University of Groningen

Skills, Early-Life Environment and Health in Adulthood

Niccodemi, Gianmaria

DOI:
[10.33612/diss.194511989](https://doi.org/10.33612/diss.194511989)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Niccodemi, G. (2022). *Skills, Early-Life Environment and Health in Adulthood*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen, SOM research school.
<https://doi.org/10.33612/diss.194511989>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Skills, Early-Life Environment and Health in Adulthood

Gianmaria Niccodemi

Publisher: University of Groningen
Groningen, the Netherlands
Printed by: Ipskamp Printing
Enschede, The Netherlands

© 2021 Gianmaria Niccodemi

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system of any nature, or transmitted in any form or by any means, electronic, mechanical, now known or hereafter invented, including photocopying or recording, without prior written permission of the publisher.

university of
 groningen

Skills, Early-Life Environment and Health in Adulthood

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 13 January 2022 at 14.30 hours

by

Gianmaria Niccodemi

born on 13 January 1990
in Montebelluna (Tv), Italy

Supervisors

Prof. R.J.M. Alessie

Prof. V. Angelini

Co-supervisor

Dr. G.E. Bijwaard

Assessment Committee

Prof. J.P. Elhorst

Prof. G. Pasini

Prof. P. Garcia Gomez

Acknowledgements

It was during my master's studies in Groningen that I asked Viola Angelini whether there was a chance to do an internship in the Netherlands on some applied research in the field of health economics. This led me to NIDI, the Netherlands Interdisciplinary Demographic Institute in The Hague, where under the supervision of both Viola Angelini and Govert Bijwaard I experienced research for the first time in my life. This wonderful experience encouraged me to apply for a PhD position in Groningen. As is often the case, my PhD has been a journey of formation, both from the human and the research point of view. It has also been a roller coaster of feelings and emotions, from satisfaction for some accomplished results to disappointment for some rejections in research, from fulfillment for feeling at home to sadness for losing some people on the way. All in all, I am grateful for such an experience and I would like to thank all the people that contributed in making this journey so unique.

First of all, I would like to thank my supervisors Rob Alessie, Viola Angelini and Govert Bijwaard. They are all admirable and wise researchers who guided me greatly through the process of writing research papers and applying for research jobs. They always encouraged me to pursue the most suitable models for the research questions we had in mind, by keeping things simple when sufficient and by not being afraid of more complex computations when needed. I learned a lot over the years from the countless meetings we had. Their humanity is as great as their ability as researchers and supervisors. No matter how down I could have felt, they always took care of me, talked to me, cheered me up and even invited me for some activities such as dinners, barbecues and birthday parties. More specifically, Rob was always there to guide me through econometric models and specifications. It often happened that we spontaneously met in his office and talked about improvements in the methodological parts of the research projects. His experience as an economist had an important impact on my formation and on the development

of this thesis. He is also one of the funniest person I met in Groningen, always ready to make jokes and have a laugh (among all the most hilarious moments we had together, a special mention goes to when he explained to me the envelope theorem on an envelope). Viola was my main point of reference over the PhD. I had innumerable conversations with her about econometrics, economics, presentations at conferences and job applications. Among other things, she taught me how important the structure and the narrative of the paper are and how to come up with interesting research questions, set-up a proper introduction, explain why and how the obtained results matter and write appropriate policy implications. From a non-academic point of view, she was always there for me in moments of emotional struggle, no matter how intense those were. Govert was the first of my supervisors who introduced me to advanced coding. His skills as a programmer and model developer taught me a lot in terms of how to exploit the power that statistical packages can offer and how to develop and estimate structural models. Unfortunately, as he is not based in Groningen it was hard for us to meet frequently. Nevertheless, he was always available for discussing projects online and having lunch at his place in Gouda.

I would also like to thank the members of my assessment committee Paul Elhorst, Giacomo Pasini and Pilar Garcia Gomez. Thank you for reading my thesis and for providing useful comments that contributed in improving its quality. I also thank my co-authors Jochen Mierau, Gerard van den Berg and Tom Wansbeek. Jochen, thank you for your contribution in structuring the papers and in giving a more definite direction to our research ideas. Gerard, thank you for improving the overall quality of our paper. Tom, thank you for your theoretical ideas and algebraic computations and for discussing one of my papers at the SOM PhD conference. It was also really nice to get to know you all outside of work. Other people in the EEF department I want to express my gratitude to are Max Groneck for assisting me in the process of applying for research jobs, Raun van Ooijen for advice and suggestions in several seminars, and Agnieszka Postepska for discussing one of my papers at the SOM PhD conference.

There are a number of friends and colleagues I want to thank as well, for accompanying me over these years, including Fabrizio Masullo, Carlotta Condello, Alfredo Blessano, Dario Pupillo, Cristina Nobile, Chiara Braidotti, Marta Tosone, Katarzyna Tych, Stephanie Weninger, Martina Gruyová, Fernanda Murbach, Tobias Grohmann, Jos Offerein, Jerry Wang, Maite Laméris, David Peng, Ziwei Rao, Stefan Pahl, Hermien Dijk, Roel Freriks, Laura Viluma and Eduard Suari-Andreu.

Finally, I want to thank my family for being so supportive and for all their sacrifices. Completing this PhD thesis would have been impossible without their love and their invaluable presence in my life. And so I thank my mother Loredana, my father Gianluca and my sister Alexandra for being my confidant, my shield and my shelter and for being the best family I could have hoped for.

Contents

1	Introduction	1
1.1	Historical background	1
1.2	About this thesis	4
1.2.1	Education, intelligence, and prescribed medications in adult age	5
1.2.2	Early-life economic determinants of type-2 diabetes	6
1.2.3	Clustered standard errors with few clusters	7
1.3	Conclusions and policy recommendations	8
2	Education, intelligence, and prescribed medications in adult age	11
2.1	Introduction	11
2.2	Causal inference of education on health	13
2.3	Data and descriptive statistics	15
2.3.1	A note on the subsample of our data linked to prescribed medications in adult age	19
2.4	Methodology	19
2.5	Results and Discussion	25
2.5.1	On the non-significant effect of intelligence on education choice	27
2.5.2	Sensitivity analysis	30
2.6	Conclusion	31
2.A	Appendix	34
2.B	Appendix	51
3	Early-life economic determinants of type-2 diabetes	53
3.1	Introduction	53
3.2	Data	56
3.2.1	The Lifelines data	56
3.2.2	Selective fertility by socioeconomic status	60

3.2.3	Unemployment	61
3.3	Empirical implementation	62
3.4	Results	64
3.5	Conclusion	69
3.A	Appendix	71
4	Clustered standard errors with few clusters	79
4.1	Introduction	79
4.2	Basic theory: CRVE, CR2VE and CR3VE	81
4.3	From CR3VE to CR3VE- λ	82
4.4	Efficient computation of CR2VE, CR3VE, CR3VE- λ with $\mathbf{H}_i$	83
4.5	Efficient computation of the Imbens and Kolesar degrees of freedom	84
4.6	Monte Carlo simulations	85
4.7	Empirical exercise	88
4.8	Computational time gain for CR2VE and CR3VE	89
4.9	Conclusion	90
4.A	Appendix	92
5	Samenvatting (Summary in Dutch)	103

List of Figures

2.1	ATEs primary to low education on medications by intelligence . .	28
2.2	ATEs primary to high education on medications by intelligence, age 59-62	29
2.A.1	ATEs primary to low education on medications by intelligence from 9-equation and 3-equation measurement system, age 59-62 .	49
2.A.2	ATEs primary to high education on medications by intelligence from 9-equation and 3-equation measurement system, age 59-62 .	50
3.1	Selective fertility by socioeconomic status over unemployment . .	60
3.2	Provincial unemployment rate series	62

List of Tables

2.1	Prescribed medications by education level, 2006	18
2.2	Average Treatment Effects in percentage points	27
2.A.1	Distribution of IQ scores by education level, 368,687 individuals .	34
2.A.2	Sample characteristics by education level, 368,687 individuals . .	35
2.A.3	Diseases and relative medications	36
2.A.4	Prescribed medications in subsample (Sub) and the Netherlands (NL), in 2007 and 2014	36
2.A.5	Sample characteristics for the 36,337 linked individuals by education	37
2.A.6	Ordered probit estimates for educational attainment in structural model	38
2.A.7	Ordered probit estimates for IQ measurement system, by educa- tion level, in structural model: Raven Test	39
2.A.8	Ordered probit estimates for IQ measurement system, by educa- tion level, in structural model: Arithmetic Test	40
2.A.9	Ordered probit estimates for IQ measurement system, by educa- tion level, in structural model: Language Test	41
2.A.10	Gompertz proportional mortality rate estimates, by education level, in structural model	42
2.A.11	Probit estimates for cardiac diseases medications, by education level, in structural model	43
2.A.12	Probit estimates for obstructive airway diseases medications, by education level, in structural model	44
2.A.13	Probit estimates for hypertension medications, by education level, in structural model	45
2.A.14	Probit estimates for hyperlipidemia medications, by education level, in structural model	46
2.A.15	Probit estimates for diabetes (type 1 and 2) medications, by edu- cation level, in structural model	47

2.A.16	Probit estimates for depression and anxiety medications, by education level, in structural model	48
3.1	Descriptive statistics	57
3.2	Sibling FE and RE estimates on <i>T2D</i>	66
3.3	Marginal effect of unemployment by age and gender on <i>T2D</i>	67
3.4	Sibling FE and RE estimates on <i>T2D</i> : siblings over age 45	68
3.A.1	Distribution of siblings by province of birth	71
3.A.2	Province FE from sibling FE and RE estimates in Table 3.2	71
3.A.3	Probit estimates on <i>T2D</i>	72
3.A.4	Sibling FE and RE estimates on <i>T2D</i> : siblings born in the same province	73
3.A.5	Sibling FE estimates on <i>T2D</i> : unemployment age interaction	74
3.A.6	Sibling FE estimates on <i>T2D</i> : alternative specifications	75
3.A.7	Sibling FE and RE estimates on self-reported <i>T2D</i>	76
3.A.8	Means-comparison <i>t</i> -test: individuals with unreported and reported type-2 diabetes	76
3.A.9	Sibling FE estimates on <i>T2D</i> : placebo regressions	77
4.1	Bias of $se(\hat{\beta})$ based on different methods: 10,000 MC simulations .	86
4.2	Bias of $se(\hat{\gamma})$ based on different methods: 10,000 MC simulations .	87
4.3	Rejection rates <i>policy</i> from 20,000 replications	88
4.4	Time for CR2VE and CR3VE using efficient and inefficient formulas	89
4.A.1	Number of observations per state - CPS 2012 dataset	93

Chapter 1

Introduction

1.1 Historical background

The impact of nature versus the influence of nurture on long-term health (and non-health) outcomes has been debated since the dawn of modern economics. The Scottish economist and philosopher Adam Smith, also known as “The Father of Economics”, believed that humans are born equal and what determines different outcomes in their life is fundamentally a different educational pattern. In *The Wealth of Nations*, Smith (1776) wrote:

“The difference between the most dissimilar characters, between a philosopher and a common street porter, for example, seems to arise not so much from nature, as from habit, custom, and education. When they came in to the world, and for the first six or eight years of their existence, they were, perhaps, very much alike, and neither their parents nor play-fellows could perceive any remarkable difference. About that age, or soon after, they come to be employed in very different occupations. The difference of talents comes then to be taken notice of, and widens by degrees, till at last the vanity of the philosopher is willing to acknowledge scarce any resemblance.”

Of course, Smith’s idea was purely based on intuition and not on systematic collection and analysis of data.

One hundred years later, the pioneering work of Sir Francis Galton (1875) suggested, for the first time, the importance of comparing identical and non-identical

twins to disentangle inherited characteristics from the influence of nurture. This, together with the Mendel's principles of genetics, originally proposed in 1866 but rediscovered only in 1900, cleared the way for assessing, in the 20th century, the contribution of nature on factors such as IQ and skills. In economics, individuals began to be seen as genetically gifted with different innate skills, which were gradually related to health outcomes later in life. With Galton's position, in favor of nature and eugenics, from the 1900s onward researchers entered into the heart of the debate on nature versus nurture.

One of the first and most passionate supporters of Galton's position was the English statistician and eugenicists Karl Pearson who believed nurture has little to no effect on health. Pearson (1913), using microdata, compared correlation indices between close relatives of several health and intelligence measures (e.g., tuberculosis between parent and offspring, deaf-mutism between siblings) to correlation indices between environmental proxies and the same health and intelligence measures (e.g., alcoholism of the parent and intelligence of the child). Based on these indices, he concluded that *"the influence of environment [on health] is not one-fifth that of heredity, and quite possibly not one-tenth of it."*

Conversely, research that put the emphasis on the importance of nurture began to be conducted. Kermack et al. (1934) wrote one of the first scientific papers on the influence of the early-life environment on health in adulthood. Using macrodata on age-specific mortality rates for England, Wales, Scotland and Sweden, the authors interpreted regularities in the drop of age-specific mortality rates within generations, at all periods of life, as evidence that *"the important factor from the point of view of the health of the individual during his whole life is his environment up to the age of say 15 years."* Overall, in the inter-war years, it was gaining ground the idea that the period *in utero* has an impact on some intrinsic characteristics that determine how individuals react to stressful events in adult life (Ben-Shlomo and Kuh, 2004).

In the following decades the interest in the link between conditions early in life and adult health decreased. Yet, from the late 1970s, Forsdahl (1977), and Barker and Osmond (1986) turned the scientific attention back to the question. Moreover, from the 1990s, Barker began to explore the so-called fetal origin hypothesis, an argument that became very popular in the following years. Essentially, Barker (1990, 1995, 1998) argued that genes and environmental, negative experiences *in utero*,

such as undernutrition or stress of the mother, affect health in adult life through their interaction, and by “programming” the organ structures and functions of the fetus. A stressful environment *in utero* can thus produce permanent metabolic and physiological modifications which may increase the risk in adulthood of lower metabolic rate and chronic conditions such as cardiovascular diseases and type-2 diabetes.

The Barker’s hypothesis has been tested and proved to be true in several papers, mainly exploiting natural experiments, such as famine (see Lumey et al., 2011, for a comprehensive literature review) or hurricanes (Sotomayor, 2013), and animal experiments (Fernandez-Twinn and Ozanne, 2006, Portha et al., 2011). More recently, the literature has partially shifted toward the estimation of the effect of less drastic circumstances, such as unemployment rates and economic downturns at birth, both on the short- and on the long-term. The main reason is that such new focus is important from the point of view of the implied health policies, as less drastic circumstances shape on a daily basis any Western society. This literature occasionally reports a positive short-term effect of poor macroeconomic conditions at birth on early-life health (Angelini and Mierau, 2014, Dehejia and Lleras-Muney, 2004). More often, findings point toward negative short- and long-term effects, such as lower birth weight (Alessie et al., 2018, De Cao et al., 2019), higher mortality in adulthood (van den Berg et al., 2009, 2006), higher risk of death and mortality from cardiovascular diseases (Alessie et al., 2019, van den Berg et al., 2011, van den Berg and Modin, 2013).

From the 2000s, contemporary economists put a special emphasis on early-life interventions for the formation and improvement of cognitive and noncognitive skills, linked to better outcomes in adult life. The main argument was that these skills, although partly genetically inherited, can be shaped in infancy to some extent. To explain such findings Cunha and Heckman (2007) developed a theoretical model of cognitive and noncognitive skills formation in childhood which emphasizes how the traditional distinction between nature and nurture is outdated, as models with additive effects of nature and nurture would not do justice to the interaction that exists between genes and environment. Specifically, the authors pointed out that cognitive and noncognitive skills have both an environmental component, especially *in utero* and early in childhood, and a genetic component. In this sense,

skills can be shaped as the manifestation of the gene is partly ruled by the surrounding environment. This model can be extended to justify the link between early-life interventions, the formation of health capabilities and adult health as the final outcome (Heckman, 2007).

Most of the contemporaneous, empirical literature, comprehensively reviewed by Almond et al. (2018), uses observational data and does not adopt randomized control trials to estimate the causal effects of early-life interventions on health outcomes in adult life, for obvious ethical reasons. However, there are few exceptions. Notable examples include the studies on the Perry Preschool Project and the Carolina Abecedarian Project (Conti et al., 2016), and on 231 *Macaca mulatta* monkeys randomly allocated at birth across different rearing conditions (Conti et al., 2012). Results show that disadvantaged children, especially males, whose environment in early-childhood was enriched, experienced better health in adult life than those disadvantaged children with no such environment. Results also show that worse rearing conditions for the *Macaca mulatta* in their childhood (e.g., lack of a mother figure) had negative long-term impact on several health outcomes such as higher probability of developing illnesses and stereotypic behavior (e.g., digit sucking, pacing, head tossing).

1.2 About this thesis

Nature and nurture and their integrated effect on health over the lifespan are currently widely studied in economics. Those important factors that have come into play over the centuries, namely education, skills, and early-life and *in utero* conditions, are still the most investigated by the literature. This thesis uses observational data to analyze all these factors and their link to objective health outcomes in adult age. First, in Chapter 2 we investigate the association between education, by intelligence level, and disease-specific medication use in adult age. Second, in Chapter 3 we estimate the causal effect of macroeconomic conditions around the time of birth on the occurrence of type-2 diabetes in adulthood. Finally, Chapter 4 is dedicated to the computation of standard errors in the case in which there are few, highly unbalanced clusters. Such computation becomes particularly relevant in the presence of variables that are constant within clusters, at the individual level, and vary

across clusters, such as policy variables or health-care reforms. This chapter has been included in this thesis because both data clustered in few, highly unbalanced clusters and variables constant within clusters are recurrent in health economics.

1.2.1 Education, intelligence, and prescribed medications in adult age

Several studies have found a positive association between education and health. Whether part of this association derives from a causal effect is still debated. The reason is that several observed and unobserved confounding factors, such as parental background and intelligence, play an important role in shaping this association by affecting both education choices and health outcomes. In this chapter we investigate the association that exists between education, intelligence and disease-specific medication use in adult age as prescribed by medical doctors.

The basic dataset we use comprehends the entire Dutch male population born between 1944-1947 who survived until the compulsory military examination for national conscription at age 18 (around 400,000 men). Besides a battery of 3 different IQ tests, that will be used to identify latent intelligence, this data includes education level, general health and psychological measures, year of birth, region and city of birth, and family background measures. A sample of this data of around 45,000 men was linked to prescribed medication records in Anatomical Therapeutic Chemical code, collected in 2006, when the individuals were from 59 to 62 years old. We consider only those medications which can be unequivocally linked to specific groups of medical conditions, including cardiac diseases, obstructive airway diseases, hypertension, hyperlipidemia, diabetes (type 1 and 2), and depression and anxiety.

Our novel contribution to the literature is the use of a structural model framework to investigate the association between education and medications prescribed in adult age by medical doctors. The structural model that we estimate analyzes the association of education with prescribed medications by modeling schooling decisions and survival up to adult age and by controlling for a series of observed characteristics and latent intelligence. The methodology we adopt allows the estimation of the effect of education on prescribed medication by level of latent intelligence. We are cautious in claiming causality in the effects that we estimate, as

some important confounding factors such as noncognitive skills are not accounted for. Yet, given the data that we use, we present some motivations that might justify an approximately causal interpretation of the main results. We find that most of the raw associations in the data are not significantly different from these estimated effects. This result suggests the presence of a strong effect of education on prescribed medications for physical diseases, especially at low levels of intelligence, but low or no effect on prescribed medication for depression and anxiety. Higher intelligence significantly decreases medication use for each considered disease, especially in individuals with primary education levels.

From a policy perspective, our results imply that higher education should be seen as a substitute of intelligence. A higher education level, especially in individuals with low intelligence, would be beneficial by decreasing healthcare costs in medication use and, indirectly, by improving the health outcomes in adult age.

1.2.2 Early-life economic determinants of type-2 diabetes

Estimating true causal effects is one of the biggest challenges in health economics. The closer a paper comes to a true causal effect estimation, the better can health policy implications be defined and implemented. In this chapter we exploit data on siblings for the estimation of the causal effect of macroeconomic conditions at the time of birth on the occurrence of type-2 diabetes in adult life.

The macroeconomic conditions that we exploit are the provincial unemployment rates in the Netherlands at the time of birth, which are in principle exogenous from the individual perspective. However, parental selective decisions on fertility might still bias the estimates as parents of low socioeconomic status might postpone conception to times of more favorable macroeconomic conditions, implying that the unemployment rates at birth are not randomly allocated over the spectrum of socioeconomic status. To account for this source of endogeneity we identify and use the siblings in our dataset and estimate sibling fixed-effects regressions.

The data that we use is from Lifelines, a cohort and biobank that covers 51,270 siblings born in the Netherlands from 1950s onward. We use siblings born in times of generous welfare systems as they are more representative of the current newborns. For a sample of 19,005 siblings genetic records are available and were used to find additional, non-reported biological relationships and discard non-biological

relationships due to adoption. To identify type-2 diabetes we use biomarkers on hemoglobin A1c concentration and fasting glucose in the blood and so we avoid the use of self-reported type-2 diabetes, which is known to be largely undiagnosed in the population. The Centers for Disease Control and Prevention (2020), through the National Diabetes Statistics Report, examine the extent of this problem in the US and report that, in 2020, 34.2 million people have diabetes (10.5% of the US population) of whom 7.3 million people (21.4%) are undiagnosed.

We find that higher unemployment at birth increases the probability of type-2 diabetes later in life in both males and females. Moreover, we show that both not accounting for selective decisions on fertility and relying on self-reported data for the identification of type-2 diabetes lead to non-significant estimates. We conclude that, even in times of generous welfare systems, worse macroeconomic conditions at birth have a significant long-term impact on the probability of contracting type-2 diabetes. In line with the Barker's hypothesis, this might be especially true if macroeconomic conditions improve in the years after birth. From a policy perspective, women who are pregnant during economic recessions should be provided with economic and psychological support and should be informed on the higher risk of type-2 diabetes for the newborns in their adult life. Later in life these newborns should be introduced to balanced diet and physical exercise by the parents to prevent type-2 diabetes. Finally, screening for type-2 diabetes should be targeted toward individuals who suffered from poor economic conditions around the time of birth (e.g., immigrants from poor economies).

1.2.3 Clustered standard errors with few clusters

Estimating linear models using data that are clustered in few clusters implies an error term that is usually correlated within clusters, due to homogeneity across observations that belong to the same cluster. This correlation makes the estimated standard errors typically downward biased. Consequently, inference becomes challenging, as the null hypothesis on the estimated parameters is overrejected by too narrow confidence intervals. This problem is particularly relevant for variables that are correlated or constant within clusters, such as policy variables that are constant at the individual level and vary across provinces, regions or states.

If the number of clusters tends to infinity, standard corrections such as the

Cluster-Robust Variance Estimator (CRVE) introduced by Liang and Zeger (1986) makes the standard errors robust to the correlation of the error term within clusters. However, if the number of clusters is small enough (i.e., lower than 20-50), the standard corrections might be insufficient to adjust for the bias and more refined methods might be necessary. These methods are of mainly two types: corrections applied to the variance-covariance matrix of the error term, such as CR2VE and CR3VE, and bootstrapping.

This chapter presents refinements of CR2VE and CR3VE, both used in the empirical literature. We introduce formulas to maximize the computational efficiency of CR2VE and CR3VE, and we develop CR3VE- λ , a new estimator that accounts for the cluster size and works particularly well with few, highly unbalanced clusters. Results from simulations show that more than three hours can be saved for the computation of CR2VE and CR3VE using a dataset with 10 clusters and 5,000 observations per cluster. Moreover, from both simulations and empirical illustrations we show that CR3VE- λ might be the optimal method to use with few, highly unbalanced clusters.

These results have empirical applications for applied researchers in health economics who use clustered data with few clusters and wish to reduce the computational time of common corrections for clustered standard errors. Any research in health economics that uses clustered data with few clusters and estimates linear models (e.g., Alessie et al. 2018, 2019) would gain computational advantage from using our formulas, especially with large cluster size. If these few clusters are also highly unbalanced, CR3VE- λ might become the best choice from an inferential point of view.

1.3 Conclusions and policy recommendations

At the time this introduction is written humanity is facing the COVID-19 crisis, the worst pandemic since the Spanish flu in 1918-1920. Besides the short-term consequences on the healthcare system and the economy, this pandemic will have long-term effects on the health of people who are currently in their childhood or adolescence, or even unborn. Pregnant women are experiencing high stress and anxiety (Preis et al., 2020, Saccone et al., 2020), due to the economic crisis and ne-

cessary measures such as social isolation. Education obtained remotely, which is typically offered during the pandemic to children or adolescents in their schooling years, is of lower quality than standard, in-presence education (Engzell et al., 2021). Moreover, children of high socioeconomic status have at their disposal a better homeschooling environment than those from disadvantaged families and the social and educational gap might further widen (Engzell et al., 2021). Children of low socioeconomic status typically experience less educational and motivational support from their parents, and lower availability of devices such as laptops, tablets and e-readers (Andrew et al., 2020, Bol, 2020, Jæger and Blaabæk, 2020). These differences in homeschooling conditions lead to differences in the effective educational achievements of children and adolescents attending the same type of school. In view of this set of problems, the results that this thesis presents lead to policy recommendations that can be contextualized with respect to the current pandemic.

Chapter 2 suggests that a lower education level, especially in children and adolescents with lower intelligence, might lead to higher medication use for physical diseases in adult life, a negative outcome from the point of view of both the individual and the healthcare system. From a policy perspective, educational achievements should compensate for lack of cognitive ability in preventing chronic diseases in adult life. As described above, the quality of the homeschooling environment is generally negatively correlated with the socioeconomic status of the family. Therefore, during the pandemic, children from disadvantaged families might be at particularly high risk of chronic diseases in adult life. To prevent such a negative outcome, an educational, tailored support should be provided to, at least, children of low socioeconomic status. This could include a tutor, made available from schools, to account for the lack of parental involvement in the educational and motivational support of these children from disadvantaged background (Bol, 2020), the possibility of borrowing free of charge, or paying in installments and at affordable prices, technological devices (e.g., laptops, tablets, e-readers) to account for differences in the availability of such necessary tools across families of different socioeconomic status (Andrew et al., 2020, Bol, 2020), and finally a space to study, available for those worse-off students who do not have their own at home (Andrew et al., 2020, Bol, 2020).

Chapter 3 shows that negative macroeconomic conditions around the time of

birth increase the risk of type-2 diabetes in adulthood, in both males and females. In line with the fetal origin hypothesis, this might be especially true in the current pandemic which, besides leading to an economic and financial crisis, is contributing to a stressful environment for pregnant women and their fetuses through factors such as social isolation and societal and political uncertainty. Pregnant women should be particularly aware of this risk and primary care should be responsible to inform parents of the unborn children. Once aware, they should pay particular attention to classic preventive measures of type-2 diabetes, such as including a healthy diet and some physical activity in the routine of their children. Such a relatively inexpensive measure should be supplemented by economic and psychological support to pregnant women who might be suffering the most during the pandemic (e.g., women of low socioeconomic background or with poor mental well-being).

To conclude, one the most important aspects of this thesis lies in the feasibility of the policy implications that can be derived from the main results. At relatively low costs, some simple actions can be implemented to protect the healthcare system and prevent long-term occurrence of chronic diseases in the adult population. The main target of such policies should be the fragile categories of our society, including pregnant women and their fetuses, as well as children and adolescents whose families are of disadvantaged socioeconomic background.

Chapter 2

Education, intelligence, and prescribed medications in adult age

2.1 Introduction

The association between education and health is well documented in the literature. Highly educated people may have higher income, savings, retirement benefits, and consequently high-quality health insurance and healthcare over their lifetime (Lochner, 2011). They may process health information methodically and adhere carefully to strict treatment regimens for diseases like diabetes and HIV (Goldman and Smith, 2002). Furthermore, they may have a balanced diet or be more likely to avoid bad habits, such as smoking or overconsumption of alcohol, being more aware of the risks of unhealthy behaviors.

The extent to which the association between education and any health outcome implies causation is not straightforward. Skilled individuals may obtain higher education levels and income, and consequently high-quality healthcare; similarly, being born in a wealthy family may have positive effects on both education and quality of diet in childhood. In other words, confounding factors such as intelligence and parental background may affect both educational attainment and health outcomes, playing an important role in shaping this association (Clark and Royer,

This chapter was written with G. Bijwaard.

2013, Fletcher, 2015, Mazumder, 2008, McCartney et al., 2013). In this chapter we investigate the association between education and an objective health outcome: medications in adult age prescribed by medical doctors. Furthermore, we investigate how this association is shaped by intelligence.

Given the absence of experimental studies on the impact of education on health, mostly for ethical reasons, estimating a true causal effect of education on any health outcomes, including prescribed medications, is nearly impossible. Mainly two approaches have been adopted by the literature to estimate a treatment effect of education on health that is as close as possible to the true causal effect: twin studies and instrumental variable (IV) models, including regression discontinuity (RD) designs. In short, data on twins or reforms and laws allow for the estimation of causal effects, but these effects are only representative of specific populations with specific characteristics that may differ from the general population.

To overcome the absence of exogenous and relevant reforms or laws or the absence of data on twins an alternative approach has been used by the literature. This approach, which we will employ, is based on structural models in which the interdependence between education, health outcomes and unobserved skills such as intelligence is explicitly modeled. From the structural model estimates the average treatment effects (ATEs) of attaining a higher educational level on these health outcomes later in life can be derived. Results from such models (Bijwaard, van Kippersluis and Veenman, 2015, Bijwaard, van Poppel et al., 2015, Conti and Heckman, 2010, Conti et al., 2010) show that at least half of the health disparities across educational groups is due to selection in higher education on individual characteristics such as intelligence and parental background.

Many studies on the impact of education on health have been conducted to date using both structural models and IV and twin studies. Structural models are mostly focused on general health outcomes early in life (Conti et al., 2010, Heckman et al., 2018, 2014) or gains in life expectancy (Bijwaard, van Kippersluis and Veenman, 2015, Bijwaard, van Poppel et al., 2015). Most of the IV and twin studies (see Galama et al., 2018, for a literature review) are focused on the impact of education on self-reported health outcomes, mortality, and main causes of diseases such as obesity and smoking behaviour.

Meghir et al. (2018) are the only researchers who investigate before us the im-

pact of education on prescribed medications, exploiting an educational reform in Sweden and finding no significant effects. To our knowledge, no studies have investigated the association between education and medications prescribed in adult age by medical doctors in a structural model framework. With the structural model that we estimate we control for observed confounding factors and latent intelligence. We then compare the ATEs estimated from the structural model to the association between education and prescribed medications, as given by the raw data, and we finally investigate the distribution of these ATEs over latent intelligence.

The estimated empirical effects of education on medication use differ with respect to medications for physical and mental diseases. We show that the association between education and prescribed medications for physical diseases, in the structural model, is almost identical to the raw differences. This suggests a strong impact of education on these prescribed medications, as the association is almost invariant after controlling for meaningful confounding factors, including unobserved intelligence and parental background. However, the empirical results suggest that hardly any effect of education on medication use for depression and anxiety exists. We find that educational gains (i.e. a lower probability of the higher educated of having the medication prescribed) decrease with intelligence for medications for physical diseases but increase with intelligence for medications for depression and anxiety.

2.2 Causal inference of education on health

In the literature mainly two approaches have been employed to determine the causal effects of education on health, using observational data. First, twin studies (e.g., Fujiwara and Kawachi, 2009) compare health outcomes of twins that differ by education level. The common assumption in such studies is that educational attainment for twins is not affected by any omitted variable that may drive education choices and health status later in life. The motivation behind this assumption is that twins share the same genes, parental background, and similar experiences in childhood. Results from such studies indicate that part of the educational differences in cause-specific mortality disappears when accounting for shared family background (Amin et al., 2015, Lundborg, 2013). However, there are some limit-

ations in this approach (see, for a discussion, Bound and Solon, 1999): examples include scarce availability of data on twins, scarce difference in the education level within pair of twins, and low representativeness of the general population, due to factors like lower weight at birth and higher probability of premature birth than singletons.

Second, another approach to estimate causal effects of education on health includes IV models and RD designs (e.g., Clark and Royer, 2013). These models usually exploit reforms or laws, namely exogenous variations in the minimum school-leaving age, creating two groups that are either affected or not by the reform. In other words, the higher educational attainment is driven by external rather than confounding factors. The estimates based on these studies point toward a small (Lleras-Muney, 2005) or even entirely absent (Arendt, 2005, Clark and Royer, 2013) causal effect of education on health outcomes. The main drawback of this approach is that only the local average treatment effect (LATE) of higher compulsory education can be estimated, since these reforms affect only those who would otherwise drop out of school at the pre-reform, minimum school-leaving age. In principle, it is not possible to formulate any conclusions on the causal effect of improving the education level from the returns to higher compulsory education estimated from IV models, since no reforms would increase the compulsory education from, say, high school to university.

A third approach, which we will employ, is based on structural models in which the interdependence between education, health outcomes and unobserved skills such as intelligence is explicitly modeled. Structural models to study the education-health association in presence of latent confounding factors require at least three components: the measurement system, the educational attainment model, and the health outcome model. The measurement system identifies the distribution of latent variables, such as intelligence, that may shape the education-health association, accounting for individual characteristics (e.g., parental background) that affect the measures, such as IQ tests, of these latent variables. The educational attainment model controls for non-random selection into education, accounting for observed and latent confounding factors. Finally, the health outcome model describes for a given education level the probability of a particular health outcome later in life, in our case prescribed medications, controlling for the same confounding factors.

An advantage of using structural models is the ability to estimate both the average treatment effects of education and the distribution of these treatment effects over latent confounding factors (Conti et al., 2010, Heckman et al., 2018, 2014). This implies that it is possible to investigate differences in the impact of education on health outcomes for segments of the population that differ by latent confounding factors (e.g., by intelligence level). A drawback of this approach is that, to control for all possible sources of endogeneity, it needs a rich number of observed, individual characteristics as well as a measurement system to identify latent confounding factors. Only few datasets have information on a battery of IQ tests needed to identify the distribution of latent intelligence. Obviously, researchers who implement these methods should be cautious in considering the treatment effects estimated from structural models as pure causal effects, as unobserved and unmeasured confounding factors might still be present. However, an advantage with respect to the twins and the IV designs is that the treatment effects estimated from the structural models do not suffer from any generalization issues.

2.3 Data and descriptive statistics

Data from a large sample from the nationwide Dutch Military Service Conscription Register for male birth cohorts 1944–1947 is analyzed. All men were called to a military service induction exam, except those living in psychiatric institutions, or in nursing institutes for the blind or for the deaf-mute. The vast majority attended the conscription examination at age 18. We have information from the military examinations for 408,015 men. At the military examination a standardized recording of demographic and socioeconomic characteristics is collected, including education, father's occupation, place and region of birth, date of birth, religion, family size, and birth order. Three standardized psychometric IQ test scores were also collected at age 18. The 28,711 individuals who attended special schools for disabled or illiterate and the remaining 10,431 individuals who did not take all the IQ tests were excluded. Family size, religion, or date of birth are not available for 186 of the remaining individuals which were also omitted from the data. The final data amounts to 368,687 individuals.

The educational attainment is observed at age 18 for each individual. The edu-

cational system in the Netherlands before the Mammoth Act (1968), the most important reform of the Dutch educational system, was characterized by a minimum school-leaving age of 14. All children from age 6 to 12 attended elementary school LO (*Lager Onderwijs*). After that, they could choose among three different options (Haak, 1964): LVO (*Lager Voortgezet Onderwijs*), MVO (*Middelbaar Voortgezet Onderwijs*), and VHMO (*Vorbereidend Hoger- en Middelbaar Onderwijs*). Children choosing LVO attended further years of education at the same elementary school, until the minimum school-leaving age, entering the job market afterwards with no final examination. Children choosing MVO attended either three or four years of education in a different school and a final examination at the end of the programme. Attending VHMO was conditional on admission. VHMO students spent either five or six years in preparation for university and were required to pass a final examination. Final examinations of both VHMO and MVO were almost entirely controlled by the government. We decided to group the individuals in three education levels. The primary level includes LO and LVO; the low level and the high level correspond to MVO and VHMO, respectively. The high level also includes those few individuals who were already attending university at age 18. We decided to group LO and LVO together as LVO required neither admission nor final examination nor changing school after the LO level.

The three IQ tests collected at the military examination are the Raven Progressive Matrices test, an abstract reasoning test which does not depend on language skills and measures fluid reasoning, an arithmetic test, and a language comprehension test. The scores are only observed in 6 ordered classes, where 1 and 6 are the lowest and the highest class, respectively. The scores correspond to the following percentile ranks: 1: 0–10; 2: 11–30; 3: 31–50; 4: 51–70; 5: 71–90; 6: 91–99. Table 2.A.1 in Appendix 2.A shows the distribution of the three IQ test scores by education among the 368,687 individuals in the data. Not surprisingly, IQ scores increase with education.

Selected demographic and socioeconomic characteristics at the time of the military examination by education level are given in Table 2.A.2 in Appendix 2.A. The general and psychological health evaluations were given by a medical doctor. The urban status of place of birth is defined in five categories: city, urban, urbanized rural, rural, and unknown (Bijwaard, van Poppel et al., 2015). Rural and unknown

are grouped together into the category “other”, due to small sample size. Father’s occupation includes four different categories ordered from the highest to the lowest level of remuneration and responsibilities: professional, white collar, skilled, and unskilled; the fifth category includes individuals whose fathers’ occupation is unknown. Religion was classified into three categories as Catholic, Protestant, and “other or none” grouped together. Famine exposure during the *Hongerwinter* represents those who were born between November 1944 and March 1946 in the most famine-exposed cities in the Netherlands, namely Amsterdam, Haarlem, Rotterdam, The Hague, Leiden, and Utrecht (Ekamper et al., 2014). The reason why famine exposure increases with education is that men born in these cities have in general a higher education level.

A subsample of 45,037 individuals of the original sample was linked to the Dutch death register, through to the end of 2014, using unique personal identification numbers. These individuals were originally sampled to study the relation between prenatal famine exposure and mortality (Ekamper et al., 2014). For this reason all the 25,283 men born in the Western Netherlands between November 1944 and March 1946 were included in the subsample. The remaining linked data is composed of a random sample of 10,667 individuals who were born in the same cities but before November 1944 or after March 1946, and a random sample of 9,087 individuals who were born in a different part of the Netherlands in 1944–1947. Those 586 individuals who emigrated after age 18 and before 2006 as well as those 4,185 individuals whose status (alive, dead, or emigrated) is unknown due to unreported emigration were excluded. The remaining 2,506 individuals who attended special schools for disabled or illiterate were also excluded. Finally, 1,423 remaining individuals were also excluded because information on IQ scores, general health, or psychological health were missing. The final subsample amounts to 36,337 individuals.

These men are also linked to administrative data on medications prescribed by medical doctors. The latter are observed over the period 2006–2014, when the youngest were age 59 (in 2006) to age 67 (in 2014) and the oldest were age 62 (in 2006) to age 70 (in 2014). No data on prescribed medications is available before 2006. Prescribed medications are actually observed for 32,946 of the 36,337 individuals since 3,391 men died before 2006: 10.0% of the individuals with primary

Table 2.1. Prescribed medications by education level, 2006

	Education level					All
	Primary	Low	Diff ^a	High	Diff ^a	
Cardiac diseases	8.54%	6.68%	-1.85	4.91%	-3.63	7.35%
Obstructive airway diseases	9.29%	6.84%	-2.45	6.45%	-2.84	8.03%
Hypertension	32.51%	30.26%	-2.24	24.93%	-7.58	30.55%
Hyperlipidemia	28.16%	24.58%	-3.58	18.98%	-9.18	25.51%
Diabetes (type 1 and 2)	10.59%	9.18%	-1.41	5.40%	-5.19	9.29%
Depression and anxiety	9.04%	8.33%	-0.71	7.79%	-1.25	8.1%
Number of individuals	16,809	10,862		5,275		32,946

^a Difference in percentage points with prescribed medications in primary education group.

education, 9.2% of the individuals with low education, and 7.5% of the individuals with high education.

Prescribed medications are observed on annual basis in Anatomical Therapeutic Chemical (ATC) code and classified in three levels: the first indicates the anatomical main group, the second the therapeutic subgroup, and the third the pharmacological subgroup. We investigate only group of medications in the three-level ATC code that can be uniquely linked to common groups of diseases (see van Ooijen et al., 2015). These diseases and the related medications are listed in Table 2.A.3 in Appendix 2.A and are cardiac diseases, obstructive airway diseases, hypertension, hyperlipidemia, diabetes (type 1 and 2),¹ and depression and anxiety.

Table 2.1 shows the prevalence of each prescribed disease-specific medication by education level in 2006. Not surprisingly, we observe that prescribed medications and education have a negative association. Table 2.A.4 in Appendix 2.A shows that prescribed medications of the linked subsample represent the Dutch population well. The prevalence of prescribed medications for both the age-group 60–63 and the age-group 67–70 in our subsample is very similar to the prevalence in the Dutch male population for similar age groups.

Selected demographic and socioeconomic characteristics, by education level, for the subsample linked to prescribed medications are given in Table 2.A.5 in Ap-

¹ We cannot distinguish between type-1 and type-2 diabetes. As type-1 diabetes usually occurs early in life there may be simultaneous impacts between diabetes and education. However, this issue should be negligible as type-2 diabetes is much more common than type-1 in adults.

pendix 2.A. The oversampling of individuals from Western Netherlands is clear from the table. Taller men and men with low birth order tend to have a higher education level. The education level is also strongly related to paternal occupation and men with the highest education tend to have fathers in professional or managerial occupations. As mentioned before, men born in big cities have higher education levels. Being overweight seems to be negatively related to the education level. General health seems only marginally related to the education level while psychological health has clearly a positive association with education.

2.3.1 A note on the subsample of our data linked to prescribed medications in adult age

Only a subsample of 45,037 men in the 1944-1947 cohort was linked to prescribed medications data in the Dutch register. The reason is that linking all the individuals would be excessively time-consuming and expensive, due to the absence of a digitized database. We decided to use both the full sample and the subsample linked to medication data in a two-step estimation framework. Specifically, as described in Section 2.4, we estimate the educational attainment model and the measurement system using the full sample, namely the whole male population born in 1944-1947 and alive and living in the Netherlands at age 18. The reason is that using the full sample leads to very precise estimates (i.e., low standard errors) of the educational attainment model, the measurement system and the distribution of the latent intelligence. These precise point estimates, including the estimated distribution of the latent intelligence, are then imported into the second step estimation, namely the disease-specific prescribed medications model and the mortality model, which both use only the subsample of 45,037 men linked to the prescribed medication data in the Dutch register.

2.4 Methodology

Our aim is to investigate the association between education and prescribed medications in adult age. We are cautious in claiming that the treatment effects we estimate are true causal effects, as we do not control for all possible confounding factors, although in Section 2.5.1 and 2.6 we present motivations that might justify

an approximately causal interpretation of the main results. By accounting for the effect of meaningful observed and latent confounding factors, our final aim is to find at least evidence of a significant impact of education in reducing the probability of having disease-specific medications prescribed in adult age. We also aim to investigate how this effect might vary over intelligence level; in other words, whether higher education is more effective in reducing the risk of prescribed medications for individuals with lower intelligence or for individuals with higher intelligence. In the former case, education would, in a way, compensate for lack of intelligence.

The methodology we use to account for possible confounding of intelligence, which might affect both educational attainment and prescribed medications later in life, is based on the structural equation framework developed by Conti et al. (2010) and Bijwaard, van Kippersluis and Veenman (2015) with some extensions to take into account mortality after the military conscription and before 2006. We only use the prescribed medication data for the year 2006, as this is the first year in which we observe such data. The structural model is composed of four parts and estimated in two steps through maximum likelihood methods. The four parts comprise: (i) an ordered probit model for educational attainment, (ii) a Gompertz mortality model for survival up to 2006, (iii) a probit model for prescribed medications in 2006, and (iv) a measurement system using IQ tests to identify latent intelligence. In the first estimation step we estimate via maximum likelihood part (i) and (iv). In the second estimation step we import the first step point estimates and we estimate part (ii) and (iii), still via maximum likelihood.

In our educational choice model we assume, as in standard discrete choice models, that individuals implicitly evaluate the expected consequences of future choices and their costs, including both monetary and psychic costs, to decide whether to continue their schooling. We are agnostic about the decision model used by the individuals and do not observe the cost of education, just like most of the treatment evaluation literature. We do not impose rational expectations. The decision is influenced by latent intelligence I . For identification of the model, we assume that conditioning on intelligence and observed confounding factors accounts for all the dependence across educational choices, mortality, and prescribed medications. Furthermore, we assume that the value of intelligence is known by the individual but not by the researcher and that it is fixed at the moment an individual makes his

schooling and behavioral choices.

After elementary school (LO) an individual has the following options: staying in the same school until the minimum school-leaving age (LVO); attending either MVO or, if admitted, VHMO in preparation for university. Each individual i gains a latent utility E_i^* from education. We assume a linear model, $E_i^* = \mathbf{X}_{1i}'\boldsymbol{\beta} + \gamma I_i + \epsilon_i$, with the net utility of each schooling level E_i^* that depends on a vector of observed individual characteristics $\mathbf{X}_{1i}$, on latent intelligence I_i and on $\epsilon_i \sim N(0, 1)$, an unobserved random variable that is assumed to be statistically independent of both $\mathbf{X}_{1i}$ and I_i . Given $\mathbf{X}_{1i}$ and I_i , the probability of an individual i choosing education level $E_i = e$, for $e = 1, 2, 3$, is an ordered probit

$$\Pr(E_i = e | \mathbf{X}_{1i}, I_i) = \Phi(b_e - \mathbf{X}_{1i}'\boldsymbol{\beta} - \gamma I_i) - \Phi(b_{e-1} - \mathbf{X}_{1i}'\boldsymbol{\beta} - \gamma I_i), \quad (2.1)$$

where b_0, b_1, b_2, b_3 are unknown cut points, $-\infty = b_0 < b_1 < b_2 < b_3 = +\infty$, and where $\Phi(\cdot)$ is the standard normal cumulative density. The different education levels refer to primary ($e = 1$, LO and LVO), low ($e = 2$, MVO), and high ($e = 3$, VHMO and university) education level.

Once the individual has decided his education level, future mortality and the probability of disease-specific medications prescribed by medical doctors in adult age are potentially causally related to this decision. More importantly, the model allows individuals to select their schooling level anticipating future mortality and differences in prescribed medications by education level. This implies that individuals select their schooling level by comparing (future) outcomes by schooling level. To deal with the issue of schooling choice based on future outcomes, we use potential outcome models in which we allow unobserved (from the point of view of the researcher, but known to the individual), latent variables and observed variables to be correlated across schooling levels, mortality rates, disease-specific medications prescribed by medical doctors. These models are also known as Roy-type models (Roy, 1951) and they are commonly applied in economics to model choices based on potential outcomes.

Since we only observe prescribed medications from 2006 and until 2014, we model the probability of surviving until 2006. The reason is that low education is associated with premature death (see Section 2.3). Ignoring possible selective attrition due to death before 2006 might bias the estimated treatment effects of edu-

cation on medications prescribed in adult age. To account for this selective attrition we include a Gompertz proportional mortality model, with a mortality rate differentiated by education level and exponentially increasing in age t , $\lambda_e(t|\mathbf{X}_{2i}, I_i, E_i = e) = \exp(\psi_e t + \mathbf{X}'_{2i}\boldsymbol{\kappa}_e + \omega_e I_i)$, with shape parameter ψ_e , and its implied survival function

$$S_e(t|\mathbf{X}_{2i}, I_i, E_i = e) = \exp\left(-\frac{1}{\psi_e}(\exp(\psi_e t) - 1)\right) \exp(\mathbf{X}'_{2i}\boldsymbol{\kappa}_e + \omega_e I_i), \quad (2.2)$$

depending on observed individual characteristics $\mathbf{X}_{2i}$ ² and latent intelligence I_i .

We model the probability of prescribed medications in 2006 through a probit model by education level. An individual i with education level $e = 1, 2, 3$ who survived up to 2006 is assumed to be affected by a particular disease if the latent utility in using the respective medication m is greater than 0, $U_{emi}^* > 0$. This latent utility linearly depends on I_i and $\mathbf{X}_{2i}$, $U_{emi}^* = \mathbf{X}'_{2i}\boldsymbol{q}_{em} + \eta_{em}I_i + \varepsilon_{emi}$, with $\varepsilon_{emi} \sim N(0, 1)$. Therefore we have that

$$\Pr(U_{emi} = 1|\mathbf{X}_{2i}, I_i, E_i = e) = \Phi(\mathbf{X}'_{2i}\boldsymbol{q}_{em} + \eta_{em}I_i), \quad (2.3)$$

where $U_{emi} = 1$ if medication m is prescribed in 2006 to individual i with education e .

The structural model is closed by a measurement system to identify latent intelligence I . The measurement system consists of nine measurement equations that link the three IQ tests $q = 1, 2, 3$, separately for each education level, to latent intelligence and observed individual characteristics.³ We assume that intelligence I is normally distributed with zero mean and unknown variance σ^2 . The continuous score M_{eqi}^* for an individual i with education level e on IQ test q is not observed; we only observe this score in 6 ordered classes, $c = 1, \dots, 6$, and therefore we choose to model our 9 measurement equations as 9 ordered probit. The latent variable M_{eqi}^* is

² Note that we use a different set of observed characteristics and different dataset for (1) education choice and the measurement system and (2) the survival and disease-specific, prescribed medications probability. The first are estimated using the whole sample of recruits but variables that might be endogenous due to simultaneous causality (i.e., overweight, general health and psychological health) are excluded ($\mathbf{X}_1$). The second are estimated using the linked sample and including these additional health indicators ($\mathbf{X}_2$).

³ We treat intelligence I as a latent variable because we aim to control for the effect of observed factors on the measures of I (the IQ scores). Note that for identification we need at least three different intelligence tests.

defined as a linear combination of observed $\mathbf{X}_{1i}$ and I_i , $M_{eqi}^* = \mathbf{X}_{1i}'\boldsymbol{\zeta}_{eq} + \zeta_{eq}I_i + \tau_{eqi}$, where $\tau_{eqi} \sim N(0, 1)$. Given I_i , $\mathbf{X}_{1i}$, and E_i , the probability of an individual i with education level e scoring $M_{eqi} = c$ on IQ test q is

$$\Pr(M_{eqi} = c | \mathbf{X}_{1i}, I_i, E_i = e) = \Phi(a_{eqc} - \mathbf{X}_{1i}'\boldsymbol{\zeta}_{eq} - \zeta_{eq}I_i) - \Phi(a_{eqc-1} - \mathbf{X}_{1i}'\boldsymbol{\zeta}_{eq} - \zeta_{eq}I_i), \quad (2.4)$$

where $a_{eq0}, \dots, a_{eq6}$ are the unknown cut points, $-\infty = a_{eq0} < \dots < a_{eq6} = +\infty$. Since I_i is unobserved, we need to establish its unit of measurement by constraining $\zeta_{11} = 1$.

We include nine ordered probit (i.e., one for each IQ test and education level) instead of three ordered probit (i.e., one for each IQ test only) to account for the effect of education on IQ scores. The reason is that IQ scores are observed at age 18, when the individuals already differ by education level. Hence, a three-equation measurement system does not control for education and it is likely to overestimate the impact of latent intelligence on IQ scores, as education is correlated with intelligence. Therefore, the impact of latent intelligence on prescribed medications would be overestimated as it would capture part of the education effect. Consequently, the estimated effect of education on prescribed medications would be downward biased. In Section 2.5.2 we discuss this bias by comparing the estimated effect of education using a three-equation measurement system and the nine-equation measurement system described above.

As already mentioned, we follow the literature (Heckman et al., 2018, 2014) and we jointly estimate the measurement system and the educational attainment model in a first step, using all the Dutch men born between 1944-1947 who were alive and attended the military examination at age 18. After importing the point estimates of the first step in the likelihood, we estimate the probit model for prescribed medications and the Gompertz mortality rate in a second step, using the linked individuals and accounting for survival up to 2006. Assuming, for identification purposes, that given $\mathbf{X}$ the interdependence among the four parts of the structural model comes from I only and averaging over the distribution of I ,⁴ the likelihood of the first step

⁴ Averaging over the distribution of I requires the computation of an integral that cannot be solved analytically. However, Gaussian quadrature can approximate this one-dimensional integral very well. Gaussian quadrature is a numerical integration method based on Hermite polynomials which provides

is

$$\mathcal{L}_1 = \prod_i \int \left\{ \prod_{e=1}^3 \left[\prod_{c=1}^6 \left[\Pr(M_{e1i} = c | \mathbf{X}_{1i}, I, E_i) \right]^{R_{ci}} \left[\Pr(M_{e2i} = c | \mathbf{X}_{1i}, I, E_i) \right]^{A_{ci}} \right. \right. \\ \left. \left. \times \left[\Pr(M_{e3i} = c | \mathbf{X}_{1i}, I, E_i) \right]^{L_{ci}} \right]^{E_{ei}} \right\} \left\{ \prod_{e=1}^3 \left[\Pr(E_i = e | \mathbf{X}_{1i}, I) \right]^{E_{ei}} \right\} dF(I), \quad (2.5)$$

where $E_{ei} = 1$ if individual i obtained education level e (i.e., $E_i = e$), and 0 otherwise; $R_{ci} = 1$ if individual i scored c in the Raven test and 0 otherwise; $A_{ci} = 1$ if individual i scored c in the arithmetic test and 0 otherwise; $L_{ci} = 1$ if individual i scored c in the language test and 0 otherwise. We do not need exclusion restrictions for identification, as we assume independence of the events conditional on $\mathbf{X}$ and I .

After importing the point estimates of the first step, including $\hat{F}(I)$, the estimated distribution of the latent intelligence (in fact, the estimated variance of the normal distributed latent intelligence), we maximize the following likelihood in the second step

$$\mathcal{L}_2 = \prod_i \int \left\{ \prod_{e=1}^3 \left[\prod_{c=1}^6 \left[\hat{\Pr}(M_{e1i} = c | \mathbf{X}_{1i}, I, E_i) \right]^{R_{ci}} \left[\hat{\Pr}(M_{e2i} = c | \mathbf{X}_{1i}, I, E_i) \right]^{A_{ci}} \right. \right. \\ \times \left[\hat{\Pr}(M_{e3i} = c | \mathbf{X}_{1i}, I, E_i) \right]^{L_{ci}} \left. \right]^{E_{ei}} \left\{ \prod_{e=1}^3 \left[\hat{\Pr}(E_i = e | \mathbf{X}_{1i}, I) \right]^{E_{ei}} \right\} \\ \times \left\{ \prod_{e=1}^3 \left[\left(\lambda_e(t_i | \mathbf{X}_{2i}, I, E_i) \right)^{\delta_i} \left(\frac{S_e(t_i | \mathbf{X}_{2i}, I, E_i)}{S_e(18 | \mathbf{X}_{2i}, I, E_i)} \right) \right]^{E_{ei}} \right\} \\ \times \left\{ \prod_{e=1}^3 \left[\left(\Pr(U_{emi} = 1 | \mathbf{X}_{2i}, I, E_i) \right)^{U_{emi}} \right. \right. \\ \left. \left. \times \left(\Pr(U_{emi} = 0 | \mathbf{X}_{2i}, I, E_i) \right)^{(1-U_{emi})} \right]^{E_{ei}(1-\delta_i)} \right\} d\hat{F}(I | T > 2005), \quad (2.6)$$

where $\delta_i = 1$ if individual i died before 2006 and where we account for differences in the distribution of I , conditioning on surviving until 2006 (i.e., $T > 2005$). The

an efficient approximation for evaluating indefinite integrals based on normal distributions. Hence, we estimate the parameters using maximum likelihood on the basis of Gaussian quadrature approximation. The STATA estimation programs are available upon request.

estimated conditional distribution of I is defined as

$$d\hat{F}(I|T > 2005) = \frac{\sum_{e=1}^3 \left[\hat{\Pr}(E_i = e|\mathbf{X}_{1i}, I) S_e(2006|\mathbf{X}_{2i}, I, E_i = e) \right] \hat{f}(I)}{\int \sum_{e=1}^3 \left[\hat{\Pr}(E_i = e|\mathbf{X}_{1i}, I) S_e(2006|\mathbf{X}_{2i}, I, E_i = e) \right] \hat{f}(I) dI}. \quad (2.7)$$

We estimate the model for each prescribed medication separately. Based on these estimates, we derive the ATEs of choosing a higher education level on prescribed medications in adult age. The ATEs are estimated averaging over the distribution of the observed individual characteristics $\mathbf{X}_2$ and the estimated unconditional distribution of latent intelligence I . The ATEs are defined as

$$ATE_{1sm} = \int \int ATE_{1sm}(\mathbf{X}_2, I) dF(\mathbf{X}_2) d\hat{F}(I), \quad (2.8)$$

where $ATE_{1sm}(\mathbf{X}_2, I) = \hat{\Pr}(U_{sm} = 1|\mathbf{X}_2, I) - \hat{\Pr}(U_{1m} = 1|\mathbf{X}_2, I)$ is the predicted gain, in terms of prescribed medication m , of going from education level 1 (primary) to $s = 2, 3$ (low and high, respectively), conditional on observed characteristics $\mathbf{X}_2$ and latent intelligence I .

Finally, we derive the distribution of the treatment effects over a significant range of latent intelligence. The ATEs conditional on latent intelligence are defined as

$$ATE_{1sm}(I) = \int ATE_{1sm}(\mathbf{X}_2, I) dF(\mathbf{X}_2), \quad (2.9)$$

where $I \in [-2; 2]$. This range represents approximately 98% of the population, according to the estimated variance of latent intelligence as reported in Table 2.A.6 in Appendix 2.A.

2.5 Results and Discussion

Before we turn to the estimated ATEs of education, we shortly discuss the parameter estimates of the structural model. The estimates of the education choice, measurement system, the Gompertz proportional mortality rate and the prescribed medication model are reported in Table 2.A.6-2.A.16 in Appendix 2.A. Intelligence positively affects all IQ scores, at each education level, but it does not significantly

affect education choice. This rather striking result is discussed in Section 2.5.1. Interestingly, intelligence has a decreasing impact on the probability of mortality and prescribed medications for physical diseases as education increases, but an increasing impact on the probability of prescribed medications for depression and anxiety as education increases. For primary educated individuals, intelligence decreases significantly the probability of prescribed medications for all the diseases that we consider; for individuals with low education, intelligence decreases significantly the probability of prescribed medications for obstructive airway diseases, hyperlipidemia, and depression and anxiety; for individuals with high education, intelligence only decreases significantly the probability of prescribed medications for depression and anxiety. Also, intelligence decreases significantly the mortality for individuals with primary and low education. Parental background, proxied by father's occupation, has some significant effects, at some education levels only, on mortality and on the probability of prescribed medications for cardiac and obstructive airway diseases, hypertension, and hyperlipidemia. It also strongly affects education choice and all IQ scores at each education level.

The ATEs defined in equation (2.8) and reported in Table 2.2 suggest that all prescribed medications for physical diseases in adult age are significantly reduced by education. However, the ATE of education on depression and anxiety medication is low and, from primary to low education, not significant. Although we miss to control for potentially relevant confounding factors the ATEs on almost all prescribed medications do not differ significantly from the raw associations between education and prescribed medications reported in Table 2.1. This suggests that higher education might causally reduce prescribed medications in adult age. The only exception is the ATEs on prescribed medications for hyperlipidemia which, from primary to high education, is significantly lower at the 95% confidence level than the raw difference.

Next, Figure 2.1 and Figure 2.2 show the ATEs of primary to low and to high education, respectively, conditional on different levels of latent intelligence, as defined in equation (2.9). For all medications for physical diseases we observe lower gains from higher education levels for highly intelligent individuals. In other words, particularly skilled individuals do not gain much from higher education levels in terms of lower chances of prescribed medications for physical diseases in adult age.

Table 2.2. Average Treatment Effects in percentage points

Medication	Prim. to Low	Std. err.	Prim. to High	Std. err.
Cardiac diseases	−1.53**	0.33	−2.90**	0.49
Obstructive airway diseases	−2.35**	0.34	−3.21**	0.49
Hypertension	−2.24**	0.57	−6.44**	0.89
Hyperlipidemia	−2.87**	0.54	−7.31**	0.83
Diabetes (type 1 and 2)	−1.34**	0.37	−4.64**	0.52
Depression and anxiety	−0.60	0.40	−1.57 ⁺	0.63

⁺ $p < 0.05$ and ** $p < 0.01$

Oppositely, for depression and anxiety medication we observe higher gains from higher education levels for highly intelligent individuals. This suggests that education is more effective in preventing medication use for depression and anxiety in adult age for individuals with higher intelligence.

Figure 2.1 and Figure 2.2 also show the 95% confidence interval of the ATEs over latent intelligence. These confidence intervals are derived from the standard errors of the ATEs conditional on latent intelligence. We compute these standard errors as described in Appendix 2.B. It is clear that the ATEs on prescribed medications for some of the physical diseases (e.g., obstructive airway diseases) are not significant for highly intelligent individuals. The ATE from primary to low education on depression and anxiety medication is never significant and the ATE from primary to high education on depression and anxiety medication is not significant for individuals with low intelligence.

2.5.1 On the non-significant effect of intelligence on education choice

As already mentioned, although we find an effect of intelligence on mortality and prescribed medications in adult age, we do not find any significant effect of intelligence on education choice. The reason is very likely that, before the Mammoth Act (1968), the educational opportunities in the Netherlands were highly unequal and very much related to social class (Frijhoff and Spies, 2004, Prick, 2006) and urbanity of the place of residence (Lauwerys and Scanlon, 2013), as we also find. One of the biggest challenges the Mammoth Act dealt with was indeed the necessity of increasing the social mobility that the education system of the time strongly impeded.

Figure 2.1. ATEs primary to low education on medications by intelligence

Figure 2.2. ATEs primary to high education on medications by intelligence, age 59-62

Before the Mammoth Act, individuals typically received a level of education that reflected the traditional and parental social niche (van Lutsenburg Maas, 1964). Remarkably, from the Middle Ages and until the Mammoth Act this pattern of very little social class mobility through educational choices had never been tackled (van Lutsenburg Maas, 1964). It is therefore likely that latent skills, including both cognitive and noncognitive skills, counted poorly in determining education choices. Of course, skills such as intelligence may have determined the success of the educational choice, but this occurred mostly for university education, where the drop out rate was nearly one half (van Lutsenburg Maas, 1964). This, of course, does not affect our results, as university education is included with VHMO in the high education level. Moreover, university access was only possible with a VHMO diploma, so no university education could in principle have occurred later in life without a VHMO diploma (van Lutsenburg Maas, 1964), excluding very few exceptions. In other words, the educational levels that we observe at age 18 are, in principle, invariant over time.

2.5.2 Sensitivity analysis

We estimate two alternative structural models as a robustness check. First, we ignore possible reverse causation of education on IQ scores and estimate a model with only a three-equation measurement system, one for each IQ test and not conditioning on the education level. Figure 2.A.1 and Figure 2.A.2 in Appendix 2.A compare the implied ATEs over latent intelligence with the ATEs estimated from our original model with a nine-equation measurement system. Overall, the ATEs on prescribed medications appear to be underestimated using the three-equation measurement system. The reason is that such a measurement system does not control for the effect of education on the IQ scores and therefore latent intelligence captures part of this education effect. Consequently, part of the impact of education on prescribed medications is captured by latent intelligence, which in the estimates also affects education choice significantly. This underlines the importance of using the nine-equation measurement system (2.4) to account for reverse causation of education on IQ scores.

Second, we include health measurements at age 18 (i.e., overweight, general and psychological health) into both the measurement system and the educational

attainment model. These early-life health measurements are already included in the mortality and the prescribed medications probability equations. In the original specification of the structural model we exclude these health measurements as they may be endogenous (i.e., they may be affected by education). The resulted ATEs from the structural model with health measurements are similar to the ATEs reported in Figure 2.1 and Figure 2.2. This suggests that, conditional on latent intelligence and observed confounding factors, there may be no remaining association between health early in life and education choice.

2.6 Conclusion

We have estimated a structural model that controls for the effect of observed confounding factors and latent intelligence on educational attainment, survival up to adult age, and prescribed medications in adult age. Our results suggest a substantial effect of education on prescribed medications for physical diseases. This is in line with recent research on the association between education and health using a structural model approach (Conti et al., 2010). However, we have obtained results that suggest, at the most, a weak impact of education on prescribed medications for depression and anxiety. To investigate the distribution of the treatment effects, we have estimated the average treatment effects of education on prescribed medications by intelligence level. These conditional ATEs suggest that individuals with higher intelligence have lower gains of education in prescribed medications for physical diseases. However, they have higher gains of education in prescribed medications for depression and anxiety. In other words, education seems to be more effective in reducing the risk of prescribed medications for physical diseases on individuals with lower intelligence and for depression and anxiety on individuals with higher intelligence.

Our study has three distinct strengths compared to previous research. First, a clear advantage is the large sample size which allows the estimation of the detailed structural model with three education levels accounting for confounding in the education attained. Second, the data is population-based and not prone to self-selection because military conscription was mandatory in the Netherlands during the 1960s. Third, our statistical method is based on a structural model in which the

education attained, mortality, and prescribed medications are modeled simultaneously; this method accounts for the confounding effect of intelligence on both mortality and prescribed medications. This enables us to investigate in depth the association between education and prescribed medication and find substantial evidence of a causal effect on prescribed medications for physical diseases. These conclusions do not suffer generalization issues inherent to using compulsory schooling reforms or twins to account for confounding.

There are also some limitations to this study. One limitation is that we do not have military examination information nor other large data containing intelligence tests for women that would allow for similar analyses. A more substantive issue is that lack of data prevents us from controlling for some potentially relevant confounding factors that may significantly shape the association between education and prescribed medications in adult age. Conditional on these confounding factors, any association left between education and health would be causal (Heckman et al., 2014). Besides latent intelligence, we account for a large set of observed early-life characteristics, including father's occupation and the urban status of place of birth. However, we do not account for noncognitive skills, differences in the local economic conditions within the Netherlands, and health early in life in the educational attainment model.

Noncognitive skills have been shown to play an important role in determining education choice (Heckman et al., 2018, 2014). Unfortunately, measures of noncognitive skills such as self-discipline, perseverance, and consciousness are not observed in our data. However, as explained in Section 2.5.1, before the Mammoth Act (1968) educational opportunities in the Netherlands were highly unequal and strictly related to the family social class and to the urbanity of the place of residence, with the majority of rural areas being culturally isolated. The estimates of the educational attainment model (see Table 2.A.6 in Appendix 2.A) provide evidence of this, since intelligence I does not affect education choices significantly. Similarly, in our data noncognitive skills might play a minor role in determining education choices.

Noncognitive skills may still have a direct effect on diseases in adult age. Accounting only for psychological health at age 18 is inadequate to capture the multidimensional heterogeneity in noncognitive skills across individuals. The latter,

rather than biasing the estimated effect of education, may lead to an overestimated effect of intelligence on health outcomes (Conti et al., 2010). The reason is that IQ scores are affected by noncognitive skills (Borghans et al., 2008) and thus our latent variable may capture not only pure intelligence but a large set of cognitive and non-cognitive skills. Whether noncognitive skills have a significant and direct impact on health in adult age is an open question: as a matter of fact, the structural-model literature has only found an indirect (through education), significant effect of non-cognitive skills on physical health and self-esteem in middle age,⁵ and not a direct effect on the same health outcomes (Heckman et al., 2014). However, noncognitive skills seem to have a significant and direct impact on smoking behavior in middle age (Heckman et al., 2014).

Differences in local or regional economic conditions at birth and/or in childhood might also affect both education choices (Heckman et al., 2014) and health in adulthood (Alessie et al., 2019, van den Berg et al., 2009, 2011).⁶ Although we control for the famine effect at birth, reliable data on proxies such as provincial unemployment rate are not available for the Netherlands before the 1950s. Consequently, we cannot account for its effect on neither education choice nor health outcomes later in life. Depending on the outcome, this effect might be rather small (Alessie et al., 2019) or more remarkable (van den Berg et al., 2009, 2011).⁷

Except for height, we do not control for health early in life in the educational attainment model. However, this should not be a major issue for two reasons. First, the literature has found no remaining association between health early in life and education choice, especially for men, after controlling for early-life environment, parental background, and latent skills (Conti and Heckman, 2010, Conti et al., 2010). Second, we estimate an alternative model which includes potentially endogenous, observed in early-life, health variables in the educational attainment model, finding no differences in the estimates from our main specification of the model. In the prescribed medications probability and mortality equations we already control for health early in life, including height, psychological and general health status at age 18, and overweight.

⁵ Which, as explained in Section 2.5.1, is likely not to apply to our data.

⁶ For additional references we refer to Chapter 3 which is entirely dedicated to the link between early-life macroeconomic conditions and health later in adulthood.

⁷ See also Chapter 3.

2.A Appendix

Additional tables

Table 2.A.1. Distribution of IQ scores by education level, 368,687 individuals

	Education level		
	Primary	Low	High
Raven:			
1 (lowest)	5.2%	1.2%	0.3%
2	14.3%	3.7%	0.9%
3	19.4%	9.3%	3.5%
4	25.3%	20.4%	11.4%
5	26.1%	38.5%	37.9%
6 (highest)	9.7%	27.0%	46.0%
Arithmetic:			
1 (lowest)	7.2%	0.3%	0.1%
2	23.7%	2.9%	0.4%
3	23.9%	9.8%	1.4%
4	23.2%	27.8%	8.0%
5	17.4%	40.7%	38.1%
6 (highest)	4.5%	18.6%	52.1%
Language:			
1 (lowest)	4.9%	0.2%	0.0%
2	21.0%	1.2%	0.2%
3	28.5%	5.4%	0.7%
4	28.8%	20.7%	6.6%
5	14.4%	49.6%	43.4%
6 (highest)	2.4%	22.9%	49.1%
Number of individuals	210,212	107,829	50,646

Table 2.A.2. Sample characteristics by education level, 368,687 individuals

	Education level		
	Primary	Low	High
Average height (cm)	176.8	178.1	179.1
Average birth order	2.8	2.4	2.2
Average family size	4.7	4.0	3.9
Father's occupation:			
Professional	8.2%	17.3%	37.7%
White collar	21.5%	37.5%	38.8%
Farm owner	12.4%	6.6%	5.1%
Skilled	32.6%	22.5%	9.7%
Unskilled	20.1%	11.8%	4.7%
Unknown	5.1%	4.3%	4.1%
Urban status of place of birth: ^a			
City	30.1%	38.0%	42.5%
Urban	19.9%	21.3%	21.9%
Urbanised rural	21.5%	18.6%	17.6%
Other	28.5%	22.0%	18.1%
Religion:			
Catholic	43.3%	38.0%	39.0%
Protestant	40.3%	43.2%	41.2%
Other or none	16.4%	18.7%	19.8%
Famine exposure	5.4%	7.6%	7.9%
Region of birth:			
West	41.9%	48.1%	49.6%
South	24.2%	20.3%	21.6%
East	19.0%	16.0%	14.4%
North	14.9%	15.6%	14.4%
Year of birth:			
1944	23.7%	24.5%	23.7%
1945	22.7%	22.2%	21.2%
1946	30.2%	30.5%	30.8%
1947	23.4%	22.8%	24.4%
Overweight ^b	7.2%	5.8%	5.0%
General health: ^c			
Fit	84.3%	83.4%	83.4%
Almost	6.6%	7.4%	7.3%
Fairly	1.9%	1.9%	1.9%
Unfit	7.2%	7.3%	7.4%
Psychological fitness: ^c			
Fit	75.7%	80.9%	82.7%
Almost	20.1%	16.3%	14.9%
Fairly	0.1%	0.1%	0.1%
Unfit	4.1%	2.8%	2.3%
Number of individuals	210,212	107,829	50,646

^a City: > 100,000 inhabitants; Urban: cities < 100,000 inhabitants; Urbanised rural: rural communities with < 20% farming population; Rural: rural communities with > 20% farming population.

^b BMI (Body Mass Index) higher than 25

^c Not available for all 368,687 individuals (we do not control for health early in life in the first step).

Table 2.A.3. Diseases and relative medications

Diseases or conditions	Medications (ATC code)
Cardiac diseases	C01, C03C
Obstructive airway diseases	R03
Hypertension	C02, C03A, C07, C08, C09A,B
Hyperlipidemia	C10
Diabetes (type 1 and 2)	A10
Depression and anxiety	N05B, N06A

Table 2.A.4. Prescribed medications in subsample (Sub) and the Netherlands (NL), in 2007 and 2014

	2007		2014	
	Sub age 60-63	NL age 60-65	Sub age 67-70	NL age 65-70
Cardiac diseases:				
C01	6.3%	6.2%	8.4%	7.2%
C03C	2.7%	2.9%	4.6%	3.9%
Obstructive airway diseases:				
R03	8.9%	9.3%	11.8%	11.7%
Hypertension:				
C02	0.8%	0.8%	1.1%	1.0%
C03A	5.6%	5.6%	9.9%	9.1%
C07	20.4%	20.5%	27.7%	25.3%
C08	9.6%	9.6%	15.9%	14.5%
C09A	13.3%	13.1%	20.0%	18.2%
C09B	2.6%	2.5%	3.6%	3.0%
Hyperlipidemia:				
C10	27.4%	27.2%	40.8%	38.0%
Diabetes (type 1 and 2):				
A10	10.0%	10.2%	15.5%	14.0%
Depression and anxiety:				
N05B	8.7%	8.0%	2.0%	1.9%
N06A	5.3%	5.3%	5.5%	5.6%

Source for the Netherlands: CBS

Table 2.A.5. Sample characteristics for the 36,337 linked individuals by education

	Education level		
	Primary	Low	High
Average height (cm)	177.0	178.1	179.3
Average birth order	2.6	2.2	2.1
Average family size	3.8	3.4	3.4
Father's occupation:			
Professional	9.6%	16.7%	37.9%
White collar	26.7%	43.3%	43.7%
Farm owner	4.9%	2.2%	1.8%
Skilled	35.0%	23.5%	9.5%
Unskilled	17.4%	9.5%	3.4%
Unknown	6.4%	4.9%	3.7%
Urban status of place of birth: ^a			
City	79.4%	86.6%	87.9%
Urban	6.1%	4.6%	4.8%
Urbanised rural	6.6%	4.3%	4.0%
Other	7.9%	4.5%	3.3%
Religion:			
Catholic	34.8%	30.3%	31.5%
Protestant	35.8%	40.0%	39.7%
Other or none	29.4%	29.7%	28.8%
Famine exposure	53.8%	60.4%	59.9%
Region of birth:			
West	83.0%	88.3%	89.2%
South	7.6%	4.9%	5.0%
East	5.5%	3.6%	3.6%
North	3.9%	3.1%	2.2%
Year of birth:			
1944	18.2%	17.5%	17.9%
1945	39.9%	41.6%	40.3%
1946	29.1%	29.3%	29.7%
1947	12.9%	11.6%	12.2%
Overweight ^b	7.6%	6.2%	5.5%
General health:			
Fit	85.0%	83.0%	83.2%
Almost	6.3%	7.4%	7.3%
Fairly	2.0%	2.3%	2.1%
Unfit	6.7%	7.5%	7.4%
Psychological fitness:			
Fit	76.0%	80.9%	82.1%
Almost	19.8%	15.8%	14.9%
Fairly	0.1%	0.1%	0.1%
Unfit	4.1%	3.2%	2.9%
Number of individuals	18,671	11,963	5,703

^a City: > 100,000 inhabitants; Urban: cities < 100,000 inhabitants; Urbanised rural: rural communities with < 20% farming population; Rural: rural communities with > 20% farming population.

^b BMI (Body Mass Index) higher than 25

Table 2.A.6. Ordered probit estimates for educational attainment in structural model

	Coeff.	Std. err.
Intelligence	0.0285	0.0184
Height	0.0191**	0.0003
Birth order	-0.0510**	0.0015
Family Size	-0.0379**	0.0012
Professional	0.4148**	0.0059
Farm Owner	-0.5681**	0.0080
Skilled	-0.6570**	0.0056
Unskilled	-0.7399**	0.0067
Unknown	-0.3788**	0.0096
City	0.0544**	0.0060
Urbanised Rural	-0.0578**	0.0063
Other	-0.1260**	0.0065
North	0.0625**	0.0065
South	0.0271**	0.0060
East	-0.0716**	0.0061
Catholic	0.0408**	0.0051
Other or No religion	-0.0349**	0.0059
Birth Year 1944	0.0318**	0.0061
Birth Year 1946	-0.0149**	0.0058
Birth Year 1947	-0.0374**	0.0062
Spring	-0.0095	0.0055
Summer	-0.0181 ⁺	0.0057
Autumn	0.0306**	0.0058
Famine Exp.	0.0152	0.0094
Cut Point 1	-0.3709**	0.0093
Cut Point 2	0.6653**	0.0094

⁺ $p < 0.05$ and ** $p < 0.01$

To identify latent intelligence I , we estimate the educational attainment model and the measurement system jointly. We estimate the variance of the normal distribution $N(0, \sigma^2)$ of I . The estimated variance is $\hat{\sigma}^2 = 0.7652$ with standard error 0.0009.

Table 2.A.7. Ordered probit estimates for IQ measurement system, by education level, in structural model: Raven Test

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	1.0000		0.8525**	0.0079	0.8406**	0.0155
Height	0.0212**	0.0005	0.0174**	0.0006	0.0122**	0.0010
Birth Order	-0.0320**	0.0021	-0.0352**	0.0032	-0.0460**	0.0052
Family Size	-0.0227**	0.0017	0.0160**	0.0026	0.0284**	0.0040
Professional	-0.0104	0.0125	0.0510**	0.0126	0.0403 ⁺	0.0147
Farm Owner	-0.4585**	0.0122	-0.0938**	0.0191	-0.2214**	0.0315
Skilled	-0.3183**	0.0098	-0.0641**	0.0130	-0.0633 ⁺	0.0231
Unskilled	-0.4467**	0.0109	-0.1603**	0.0158	-0.1374**	0.0311
Unknown	-0.3092**	0.0148	-0.0250	0.0206	-0.0957 ⁺	0.0311
City	0.0930**	0.0092	0.0243	0.0119	0.0257	0.0177
Urbanised Rural	-0.1111**	0.0093	-0.0819**	0.0129	-0.0452	0.0200
Other	-0.1219**	0.0093	-0.1230**	0.0133	-0.1401**	0.0213
North	-0.1568**	0.0097	-0.2117**	0.013	-0.1564**	0.0204
South	0.0384**	0.0089	0.0888**	0.0125	0.0257	0.0185
East	-0.1672**	0.0089	-0.0824**	0.0126	-0.0583 ⁺	0.0199
Catholic	0.0669**	0.0076	0.0075	0.0103	-0.1121**	0.016
Other or No religion	0.0035	0.0091	0.0154	0.0115	0.0506 ⁺	0.0172
Birth Year: 1944	-0.0528**	0.0092	-0.0758**	0.0126	-0.0218	0.0198
Birth Year: 1946	-0.0395**	0.0086	-0.0171	0.012	-0.0046	0.0187
Birth Year: 1947	-0.5038**	0.0091	-0.3846**	0.0126	-0.3508**	0.0196
Spring	0.0259 ⁺	0.0083	0.0388**	0.0111	0.0394	0.0172
Summer	0.0460**	0.0084	0.0641**	0.0114	0.0440	0.0178
Autumn	-0.0430**	0.0087	-0.0455**	0.0118	-0.0323	0.0177
Famine Exp.	0.0446 ⁺	0.0153	0.0198	0.0184	0.0287	0.0279
Cut Point 1	-2.8750**	0.0207	-3.0144**	0.0228	-3.6347**	0.0507
Cut Point 2	-1.8312**	0.0201	-2.2575**	0.0205	-2.9606**	0.0411
Cut Point 3	-1.0273**	0.0199	-1.5075**	0.0193	-2.1982**	0.0361
Cut Point 4	-0.1296**	0.0198	-0.6405**	0.0187	-1.3183**	0.0336
Cut Point 5	1.1568**	0.0201	0.6394**	0.0187	0.0591	0.0324

⁺ $p < 0.05$ and ** $p < 0.01$

To identify latent intelligence I , we estimate the educational attainment model and the measurement system jointly. We estimate the variance of the normal distribution $N(0, \sigma^2)$ of I . The estimated variance is $\hat{\sigma}^2 = 0.7652$ with standard error 0.0009.

Table 2.A.8. Ordered probit estimates for IQ measurement system, by education level, in structural model: Arithmetic Test

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	2.0094**	0.0172	1.6023**	0.0207	1.2326**	0.0292
Height	0.0338**	0.0008	0.0250**	0.0009	0.0152**	0.0012
Birth Order	-0.0486**	0.0032	-0.0541**	0.0046	-0.0652**	0.0062
Family Size	-0.0577**	0.0027	0.0212**	0.0036	0.0285**	0.0049
Professional	-0.0073	0.0194	0.0414	0.0180	-0.0634**	0.0183
Farm Owner	-0.4529**	0.0196	0.3153**	0.0277	0.1385**	0.0380
Skilled	-0.6208**	0.0165	-0.1347**	0.0202	-0.0507	0.0288
Unskilled	-0.8591**	0.0186	-0.2999**	0.0244	-0.1509**	0.0378
Unknown	-0.6982**	0.0228	-0.1732**	0.0289	-0.2845**	0.0377
City	0.0141	0.0139	-0.1024**	0.0165	0.0043	0.0217
Urbanised Rural	-0.0966**	0.0141	-0.0555 ⁺	0.0180	-0.0407	0.0245
Other	-0.0798**	0.0142	-0.0428	0.0185	-0.1380**	0.0261
North	0.0745**	0.0147	-0.0028	0.0179	0.1472**	0.0255
South	0.2683**	0.0135	0.3437**	0.0174	0.0284	0.0224
East	-0.0005	0.0134	0.0738**	0.0177	0.0011	0.0242
Catholic	-0.0401**	0.0115	0.0425 ⁺	0.0143	-0.3123**	0.0199
Other or No religion	-0.2102**	0.0136	-0.0749**	0.0156	-0.0068	0.0210
Birth Year: 1944	0.0967**	0.0139	-0.0050	0.0173	-0.0082	0.0239
Birth Year: 1946	-0.1119**	0.0130	-0.0976**	0.0164	-0.0465	0.0226
Birth Year: 1947	-0.5113**	0.0139	-0.4240**	0.0177	-0.2184**	0.0241
Spring	-0.0209	0.0124	0.0179	0.0152	0.0010	0.0210
Summer	-0.0167	0.0126	0.0533**	0.0157	0.0235	0.0216
Autumn	-0.0464**	0.0131	-0.0594**	0.0163	-0.0308	0.0217
Famine Exposure	0.0253	0.0228	0.0474	0.0246	0.0526	0.0335
Cut Point 1	-4.0855**	0.0405	-5.0297**	-0.0552	-5.1710**	0.1138
Cut Point 2	-2.0925**	0.0364	-3.3975**	-0.0385	-4.0634**	0.0736
Cut Point 3	-0.7894**	0.0352	-2.0762**	-0.0310	-3.2626**	0.0602
Cut Point 4	0.5811**	0.0350	-0.4839**	-0.0261	-2.0546**	0.0489
Cut Point 5	2.5064**	0.0375	1.4682**	-0.0283	-0.2099**	0.0415

⁺ $p < 0.05$ and ** $p < 0.01$

To identify latent intelligence I , we estimate the educational attainment model and the measurement system jointly. We estimate the variance of the normal distribution $N(0, \sigma^2)$ of I . The estimated variance is $\hat{\sigma}^2 = 0.7652$ with standard error 0.0009.

Table 2.A.9. Ordered probit estimates for IQ measurement system, by education level, in structural model: Language Test

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	1.4201**	0.0087	0.9650**	0.0092	0.6296**	0.0111
Height	0.0281**	0.0006	0.0195**	0.0007	0.0084**	0.0009
Birth Order	-0.0667**	0.0025	-0.0463**	0.0034	-0.0645**	0.0049
Family Size	-0.0334**	0.0021	0.0101**	0.0027	0.0262**	0.0038
Professional	0.0483 ⁺	0.0149	-0.0357 ⁺	0.0136	-0.0816**	0.0139
Farm Owner	-0.3879**	0.0149	0.1134**	0.0211	-0.0172	0.0299
Skilled	-0.5160**	0.0125	-0.0674**	0.0141	-0.0444	0.0217
Unskilled	-0.6342**	0.0140	-0.0990**	0.0170	0.0209	0.0293
Unknown	-0.4798**	0.0178	-0.1231**	0.0218	-0.2363**	0.0286
City	0.0500**	0.0112	-0.0201	0.0128	0.0133	0.0170
Urbanised Rural	-0.0750**	0.0113	-0.0172	0.0140	-0.0695**	0.0193
Other	-0.0788**	0.0113	-0.0442 ⁺	0.0144	-0.0970**	0.0203
North	0.0457**	0.0115	-0.1108**	0.0139	-0.0615 ⁺	0.0196
South	0.1182**	0.0108	0.1523**	0.0134	0.0084	0.0174
East	-0.0734**	0.0106	0.0332	0.0137	-0.0111	0.0191
Catholic	-0.1119**	0.0091	-0.2012**	0.0110	-0.2305**	0.0152
Other or No religion	-0.1377**	0.0107	-0.1075**	0.0120	-0.0229	0.0164
Birth Year: 1944	-0.0484**	0.0111	-0.1135**	0.0134	-0.0956**	0.0185
Birth Year: 1946	-0.0758**	0.0104	-0.0765**	0.0127	-0.0465 ⁺	0.0177
Birth Year: 1947	-0.6106**	0.0110	-0.4157**	0.0136	-0.3043**	0.0187
Spring	-0.0331**	0.0099	0.0109	0.0118	0.0035	0.0163
Summer	-0.0198	0.0100	0.0440**	0.0121	0.0470 ⁺	0.0169
Autumn	-0.1387**	0.0104	-0.1046**	0.0126	-0.0664**	0.0168
Famine Exposure	-0.0160	0.0182	-0.0105	0.0191	0.0576	0.0260
Cut Point 1	-3.7284**	0.0286	-4.3022**	0.0387	-4.2774**	0.0857
Cut Point 2	-2.0512**	0.0266	-3.2788**	0.0259	-3.6243**	0.0510
Cut Point 3	-0.7942**	0.0260	-2.2866**	0.0222	-2.9929**	0.0382
Cut Point 4	0.6234**	0.0260	-1.0634**	0.0204	-1.8897**	0.0315
Cut Point 5	2.3057**	0.0276	0.7052**	0.0201	-0.1924**	0.0296

⁺ $p < 0.05$ and ^{**} $p < 0.01$

To identify latent intelligence I , we estimate the educational attainment model and the measurement system jointly. We estimate the variance of the normal distribution $N(0, \sigma^2)$ of I . The estimated variance is $\hat{\sigma}^2 = 0.7652$ with standard error 0.0009.

Table 2.A.10. Gompertz proportional mortality rate estimates, by education level, in structural model

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	-0.1974**	0.0313	-0.1317**	0.0412	-0.1040	0.0775
General Health: Unfit	0.1107	0.0631	0.2649**	0.0753	0.0847	0.1287
Psych. Health: Unfit	0.0785	0.0558	0.2262**	0.0734	-0.0790	0.1311
Overweight	0.1935 ⁺	0.0807	0.3257**	0.1094	0.2129	0.1942
Height	-0.0092 ⁺	0.0037	-0.0037	0.0048	0.0011	0.0078
Birth Order	-0.0209	0.0172	0.0340	0.0261	0.0386	0.0436
Family Size	0.0021	0.0182	-0.0307	0.0256	-0.0734	0.0421
Professional	-0.0759	0.0906	-0.0832	0.0905	0.1668	0.1106
Farm Owner	-0.0889	0.1263	0.1050	0.2147	0.5121	0.3436
Skilled	0.0132	0.0610	0.0730	0.0770	0.2914	0.1629
Unskilled	0.0858	0.0716	0.1014	0.1059	-0.0224	0.2895
Unknown	0.2694**	0.0937	0.1864	0.1339	0.3333	0.2374
City	0.0866	0.1201	0.1111	0.1834	0.2242	0.2895
Urbanised Rural	0.0381	0.1351	0.0339	0.2148	0.0157	0.3595
Other	0.1601	0.1313	0.0948	0.2189	0.3900	0.3600
North	0.2393	0.1326	-0.1341	0.2209	-0.3611	0.4113
South	-0.1493	0.1180	-0.0550	0.1741	0.0876	0.2520
East	0.1939	0.1203	0.1628	0.1934	-0.1167	0.3306
Catholic	0.0477	0.0593	-0.0196	0.0778	0.2824 ⁺	0.1247
Other or No religion	0.1019	0.0592	0.0222	0.0741	0.1893	0.1223
Birth Year: 1944	0.0187	0.0733	0.1179	0.0970	0.0744	0.1594
Birth Year: 1946	-0.1041	0.0679	-0.2082 ⁺	0.0889	0.0964	0.1388
Birth Year: 1947	-0.0157	0.0968	-0.0652	0.1347	-0.1882	0.2197
Spring	-0.0032	0.0656	0.0081	0.0843	0.1118	0.1371
Summer	-0.0051	0.0688	-0.1062	0.0935	-0.1242	0.1543
Autumn	0.0634	0.0669	-0.0674	0.0884	-0.0812	0.1406
Famine Exposure	-0.0338	0.0754	-0.0159	0.0991	-0.1115	0.1542
Constant	-10.1426**	0.2019	-10.1978**	0.2765	-10.3806**	0.4496
Shape	0.0881**	0.0025	0.0889**	0.0033	0.0860**	0.0052

⁺ $p < 0.05$ and ** $p < 0.01$

Table 2.A.11. Probit estimates for cardiac diseases medications, by education level, in structural model

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	-0.0854**	0.0180	-0.0371	0.0256	-0.0715	0.0474
General Health: Unfit	0.0340	0.0390	0.0249	0.0501	0.1360	0.0766
Psych. Health: Unfit	0.0894**	0.0329	0.0809	0.0475	0.0740	0.0762
Overweight	0.1223 ⁺	0.0504	0.3770**	0.0665	0.2350 ⁺	0.1157
Height	-0.0045 ⁺	0.0022	-0.0037	0.0030	-0.0027	0.0049
Birth Order	0.0110	0.0100	0.0507**	0.0163	-0.0066	0.0276
Family Size	0.0103	0.0110	-0.0488**	0.0160	-0.0406	0.0257
Professional	0.0174	0.0524	0.0039	0.0540	0.1168	0.0681
Farm Owner	-0.1572	0.0805	-0.1327	0.1502	0.6444**	0.2049
Skilled	0.0540	0.0361	0.0389	0.0478	0.1712	0.1048
Unskilled	0.0490	0.0433	0.0575	0.0661	0.3413 ⁺	0.1451
Unknown	0.0659	0.0610	0.0559	0.0875	0.1959	0.1527
City	-0.0721	0.0694	0.0678	0.1105	0.2042	0.1880
Urbanised Rural	-0.0914	0.0790	0.0607	0.1311	0.2395	0.2152
Other	-0.1173	0.0798	0.0955	0.1361	0.0637	0.2460
North	0.0080	0.0879	-0.1365	0.1345	-0.3427	0.2716
South	-0.0078	0.0679	-0.0180	0.1029	0.0223	0.1607
East	-0.0045	0.0766	-0.1260	0.1278	-0.0914	0.2063
Catholic	0.0254	0.0354	0.0141	0.0477	-0.0467	0.0752
Other or No religion	0.0592	0.0356	-0.0173	0.0462	-0.1554 ⁺	0.0752
Birth Year: 1944	-0.0075	0.0458	0.1128	0.0620	-0.0309	0.0956
Birth Year: 1946	-0.0669	0.0400	-0.0400	0.0536	-0.1741 ⁺	0.0875
Birth Year: 1947	-0.1632**	0.0587	-0.0429	0.0811	-0.2587 ⁺	0.1321
Spring	0.0230	0.0394	-0.0628	0.0538	-0.0046	0.0868
Summer	0.0045	0.0414	-0.0446	0.0569	-0.0960	0.0943
Autumn	-0.0363	0.0412	-0.0958	0.0550	0.0664	0.0833
Famine Exposure	0.0245	0.0456	-0.0290	0.0609	-0.0648	0.0971
Constant	-1.4256**	0.0934	-1.4589**	0.1355	-1.6659**	0.2308

⁺ $p < 0.05$ and ** $p < 0.01$

Table 2.A.12. Probit estimates for obstructive airway diseases medications, by education level, in structural model

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	-0.1276**	0.0178	-0.0720**	0.0251	-0.0593	0.0433
General Health: Unfit	0.2355**	0.0359	0.3039**	0.0459	0.4171**	0.0651
Psych. Health: Unfit	0.0608	0.0322	0.0527	0.0470	0.0417	0.0696
Overweight	0.0548	0.0503	-0.0578	0.0798	0.0859	0.1136
Height	-0.0040	0.0022	-0.0008	0.0029	-0.0127**	0.0045
Birth Order	0.0124	0.0099	-0.0069	0.0163	0.0070	0.0257
Family Size	0.0030	0.0108	0.0148	0.0153	-0.0364	0.0238
Professional	0.1147 ⁺	0.0498	0.0232	0.0537	-0.0327	0.0606
Farm Owner	-0.0063	0.0736	-0.1312	0.1458	0.1214	0.2200
Skilled	0.0062	0.0358	0.0683	0.0475	-0.2192 ⁺	0.1087
Unskilled	0.0898 ⁺	0.0423	0.1105	0.0647	-0.2605	0.1735
Unknown	0.1450 ⁺	0.0584	0.0870	0.0873	-0.0191	0.1468
City	0.0933	0.0697	0.0301	0.1093	0.1042	0.1653
Urbanised Rural	0.0162	0.0782	-0.0631	0.1334	0.2714	0.1903
Other	-0.0410	0.0791	0.1673	0.1305	0.0515	0.2207
North	0.0262	0.0850	-0.0090	0.1234	-0.2286	0.2420
South	-0.0039	0.0670	-0.1809	0.1082	0.1687	0.1457
East	0.1005	0.0728	-0.1788	0.1270	0.0001	0.1837
Catholic	-0.0677 ⁺	0.0343	-0.0353	0.0474	-0.0114	0.0712
Other or No religion	-0.0641	0.0349	-0.0220	0.0459	0.0386	0.0675
Birth Year: 1944	-0.0473	0.0457	0.0278	0.0618	0.0074	0.0915
Birth Year: 1946	-0.0436	0.0391	-0.0616	0.0538	-0.0753	0.0787
Birth Year: 1947	-0.1603**	0.0571	-0.0970	0.0799	-0.0739	0.1181
Spring	0.0526	0.0383	0.0306	0.0535	-0.0591	0.0800
Summer	0.0086	0.0404	0.0320	0.0567	-0.0289	0.0831
Autumn	-0.0896 ⁺	0.0408	-0.0249	0.0549	-0.1192	0.0799
Famine Exposure	-0.0335	0.0446	-0.0601	0.0605	0.0660	0.0899
Constant	-1.4401**	0.0935	-1.5812**	0.1346	-1.5359**	0.2087

⁺ $p < 0.05$ and ** $p < 0.01$

Table 2.A.13. Probit estimates for hypertension medications, by education level, in structural model

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	-0.0317 ⁺	0.0130	-0.0277	0.0174	-0.0071	0.0302
General Health: Unfit	-0.0039	0.0287	0.0305	0.0348	0.0299	0.0514
Psych. Health: Unfit	0.0061	0.0246	0.0210	0.0333	0.0567	0.0500
Overweight	0.2703**	0.0377	0.4471**	0.0513	0.4515**	0.0784
Height	-0.0084**	0.0016	-0.0084**	0.0020	-0.0083**	0.0031
Birth Order	-0.0041	0.0075	-0.0111	0.0112	0.0089	0.0173
Family Size	-0.0168 ⁺	0.0080	-0.0287**	0.0107	-0.0605**	0.0163
Professional	-0.0354	0.0378	0.0034	0.0367	0.0228	0.0425
Farm Owner	-0.2061**	0.0545	-0.1853	0.0989	-0.0166	0.1606
Skilled	-0.0123	0.0260	0.0369	0.0327	0.1736**	0.0666
Unskilled	-0.0285	0.0314	0.0291	0.0457	-0.0491	0.1075
Unknown	0.0498	0.0446	0.1601**	0.0601	0.0683	0.1017
City	0.0750	0.0512	0.0309	0.0732	0.0947	0.1091
Urbanised Rural	0.0610	0.0577	0.0016	0.0867	0.1632	0.1301
Other	0.0848	0.0576	-0.0571	0.0902	0.0140	0.1454
North	-0.0269	0.0620	0.1066	0.0866	0.1112	0.1412
South	0.0444	0.0484	0.0781	0.0701	0.0155	0.1042
East	0.0048	0.0543	0.0229	0.0835	-0.0340	0.1244
Catholic	-0.0381	0.0255	0.0229	0.0325	0.0841	0.0482
Other or No religion	-0.0294	0.0258	-0.0072	0.0316	-0.0508	0.0475
Birth Year: 1944	-0.0068	0.0332	0.0497	0.0425	-0.0598	0.0629
Birth Year: 1946	-0.1528**	0.0290	-0.1612**	0.0367	-0.1270 ⁺	0.0550
Birth Year: 1947	-0.1588**	0.0413	-0.1092 ⁺	0.0549	-0.2224**	0.0817
Spring	0.0226	0.0286	0.0112	0.0367	0.0319	0.0554
Summer	-0.0397	0.0301	0.0091	0.0390	-0.0144	0.0587
Autumn	-0.0713 ⁺	0.0296	-0.0798 ⁺	0.0376	0.0409	0.0543
Famine Exposure	-0.0205	0.0329	0.0474	0.0420	-0.0098	0.0619
Constant	-0.3469**	0.0682	-0.4591**	0.0913	-0.5763**	0.1400

⁺ $p < 0.05$ and ** $p < 0.01$

Table 2.A.14. Probit estimates for hyperlipidemia medications, by education level, in structural model

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	-0.0492**	0.0134	-0.0407 ⁺	0.0181	-0.0480	0.0321
General Health: Unfit	0.0034	0.0294	0.0262	0.0362	-0.0401	0.0554
Psych. Health: Unfit	-0.0242	0.0252	-0.0552	0.0350	-0.0459	0.0541
Overweight	0.1189**	0.0389	0.2940**	0.0529	0.2194**	0.0843
Height	-0.0129**	0.0017	-0.0190**	0.0021	-0.0184**	0.0033
Birth Order	0.0050	0.0076	-0.0002	0.0116	0.0337	0.0181
Family Size	0.0045	0.0082	-0.0108	0.0111	-0.0162	0.0173
Professional	-0.0352	0.0391	-0.0160	0.0380	-0.0091	0.0452
Farm Owner	-0.2069**	0.0568	-0.2992**	0.1069	-0.0339	0.1718
Skilled	0.0452	0.0267	-0.0267	0.0340	0.0530	0.0717
Unskilled	0.0678 ⁺	0.0320	-0.0085	0.0475	-0.0351	0.1134
Unknown	0.0907 ⁺	0.0457	0.0845	0.0625	0.1315	0.1060
City	-0.0071	0.0519	0.0156	0.0768	0.2034	0.1176
Urbanised Rural	0.0027	0.0584	0.0608	0.0901	0.2086	0.1384
Other	-0.0527	0.0589	0.0419	0.0943	-0.0501	0.1581
North	0.0873	0.0630	-0.0600	0.0928	0.0153	0.1584
South	0.0559	0.0493	0.0481	0.0730	0.1197	0.1102
East	-0.0335	0.0561	0.0023	0.0874	0.2408	0.1282
Catholic	-0.0048	0.0262	0.0284	0.0339	-0.0080	0.0510
Other or No religion	0.0236	0.0265	0.0667 ⁺	0.0328	-0.1070 ⁺	0.0509
Birth Year: 1944	0.0187	0.0340	0.0268	0.0441	-0.0078	0.0664
Birth Year: 1946	-0.0768**	0.0297	-0.0814 ⁺	0.0381	-0.1008	0.0578
Birth Year: 1947	-0.1110**	0.0425	-0.0960	0.0574	-0.2906**	0.0884
Spring	0.0193	0.0294	0.0265	0.0383	-0.0664	0.0590
Summer	0.0162	0.0307	0.0120	0.0407	-0.0128	0.0617
Autumn	-0.0599 ⁺	0.0304	-0.0186	0.0389	-0.0944	0.0580
Famine Exposure	0.0175	0.0337	0.0516	0.0437	-0.0002	0.0657
Constant	-0.6105**	0.0697	-0.6991**	0.0954	-0.9000**	0.1502

⁺ $p < 0.05$ and ** $p < 0.01$

Table 2.A.15. Probit estimates for diabetes (type 1 and 2) medications, by education level, in structural model

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	-0.0728**	0.0171	-0.0383	0.0232	-0.0347	0.0461
General Health: Unfit	-0.0206	0.0374	0.0957 ⁺	0.0451	0.0727	0.0755
Psych. Health: Unfit	0.0313	0.0316	0.0171	0.0443	0.1248	0.0730
Overweight	0.4187**	0.0437	0.5138**	0.0595	0.4825**	0.1017
Height	-0.0061**	0.0021	-0.0127**	0.0027	-0.0071	0.0047
Birth Order	-0.0298**	0.0100	-0.0167	0.0152	-0.0066	0.0274
Family Size	0.0083	0.0103	-0.0110	0.0143	-0.0400	0.0251
Professional	0.0127	0.0486	0.0017	0.0494	0.0340	0.0656
Farm Owner	-0.1326	0.0755	-0.1666	0.1476	0.1812	0.2492
Skilled	-0.0105	0.0339	0.0391	0.0435	0.1145	0.1012
Unskilled	0.0284	0.0404	0.0724	0.0598	0.1743	0.1496
Unknown	0.0064	0.0581	0.0346	0.0816	0.2667	0.1397
City	-0.0615	0.0652	-0.1240	0.0965	-0.0075	0.1667
Urbanised Rural	-0.1465	0.0757	-0.0744	0.1133	-0.4859 ⁺	0.2452
Other	-0.2154**	0.0770	-0.3547**	0.1283	-0.1194	0.2233
North	0.1010	0.0801	0.1392	0.1181	0.1337	0.2181
South	-0.0895	0.0662	-0.0152	0.0974	-0.0640	0.1732
East	-0.0414	0.0728	0.0919	0.1121	0.1020	0.1948
Catholic	-0.0207	0.0335	0.0336	0.0441	-0.0133	0.0742
Other or No religion	0.0343	0.0332	0.0870 ⁺	0.0418	-0.0305	0.0711
Birth Year: 1944	-0.0269	0.0429	0.0112	0.0573	-0.0719	0.0969
Birth Year: 1946	-0.1697**	0.0381	-0.0772	0.0492	-0.0969	0.0841
Birth Year: 1947	-0.2204**	0.0551	-0.0963	0.0750	-0.3045 ⁺	0.1333
Spring	-0.0100	0.0371	-0.0178	0.0495	0.0235	0.0844
Summer	-0.0217	0.0389	0.0415	0.0518	0.0073	0.0893
Autumn	-0.0265	0.0383	-0.0367	0.0506	-0.0440	0.0842
Famine Exposure	-0.0599	0.0428	0.0777	0.0568	-0.0170	0.0956
Constant	-1.0832**	0.0871	-1.2808**	0.1213	-1.4567**	0.2135

⁺ $p < 0.05$ and ** $p < 0.01$

Table 2.A.16. Probit estimates for depression and anxiety medications, by education level, in structural model

	Primary Educ.		Low Educ.		High Educ.	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Intelligence	-0.0730**	0.0162	-0.1007**	0.0217	-0.1216**	0.0372
General Health: Unfit	0.0905**	0.0345	0.1201**	0.0418	0.1108	0.0613
Psych. Health: Unfit	0.1343**	0.0294	0.1553**	0.0397	0.0990	0.0601
Overweight	-0.0991 ⁺	0.0496	-0.0277	0.0673	0.0388	0.1015
Height	-0.0040 ⁺	0.0020	-0.0070**	0.0026	0.0014	0.0038
Birth Order	0.0154	0.0091	0.0085	0.0138	0.0495 ⁺	0.0213
Family Size	-0.0107	0.0099	0.0087	0.0133	-0.0398	0.0203
Professional	-0.0099	0.0472	0.0388	0.0448	0.0375	0.0519
Farm Owner	-0.0661	0.0680	-0.1421	0.1249	-0.2469	0.2230
Skilled	0.0439	0.0322	-0.0456	0.0412	-0.0451	0.0862
Unskilled	-0.0048	0.0393	-0.0223	0.0574	0.1217	0.1256
Unknown	0.0889	0.0543	-0.0870	0.0789	-0.2773	0.1459
City	-0.0063	0.0619	-0.0250	0.0913	-0.1855	0.1286
Urbanised Rural	-0.0950	0.0710	-0.1485	0.1126	-0.0733	0.1558
Other	-0.0789	0.0712	0.0369	0.1121	-0.1171	0.1768
North	-0.0398	0.0788	-0.1274	0.1112	-0.4451 ⁺	0.2117
South	0.0186	0.0593	-0.0684	0.0900	0.0296	0.1269
East	-0.0509	0.0682	-0.1525	0.1079	-0.1023	0.1505
Catholic	0.0010	0.0313	-0.0746	0.0411	-0.0341	0.0605
Other or No religion	-0.0767 ⁺	0.0323	0.0430	0.0390	0.0581	0.0578
Birth Year: 1944	-0.0114	0.0416	-0.0689	0.0546	-0.0444	0.0803
Birth Year: 1946	0.0131	0.0361	-0.0387	0.0453	-0.0364	0.0671
Birth Year: 1947	-0.0295	0.0512	-0.0364	0.0682	0.0198	0.0982
Spring	0.0667	0.0356	-0.0519	0.0455	-0.0653	0.0678
Summer	0.0458	0.0373	-0.0860	0.0488	-0.1029	0.0719
Autumn	-0.0115	0.0370	-0.1320**	0.0472	-0.1655 ⁺	0.0682
Famine Exposure	-0.0182	0.0407	-0.0545	0.0523	-0.0156	0.0769
Constant	-1.1727**	0.0842	-1.1123**	0.1140	-0.9440**	0.1680

⁺ $p < 0.05$ and ** $p < 0.01$

Additional figures

Figure 2.A.1. ATEs primary to low education on medications by intelligence from 9-equation and 3-equation measurement system, age 59-62

Figure 2.A.2. ATEs primary to high education on medications by intelligence from 9-equation and 3-equation measurement system, age 59-62

2.B Appendix

Computation of the standard errors of the Average Treatment Effects

We compute the standard errors of the $ATE_{1sm}(I)$ for each medication m , where $s = 2, 3$ and $I \in [-2; 2]$. Define the estimated (28×28) covariance matrix V_{em} of the estimated (28×1) vector of parameters $\mathbf{b}_{em} = \begin{pmatrix} \mathbf{q}_{em}, & \eta_{em} \end{pmatrix}'$ of the probit model (2.3) for prescribed medication m and education level $e = 1, 2, 3$. We compute the Cholesky decomposition C_{em} of the matrix V_{em} , for each e and m . We generate a random (28×1) vector $\mathbf{s}_{em}$ of normally distributed components with mean 0 and variance 1, for each e and m . We compute the (28×1) vectors $\mathbf{g}_{em} = C_{em}\mathbf{s}_{em}$ and $\mathbf{m}_{em} = \mathbf{b}_{em} + \mathbf{g}_{em}$, for each e and m . Using the vectors $\mathbf{m}_{em}$ instead of the vectors $\mathbf{b}_{em}$ we compute the first $\widetilde{ATE}_{1sm}^{(1)}(I)$ for $I \in [-2; 2]$, and for $s = 2, 3$ and for each m . We repeat this procedure 100 times and we obtain an empirical distribution of the $\widetilde{ATE}_{1sm}^{(p)}(I)$, $p = 1, \dots, 100$, for $I \in [-2; 2]$, and for $s = 2, 3$ and for each m . Finally, we compute the standard deviations of these $\widetilde{ATE}_{1sm}^{(p)}(I)$ for $I \in [-2; 2]$, which correspond to the standard errors of the $ATE_{1sm}(I)$. Based on these standard errors we construct the confidence intervals of the $ATE_{1sm}(I)$ for $I \in [-2; 2]$. The procedure to compute the standard errors of the unconditional ATE_{1sm} is analogous.

Chapter 3

Early-life economic determinants of type-2 diabetes

3.1 Introduction

It is well-established that adverse conditions *in utero* and shortly after birth negatively affect individual health later in life (Barker 1995, 1998). Fetuses and newborns are sensitive to stressful environments and are “programmed” to deal with survival in similar environments after birth. A mismatch between the early-life environment and the environment later in life may then negatively impact health outcomes over the life-cycle.

Early-life conditions can be measured by indicators of the parental living circumstances at birth. However, any association between such indicators and health in adulthood might not imply a causal effect as unobserved factors may affect both. For example, parental genetic predisposition to some diseases may decrease the parental socioeconomic status at birth and worsen the health outcomes in adulthood. To overcome this endogeneity issue, a growing literature exploits natural experiments, using exogenous variation in contextual (macro) circumstances at birth that are randomly assigned to individuals as proxies for economic conditions early in life. Landmark examples of this approach exploit exposure to famines or recessions (see the overview by Almond and Currie (2011)). In the famine-based literature, the late-life outcome that seems most responsive to early-life malnutri-

tion is type-2 diabetes. The comprehensive overview by Lumey et al. (2011) shows that this adverse effect of early-life malnutrition has been found across a range of different famines. These findings are confirmed in animal experiments (Fernandez-Twinn and Ozanne, 2006, Portha et al., 2011) as well as in other natural experiments such as hurricanes (Sotomayor, 2013) (see also Calkins and Devaskar, 2011, for another review).

Famines are catastrophic events, typically dating back a long time and having occurred in non-advanced economies. In this paper we expand on the existing evidence by exploiting milder fluctuations in early-life conditions in more recent eras and in more advanced economies, making the results that we obtain more interesting from a policy perspective. The reason is that such milder fluctuations shape on a daily basis any modern Western society. The prevalence of type-2 diabetes has been increasing over time and knowledge on determinants of type-2 diabetes is of high interest for the policy maker, as the economic burden of this disease is high and is expected to further increase in the next decades worldwide. For instance, in the Netherlands the total economic burden of type-2 diabetes was 6.8 billion euro in 2016 (Peters et al., 2017).

We exploit spatial and temporal variation in unemployment rates at birth as contextual variation in early-life economic conditions. Effectively, we compare type-2 diabetes outcomes of those born in recessions to those born in adjacent boom years, in the Netherlands in post-1950 birth cohorts. The paper makes a number of innovative methodological contributions. The first of these concerns the observability and measurement of type-2 diabetes. Type-2 diabetes is known to be severely underdiagnosed in the population. Many patients may be unaware of having this condition, leading to underreporting in survey data. Those with lower socioeconomic status may consult doctors less systematically, and the disease may be asymptomatic in younger and prime-aged individuals. For example, Whicher et al. (2020) report that in the UK in 2019 six out of ten individuals did not have any symptoms at the time of the diagnosis and that nearly 1 million people had undiagnosed type-2 diabetes. We solve this by identifying type-2 diabetes through biomarkers. In accordance with the diagnosis criteria of the American Diabetes Association (2020), we consider the individuals to be affected by type-2 diabetes if the hemoglobin A1c (HbA1c) concentration in the blood is equal to or higher than 6.5%

or if the fasting glucose level is equal to or higher than 7 mmol/L. We check the robustness of the results with respect to deviations in the threshold levels. We also examine to which extent the outcome is related to self-reported type-2 diabetes, in the same sample of individuals. More importantly, we compare the empirical findings to those if we only rely on self-reported diabetes.

The recent literature has found some evidence of selective fertility depending on the unemployment rate at birth (see van den Berg et al., 2020, for an overview). For example, Dehejia and Lleras-Muney (2004) find that the socioeconomic status affects how fertility changes when the unemployment rate is high. To account for selective fertility across families we adopt a sibling study design, effectively comparing outcomes of siblings born at different stages of the business cycle. This approach follows other studies, such as van den Berg et al. (2020) and De Cao et al. (2019) for outcomes in infancy. Note that the sibling design also controls for potential associations between parental genetic predisposition to type-2 diabetes and fertility. It has been estimated that the heritability of type-2 diabetes is of at least 20% (Ali, 2013). Finally, the sibling design also accounts for common family characteristics, such as the family background and parental education.

We use the Lifelines data which is a cohort study of over 160,000 individuals, consisting of longitudinal surveys and a biobank based on a multigenerational sampling design. Most individuals reside in the Northern part of the Netherlands. The data include over 51,000 siblings born between 1950 and 1995. Sibling-sibling relationships are self-reported or identified through questionnaire information such as coded surname, number of siblings and year of birth of the siblings. These relationships were all checked and, if necessary, corrected by using linked administrative register data. Interestingly, a sub-sample of the Lifelines data, called Lifelines-UGLI, contains genetic records that enable us to identify sibling-sibling relationships in a completely different way, namely by using genetic similarity. Exploiting both types of sibling identifier opens up exciting possibilities for the study of health determinants. As it happens, the sub-sample of biological siblings identified uniquely through genetic records is currently too small for quantitative inference, beyond a simple categorization of differences in terms of covariates. The lack of sample size also prevents separate analysis with non-biological sibling-sibling relationships due to adoption.

The outline of the paper is as follows. In Section 3.2 we discuss the data. In Section 3.3 we discuss the sibling fixed-effects econometric model that we estimate; in Section 3.4 we present and discuss the main results as well as several sensitivity analysis; in Section 3.5 we discuss some policy suggestions and conclude the paper.

3.2 Data

3.2.1 The Lifelines data

Lifelines is a multi-disciplinary prospective population-based cohort study examining health and health-related behaviors in a unique three-generation design. Recruitment took place in the three provinces in the Northern Netherlands (Groningen, Friesland and Drenthe). The sample consists of more than 167,000 participants. The Lifelines study collects biomedical, socioeconomic, demographic, physiological and psychological information. The participants, usually recruited by the general practitioner, were often asked to invite their family members. Consequently the number of siblings in the cohort study is relatively high. Initially, siblings could only be identified if at least one of their parents also participated in Lifelines. Recently, however, through the so-called Pedigree file, all siblings could be identified. This file links individuals to a unique identifier for both the mother and the father, regardless the presence of the mother and the father in the database. It was constructed mainly based on parent-child and sibling-sibling relationships declared by Lifelines participants. However, other sources were used to complete the file, as siblings who participated in the study without their parents may have not declared their blood-relationship. Additional information used to identify further sibling-sibling relationships were found in the Lifelines questionnaire, including questions on number of siblings, year of birth of the siblings and coded surnames. All the sibling-sibling relationships were checked and corrected through administrative data.

Having a unique identifier for both the mother and the father enables us to define as siblings those who have the same mother and father. This is an advantage compared to data that only allow them to be defined by a common mother. Lifelines-UGLI, a sub-sample of the original Lifelines data for which genetic records are also available, was used to identify genetically defined sibling-sibling

Table 3.1. Descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
<i>Full sample</i>					
T2D in females (%)	25,895	1.27	11.18	0	100
T2D in males (%)	17,236	1.76	13.16	0	100
Unemployment rate (%)	43,131	5.22	3.17	0.5	13.8
Age	43,131	41.2	10.5	18	63
First-born	43,131	0.4	0.5	0	1
Low education	42,394	0.61	0.49	0	1
<i>Genetically-confirmed biological siblings</i>					
T2D in females (%)	9,795	1.03	10.10	0	100
T2D in males (%)	6,580	1.49	12.11	0	100
Unemployment rate (%)	16,375	5.77	3.38	0.5	13.8
Age	16,375	39.26	11.26	18	63
First-born	16,375	0.41	0.49	0	1
Low education	16,117	0.57	0.49	0	1
<i>Biological siblings identified only via genetic records</i>					
T2D in females (%)	35	5.71	23.55	0	100
T2D in males (%)	33	6.06	24.23	0	100
Unemployment rate (%)	68	5.36	3.12	1	13.8
Age	68	44.22	11.18	20	61
First-born	68	0.35	0.48	0	1
Low education	67	0.71	0.46	0	1
<i>Non-biological siblings identified via genetic records</i>					
T2D in females (%)	12	0	0	0	0
T2D in males (%)	0	0	0	0	0
Unemployment rate (%)	12	7.00	4.08	2.5	13.2
Age	12	32.5	8.34	19	44
First-born	12	0	0	0	0
Low education	12	0.50	0.52	0	1

T2D is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Low education is not controlled for in the analysis and it is not observed for all the individuals in our sample but it is included for completeness. Genetically-confirmed biological siblings include those siblings who were biologically related to their mother and father in the data. Biological siblings identified only via genetic records is the subsample of the genetically-confirmed biological siblings whose biological relationship to their father and mother was not reported nor identified through questionnaire information. Non-biological siblings identified via genetic records include those siblings who were adopted by the parents and so were excluded by the main analysis.

relationships as well as non-biological (due to adoption) sibling-sibling relation-

ships.¹ The genetic records allow us to confirm sibling relationships for all 19,005 siblings for which genetic data is available. They uncovered 110 additional siblings and 13 were identified as non-biological adopted siblings, which were dropped. The total number of siblings aged 18 or above is 51,270.

We use data provided by Van Duijn (1975) for the provincial unemployment rates between 1950-1972, which was obtained by the author from CBS (Statistics Netherlands), while unemployment data from 1973 onward is obtained directly from CBS.

Some siblings could not be used as they could not be linked to provincial unemployment data. These include those who were born abroad (376) or in homonymous Dutch cities that belong to different provinces (387). Moreover, because provincial unemployment data is not available from before 1950 we drop individuals born earlier than that (3,484). Flevoland, which became an independent province in 1986, only has unemployment data since 1973, leading to the exclusion of 199 individuals. We further exclude from the data 1,107 individuals for whom neither the HbA1c concentration in the blood nor the fasting glucose level in the blood used for the diagnosis of type-2 diabetes was collected. Finally, we exclude those individuals who declare to have type-1 diabetes (88), other forms of diabetes (109) or who do not know which type of diabetes they have (47). Obviously, excluding individuals implies that for some families only one sibling is still present in the data, so we finally exclude these siblings as fixed-effects cannot be computed in such cases (2,342).

The final number of siblings used for our study is 43,131. Of these individuals 25,895 are females and 17,236 are males. After the exclusions the number of siblings for which the genetic records are available is 16,375, of whom 68 are the biological siblings linked to their parents only through the genetic data and 12 are the adopted siblings excluded from the data. We report the distribution of the 43,131 siblings by province of birth in Table 3.A.1 in Appendix 3.A. As expected, the vast majority of siblings were born in the three provinces in the Northern Netherlands where Lifelines recruitment took place (Groningen, Friesland and Drenthe).

According to the standard diagnosis criteria defined by the American Diabetes

¹ See Lopera Maya et al. (2020) for a description of the genotyping data. Unfortunately, genetic records for all Lifelines participants will only become available in 2022.

Association (2020), we define the binary dependent variable *T2D* to be equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level in the blood is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Both the HbA1c concentration and the fasting glucose level in the blood are commonly used for the diagnosis of type-2 diabetes. The HbA1c concentration in the blood is a measure of *chronic* glucose exposure as it reflects the average glucose concentration in the previous 3 months and is relatively stable over time. Consequently, measuring the HbA1c concentration in the blood does not require the individuals to fast before the test, while the actual glucose level in the blood varies with food digestion and so requires the individual to fast for 8 hours beforehand. In our sample, the correlation between HbA1c and the fasting glucose level is 0.60, while the correlation between the binary versions based on their diabetes threshold values is 0.71.

Descriptive statistics are reported in Table 3.1. The prevalence of type-2 diabetes in our sample is lower than in the Dutch population, where for 2013, this is estimated to be 5.13% (Kleefstra et al., 2016). This is explained by the age distribution in our sample which does not include individuals over 63.

The binary variable *First-born* is defined to equal 1 for the first-born sibling and to equal 0 otherwise as well as for twins.² Age is measured in years at baseline. The provincial unemployment rate is measured in percentages.

Table 3.1 also reports descriptive statistics of all the siblings who have been checked through genetic records, of those who were found to have a non-biological sibling relationship (i.e., adopted siblings) and of those for whom the sibling-sibling relationship was found solely through the genetic records. Interestingly, biological siblings identified only via genetic records are older, have lower socio-economic status (i.e., lower education level) and much higher *T2D* prevalence than the rest of the full sample (where the latter may be due to the former). Although this group is quantitatively negligible in our setting, there is a wider relevance in the fact that certain types of siblings are overrepresented in it. Sibling study designs using data with self-reported sibling status may be selective and hence findings in them may not carry over to the wider population of siblings.

²Twins are only included if they have further siblings, otherwise they are excluded due to lack of between-sibling variation in early-life conditions.

3.2.2 Selective fertility by socioeconomic status

Figure 3.1. Selective fertility by socioeconomic status over unemployment

Any relationship between early-life socioeconomic conditions and health outcomes later in life may be driven by selective fertility whereby certain groups are more likely or less likely to have children during economic downturns. We use the education level of the parents in a first attempt to investigate selective fertility over different unemployment levels. The education level of the parents is not reported by the siblings, so we have this data only for the parents who participated in the study. Specifically, we have information on the completed education level of the mothers of 13,685 siblings and the fathers of 9,949 siblings. For 8,102 siblings we have information on the completed education level of both the mother and the father. We define as high education a level of senior general secondary education³ or higher and as low education all the other levels.

It turns out that 2,595 siblings have mothers with high education, 11,090 siblings have mothers with low education, 2,757 siblings have fathers with high education, and 7,192 siblings have fathers with low education. Finally, among 1,068 siblings both parents have high education, among 5,134 siblings both have low education, and among 1,900 siblings one parent has a high education and one a low education.

Figure 3.1 depicts the kernel estimates of the probability density functions of unemployment rates at birth, by education level of the parents (with bandwidths

³Senior general secondary education includes pre-university education such as HAVO, VWO, Atheneum, Gymnasium, HBS and MMS.

of 0.8 percentage points throughout). Specifically, in Figure 3.1a we report the birth distribution for both parents with low education, both parents with high education and one parent with high education and one parent with low education. It can be seen that as the socioeconomic status (proxied by the education level) decreases individuals tend to become parents during favourable unemployment phases. We produce the same graph based on education of the mother only (Figure 3.1b) and we observe a very similar pattern. Although we only observe individuals who become parents and so we are unable to compute the fertility rates by socioeconomic status, these graphs suggest that unemployment rates are not randomly assigned to newborns in the sense that they are somewhat related to the parental socioeconomic status. This in turn suggests that there may also be systematic fertility differences in unobserved parental characteristics, strengthening the case for a sibling fixed-effects approach when estimating long-run effects of conditions around birth.⁴

3.2.3 Unemployment

We may distinguish between two phases in the economic development of the Netherlands in the era after World War II. In the so-called Golden Age (1950-1973) the Dutch economy experienced high growth, led by the liberalization of trade with Western European countries and the growth of the petro-chemical and agricultural sector (van Zanden, 2005). After 1973 economic growth slowed down, leading to an economic crisis in the early 1980s due to the collapse of the Bretton Woods system (1971), the two oil crisis (1973 and 1979) and the expansion of the public sector (Smits et al., 1999, van Zanden, 2005). The manufacturing sector declined and unemployment increased substantially. In many respects, these developments were mirrored in other West-European countries.

The two phases are clearly visible in Figure 3.2, with the unemployment rates decreasing until the early 1970s and then increasing during the various economic crises leading up to the early 1980s. The three provinces in the Northern Netherlands experienced higher unemployment rates than most other provinces in the Netherlands. To gauge the relevance of the years depicted on the horizontal axis

⁴ As we shall see below, unemployment rates were highest in the most recent birth years. To the extent that parental education increases in the year of birth, some of the effects visualized in this subsection may be attributed to this. However, the relation between the unemployment rate and birth year is actually U-shaped. Moreover, the correlations between paternal and maternal education on the one hand and birth year on the other hand are only 0.15 and 0.25, respectively.

Figure 3.2. Provincial unemployment rate series

of Figure 3.2 it should be considered that over 91% of the individuals in our data with type-2 diabetes were born before 1973. This reflects the fact that prevalence increases with current age.

In the absence of reliable data on regional GDP, we use the provincial unemployment rate as a proxy for transitory contextual economic conditions. In the birth years in the data, some sort of unemployment benefits and welfare was typically available for unemployed workers. However, unemployment did lead to income loss and uncertainty about future work opportunities. Also, in recessions, families that did not encounter unemployment were nevertheless affected by the risk of future unemployment. In all this it is worth noting that female labor force participation in the period 1950-1973 was low by international standards and most household had at most one breadwinner. It should be pointed out that the Lifelines data do not provide detailed information on individual labor market outcomes over the years 1950-1995.

3.3 Empirical implementation

We define the model

$$T2D_i = \beta_1 g_i \cdot u_{p(i)} + \beta_2 (1 - g_i) \cdot u_{p(i)} + x_i' \gamma + \alpha_{j(i)} + \varepsilon_i \quad (3.1)$$

where the indices i, j, p denote individual, family and province, respectively. $T2D_i$ is a binary variable equal to 1 if the individual is affected by type-2 diabetes in the year that this was measured, and equal to 0 otherwise. To avoid misunderstandings, note that we only observe $T2D$ once per individual. In the notation we suppress that $T2D_i$ and covariates x_i are measured at the Lifelines baseline collection date of individual i . It is important to point out that the baseline collection date is not identical across individuals, so $T2D_i$ may be measured at different points in time for different individuals i . The binary gender indicator $g_i = 1$ if the individual is female. Next, $u_{p(i)}$ is the unemployment rate in the birth province and birth year. The coefficients β_1 and β_2 are the parameters of interest. The residual term in (3.1) is the sum of a component $\alpha_{j(i)}$ shared among siblings and an idiosyncratic component ε_i . To account for selective fertility and other family-shared determinants of diabetes, we control for dependence between $\alpha_{j(i)}$ on the one hand and $u_{p(i)}, g_i$ and x_i on the other hand, by estimating sibling fixed-effects regressions.

The vector x_i contains covariates that may differ between siblings, such as gender, being first-born, current age at baseline, the birth province (via “fixed-effects” ι_p) and the birth year (via “fixed-effects” δ_t for birth years t). The latter warrants some discussion. In a cross-sectional sample collected in a given year, current age and birth year (as parameterized by δ_t) are fully collinear. In our sampling design, their effects are separately identified because siblings may have had their Lifelines baseline date at different points in time, somewhere in a window of 8 years covering 2006-2013. Clearly, this is a fragile way of identifying these effects. Moreover, if we do not impose any functional form on the δ_t sequence then the estimated sequence will capture the national business cycle effect so that the effects of $u_{p(i)}$ are only identified by variation across provinces within birth years. Clearly, variation in the national business cycle over time quantitatively dominates cross-sectional variation across provincial business cycles. We therefore also estimate model specifications where the δ_t parameters are absent and where current age is included by way of a flexible polynomial.⁵

In addition, we estimate the model 3.1 with sibling random-effects in order to subsequently test it against the fixed-effects specification, using the so-called

⁵ Similarly, the ι_p parameters are identified because some siblings who belong to the same family were born in different provinces.

Mundlak approach (Mundlak, 1978). As unemployment at birth is exogenous from the individual perspective, rejecting random-effects in favor of fixed-effects implies that parental selective fertility should be accounted for. In other words, if unemployment at birth is randomly assigned to newborns and so there is no significant difference in the coefficient estimates between the random- and fixed-effects specifications, the Mundlak test rejects fixed-effects in favor of the more efficient random-effects estimator.

3.4 Results

The sibling fixed-effects estimates based on model (3.1) are reported in Table 3.2. We find a significant effect of the unemployment rate at birth on type-2 diabetes: 1 percentage point higher unemployment at birth increases by 0.3 percentage points the probability of having type-2 diabetes in both females and males. Thus, there is a causal effect of adverse economic conditions around birth on the development of type-2 diabetes later in life. The random-effects estimates are also reported in Table 3.2. It can be seen that the effect of the unemployment rate at birth is not significant. The Mundlak test marginally rejects the sibling random-effects in favor of the sibling fixed-effects specification. The province fixed-effects estimates, omitted from Table 3.2, are reported in Table 3.A.2 in Appendix 3.A. It can be seen that the province fixed-effects play a limited role as they are only identified by those few siblings who were born in different provinces (3,881).

A potential issue in our data is that the outcome distribution is skewed and so the linear probability model (3.1) may not be ideal. Therefore we also estimate a Chamberlain's correlated random-effects probit model, controlling for the family mean of the covariates (see Chamberlain, 1980, Wooldridge, 2010). The estimates, as well as the marginal effects of the unemployment rate, are reported in Table 3.A.3 in Appendix 3.A. Simple probit estimates and marginal effects are also reported for comparison. It can be seen that the marginal effects of unemployment in the Chamberlain's probit model are similar to the fixed-effects estimates in the linear probability model, although the result for females is not significant.

To ascertain the robustness of our results we perform a series of sensitivity ana-

lyses.⁶ These are discussed by topic in the remainder of this section:

Sibling status. The estimates reported in Table 3.2 are virtually identical to those based on data that are not corrected through the genetic records, namely including adopted siblings and excluding siblings whose blood-relationship is not reported or identified through questionnaire information. The same applies if we define as siblings those who have the mother in common but not necessarily the father (43,247 in total).

As an alternative to the sibling design one may sample cousin-cousin relationships in order to run cousin fixed-effects. The advantage of using cousins instead of siblings is that they can be born at the same point in time and in different provinces, so more variation of unemployment rate at birth can be exploited. Unfortunately, we can only identify these relationships if the sibling-sibling relationships of the respective parents are identified in Lifelines, so most of the cousin-cousin relationships remain unidentified. This severely reduces the sample size compared to our sibling sample. Moreover, the cousins thus identified are on average younger than the siblings in our data, implying that type-2 diabetes is rarer. Additional challenges are (i) that cousins can give rise to chains of relationships where two individuals that are not cousins of each other are each cousins of a third individual, and (ii) that an individual can have cousins that are siblings of each other. Any simple econometric approach would therefore require the cousin sample size to be reduced even further. In conclusion, a cousin study design is infeasible and we do not pursue this further.

Unemployment, province of birth, and age in the model specification. To check whether the results are driven by inter-province mobility due to differences in unemployment rates, we estimate the model omitting all those 3,881 siblings that, within a family, were born in different provinces. Obviously, we now exclude the province fixed-effects ι_p as they cannot be identified in regression. These estimates are reported in Table 3.A.4 in Appendix 3.A. We obtain very similar results to the estimates reported in Table 3.2, although the effect in males is now only marginally significant. It would be also interesting to estimate the same model for only those

⁶ When not reported, results of these analyses are available upon request.

Table 3.2. Sibling FE and RE estimates on *T2D*

	FE	RE
Unem. $\times$ female	0.0030* (0.0016)	0.0009 (0.0009)
Unem. $\times$ male	0.0032** (0.0016)	0.0008 (0.0009)
Female	−0.0044 (0.0030)	−0.0058** (0.0022)
Age	0.0007 (0.0008)	0.0009** (0.0004)
First born	0.0049*** (0.0018)	0.0036*** (0.0011)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test <i>p</i> -value		0.0620
$\hat{\sigma}_\alpha^2$		0.0006
$\hat{\sigma}_\epsilon^2$		0.0136
Obs.	43,131	43,131

T2D is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Standard errors (in parentheses) are clustered at the family level. * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

siblings who were born in different provinces, but the sample size is too small to detect any effect as our model specification is quite demanding.

Next, we turn to the age variable. It is well known that age is a risk factor for the occurrence of type-2 diabetes. The American Diabetes Association (2020) states that testing individuals for high diabetes risk should begin at age 45. The estimated age effect may be conflated with the effects of parity and, in particular, birth-year fixed-effects, as discussed in the previous section. We proceed in a number of ways. First, we interact the unemployment effects with age. The estimates are reported in Table 3.A.5 in Appendix 3.A while the marginal effects at different values of age are reported in Table 3.3. It can be seen that the effect of unemployment manifests with age with a significant impact in males after the age of 40. Second, we estimate

Table 3.3. Marginal effect of unemployment by age and gender on *T2D*

Age	Female	Male
20	0.0011 (0.0017)	0.0009 (0.0017)
30	0.0016 (0.0013)	0.0020 (0.0014)
40	0.0021 (0.0014)	0.0031** (0.0014)
50	0.0026 (0.0018)	0.0042** (0.0019)
60	0.0031 (0.0025)	0.0054** (0.0025)

* $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$.

the gender-specific average effect of unemployment at birth for the population of siblings over the age of 45. The estimates are reported in Table 3.4. As expected, the average effect of unemployment at birth is much higher for this sub-population of siblings. Finally, we estimate the main model excluding the birth-year fixed-effects δ_t while including age as well as age squared. We compute the marginal effect of age and we find that this is significant and increases with age, as expected. If year fixed-effects are also included, although the point estimates of the marginal effect of age do not change relevantly the standard errors increase dramatically, again as expected. The estimates of both these specifications are reported in Table 3.A.6 in Appendix 3.A.

It should also be pointed out that the results are robust with respect to the birth cohort interval. In particular, larger intervals lead to virtually identical results.

Operationalization of type-2 diabetes. In Table 3.A.7 in Appendix 3.A we also report the estimates based on self-reported type-2 diabetes as dependent variable. It can be seen that the estimates of the parameters of interest are not significant. This is likely due to underdiagnosis of the disease, suggesting that self-reported data on type-2 diabetes is not reliable. In principle, measurement error in the dependent variable should not bias the estimates as long as this measurement error is randomly distributed. To investigate whether this is the case, we divide individuals who have type-2 diabetes based on the biomarkers in two groups. Indi-

Table 3.4. Sibling FE and RE estimates on *T2D*: siblings over age 45

	FE	RE
Unem. $\times$ female	0.0059* (0.0033)	0.0018 (0.0023)
Unem. $\times$ male	0.0084** (0.0034)	0.0037 (0.0024)
Female	0.0023 (0.0074)	−0.0014 (0.0057)
Age	0.0009 (0.0017)	0.0007 (0.0010)
First born	0.0077** (0.0039)	0.0073*** (0.0027)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test <i>p</i> -value		0.0948
$\hat{\sigma}_\alpha^2$		0.0018
$\hat{\sigma}_\epsilon^2$		0.0247
Obs.	15,617	15,617

T2D is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Standard errors (in parentheses) are clustered at the family level. * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

viduals in the first group report having the disease while individuals in the second group report not having the disease. In Table 3.A.8 in Appendix 3.A we display descriptive statistics of the two groups regarding age, gender and education and we run a *t*-test on the mean-comparisons between the two samples. A number of things stand out. First, more than 50% of the individuals who, based on the biomarkers, have type-2 diabetes do not report the disease. Second, the two groups differ by gender (although this difference is not significant), education and age, with younger individuals more likely to have undiagnosed type-2 diabetes. The latter may be because the disease is asymptomatic at early stages and because, as already mentioned, recommendations point toward testing after the age of 45. The difference in the education level might be explained by the age difference between

the two samples, with younger individuals more likely to have higher education.

Placebo regressions Macroeconomic conditions before conception should not have any impact on the probability of contracting type-2 diabetes, otherwise our main results might be spurious rather than true causal effects. Therefore we test the robustness of our main results by regressing type-2 diabetes on unemployment rate two years before birth (so at least one year before conception). For this regression we exclude those who were born before 1952 or in the Flevoland province before 1975 from the data, as we do not have unemployment data for those years. Results are reported in Table 3.A.9 in Appendix 3.A. Clearly, there is no significant effect of the unemployment rate on the occurrence of type-2 diabetes in adulthood.

3.5 Conclusion

We find that among males, a one percentage point higher provincial unemployment rate at birth increases the probability of developing type-2 diabetes in adulthood by 0.3 percentage points. This increases up to 0.8 percentage points for the sub-population of individuals older than 45. The estimated effects among females are similar in magnitude but only marginally significant. In our explanatory framework, this implies that at least among men, adverse economic conditions *in utero* and shortly after birth increase the probability of type-2 diabetes later in life.

The fact that diabetes in modern societies responds to modest changes in economic conditions around birth means that the landmark findings on long-run effects of famines on diabetes can be generalized from extreme historical settings to common contemporaneous settings. Our study therefore extends the relevance of the findings in the famine studies, towards modern economies.

We find that analyses that solely rely on the self-reported prevalence of type-2 diabetes produce systematically different results.⁷ This suggests that self-reported data is not reliable. Individuals who do not report diabetes but have diabetes according to the official diagnostic guidelines may be expected to suffer from long-term health damage if no medical help is requested and/or if the diabetes goes undetected for many years.

⁷ In particular, they give insignificant estimates of the long-run effects, presumably because self-reported prevalence is much lower than diagnosed.

Our study shows that not accounting for selective fertility leads to insignificant and hence erroneous estimates for the effects of economic conditions at birth on type-2 diabetes. One may argue that in studies like ours, where individuals have a high degree of control over fertility, the potential for endogenous fertility selection is larger than in historical studies. The fact that it matters a great deal if selection is properly dealt with is therefore of importance for the wider literature using relatively recent birth cohorts.

The data enable a comparison between self-reported sibship and sibship status based on genetic identifiers. The sub-sample of biological but unreported siblings is currently too small for quantitative inference, but we do observe systematic differences in terms of covariates. In the future, the availability of both types of sibling identifier may open up exciting research opportunities for the study of health determinants.

The results of the paper have some obvious policy implications. First of all, note that screening of individuals on diabetes is important because, as we have seen, self-reported diabetes underestimates clinically diagnosed diabetes. When targeting individuals for screening it would be efficient to take into account if individuals experienced adverse conditions around birth (e.g. immigrants who were born in a poor country of origin) as their susceptibility is higher. Similarly, it would be efficient to focus on such groups when attempting to raise awareness of type-2 diabetes. All this is particularly relevant for type-2 diabetes as a health hazard because, if diagnosed early, it may be fully reversed for at least some groups of individuals, by way of specific weight-loss programs (see Lean et al., 2018 and Taylor, 2019). A more down-to-earth policy recommendation would be to inform and support pregnant women and parents of newborns living in poor economic conditions. In this sense, prevention of type-2 diabetes may start before birth.

3.A Appendix

Table 3.A.1. Distribution of siblings by province of birth

Drenthe	9,746	Gelderland	469	Noord-Brabant	241	Utrecht	293
Flevoland	84	Groningen	11,882	Noord-Holland	599	Zeeland	30
Friesland	17,885	Limburg	66	Overijssel	1,154	Zuid-Holland	682

Table 3.A.2. Province FE from sibling FE and RE estimates in Table 3.2

	FE	RE
Flevoland	0.0027 (0.0059)	−0.0046*** (0.0013)
Friesland	−0.0001 (0.0075)	−0.0037** (0.0016)
Gelderland	0.0001 (0.0108)	−0.0046 (0.0052)
Groningen	0.0081 (0.0060)	0.0002 (0.0015)
Limburg	0.0434 (0.0434)	−0.0036 (0.0161)
Noord-Brabant	0.0021 (0.0113)	−0.0054 (0.0062)
Noord-Holland	0.0059 (0.0141)	−0.0019 (0.0055)
Overijssel	0.0112 (0.0119)	0.0032 (0.0039)
Utrecht	0.0351* (0.0182)	−0.0004 (0.0071)
Zeeland	−0.0474 (0.0513)	−0.0247*** (0.0040)
Zuid-Holland	−0.0027 (0.0130)	0.0021 (0.0064)

Standard errors (in parentheses) are clustered at the family level. * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.3. Probit estimates on *T2D*

	Probit	Probit (Chamberlain)
Unem. × female	0.0033 (0.0222)	0.0652 (0.0401)
Unem. × male	0.0130 (0.0221)	0.0878** (0.0408)
Female	−0.1083 (0.0672)	−0.0605 (0.0940)
Age	0.0277** (0.0113)	0.0205 (0.0221)
First born	0.1045*** (0.0337)	0.1150** (0.0537)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Obs.	43,131	43,131
Average marginal effect Unem. × female	0.0001 (0.0008)	0.0023 (0.0014)
Average marginal effect Unem. × male	0.0005 (0.0008)	0.0031** (0.0015)

T2D is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Probit (Chamberlain) also includes the family mean of the covariates (estimates not reported). Standard errors (in parentheses) are clustered at the family level.

* $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.4. Sibling FE and RE estimates on *T2D*: siblings born in the same province

	FE	RE
Unem. × female	0.0034* (0.0019)	0.0011 (0.0008)
Unem. × male	0.0034* (0.0020)	0.0009 (0.0008)
Female	−0.0064** (0.0032)	−0.0069*** (0.0023)
Age	0.0010 (0.0009)	0.0010* (0.0004)
First born	0.0057*** (0.0018)	0.0038*** (0.0012)
Year of birth FE	Yes	Yes
Mundlak test <i>p</i> -value		0.0561
$\hat{\sigma}_\alpha^2$		0.0007
$\hat{\sigma}_\epsilon^2$		0.0135
Obs.	39,250	39,250

T2D is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Standard errors (in parentheses) are clustered at the family level. * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

Table 3.A.5. Sibling FE estimates on *T2D*: unemployment age interaction

	FE	RE
Unem. $\times$ female	0.0000 (0.0030)	−0.0001 (0.0017)
Unem. $\times$ male	−0.0013 (0.0031)	−0.0017 (0.0018)
Unem. $\times$ female $\times$ age	0.0001 (0.0001)	0.0000 (0.0000)
Unem. $\times$ male $\times$ age	0.0001 (0.0001)	0.0001 (0.0000)
Female	−0.0009 (0.0037)	−0.0023 (0.0029)
Age	0.0004 (0.0009)	0.0008 (0.0005)
First born	0.0049*** (0.0018)	0.0036*** (0.0011)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test p -value		0.0494
$\hat{\sigma}_\alpha^2$		0.0006
$\hat{\sigma}_\epsilon^2$		0.0136
Obs.	43,131	43,131

T2D is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. The unemployment variable is the unemployment rate of the province of birth 2 years before birth. Standard errors (in parentheses) are clustered at the family level. * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

Table 3.A.6. Sibling FE estimates on *T2D*: alternative specifications

Unem. × female	−0.0003 (0.0004)	0.0029* (0.0016)
Unem. × male	−0.0001 (0.0005)	0.0032* (0.0016)
Female	−0.0045 (0.0030)	−0.0044 (0.0030)
Age	−0.0049*** (0.0010)	−0.0037 (0.0033)
Age ²	0.0001*** (0.0000)	0.0001 (0.0000)
First born	0.0052*** (0.0017)	0.0049*** (0.0018)
Province FE	Yes	Yes
Year of birth FE	No	Yes
Obs.	43,131	43,131
Marginal effect age at:		
age=20	−0.0021*** (0.0006)	−0.0017 (0.0017)
age=40	0.0006** (0.0003)	0.0004 (0.0007)
age=60	0.0033*** (0.0005)	0.0024 (0.0020)

T2D is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. Standard errors (in parentheses) are clustered at the family level. * $p < 0.1$
 ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.7. Sibling FE and RE estimates on self-reported *T2D*

	FE	RE
Unem. $\times$ female	0.0006 (0.0014)	0.0007 (0.0007)
Unem. $\times$ male	0.0004 (0.0014)	0.0007 (0.0007)
Female	−0.0025 (0.0025)	−0.0018 (0.0017)
Age	0.0027*** (0.0007)	0.0014*** (0.0003)
First born	0.0005 (0.0015)	0.0012 (0.0009)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test <i>p</i> -value		0.0002
$\hat{\sigma}_\alpha^2$		0.0003
$\hat{\sigma}_\epsilon^2$		0.0088
Obs.	43,131	43,131

Self-reported *T2D* is a dummy variable equal to 1 if the individual self-reports to have type-2 diabetes, and equal to 0 otherwise. Standard errors (in parentheses) are clustered at the family level. * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

Table 3.A.8. Means-comparison *t*-test: individuals with unreported and reported type-2 diabetes

	Unreported <i>T2D</i>			Reported <i>T2D</i>			<i>t</i> -Test (<i>p</i> -value)
	Obs.	Mean	Std. dev.	Obs.	Mean	Std. dev.	
Female	338	0.49	0.50	294	0.55	0.50	0.180
Age	338	47.69	9.05	294	51.10	6.64	0.000
Low education	333	0.75	0.43	285	0.82	0.38	0.033

We test the null hypothesis $\mu_x = \mu_z$ against the alternative $\mu_x \neq \mu_z$. The *t*-Test is given by $t = \frac{\bar{x} - \bar{z}}{(\frac{s_x^2}{n_x} + \frac{s_z^2}{n_z})^{1/2}}$ and is distributed as a *t*-Student with $t = \frac{(\frac{s_x^2}{n_x} + \frac{s_z^2}{n_z})^2}{(\frac{s_x^2}{n_x})^2 + (\frac{s_z^2}{n_z})^2}$ degrees of freedom.

Table 3.A.9. Sibling FE estimates on *T2D*: placebo regressions

	2 years before birth	
	FE	RE
Unem. $\times$ female	−0.0002 (0.0016)	0.0001 (0.0008)
Unem. $\times$ male	0.0002 (0.0015)	0.0001 (0.0008)
Female	−0.0039 (0.0029)	−0.0050** (0.0021)
Age	0.0006 (0.0008)	0.0009** (0.0004)
First born	0.0040** (0.0018)	0.0031*** (0.0011)
Province FE	Yes	Yes
Year of birth FE	Yes	Yes
Mundlak test p -value		0.1041
$\hat{\sigma}_\alpha^2$		0.0006
$\hat{\sigma}_\epsilon^2$		0.0128
Obs.	41,756	41,756

T2D is a dummy variable equal to 1 if the HbA1c concentration in the blood is equal to or higher than 6.5% or if the fasting glucose level is equal to or higher than 7 mmol/L, and equal to 0 otherwise. The unemployment variable is the unemployment rate of the province of birth 2 years before birth. Standard errors (in parentheses) are clustered at the family level. * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. α and ϵ are the family-specific component and the idiosyncratic component of the error term, respectively.

Chapter 4

Clustered standard errors with few clusters

4.1 Introduction

In linear regressions with data grouped into clusters it is common practice to estimate the variance of the estimated parameters with the Cluster-Robust Variance Estimator (CRVE from hereon) introduced by Liang and Zeger (1986) as a generalization of the White's (1980) heteroskedastic-robust estimator. The justification is asymptotic, with number of clusters tending to infinity. Bell and McCaffrey (2002) show that in a finite context, with few clusters¹ and error terms that are correlated within clusters, CRVE leads to severely downward-biased standard errors and thus to misleading inference on the estimated parameters. Moulton (1986, 1990) and Cameron and Miller (2015) point out that this issue is relevant for regressors that are correlated or constant within clusters, such as policy variables that are only implemented in certain regions or states. An additional issue for inference on the estimated parameters is that, under the null hypothesis and with few clusters, the distribution of the test statistic is unknown and not approximately normal.

Following the idea of Bell and McCaffrey (2002), inference on the estimated parameters can be improved by (i) reducing the bias of CRVE with either CR2VE, also known as BRL (bias reduced linearization), or the jackknife estimator CR3VE,

This chapter was written with R. Alessie, V. Angelini, J. Mierau and T. Wansbeek except the parts on CR3VE- λ which were written with T. Wansbeek only.

¹ Less than 20-50, according to Cameron and Miller (2015).

both based on transformed OLS residuals, and by (ii) approximating the distribution of the test statistic with the t -distribution with an extension of the Satterthwaite (1946) degrees of freedom (DOF) that are data-determined and regressor-specific. Imbens and Kolesar (2016) develop the IK, a more refined version of the data-determined regressor-specific DOF used by Bell and McCaffrey (2002).

These methods have drawbacks that are particularly relevant for empirical research. First, Bell and McCaffrey (2002) show that CR3VE tends to overestimate the standard errors. Second, CR2VE, CR3VE and the IK can be computationally demanding as they are based on the computation of the inverse (CR3VE) or the inverse square root (CR2VE and the IK) of square matrices of order equal to the size of the clusters.

In view of these issues, this chapter presents some results that are meant for empirical researchers who wish to estimate a linear model on cross-sectional data clustered in few clusters. First, we introduce CR3VE- λ , a cluster-robust variance estimator that is identical to CR3VE in case of balanced clusters but, with unbalanced clusters, takes the difference in cluster sizes into account so that the computed standard errors are less conservative and unbiased under more general conditions. Second, we show how to compute CR2VE, CR3VE, CR3VE- λ and the IK efficiently, regardless of the size of the clusters, by inverting matrices of order equal to the number of regressors only.

Through Monte Carlo simulations we compare the performance of CRVE, CR2VE, CR3VE and CR3VE- λ and we show that, with few highly unbalanced clusters, CR3VE- λ might be the best choice from an inferential point of view. Moreover, we show that our efficient formulas produce high gains in terms of computational time: for example, more than three hours can be saved for the computation of CR2VE and CR3VE on a standard machine using a dataset with 10 clusters and 5,000 observations per cluster.

This chapter is organized as follows. In Section 4.2 we discuss basic theory on CRVE, CR2VE and CR3VE. In Section 4.3 we introduce CR3VE- λ . In Section 4.4 and 4.5 we introduce the efficient computation of CR2VE, CR3VE and CR3VE- λ , and the IK, respectively. In Section 4.6 we illustrate and test the performance of CRVE, CR2VE, CR3VE and CR3VE- λ to compute standard errors with few clusters by Monte Carlo simulations based on simulated data. In Section 4.7 we use real data

and illustrate an empirical exercise to compare the performance of CRVE, CR2VE, CR3VE and CR3VE- λ with few highly unbalanced clusters. In Section 4.8 we explore the computational time that is saved by using our efficient formulas for the computation of CR2VE and CR3VE. Section 4.9 concludes the chapter.

This chapter comes with a Stata do-file that can be used with any cross-sectional dataset for the computation of the standard errors based on the discussed methods and with the Stata do-files to replicate the experiments and the simulated datasets.²

4.2 Basic theory: CRVE, CR2VE and CR3VE

Consider the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with observations that can be grouped into c clusters of size $n_1, \dots, n_c$, where $\sum_i n_i = n$. We write, for the i -th cluster, $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$, with $E(\boldsymbol{\varepsilon}_i) = \mathbf{0}$ and $\text{var}(\boldsymbol{\varepsilon}_i) = \mathbf{V}_i$. The $\mathbf{V}_i$'s are collected in the block-diagonal matrix $\mathbf{V}$. After estimating the regression model with OLS we define

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_i \mathbf{X}_i'\mathbf{V}_i\mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1}. \quad (4.1)$$

The cluster-robust variance estimator (CRVE) based on OLS residuals per cluster $\hat{\boldsymbol{\varepsilon}}_i$ is

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_i \mathbf{X}_i'\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1}. \quad (4.2)$$

This estimator, which is introduced by Liang and Zeger (1986) and generalizes the heteroskedastic-robust estimator defined by White (1980), is consistent when the number of clusters goes to infinity. In case of few clusters, asymptotics will be a poor guide. Therefore we consider its bias instead.

Let $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, let $\mathbf{S}_i$ be the $n \times n_i$ matrix that selects the columns of $\mathbf{M}$ corresponding to cluster i , let $\mathbf{L}_i \equiv \mathbf{M}\mathbf{S}_i$ and let

$$\mathbf{H}_i \equiv \mathbf{S}_i'\mathbf{M}\mathbf{S}_i = \mathbf{I}_{n_i} - \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i'.$$

There holds $\mathbf{H}_i = \mathbf{L}_i'\mathbf{L}_i$ since $\mathbf{M}$ is idempotent and symmetric. With $\hat{\boldsymbol{\varepsilon}} = \mathbf{M}\boldsymbol{\varepsilon}$ and

² The Stata do-files are available upon request.

$\hat{\varepsilon}_i = \mathbf{L}_i' \varepsilon$, we have $E(\hat{\varepsilon}_i \hat{\varepsilon}_i') = \mathbf{L}_i' \mathbf{V} \mathbf{L}_i \neq \mathbf{V}_i$, so

$$E[\widehat{\text{var}}(\hat{\beta})] = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{L}_i' \mathbf{V} \mathbf{L}_i \mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1} \neq \text{var}(\hat{\beta}).$$

To reduce the bias, consider a variance estimator based on transformed residuals $\tilde{\varepsilon}_i \equiv \mathbf{A}_i \hat{\varepsilon}_i$, for some $\mathbf{A}_i$ to be chosen. Then

$$E[\widehat{\text{var}}(\hat{\beta})] = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{A}_i \mathbf{L}_i' \mathbf{V} \mathbf{L}_i \mathbf{A}_i' \mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1}.$$

From (4.1), unbiasedness requires the $\mathbf{A}_i$ to be such that $\mathbf{A}_i \mathbf{L}_i' \mathbf{V} \mathbf{L}_i \mathbf{A}_i' = \mathbf{V}_i$ for all i uniformly in the $\mathbf{V}_i$. This is infeasible and therefore we consider two second-best solutions.

The first one is to consider the case of no cluster effects, $\mathbf{V}_i = \sigma^2 \mathbf{I}_{n_i}$ for all i , and make the estimator unbiased for this case. Then $E(\hat{\varepsilon}_i \hat{\varepsilon}_i') = \mathbf{L}_i' \mathbf{V} \mathbf{L}_i = \sigma^2 \mathbf{L}_i' \mathbf{L}_i = \sigma^2 \mathbf{H}_i$ and consequently

$$E[\widehat{\text{var}}(\hat{\beta})] = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{A}_i \mathbf{H}_i \mathbf{A}_i' \mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1}. \quad (4.3)$$

The variance estimator is unbiased if $\mathbf{A}_i \mathbf{H}_i \mathbf{A}_i' = \mathbf{I}_{n_i}$ and so we choose $\mathbf{A}_i = \mathbf{H}_i^{-1/2}$. This estimator, introduced by Bell and McCaffrey (2002), is known as both CR2VE and BRL.

The second one is based on the idea that the elements in $\mathbf{M}$ outside the blocks on the diagonal may be small. Then $\mathbf{L}_i$ can be approximated by a matrix with $\mathbf{H}_i$ as its i th block and zeros outside this block. Then $\mathbf{L}_i' \mathbf{V} \mathbf{L}_i = \mathbf{H}_i \mathbf{V}_i \mathbf{H}_i$ and choosing $\mathbf{A}_i = \mathbf{H}_i^{-1}$ leads, when scaled by a factor $(c-1)/c$, to an estimator that is approximately unbiased when there are no cluster effects. In line with Cameron and Miller (2015), we call this estimator with the jackknife correction CR3VE.

4.3 From CR3VE to CR3VE- λ

To analyze the bias of CR3VE we scale (4.3) by $(c-1)/c$ and use

$$\mathbf{A}_i \mathbf{H}_i \mathbf{A}_i' = \mathbf{H}_i^{-1} = \mathbf{I}_{n_i} + \mathbf{X}_i (\mathbf{X}'\mathbf{X} - \mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i'$$

to obtain

$$E[\widehat{\text{var}}(\hat{\beta})] = \frac{c-1}{c}\sigma^2 \left((\mathbf{X}'\mathbf{X})^{-1} + \sum_i (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_i \mathbf{X}_i (\mathbf{X}'\mathbf{X} - \mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{X}_i (\mathbf{X}'\mathbf{X})^{-1} \right). \quad (4.4)$$

When clusters are balanced and have the same covariance structure there holds $\mathbf{X}'_i \mathbf{X}_i = \mathbf{X}'\mathbf{X}/c$ for all i , and (4.4) reduces to $E[\widehat{\text{var}}(\hat{\beta})] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Therefore, in case of balanced clusters having the same covariance structure, CR3VE with the correction factor $(c-1)/c$ is unbiased.

We propose a different scaling factor than $(c-1)/c$ for CR3VE in the more general case of unbalanced clusters that still have the same covariance structure. Define $\pi_i \equiv n_i/n$ for cluster i . Then $\mathbf{X}'_i \mathbf{X}_i = \pi_i \mathbf{X}'\mathbf{X}$ and the expression in parentheses in (4.4) becomes $\lambda(\mathbf{X}'\mathbf{X})^{-1}$, with

$$\lambda \equiv 1 + \sum_i \frac{\pi_i^2}{1 - \pi_i}.$$

There holds $\lambda \geq c/(c-1)$, with equality in case of balanced clusters. To see this, let $\boldsymbol{\pi} \equiv (\pi_1, \dots, \pi_c)'$, $\mathbf{P} \equiv \text{diag}(\boldsymbol{\pi})$, $\mathbf{a} \equiv (\mathbf{I}_c - \mathbf{P})^{-1/2} \boldsymbol{\pi}$ and $\mathbf{b} \equiv (\mathbf{I}_c - \mathbf{P})^{1/2} \boldsymbol{\iota}_c$, $\mathbf{a}'\mathbf{a} = \boldsymbol{\pi}'(\mathbf{I}_c - \mathbf{P})^{-1} \boldsymbol{\pi}$, $\mathbf{b}'\mathbf{b} = \boldsymbol{\iota}'_c(\mathbf{I}_c - \mathbf{P}) \boldsymbol{\iota}_c$, and $\mathbf{a}'\mathbf{b} = 1$. Since $(\mathbf{a}'\mathbf{b})^2 \leq \mathbf{a}'\mathbf{a} \mathbf{b}'\mathbf{b}$ there holds

$$\sum_i \frac{\pi_i^2}{1 - \pi_i} = \boldsymbol{\pi}'(\mathbf{I}_c - \mathbf{P})^{-1} \boldsymbol{\pi} \geq \frac{1}{\boldsymbol{\iota}'_c(\mathbf{I}_c - \mathbf{P}) \boldsymbol{\iota}_c} = \frac{1}{c-1},$$

so $\lambda - 1 \geq 1/(c-1)$ or $\lambda \geq c/(c-1)$. This suggests that $1/\lambda$ may be a better scaling factor than $(c-1)/c$. We denote this estimator, which is unbiased under more general conditions than CR3VE, by CR3VE- λ .

4.4 Efficient computation of CR2VE, CR3VE, CR3VE- λ with $\mathbf{H}_i$

CR2VE and CR3VE are based on $\mathbf{H}_i^a$, with $a = -1/2$ and $a = -1$, respectively. Especially with large n_i it is desirable to exploit the structure of $\mathbf{H}_i$ for the computations. We do so through the following result, that allows for reducing the computing and storage requirements to be just $O(n_i)$ instead of $O(n_i^2)$ for storage and $O(n_i^3)$ for

inversion.³ Let $\mathbf{R}$ be a matrix of “large” number of rows ℓ and “small” number of columns s , $\ell \geq s$, and let $\mathbf{R}$ have full column rank and satisfy $\mathbf{R}'\mathbf{R} \leq \mathbf{I}_s$. Then

$$\mathbf{R}'(\mathbf{I}_\ell - \mathbf{R}\mathbf{R}')^a = (\mathbf{I}_s - \mathbf{R}'\mathbf{R})^a \mathbf{R}' \quad (4.5)$$

for any a . To see this, take the singular value decomposition $\mathbf{R} = \mathbf{U}\mathbf{J}\mathbf{T}'$, with $\mathbf{J}$ diagonal, $\mathbf{T}$ square orthonormal, and $\mathbf{U}$ having orthonormal columns. Then both sides of (4.5) appear to be equal to $\mathbf{T}\mathbf{J}(\mathbf{I} - \mathbf{J}^2)^a \mathbf{U}'$.

Now write the right-hand side of (4.2) as $(\mathbf{X}'\mathbf{X})^{-1}[\sum_i \hat{\mathbf{s}}_i \hat{\mathbf{s}}_i'](\mathbf{X}'\mathbf{X})^{-1}$, where $\hat{\mathbf{s}}_i \equiv \mathbf{X}'_i \hat{\boldsymbol{\varepsilon}}_i$. With CR2VE ($a = -1/2$), CR3VE ($a = -1$) and CR3VE- λ ($a = -1$), $\hat{\mathbf{s}}_i$ has to be replaced by $\tilde{\mathbf{s}}_i \equiv \mathbf{X}'_i \mathbf{H}_i^a \hat{\boldsymbol{\varepsilon}}_i$, still with scaling to be added for CR3VE and CR3VE- λ . Define

$$\mathbf{R}_i \equiv \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1/2}, \quad (4.6)$$

so $\mathbf{X}'_i = (\mathbf{X}'\mathbf{X})^{1/2} \mathbf{R}'_i$ and $\mathbf{H}_i = \mathbf{I}_i - \mathbf{R}_i \mathbf{R}'_i$. Then from (4.5)

$$\begin{aligned} \tilde{\mathbf{s}}_i &= (\mathbf{X}'\mathbf{X})^{1/2} \mathbf{R}'_i (\mathbf{I}_i - \mathbf{R}_i \mathbf{R}'_i)^a \hat{\boldsymbol{\varepsilon}}_i \\ &= (\mathbf{X}'\mathbf{X})^{1/2} (\mathbf{I}_k - \mathbf{R}'_i \mathbf{R}_i)^a (\mathbf{X}'\mathbf{X})^{-1/2} \hat{\mathbf{s}}_i. \end{aligned}$$

So the computations to obtain $\tilde{\mathbf{s}}_i$ involve only matrices of order $k \times k$, which is $O(1)$ in n_i given $\mathbf{R}'_i \mathbf{R}_i$, $\mathbf{X}'\mathbf{X}$ and $\hat{\mathbf{s}}_i$; all three are computable in $O(n_i)$. This essentially simplifies the computation of CR2VE, CR3VE and CR3VE- λ . In Appendix 4.A we summarize the efficient formulas for CR2VE, CR3VE and CR3VE- λ .

4.5 Efficient computation of the Imbens and Kolesar degrees of freedom

Define $\hat{\beta}_r$ the estimated coefficient of the r th regressor, $r = 1, \dots, k$. With few clusters the distribution under the null of the test statistic for inference on $\hat{\beta}_r$ is unknown and not approximable to $N(0, 1)$. It is common practice in empirical research to use the t -distribution with $(c - 1)$ DOF or, more recently, with the IK developed by Imbens and Kolesar (2016) and based on $\mathbf{H}_i^{-1/2}$.

³ Le Gall (2014) gives the best-known lower bound of $O(n^{2.373})$. This is mainly of theoretical value and it holds for the optimized Coppersmith-Winograd algorithm.

Define the $n \times c$ matrix $\mathbf{F}_r$ with i th column equal to

$$\mathbf{F}_{ri} = \mathbf{G}_i \mathbf{e}_r, \quad (4.7)$$

where $\mathbf{G}_i = \mathbf{L}_i \mathbf{H}_i^{-1/2} \mathbf{X}_i (\mathbf{X}' \mathbf{X})^{-1}$ and $\mathbf{e}_r$ is a k -vector with r th element equal to 1 and any other elements equal to 0. Consider the random effect parametrization of $\mathbf{V} = \sigma^2 \mathbf{I}_n + \theta^2 \mathbf{D} \mathbf{D}'$, where $\mathbf{D} \equiv \text{diag}(\boldsymbol{\iota}_i)$. Then the IK for regressor r are

$$\text{IK}_r = \frac{(\sum_i \kappa_i)^2}{\sum_i \kappa_i^2}, \quad (4.8)$$

where κ_i are the eigenvalues of $\mathbf{F}_r' \hat{\mathbf{V}} \mathbf{F}_r \equiv \hat{\sigma}^2 \mathbf{F}_r' \mathbf{F}_r + \hat{\theta}^2 \mathbf{F}_r' \mathbf{D} \mathbf{D}' \mathbf{F}_r$ and $\hat{\sigma}^2$ and $\hat{\theta}^2$ can be obtained from a random effect estimation.

Based on (4.5) and with $\mathbf{R}$ as defined in (4.6) we derive the efficient formula for $\mathbf{G}_i$ as

$$\begin{aligned} \mathbf{G}_i &= \mathbf{L}_i \mathbf{H}_i^{-1/2} \mathbf{X}_i (\mathbf{X}' \mathbf{X})^{-1} \\ &= \mathbf{L}_i (\mathbf{I}_i - \mathbf{X}_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i')^{-1/2} \mathbf{X}_i (\mathbf{X}' \mathbf{X})^{-1} \\ &= \mathbf{L}_i (\mathbf{I}_i - \mathbf{R}_i \mathbf{R}_i')^{-1/2} \mathbf{R}_i (\mathbf{X}' \mathbf{X})^{-1/2} \\ &= \mathbf{L}_i \mathbf{R}_i (\mathbf{I}_k - \mathbf{R}_i' \mathbf{R}_i)^{-1/2} (\mathbf{X}' \mathbf{X})^{-1/2} \\ &= [\mathbf{S}_i - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i'] \mathbf{X}_i (\mathbf{X}' \mathbf{X})^{-1/2} [\mathbf{I}_k - (\mathbf{X}' \mathbf{X})^{-1/2} \mathbf{X}_i' \mathbf{X}_i (\mathbf{X}' \mathbf{X})^{-1/2}]^{-1/2} (\mathbf{X}' \mathbf{X})^{-1/2} \\ &\equiv \mathbf{S}_i \mathbf{X}_i \mathbf{W}_i - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i' \mathbf{X}_i \mathbf{W}_i, \end{aligned}$$

where $\mathbf{W}_i = (\mathbf{X}' \mathbf{X})^{-1/2} [\mathbf{I}_k - (\mathbf{X}' \mathbf{X})^{-1/2} \mathbf{X}_i' \mathbf{X}_i (\mathbf{X}' \mathbf{X})^{-1/2}]^{-1/2} (\mathbf{X}' \mathbf{X})^{-1/2}$ and where $\mathbf{S}_i \mathbf{X}_i \mathbf{W}_i$ is the $n \times k$ matrix with block that corresponds to cluster i equal to $\mathbf{X}_i \mathbf{W}_i$ and all the other rows equal to $\mathbf{0}$.

4.6 Monte Carlo simulations

We run several sets of Monte Carlo (MC) simulations and compare the bias of the standard errors based on unclustered standard errors (UN), CR1VE, CR2VE and CR3VE with the bias of the standard errors based on CR3VE- λ . In each simulation we generate randomly N unbalanced clusters with number of observations per cluster $n_i \sim U\{1000 - g, 1000 + g\}$, where g is different in each set of simulations. By doing so we keep constant the average number of observations

Table 4.1. Bias of $se(\hat{\beta})$ based on different methods: 10,000 MC simulations

	Std. deviation cluster size				
	Balanced	145	289	520	572
4 clusters					
$E(\hat{\beta})$	0.9977	0.9938	0.9998	1.0003	0.9984
$E[sd(\hat{\beta})]$	0.1975	0.1970	0.1928	0.1778	0.1740
Bias $se_{UN}(\hat{\beta})$	-0.1817	-0.1812	-0.1769	-0.1615	-0.1575
Bias $se_{CR1VE}(\hat{\beta})$	-0.1297	-0.1273	-0.1201	-0.1060	-0.1034
Bias $se_{CR2VE}(\hat{\beta})$	-0.0678	-0.0663	-0.0635	-0.0601	-0.0595
Bias $se_{CR3VE}(\hat{\beta})$	0.0170	0.0196	0.0199	0.0162	0.0154
Bias $se_{CR3VE-\lambda}(\hat{\beta})$	0.0170	0.0188	0.0169	0.0065	0.0036
6 clusters					
$E(\hat{\beta})$	0.9989	1.0020	0.9989	0.9968	1.0009
$E[sd(\hat{\beta})]$	0.1837	0.1838	0.1828	0.1817	0.1811
Bias $se_{UN}(\hat{\beta})$	-0.1708	-0.1709	-0.1698	-0.1685	-0.1678
Bias $se_{CR1VE}(\hat{\beta})$	-0.0774	-0.0778	-0.0783	-0.0845	-0.0875
Bias $se_{CR2VE}(\hat{\beta})$	-0.0302	-0.0306	-0.0312	-0.0381	-0.0420
Bias $se_{CR3VE}(\hat{\beta})$	0.0196	0.0200	0.0214	0.0214	0.0184
Bias $se_{CR3VE-\lambda}(\hat{\beta})$	0.0196	0.0195	0.0197	0.0153	0.0110

per cluster but we vary the unbalancedness. We generate our dependent variable $y_{hi} = \alpha + \beta x_{hi} + \gamma d_i + e_{hi}$, where h identifies the single observation (e.g., household) and i identifies the c clusters of size $n_i = n_1, \dots, n_c$, and where $x_{hi} = q_{hi} + z_i$ and $e_{hi} = w_{hi} + u_i$. Moreover, q_{hi}, z_i, w_{hi}, u_i are independently drawn from $N(0, 1)$, $\alpha = 0$ and $\beta = \gamma = 1$, and d_i is a dummy variable constant within cluster and randomly constrained, in each simulation, to be equal to 1 in half of the randomly generated clusters. The simulation set-up is very similar to the one in Cameron, Gelbach and Miller (2008). As pointed out by Cameron and Miller (2015), unclustered standard errors and CRVE are likely to be severely biased if the cluster effect and the correlation of the regressors within cluster are different from zero. Therefore we set-up experiments that allow to be correlated within cluster both e_{hi} and the regressors, including the extreme case of d_i , a dummy variable that is constant within clusters. The presence of a cluster effect and of regressors correlated within cluster implies that the assumption under which CR3VE and CR3VE- λ are unbiased are not met. Yet CR3VE- λ takes into account the difference in cluster size and, as this difference increases, CR3VE- λ is expected to be more unbiased than

Table 4.2. Bias of $se(\hat{\gamma})$ based on different methods: 10,000 MC simulations

	Std. deviation cluster size				
	Balanced	145	289	520	572
4 clusters					
$E(\hat{\gamma})$	1.0042	1.0216	0.9941	1.0013	0.9978
$E[sd(\hat{\gamma})]$	1.0212	1.0246	1.0371	1.0849	1.1078
Bias $se_{UN}(\hat{\gamma})$	-0.9807	-0.9839	-0.9960	-1.0420	-1.0634
Bias $se_{CR1VE}(\hat{\gamma})$	-0.4723	-0.4774	-0.5024	-0.6102	-0.6531
Bias $se_{CR2VE}(\hat{\gamma})$	-0.1908	-0.1926	-0.2156	-0.3241	-0.3684
Bias $se_{CR3VE}(\hat{\gamma})$	0.0941	0.1044	0.1063	0.1000	0.1132
Bias $se_{CR3VE-\lambda}(\hat{\gamma})$	0.0941	0.1004	0.0897	0.0364	0.0276
6 clusters					
$E(\hat{\gamma})$	1.0006	0.9899	0.9991	0.9944	1.0178
$E[sd(\hat{\gamma})]$	0.8303	0.8356	0.8504	0.9052	0.9278
Bias $se_{UN}(\hat{\gamma})$	-0.7962	-0.8014	-0.8160	-0.8699	-0.8921
Bias $se_{CR1VE}(\hat{\gamma})$	-0.2461	-0.2543	-0.2763	-0.3583	-0.3950
Bias $se_{CR2VE}(\hat{\gamma})$	-0.0820	-0.0877	-0.1029	-0.1598	-0.1891
Bias $se_{CR3VE}(\hat{\gamma})$	0.0537	0.0537	0.0547	0.0627	0.0616
Bias $se_{CR3VE-\lambda}(\hat{\gamma})$	0.0537	0.0518	0.0469	0.0327	0.0228

CR3VE.

We run 10,000 simulations for each MC set and each MC set differs with respect to the number of clusters N and g . We show results for $N = 4$ and $N = 6$, and for $g = 0$ (i.e., balanced clusters), $g = 250$, $g = 500$, $g = 900$ and $g = 990$, with standard deviation of the cluster size equal to 0, 145, 289, 520 and 572, respectively. For each simulation: (i) we compute the true standard deviation of $\hat{\beta}$, $sd(\hat{\beta})$, based on

$$\text{var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_i \mathbf{X}'_i \mathbf{V}_i \mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1},$$

where

$$\mathbf{V}_i = \mathbf{I}_{n_i} + \boldsymbol{\iota}_i \boldsymbol{\iota}'_i,$$

and where $\boldsymbol{\beta} = (\alpha, \beta, \gamma)$, (ii) we compute the standard errors of $\hat{\beta}$ and of $\hat{\gamma}$ based on the different methods se_{UN} , se_{CR1VE} , se_{CR2VE} , se_{CR3VE} and $se_{CR3VE-\lambda}$, (iii) we

compute the difference between the true standard deviations $\text{sd}(\hat{\beta})$, $\text{sd}(\hat{\gamma})$ and the standard errors based on the different methods. Finally, for each MC set we compute the expected value of this difference (i.e., the bias) for each method to compute the standard errors. From Table 1-2 we can see that CR3VE- λ always leads to the least biased standard errors, with bias always close to zero. Moreover, it remarkably reduces the bias of CR3VE with high unbalancedness. This is especially true for the dummy variable d_i .

4.7 Empirical exercise

Table 4.3. Rejection rates *policy* from 20,000 replications

		No. of states			
Method	Distribution	6	6hu ₁	6hu ₂	6hu ₃
Unclustered s.e.	$t(\text{IK})$	21.8	23.0	47.4	78.5
CRVE	$t(\text{IK})$	9.6	8.8	15.8	50.8
CR2VE	$t(\text{IK})$	5.3	3.4	4.4	12.3
CR3VE	$t(\text{IK})$	3.3	2.0	1.5	0.9
CR3VE- λ	$t(\text{IK})$	3.8	2.8	3.2	4.6
	mean(IK)	3.3	2.2	2.5	3.1
	$1/\lambda$		0.69	0.61	0.41

Rejection rates, in percentage, of the true null hypothesis on the fake *policy* variable from 20,000 replications for different methods to compute standard errors. Ideal rejection rate is 5%. 20% observations within sampled states are randomly sampled with replacement. The 6 states are randomly sampled with replacement. For all replications, the 6 highly unbalanced hu₁, hu₂ and hu₃ states are the 3 with most observations and the 3 with least observations, the 2 with most observations and the 4 with least observations, and the 1 with most observations and the 5 with least observations, respectively. The variance components for computing the IK DOF are estimated with restricted maximum likelihood (`mixed, reml` or `xtmixed, reml` command in Stata).

For the empirical illustration we use the same dataset CPS 2012 as in Cameron and Miller (2015). The dataset consists of 51 clusters, namely the 50 American States and the District of Columbia. We define the same model for individual h in the sampled cluster i

$$\ln(\text{wage})_{hi} = \beta_0 + \beta_1 \text{educ}_{hi} + \beta_2 \text{age}_{hi} + \beta_3 \text{age}_{hi}^2 + \beta_4 \text{policy}_i + \varepsilon_{hi}, \quad (4.9)$$

where *policy* is a fake policy variable randomly assigned to half of the sampled clusters and constant within each cluster. The clusters are unbalanced and the number of observations per cluster is reported in Table 4.A.1 in 4.A.

We run 4 sets of 20,000 replications. In the first set we randomly sample with replacement 6 states. In the other sets we fix the 6 states to be the 3 with most observations and the 3 with least observations ($6hu_1$), the 2 with most observations and the 4 with least observations ($6hu_2$), and the 1 with most observations and the 5 with least observations ($6hu_3$). In each set we randomly sample with replacement 20% of the observations within each state and we test the true null hypothesis $H_0 : \beta_4 = 0$ at the 5% level and thus we expect the standard errors of *policy* to lead to rejection of the true null hypothesis H_0 in 5% of the replications. Rejection rates using for inference the $t(\text{IK})$ distribution are reported in Table 4.3, where IK are the Imbens and Kolesar (2016) DOF. As expected the improvement based on CR3VE- λ is particularly relevant with higher unbalancedness of the sampled states.

4.8 Computational time gain for CR2VE and CR3VE

Table 4.4. Time for CR2VE and CR3VE using efficient and inefficient formulas

		No. of observations per cluster				
CR2VE+CR3VE	No. of clusters	1000	2000	3000	4000	5000
Efficient	6	1	1	1	1	1
Inefficient	6	44	368	1371	3569	6943
Efficient	10	1	1	1	1	1
Inefficient	10	67	599	2214	5630	11568

Total computational time in seconds of CR2VE and CR3VE using the formulas reported in Section 4.4 (efficient) and in A. Cameron and Miller (2015) (inefficient). The computations are run using Stata/SE 15.0 on the following machine: Intel(R) Core(TM) i3-4130 CPU @ 3.40GHz, RAM: 8,00 GB, Windows 7.

We report in Table 4.4 the computational time of our efficient formulas for CR2VE and CR3VE and of the equivalent, but inefficient, CR2VE and CR3VE as introduced by Bell and McCaffrey (2002). This computational time refers to CR2VE and CR3VE estimated together on a standard machine. We run these computations on simulated data with 51 balanced clusters and 5000 observations per cluster. The data generating process is $\ln(wage)_{hi} = 0.7495 + 0.0844age_{hi} - 0.0009age_{hi}^2 + u_i + e_{hi}$,

where $age \sim U\{18, 65\}$, $u_i \sim N(0, \theta^2)$ is constant within cluster i and $e_{hi} \sim N(0, \sigma^2)$, and where age_{hi} , u_i and e_{hi} are mutually independent. We set $\theta^2 = 9.5818 \times 10^{-3}$ and $\sigma^2 = 0.3489$. The parameters of the data generating process and of the u_i and e_{hi} distributions are chosen from a random effects regression on the CPS 2012 dataset, using all the 51 clusters and all the observations. Based on this data we define the model

$$\ln(wage)_{hi} = \beta_0 + \beta_1 age_{hi} + \beta_2 age_{hi}^2 + \beta_3 policy_i + \varepsilon_{hi}, \quad (4.10)$$

where *policy* is a fake policy variable randomly assigned to half of the clusters.

We sample 6 and 10 clusters, and 1000, 2000, 3000, 4000 and 5000 observations within each cluster from this simulated data. We compute the clustered standard errors based on CR2VE and CR3VE with these different samples. The computations of the inefficient CR2VE and CR3VE take up to more than three hours for 10 clusters. This depends on the fact that the inefficient formulas invert matrices of order $n_i \times n_i$ and thus the computational time increases with cluster size n_i . Oppositely, as shown in Section 4.4 the efficient formulas invert matrices of order $k \times k$ that does not depend on the cluster size n_i , where $k = 4$ is the number of regressors in model (4.10).

4.9 Conclusion

In this chapter we first introduce CR3VE- λ , an estimator for clustered standard errors that refines the jackknife estimator CR3VE and is unbiased under more general conditions, in case of few unbalanced clusters. In our simulations CR3VE- λ reduces the bias of CR3VE as the cluster unbalancedness and the correlation of the regressor within cluster increase. Our empirical exercise also suggests that CR3VE- λ might be the optimal choice with highly unbalanced clusters. Second, in this chapter we also propose efficient formulas to compute CR2VE, CR3VE and CR3VE- λ . These formulas reduce to 1 second the computational time of CR2VE, CR3VE and CR3VE- λ , regardless the size of the clusters. In our simulations more than three hours can be saved for the computation of CR2VE and CR3VE on a standard machine using a dataset with 10 clusters and 5,000 observations per cluster. We conclude that empirical researchers should always prefer CR3VE- λ to CR3VE with few unbal-

anced clusters and should always use our efficient formulas for the computation of CR2VE, CR3VE, CR3VE- λ and the IK.

4.A Appendix

CRVE, CR2VE, CR3VE and CR3VE- λ in a nutshell

Define the matrix of observations $\mathbf{X}$ of order $n \times k$ and the linear model for cluster $i = 1, \dots, c$

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

where $\mathbf{X}_i$ is a matrix of order $n_i \times k$, and where $E(\boldsymbol{\varepsilon}_i) = \mathbf{0}$ and $\text{var}(\boldsymbol{\varepsilon}_i) = \mathbf{V}_i$. Define the OLS residuals $\hat{\boldsymbol{\varepsilon}}_i$. The general expression for the cluster-robust estimator of $\text{var}(\hat{\boldsymbol{\beta}})$ is

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_i \mathbf{X}_i' \tilde{\boldsymbol{\varepsilon}}_i \tilde{\boldsymbol{\varepsilon}}_i' \mathbf{X}_i \right) (\mathbf{X}'\mathbf{X})^{-1},$$

where $\tilde{\boldsymbol{\varepsilon}}_i$ are a transformation of OLS residuals to be specified. CRVE simply uses $\tilde{\boldsymbol{\varepsilon}}_i = \hat{\boldsymbol{\varepsilon}}_i$. CR2VE uses $\tilde{\boldsymbol{\varepsilon}}_i = (\mathbf{I}_i - \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i')^{-1/2}\hat{\boldsymbol{\varepsilon}}_i$, while CR3VE and CR3VE- λ use $\tilde{\boldsymbol{\varepsilon}}_i = g[(\mathbf{I}_i - \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i')^{-1}\hat{\boldsymbol{\varepsilon}}_i]$, where $g = [(c-1)/c]^{1/2}$ for CR3VE and $g = \{1 + [\sum_i (n_i/n)^2 / (1 - n_i/n)]\}^{-1/2}$ for CR3VE- λ . In case of balanced clusters CR3VE and CR3VE- λ are identical. Only CRVE requires $c \rightarrow \infty$ which, in empirical applications, means that the number of clusters has to be sufficiently large.

CR2VE, CR3VE and CR3VE- λ can be computed efficiently with the inversion of matrices of order $k \times k$ instead of $n_i \times n_i$. Define $\hat{\mathbf{s}}_i = \mathbf{X}_i' \hat{\boldsymbol{\varepsilon}}_i$, $\mathbf{R}_i' = \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1/2}$ and the cluster robust variance estimator

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_i \tilde{\mathbf{s}}_i \tilde{\mathbf{s}}_i' \right) (\mathbf{X}'\mathbf{X})^{-1}.$$

Then to compute CR2VE we use $\tilde{\mathbf{s}}_i = [(\mathbf{X}'\mathbf{X})^{1/2}(\mathbf{I}_k - \mathbf{R}_i'\mathbf{R}_i)^{-1/2}(\mathbf{X}'\mathbf{X})^{-1/2}] \hat{\mathbf{s}}_i$, and to compute CR3VE and CR3VE- λ we use $\tilde{\mathbf{s}}_i = g[(\mathbf{X}'\mathbf{X})^{1/2}(\mathbf{I}_k - \mathbf{R}_i'\mathbf{R}_i)^{-1}(\mathbf{X}'\mathbf{X})^{-1/2}] \hat{\mathbf{s}}_i$.

Additional tables

Table 4.A.1. Number of observations per state - CPS 2012 dataset

Alabama	680	Kentucky	955	North Dakota	862
Alaska	712	Louisiana	560	Ohio	1504
Arizona	839	Maine	1039	Oklahoma	798
Arkansas	594	Maryland	1824	Oregon	803
California	5866	Massachusetts	971	Pennsylvania	1883
Colorado	1546	Michigan	1349	Rhode Island	1010
Connecticut	1457	Minnesota	1729	South Carolina	765
Delaware	1055	Mississippi	546	South Dakota	1012
District of Columbia	1009	Missouri	971	Tennessee	859
Florida	2630	Montana	519	Texas	3945
Georgia	1414	Nebraska	1207	Utah	827
Hawaii	1183	Nevada	1015	Vermont	949
Idaho	661	New Hampshire	1368	Virginia	1539
Illinois	2115	New Jersey	1376	Washington	1035
Indiana	962	New Mexico	538	West Virginia	590
Iowa	1343	New York	2842	Wisconsin	1259
Kansas	956	North Carolina	1290	Wyoming	924

The 51 clusters in the CPS 2012 dataset correspond to the 50 American states and the District of Columbia. The average number of observations per cluster is 1,288.

References

- Alessie, R., Angelini, V., Mierau, J. O. and Viluma, L. (2018). Economic downturns and infant health. *Economics and Human Biology*, 30, 162–171.
- Alessie, R., Angelini, V., van den Berg, G. J., Mierau, J. O. and Viluma, L. (2019). Economic conditions at birth and cardiovascular disease risk in adulthood: Evidence from post-1950 cohorts. *Social Science and Medicine*, 224, 77–84.
- Ali, O. (2013). Genetics of type 2 diabetes. *World Journal of Diabetes*, 4, 114–123.
- Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, 25, 153–172.
- Almond, D., Currie, J. and Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature*, 56, 1360–1446.
- American Diabetes Association. (2020). Classification and diagnosis of diabetes: Standards of medical care in diabetes in 2020. *Diabetes Care*, 43, S14–S31.
- Amin, V., Behrman, J. R. and Kohler, H. P. (2015). Schooling has smaller or insignificant effects on adult health in the US than suggested by cross-sectional associations: New estimates using relatively large samples of identical twins. *Social Science and Medicine*, 127, 181–189.
- Andrew, A., Cattan, S., Costa Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., ... Sevilla, A. (2020). Inequalities in children's experiences of home learning during the COVID-19 lockdown in England. *Fiscal Studies*, 41, 653–683.
- Angelini, V. and Mierau, J. O. (2014). Born at the right time? Childhood health and the business cycle. *Social Science and Medicine*, 109, 35–43.
- Arendt, J. N. (2005). Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review*, 24, 149–160.

- Barker, D. J. (1990). The fetal and infant origins of adult disease. *British Medical Journal*, 301, 1111.
- Barker, D. J. (1995). Fetal origins of coronary heart disease. *British Medical Journal*, 311, 171–174.
- Barker, D. J. (1998). In utero programming of chronic disease. *Clinical Science*, 95, 115–128.
- Barker, D. J. and Osmond, C. (1986). Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *The Lancet*, 327, 1077–1081.
- Bell, R. and McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28, 169–179.
- Ben-Shlomo, Y. and Kuh, D. (2004). *A Life Course Approach to Chronic Disease Epidemiology*. Oxford University Press.
- Bijwaard, G. E., van Kippersluis, H. and Veenman, J. (2015). Education and health: The role of cognitive ability. *Journal of Health Economics*, 42, 29–43.
- Bijwaard, G. E., van Poppel, F., Ekmamp, P. and Lumey, L. (2015). Gains in life expectancy associated with higher education in men. *PloS one*, 10, e0141200.
- Bol, T. (2020). Inequality in homeschooling during the corona crisis in the Netherlands. First results from the LISS panel. *Mimeo University of Amsterdam*.
- Borghans, L., Meijers, H. and Ter Weel, B. (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, 46, 2–12.
- Bound, J. and Solon, G. (1999). Double trouble: On the value of twins-based estimation of the return to schooling. *Economics of Education Review*, 18, 169–182.
- Calkins, K. and Devaskar, S. U. (2011). Fetal origins of adult disease. *Current Problems in Pediatric and Adolescent Health Care*, 41, 158–176.
- Cameron, A. and Miller, D. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50, 317–372.
- Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90, 414–427.

- Centers for Disease Control and Prevention. (2020). National diabetes statistics report, 2020. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services.
- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47, 225–238.
- Clark, D. and Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, 103, 2087–2120.
- Conti, G., Hansman, C., Heckman, J. J., Novak, M. F., Ruggiero, A. and Suomi, S. J. (2012). Primate evidence on the late health effects of early-life adversity. *Proceedings of the National Academy of Sciences*, 109, 8866–8871.
- Conti, G. and Heckman, J. J. (2010). Understanding the early origins of the education-health gradient: A framework that can also be applied to analyze gene-environment interactions. *Perspectives on Psychological Science*, 5, 585–605.
- Conti, G., Heckman, J. J. and Pinto, R. (2016). The effects of two influential early childhood interventions on health and healthy behaviour. *The Economic Journal*, 126, F28–F65.
- Conti, G., Heckman, J. J. and Urzua, S. (2010). The education-health gradient. *American Economic Review*, 100, 234–238.
- Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97, 31–47.
- Currie, J. and Rossin-Slater, M. (2013). Weathering the storm: Hurricanes and birth outcomes. *Journal of Health Economics*, 32, 487–503.
- De Cao, E., McCormick, B. and Nicodemo, C. (2019). Does unemployment worsen babies' health? A tale of siblings, maternal behaviour and selection. *IZA Discussion Paper No. 12568*.
- Dehejia, R. and Lleras-Muney, A. (2004). Booms, busts, and babies' health. *The Quarterly Journal of Economics*, 119, 1091–1130.
- Ekamper, P., van Poppel, F., Stein, A. and Lumey, L. (2014). Independent and additive association of prenatal famine exposure and intermediary life conditions with adult mortality between age 18–63 years. *Social Science and Medicine*, 119, 232–239.

- Engzell, P., Frey, A. and Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, 118.
- Fernandez-Twinn, D. and Ozanne, S. (2006). Mechanisms by which poor early growth programs type-2 diabetes, obesity and the metabolic syndrome. *Physiology and Behavior*, 88, 234–243.
- Fletcher, J. M. (2015). New evidence of the effects of education on health in the US: Compulsory schooling laws revisited. *Social Science and Medicine*, 127, 101–107.
- Forsdahl, A. (1977). Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease? *Journal of Epidemiology and Community Health*, 31, 91–95.
- Frijhoff, W. and Spies, M. (2004). *Dutch Culture in a European Perspective: 1950, Prosperity and Welfare*. Uitgeverij Van Gorcum.
- Fujiwara, T. and Kawachi, I. (2009). Is education causally related to better health? A twin fixed-effect study in the USA. *International Journal of Epidemiology*, 38, 1310–1322.
- Galama, T. J., Lleras-Muney, A. and van Kippersluis, H. (2018). The effect of education on health and mortality: A review of experimental and quasi-experimental evidence. In *Oxford Research Encyclopedia of Economics and Finance* (pp. 1–96). Oxford University Press.
- Galton, F. (1875). The history of twins, as a criterion of the relative powers of nature and nurture. *Fraser's Magazine*, 12, 566–576.
- Goldman, D. P. and Smith, J. P. (2002). Can patient self-management help explain the SES health gradient? *Proceedings of the National Academy of Sciences*, 99, 10929–10934.
- Haak, W. (1964). Education and training of chemists and technicians in the Netherlands. *Journal of the Royal Institute of Chemistry*, 88, 310–314.
- Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the national Academy of Sciences*, 104, 13250–13255.

- Heckman, J. J., Humphries, J. E. and Veramendi, G. (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy*, 126, S197–S246.
- Heckman, J. J., Humphries, J. E., Veramendi, G. and Urzua, S. S. (2014). Education, health and wages. *NBER Working Paper No. 19971*.
- Imbens, G. W. and Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, 98, 701–712.
- Jæger, M. M. and Blaabæk, E. H. (2020). Inequality in learning opportunities during COVID-19: Evidence from library takeout. *Research in Social Stratification and Mobility*, 68, 100524.
- Kermack, W. O., McKendrick, A. G. and McKinlay, P. L. (1934). Death-rates in Great Britain and Sweden. Some general regularities and their significance. *The Lancet*, 698–703.
- Kleefstra, N., Landman, G. W., Van Hateren, K. J., Meulepas, M., Romeijnders, A., Rutten, G. E., ... Bilo, H. J. (2016). Dutch diabetes prevalence estimates (DUDE-1). *Journal of Diabetes*, 8, 863–865.
- Lauwerys, J. A. and Scanlon, D. G. (2013). *World Yearbook of Education 1969: Examinations*. Routledge.
- Lean, M. E., Leslie, W. S., Barnes, A. C., Brosnahan, N., Thom, G., McCombie, L., ... Taylor, R. (2018). Primary care-led weight management for remission of type 2 diabetes (DiRECT): An open-label, cluster-randomised trial. *The Lancet*, 391, 541–551.
- Le Gall, F. (2014). Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation* (pp. 296–303). New York ACM Press.
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.
- Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies*, 72, 189–221.
- Lochner, L. (2011). Non-production benefits of education: Crime, health, and good citizenship. In *Handbook of the Economics of Education* (Vol. 4, pp. 183–282). Elsevier.

- Lopera Maya, E. A., van der Graaf, A., Lanting, P., van der Geest, M., Fu, J., Swertz, M., ... Sanna, S. (2020). Lack of association between genetic variants at ACE2 and TMPRSS2 genes involved in SARS-CoV-2 infection and human quantitative phenotypes. *Frontiers in Genetics*, 11, 613.
- Lumey, L. H., Stein, A. D. and Susser, E. (2011). Prenatal famine and adult health. *Annual Review of Public Health*, 32, 237–262.
- Lundborg, P. (2013). The health returns to schooling: What can we learn from twins? *Journal of Population Economics*, 26, 673–701.
- Mazumder, B. (2008). Does education improve health: A reexamination of the evidence from compulsory schooling laws. *Federal Reserve Bank of Chicago Economic Perspectives*, 33, 2–16.
- McCartney, G., Collins, C. and Mackenzie, M. (2013). What (or who) causes health inequalities: Theories, evidence and implications? *Health Policy*, 113, 221–227.
- Meghir, C., Palme, M. and Simeonova, E. (2018). Education and mortality: Evidence from a social experiment. *American Economic Journal: Applied Economics*, 10, 234–256.
- Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics*, 32, 385–397.
- Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro unit. *Review of Economics and Statistics*, 72, 334–338.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46, 69–85.
- Pearson, K. (1913). *Nature and Nurture, the Problem of the Future*. Dulau.
- Peters, M., Huisman, E., Schoonen, M. and Wolffenbuttel, B. (2017). The current total economic burden of diabetes mellitus in the Netherlands. *The Netherlands Journal of Medicine*, 75, 281–297.
- Portha, B., Chavey, A. and Movassat, J. (2011). Early-life origins of type 2 diabetes: fetal programming of the beta-cell mass. *Experimental Diabetes Research*, 2011, 105076.

- Preis, H., Mahaffey, B., Heiselman, C. and Lobel, M. (2020). Pandemic-related pregnancy stress and anxiety among women pregnant during the coronavirus disease 2019 pandemic. *American Journal of Obstetrics and Gynecology* $\diamond$ *Maternal-Fetal Medicine*, 2.
- Prick, L. G. M. (2006). *Drammen, Dreigen, Draaien: Hoe het Onderwijs Twintig Jaar Lang Vernieuwd Werd*. Mets and Schilt.
- Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3, 135–146.
- Saccone, G., Florio, A., Aiello, F., Venturella, R., De Angelis, M. C., Locci, M., ... Sardo, A. D. S. (2020). Psychological impact of coronavirus disease 2019 in pregnant women. *American Journal of Obstetrics and Gynecology*, 223, 293–295.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110–114.
- Smith, A. (1776). *The Wealth of Nations*. Penguin Books.
- Smits, J.-P., de Jong, H. and van Ark, B. (1999). Three phases of Dutch economic growth and technological change, 1815–1997. *Mimeo University of Groningen*.
- Sotomayor, O. (2013). Fetal and infant origins of diabetes and ill health: Evidence from Puerto Rico's 1928 and 1932 hurricanes. *Economics and Human Biology*, 11, 281–293.
- Taylor, R. (2019). Calorie restriction for long-term remission of type 2 diabetes. *Clinical Medicine*, 19, 37–42.
- van den Berg, G. J., Doblhammer-Reiter, G. and Christensen, K. (2009). Exogenous determinants of early-life conditions, and mortality later in life. *Social Science and Medicine*, 68, 1591–1598.
- van den Berg, G. J., Doblhammer-Reiter, G. and Christensen, K. (2011). Being born under adverse economic conditions leads to a higher cardiovascular mortality rate later in life: Evidence based on individuals born at different stages of the business cycle. *Demography*, 48, 507–530.
- van den Berg, G. J., Lindeboom, M. and Portrait, F. (2006). Economic conditions early in life and individual mortality. *American Economic Review*, 96, 290–302.

- van den Berg, G. J. and Modin, B. (2013). Economic conditions at birth, birth weight, ability, and the causal path to cardiovascular mortality. *IZA Discussion Paper No. 7605*.
- van den Berg, G. J., Paul, A. and Reinhold, S. (2020). Economic conditions and the health of newborns: Evidence from comprehensive register data. *Labour Economics*, 63, 101795.
- Van Duijn, J. (1975). The cyclical sensitivity to unemployment of Dutch provinces, 1950–1972. *Regional Science and Urban Economics*, 5, 107–132.
- van Lutsenburg Maas, J. (1964). The “Mammoth Law” reforms in Dutch education. *Comparative Education Review*, 7, 279–285.
- van Ooijen, R., Alessie, R. and Knoef, M. (2015). Health status over the life cycle. *Netspar Discussion Paper No. 10/2015-062*.
- van Zanden, J. L. (2005). *The Economic History of the Netherlands 1914-1995: A Small Open Economy in the “Long” Twentieth Century*. Routledge.
- Whicher, C., O’Neill, S. and Holt, R. G. (2020). Diabetes in the UK: 2019. *Diabetic Medicine*, 37, 242–247.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Hoofdstuk 5

Samenvatting

Binnen het vakgebied economie wordt momenteel veel onderzoek gedaan naar genetische aanleg en opvoeding, en het geïntegreerde effect van beide op de gezondheid gedurende de hele levensloop. In de huidige literatuur wordt nog steeds vooral onderzoek gedaan naar de factoren die na eeuwen onderzoek naar voren zijn gekomen, te weten opvoeding, vaardigheden en de omstandigheden in utero. Dit proefschrift maakt gebruik van observatiegegevens om al deze factoren en het effect ervan op de objectieve gezondheid op volwassen leeftijd te analyseren.

In hoofdstuk 2 onderzoeken we het verband tussen opvoeding, intelligentie en gebruik van medicatie voor specifieke aandoeningen op volwassen leeftijd zoals voorgeschreven door medische professionals. Hoewel uit verschillende studies al is gebleken dat er een positief verband bestaat tussen opvoeding en gezondheid, is het nog steeds onduidelijk of een deel van dit verband causaal is. De reden hiervoor is dat verschillende waargenomen en niet waargenomen, voor vertekening zorgende factoren, zoals achtergrond van de ouders en intelligentie, een belangrijke rol spelen in het vormen van dit verband doordat ze van invloed zijn op zowel de gekozen opleiding als de gezondheid.

De basisdataset die we gebruiken omvat alle Nederlandse mannen die tussen 1944 en 1947 zijn geboren en die nog in leven waren op 18-jarige leeftijd, het moment waarop ze de verplichte medische keuring voor de dienstplicht ondergingen (ongeveer 400.000 mannen). Naast een reeks van 3 verschillende IQ-testen, die worden gebruikt om latente intelligentie te identificeren, bevat deze dataset informatie over het opleidingsniveau en andere sociaal-economische en demografische gege-

vens. Een steekproef van deze gegevens is gekoppeld aan registraties van voorgeschreven medicijnen die in 2006 zijn verzameld, toen de personen tussen de 59 en 62 jaar oud waren. Het betrof onder andere medicatie voor hartaandoeningen, obstructieve luchtwegaandoeningen, hypertensie, hyperlipidemie, diabetes (type 1 en 2) en depressie en angst.

Het structurele model dat we calculeren, analyseert het verband tussen opleiding en voorgeschreven medicatie door de gekozen opleidingen en het behalen van de volwassen leeftijd te modelleren, waarbij rekening wordt gehouden met een aantal geobserveerde kenmerken en latente intelligentie. De methodologie die we gebruiken, biedt de mogelijkheid om de invloed van opleiding op de voorgeschreven medicatie te calculeren in verhouding tot het niveau van latente intelligentie. Gezien de gegevens die we gebruiken, presenteren we een aantal motivaties die bij benadering een causale interpretatie van de belangrijkste resultaten zouden kunnen rechtvaardigen. We constateren dat de onuitgewerkte verbanden in de gegevens niet significant afwijken van de gecalculeerde effecten. Dit resultaat suggereert dat opleiding een grote invloed heeft op voorgeschreven medicatie voor fysieke aandoeningen, en dan met name bij een laag intelligentieniveau, maar geen of vrijwel geen invloed op voorgeschreven medicatie voor depressie en angst.

Vanuit het oogpunt van beleid impliceren onze resultaten dat een hoog opleidingsniveau moet worden gezien als vervanging voor intelligentie. Een hoog opleidingsniveau, vooral bij personen met lage intelligentie, zou bijdragen aan het verminderen van medicijngebruik en, indirect, aan het verbeteren van de gezondheid op volwassen leeftijd.

In hoofdstuk 3 gebruiken we gegevens over broers en zussen om te calculeren in hoeverre macro-economische invloeden rond de geboorte een causaal effect hebben op de kans op diabetes type 2 op volwassen leeftijd. De macro-economische omstandigheden die we gebruiken, bestaan uit de provinciale werkloosheidscijfers in Nederland ten tijde van de geboorte. Deze cijfers zijn in principe exogeen aan het individuele principe. Selectieve beslissingen van de ouders op het gebied van voortplanting kunnen de calculatie echter nog steeds beïnvloeden, omdat ouders met een lage sociaal-economische status de conceptie mogelijk uitstellen tot een moment waarop de macro-economische omstandigheden gunstiger zijn. Dit impliceert dat de werkloosheidscijfers ten tijde van de geboorte niet willekeurig verdeeld

zijn over het spectrum van sociaal-economische status. We houden rekening met deze bron van endogeniteit door de gegevens van broers en zussen in onze dataset te identificeren en te gebruiken om fixed-effect regressies voor broers en zussen te calculeren.

De gegevens die we gebruiken zijn afkomstig uit Lifelines, een cohort en biobank met gegevens van 51.270 broers en zussen die na 1950 in Nederland zijn geboren. We gebruiken de gegevens van broers en zussen die zijn geboren in periodes waarin er ruime sociale zekerheid was, omdat deze periodes representatiever zijn voor de huidige pasgeborenen. Om diabetes type 2 te identificeren gebruiken we biomarkers voor de hemoglobine A1c-concentratie en de nuchtere glucosewaarde in het bloed, waarmee we door personen zelf gerapporteerde diabetes type 2, waarvan bekend is dat het grotendeels niet gediagnosticeerd is in de populatie, vermijden.

We concluderen dat een hogere werkloosheid ten tijde van de geboorte zowel bij mannen als bij vrouwen tot een grotere kans op diabetes type 2 op latere leeftijd leidt. Daarnaast tonen we aan dat de calculaties niet significant zijn als er geen rekening wordt gehouden met selectieve beslissingen met betrekking tot voortplanting en als er wordt vertrouwd op zelf gerapporteerde gegevens voor het identificeren van diabetes type 2. We concluderen dat, zelfs in periodes met ruime sociale zekerheid, slechtere macro-economische omstandigheden ten tijde van de geboorte op de lange termijn significante invloed hebben op de kans op diabetes type 2. Vanuit het oogpunt van beleid zouden vrouwen die zwanger zijn tijdens economische recessies economische en psychologische hulp moeten krijgen en ze zouden moeten worden geïnformeerd dat hun pasgeborenen als volwassenen een groter risico lopen om diabetes type 2 te krijgen. Ouders moeten deze pasgeborenen op latere leeftijd leren een evenwichtig dieet te volgen en veel te bewegen om te voorkomen dat ze diabetes type 2 krijgen. Ten slotte zou screening op diabetes type 2 gericht moeten zijn op personen die zijn geboren in slechte economische omstandigheden (bijvoorbeeld immigranten uit arme economieën).

Hoofdstuk 4 gaat over het berekenen van standaardafwijkingen in gevallen waarin sprake is van heel kleine, zeer onevenwichtige clusters. Omdat de foutwaarde meestal gecorreleerd is binnen clusters als gevolg van homogeniteit tussen observaties die tot hetzelfde cluster behoren, zijn de gecalculeerde standaardafwij-

kingen normaal gesproken te laag. Als het aantal clusters naar oneindigheid neigt, zorgen standaardcorrecties zoals de Cluster-Robust Variance Estimator (CRVE) ervoor dat de standaardafwijkingen niet worden beïnvloed door de correlatie van de foutwaarde binnen clusters. Als het aantal clusters echter klein genoeg is (lager dan 20 tot 50), dan zijn de standaardcorrecties mogelijk ontoereikend om de vertekening te corrigeren en moeten eventueel meer verfijnde methoden worden toegepast, zoals de zogenaamde CR2VE en CR3VE.

Dit hoofdstuk biedt de verfijningen van CR2VE en CR3VE, methoden die beide worden gebruikt in de empirische literatuur. Omdat deze methoden rekenintensief zijn, hebben we formules geïntroduceerd om de berekeningen zo efficiënt mogelijk uit te voeren. Daarnaast hebben we CR3VE- λ ontwikkeld, een nieuwe calculatie die rekening houdt met de grootte van het cluster en die vooral goed werkt met kleine aantallen clusters die zeer onevenwichtig zijn. Resultaten uit simulaties laten zien dat het berekenen van CR2VE en CR3VE met een dataset van 10 clusters en 5.000 observaties per cluster, drie uur minder kost. Bovendien tonen we op basis van beide simulaties en empirische illustraties aan dat CR3VE- λ mogelijk de optimale methode is voor kleine aantallen clusters die zeer onevenwichtig zijn.

Deze resultaten kunnen empirisch worden gebruikt voor toegepast onderzoek op het gebied van gezondheidseconomie waarbij gebruik wordt gemaakt van geclusterde data met een klein aantal clusters, waarbij de onderzoeker de rekentijd van veelvoorkomende correcties voor geclusterde standaardafwijkingen wil beperken. Elk onderzoek op het gebied van gezondheidseconomie dat gebruikmaakt van geclusterde data met een klein aantal clusters en lineaire modellen calculeert, kan met onze formule tijdwinst behalen in het berekenen, met name voor grote clusters. Als dit kleine aantal clusters ook nog eens zeer onevenwichtig is, zou CR3VE- λ de beste keuze kunnen zijn vanuit inferentieel oogpunt.