidemiology*, 43(2):576–585.

Sood, N. and Ghosh, A. (2007). The short and long run effects of daylight saving time on fatal automobile crashes. *The BE Journal of Economic Analysis & Policy*, 7(1).

Sotsiaalkindlustusamet (2019). Vanus on väärtus. `https://pension.sotsiaalkindlustusamet.ee/personaalne-pensioniplaan`.

Statistisches Bundesamt (2016). Insolvenzverfahren. Fachserie 2 Reihe 4.1.

Stenholm, S., Tiainen, K., Rantanen, T., Sainio, P., Heliövaara, M., Impivaara, O., and Koskinen, S. (2012). Long-term determinants of muscle strength decline: Prospective evidence from the 22-year mini-Finland follow-up survey. *Journal of the American Geriatrics Society*, 60(1):77–85.

Steptoe, A., Breeze, E., Banks, J., and Nazroo (2003). Cohort profile: the English Longitudinal Study of Ageing. *International Journal of Epidemiology*, 42(6):1640–1648.

Stewart, J. M. (2001). The impact of health status on the duration of unemployment spells and the implications for studies of the impact of unemployment on health status. *Journal of Health Economics*, 20(5):781–796.

Sullivan, J. M. and Flannagan, M. J. (2002). The role of ambient light level in fatal crashes: Inferences from daylight saving time transitions. *Accident Analysis & Prevention*, 34(4):487–498.

Sun, L. (2020). eventstudyweights: Stata module to estimate the implied weights on the cohortspecific average treatment effects on the treated (CATTs) (event study specifications).

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225.

Sütterlin, S., Hoßmann, I., and Klingholz, R. (2011). Demenz-Report: wie sich die Regionen in Deutschland, Österreich und der Schweiz auf die Alterung der Gesellschaft vorbereiten können. `https://nbn-resolving.org/urn:nbn:de:0168-ssoar-321483`.

Toro, W., Tigre, R., and Sampaio, B. (2015). Daylight saving time and incidence of myocardial infarction: Evidence from a regression discontinuity design. *Economics Letters*, 136:1–4.

United States Congress (2019). Sunshine protection act of 2019. `https://www.congress.gov/bill/116th-congress/senate-bill/670/text`.

van de Mheen, H., Stronks, K., Schrijvers, C. T. M., and Mackenbach, J. P. (1999). The influence of adult ill health on occupational class mobility and mobility out of and into employment in the Netherlands. *Social Science & Medicine*, 49(9):509–518.

van den Berg, G. J., Deeg, D. J., Lindeboom, M., and Portrait, F. (2010). The role of early-life conditions in the cognitive decline due to adverse events later in life. *Economic Journal*, 120(548):411–428.

van den Berg, G. J., Lindeboom, M., and Portrait, F. (2011). Conjugal bereavement effects on health and mortality at advanced ages. *Journal of Health Economics*, 30(4):774–794.

Varughese, J. and Allen, R. P. (2001). Fatal accidents following changes in daylight savings time: The American experience. *Sleep Medicine*, 2(1):31–36.

Vilagut, G., Forero, C. G., Pinto-Meza, A., Haro, J. M., de Graaf, R., Bruffaerts, R., Kovess, V., de Girolamo, G., Matschinger, H., Ferrer, M., and Alonso, J. (2013). The mental component of the short-form 12 health survey (sf-12) as a measure of depressive disorders in the general population: Results with three alternative scoring methods. *Value in Health*, 16(4):564 – 573.

Vom Berge, P., Burghardt, A., and Trenkle, S. (2013). Sample of integrated labour market biographies regional file 1975-2010 (SIAB-R 7510). FDZ Data Report 09/2013.

Von Wachter, T. (2020). The persistent effects of initial labor market conditions for young adults and their sources. *Journal of Economic Perspectives*, 34(4):168–94.

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP). Scope, evolution and enhancements. *Journal of Applied Social Science Studies (Schmollers Jahrbuch)*, 127(1):139–169.

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

Warr, P. B. (1987). *Work, Unemployment and Mental Health*. Clarendon Press, Oxford, New York.

WHO (2000). The asia-pacific perspective. Redefining obesity and its treatment. `http://www.wpro.who.int/nutrition/documents/docs/Redefiningobesity.pdf?ua=1`.

Wolff, H. and Makino, M. (2012). Extending becker's time allocation theory to model continuous time blocks: Evidence from daylight saving time. IZA Discussion Paper 6787.

Wyatt, S. B., Winters, K. P., and Dubbert, P. M. (2006). Overweight and obesity: Prevalence, consequences, and causes of a growing public health problem. *The American Journal of the Medical Sciences*, 331(4):166–174.

Xue Q, Walston JD, Fried LP, and Beamer BA (2011). Prediction of risk of falling, physical disability, and frailty by rate of decline in grip strength: The women's health and aging study. *Archives of Internal Medicine*, 171(12):1119–1121.

ZPIZ (2019). Pokojnine. `https://www.zpiz.si/cms/content2019/pokojnine#Pred%C4%8Dasna%20pokojna`.

Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Datum   Unterschrift