University of Groningen

Economics of Youth Mental Health

Freriks, Roel

DOI:
[10.33612/diss.165767790](https://doi.org/10.33612/diss.165767790)

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:
2021

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

Citation for published version (APA):
Freriks, R. (2021). *Economics of Youth Mental Health: essays on Juvenile Delinquency, Education and Policy*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen, SOM research school.
<https://doi.org/10.33612/diss.165767790>

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.

Economics of Youth Mental Health

Essays on Juvenile Delinquency, Education and Policy

Roel Donald Freriks

Publisher: University of Groningen, Groningen, The Netherlands

Printed by: Ipskamp Printing
P.O. Box 333
7500 AH Enschede
The Netherlands

© 2021 R.D. Freriks

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, electronical, mechanical, now known or hereafter invented, including photo-copying or recording, without prior written permission of the publisher.

university of
 groningen

Economics of Youth Mental Health

Essays on Juvenile Delinquency, Education and Policy

PhD thesis

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

This thesis will be defended in public on
Monday 12 April 2021 at 14.30 hours

by

Roel Donald Freriks

born on 6 June 1992
in Doetinchem

Promotores

Prof. E. Buskens

Prof. J.O. Mierau

Assessment Committee

Prof. M.R.J. Knapp

Prof. G.J. van den Berg

Prof. B.J. van den Hoofdakker

Thesis in 10 (+2) Core Aspects!

J I N C O P S U O M G E M Q N U Z N
Q W O U Y A P F X M C W O U P G N H
T X N O M R R C N N R X I O A M H T
M B U O U B W X E P O G I F C R E L
E T H K C S P I L O W I P X A B A A
H T S Q A G C G I L E I I W D W L E
O I T Z D S F E R I N W Q O E Y T H
C N A E V E F D W C A F K R M B H L
O L T X Z D Y M F Y C I R C I N E A
C T I P X U S I E U E R O M A K C T
N G S E C C P J K V T I G H I B O N
U Q T R L A Y L B L P V S O W E N E
U N I I T T V N E G N I N O R G O M
C B C M K I A F H H K U Q A Y L M I
T Q S E Y O S A Y W J Y E J S P I D
U W B N C N E Q X N R S N F T P C Y
N I R T V W I R X P I B Y K K B S H
V S S V F W Z K W N G M S I Q F E G

GRONINGEN

HEALTH ECONOMICS

POLICY

ACADEMIA

MENTAL HEALTH

EDUCATION

SCIENCE

STATISTICS

YOUTH

ROWENA

EXPERIMENT

THOMAS

Acknowledgements

The journey started in Platanos, a small village on the Greek island Samos during the summer of 2015. While drinking an ice coffee with my other *brain hemisphere* (hereafter Rowena) and discussing our future we came to the conclusion that I had to quite my parttime job at the ING bank and focus more on *economics* and *health*.

After one of the most detailed, and redundant, application letters for a student-assistant (SA) position in the university's history, I had the opportunity to join Kees Ahaus in the ambition of Healthwise (> Centre for Public Health in Economics and Business) to develop a master programme in healthcare management and economics (> MSc BA Health). Still thankful for our coffees, especially remember your personal touch and societal-driven vision towards academia. No need to say who else from Healthwise supervised my SA job, my master's thesis and actually my entire mind map towards starting a PhD. Jochen Mierau, a.k.a. *Dr. Vater*, I'll be always grateful for your role in shaping my early career in academia. We've a dynamic, and effective, interplay with an humorous nod, leading us to jointly writing and publishing papers, developing a master course, international summer school, what not (to end the sum-up in Jochen style). You also let me swim outside the walls of the Groningen pool, but always with *mentoring water wings*. After the ceremony I'll return these to you knowing our waters will always be parallel.

Searching for a master's thesis subject in economics and health led to a research meeting in the Neurology department of the UMCG in the spring of 2016. They gave me the opportunity to add a small piece to their interdisciplinary stroke puzzle. Special thanks to the chief of puzzle, bonus supervisor Maarten Lahr. The master's thesis was also the first time that paths crossed with my other promotor, Erik Buskens, who has introduced me to the world of *health technology assessment* (HTA). Grateful for your guidance and mentoring advice throughout the years. I've learned a lot from your overarching, and connecting, view on many aspects in academia. Together we've also developed the first summer school of the Aletta Jacobs School of Public Health (> Data-Driven Prevention Policy in Public Health). Looking back on a great period that summer.

As *health economics* is by name already divided between two faculties, I had the privilege to collaborate with many more colleagues from both the Faculty of Economics and Business (FEB) and Medicine. It would be a shame to not start with my FEB's *academic stepsister*, roommate, conference buddy, fellow explorer in econometrics and youth mental health, and of course paranymp, Hermien. I've had the privilege to share this journey with you on the same PhD project - true economies of scale. Also many thanks to my FEB buddies Gerben and Jan-Eise for the endless hours spent together playing table football, having philosophical discussions at *our bench* outside, or walking around the faculty river while grumbling about our programming bugs.

Also thankful to all other Groningen colleagues from the 'ninth floor', EEF department, HTA unit, Accare group and Aletta team. Specials thanks to my other EEF roommates Yang and Jos, UMCG coffee mates Kaying, Steef and Jurjen, cycling coach Daan, psychiatry sounding board Annabeth and Pieter, HTA mentors Qi and Maarten Postma, Aletta's senior namesake Roel, and my econometrics helpdesk Rob. Without your support, feedback, coffees and beers, the journey wouldn't have been the same.

My promotors also encouraged me to explore the academic world outside Groningen. I would like to thank my fellow board members at the PhD Network Netherlands (PNN) and Vereniging voor Gezondheidseconomie (VGE). The PNN taught me the politics in academia and the value of a PhD thesis. The VGE showed me that health economics is more than only an academic discipline. Special thanks to the chairman, Richard van Kleef, for giving me the opportunity to focus on building a bridge between academia and industry for early career researchers. We've organized the first Dutch health economics job market in the spring of 2020. While organizing this event with no personal intentions, my academic mind made a backflip after one of the company owners present, Evgeni Dvortsin, asked me to join the management team of his health economics and market access consultancy *Asc Academics*. Still extremely happy that I've accepted this new adventure, working daily together in *super saiyan* mode to bring Asc Academics to the next level.

Let's drive 162.2 km south of Groningen to the area where the whole journey actually started in the summer of 1992, the Achterhoek, specifically Doetinchem. I've had the privilege to grow up in a family with love, safety and support. I'm extremely grateful to my parents for giving my sister and me a *youth full of opportunities*, laying the foundation for (academic) achievements later in life. Previous literature shows ambiguous peer effects on academic achievement. In my case this turned out positive, as peers turned into *lifelong friends*. Wout, Joris, Tom and Daan, journeys with you are like writing an article; at the start you think you're in total control, then you lose almost all focus, but at the end you're in a next level of happiness. The most sustainable peer effect is definitely caused by Wout. Already more than 20 years ago at the start of primary school you've been at my side. Thanks for standing there again as paranymp during my defense. Driving 11.6 km back North brings us the village of my family in law and what has become over the years my second home town, Hengelo (Gld.). Special thanks to my brothers in law Imro and Dion for always providing a place to work with a sober mind during the day and then forget all progress after a bbq with beer in the evenings. Jochen always asked for an estimation of my days of recovery before the start of a weekend *Bar Gezellig*.

Family got a new dimension with you. Rowena, my soulmate, lovely wife, and of course the mother of our babyboy Thomas, you've been the *cornerstone* of my life for already more than ten years. Without you I wouldn't reach the intrinsic motivation to get to the doctoral finish line. By listening endless days and nights to my research ideas, failures, statistical issues, health economics acronyms, ..., you were always there as my other brain hemisphere when others couldn't keep me on track. Without a doubt you've turned my life into the epic adventure it is today. Already excited for the next chapter with you!

Bella, loyal sweet dog, walking buddy, you've always kept me with *four feet* on the ground!

Thomas, amazing son, little best friend, the new steps on the *horizon* will be yours and I'll always be right behind you!

Roel Freriks

18 February 2021, Groningen

Table of contents

1.	Introduction.....	1
1.1	Economics of Youth Mental Health	1
1.2	Youth Policy Evaluation in Observational Settings.....	4
2.	The Persistence of Child and Adolescent Mental Healthcare: Results from Registry Data	6
2.1	Background.....	6
2.2	Data.....	8
2.3	Empirical Strategy	10
2.4	Results	11
2.5	Supplementary Analyses	13
2.5.1	Policy Reforms	13
2.5.2	Definitions of Care	14
2.5.3	Decomposition by Diagnoses and Distribution	14
2.6	Concluding Remarks	14
	Appendix.....	17
A2.1	Figures & Tables.....	17
A2.2	List of Abbreviations	18
3.	Cost-Effectiveness of Treatments in Children with Attention-Deficit/Hyperactivity Disorder: A Continuous-Time Markov Modelling Approach ...	19
3.1	Background.....	19
3.2	Data.....	21
3.2.1	MTA Study	21
3.2.2	Delinquency Outcome.....	23
3.3	Empirical Strategy	23
3.3.1	Simulation Model.....	23
3.3.2	Net-Monetary Benefit Framework.....	25
3.3.3	Robustness Checks.....	25
3.4	Results	26
3.4.1	Model Validation.....	26
3.4.2	Economic Evaluation	26
3.5	Supplementary Analyses	29
3.5.1	WTP Threshold	29
3.5.2	Sampling Assumptions	30
3.6	Concluding Remarks	30
	Appendix.....	32

A3.1	Figures & Tables.....	32
A3.2	List of Abbreviations	32
4.	The Effect of Monitoring Compliance with Compulsory Education on Test Scores: A Non-Parametric Bounds Analysis	33
4.1	Background.....	33
4.1.1	Compulsory Education & Child Development.....	33
4.1.2	Dutch Education System.....	34
4.1.3	Monitoring Compulsory Education	36
4.2	Data.....	37
4.2.1	COOL ⁵⁻¹⁸	37
4.2.2	Demographic Characteristics.....	37
4.2.3	Sample Selection	38
4.3	Empirical Strategy	39
4.3.1	Ordinary Least Squares	39
4.3.2	Non-Parametric Bounds	40
4.4	Results	41
4.5	Supplementary Analyses	43
4.5.1	Tightening Bounds.....	43
4.5.2	Effect Heterogeneity	43
4.6	Concluding Remarks	44
Appendix.....		45
A4.1	Figures & Tables.....	45
A4.2	List of Abbreviations	47
5.	Long-run Returns to Government Expenditures in Special Needs Education: Evidence from the Netherlands.....	48
5.1	Background.....	48
5.1.1	Government Expenditures & Education.....	48
5.1.2	Special Needs Funding.....	49
5.2	Data.....	51
5.2.1	TRAILS.....	51
5.2.2	Mental Health	51
5.2.3	Demographic Characteristics.....	52
5.2.4	Indication Process.....	52
5.3	Empirical Strategy	53
5.4	Results	54
5.4.1	Mental Health	54

5.4.2	School Indicators	57
5.4.3	Cost/Benefit-ratio	57
5.5	Supplementary Analyses	60
5.5.1	Attrition Bias	60
5.5.2	Propensity Reweighting	60
5.5.3	Broad-band Syndromes	61
5.5.4	Effect Heterogeneity	62
5.6	Concluding Remarks	62
	Appendix.....	64
A5.1	Figures & Tables.....	64
A5.2	List of Abbreviations	67
A5.3	TRAILS Data Statement	67
6.	Policy Notes	68
	Dutch Summary.....	71
	Bibliography.....	74

Chapter |1|

1. Introduction

1.1 Economics of Youth Mental Health

The World Health Organization (WHO) has categorised mental health problems as among the most disabling clinical diagnoses in the world (WHO, 2019). Mental health in ten facts:

1. Mental, neurological and substance use disorders make up 10% of the global burden of disease and 30% of non-fatal disease burden.
2. Around 1 in 5 of the world's children and adolescents have a mental disorder.
3. Depression is one of the leading causes of disability, affecting 264 million people.
4. About half of mental disorders begin before the age of 14.
5. About 800,000 people die by suicide every year; 1 person dies from suicide every 40 seconds. Suicide is the second leading cause of death in individuals aged 15-29 years.
6. Around 1 in 9 people in settings affected by conflict have a moderate or severe mental disorder.
7. People with severe mental disorders die 10 to 20 years earlier than the general population.
8. Rates of mental health workers vary from below 2 per 100,000 population in low-income countries to over 70 per 100,000 in high-income countries.
9. Less than half of the 139 countries that have mental health policies and plans report having these aligned with human rights conventions.
10. The global economy loses about US\$ 1 trillion per year in productivity due to depression and anxiety.

Figure 1 illustrates the latest estimates of mental health disorder prevalence and the associated disease burden, produced by the Institute for Health Metrics and Evaluation (IHME) and reported in their Global Burden of Disease (GBD) study (2017) in accordance with WHO's International Classification of Diseases (ICD-10). Worldwide, one in ten people (10.7%) in 2017 lived with a mental disorder. Using whole-population health insurance data Vektis Intelligence (2017) has demonstrated that 1.1 million people are registered in mental healthcare in the Netherlands – more than one in seven (15.4%).

Figure 1 Share of population with mental health and substance use disorders

The corresponding disease burden results in significant societal costs (OECD, 2014). The direct and indirect costs of mental health problems can exceed 4% of Gross Domestic Product (GDP) (OECD, 2014). The direct costs include healthcare expenditures caused by utilisation of hospital and long-term care. **Figure 2** illustrates that mental healthcare expenditures in the Netherlands has increased since 1998 and accounted in 2018 for 1.2% of GDP, which equalled 6.8% of total healthcare expenditures (CBS, 2019, 2020). However, the

Figure 2 Trend in mental healthcare expenditures in the Netherlands (1998-2018)

true share of healthcare expenditures associated with mental health problems is larger, as poor mental health also drives up the cost of treating other physical health problems. For example, it is more expensive to treat diabetes when the patient is also suffering from depression (Ducat *et al.*, 2014), and people with mental disorders are more likely to also suffer from cancer and cardiovascular diseases (Nakash *et al.*, 2014; Fenton & Stover, 2006).

Mental health problems have their origin in youth (Kessler *et al.*, 2007). Life events often act as trigger of the onset, such as moving home or school or the birth of a new sibling. This manifests itself in the primary school age period in deviant externalizing behaviour, which cause impairment or interference in life functioning (Bayer *et al.*, 2011). Norm-deviated externalizing behaviour frequently leads to the diagnosis Attention-Deficit/Hyperactivity Disorder (ADHD) (Faraone *et al.*, 2015). Adolescents experience emotional turmoil as their minds and bodies develop. These feelings frequently internalize, potentially leading to mental disorders such as depression (Bayer *et al.*, 2011). Despite the relevance of mental health as a leading cause of health-related disability in youth, 70% of children and adolescents who experience a mental health problem have not had appropriate interventions at the time they needed it (Kieling *et al.*, 2011; Aydin *et al.*, 2020).

The burden of mental health problems in youth yields substantial indirect costs, as mental health is a pivotal element of human capital formation (Becker, 1964).

“Human capital is the stock of habits, knowledge, social and personality attributes embodied in the ability to perform labour so as to produce economic value”.

Becker (1964)

For example, Currie and Stabile (2006) examined US and Canadian children with symptoms of ADHD and they found large negative effects on test scores and schooling attainment. De Zeeuw *et al.* (2017) found a similar pattern between ADHD and educational achievement using data from the Dutch Twin Register. Furthermore, OECD data suggest that mental ill-health reduces employment prospects, productivity and wages (OECD, 2014). Previous research demonstrated that a one-standard-deviation decline in mental health score reduces employment by 30 percentage points (Frijters *et al.*, 2014). People with mental health problems are also absent from work more often, and suffer more from reduced productivity at work. Additionally, indirect costs also include informal care provided by family members, the cost of increased homelessness and juvenile delinquency (OECD, 2014). Consequently,

the gains over the life-cycle of efficient youth policies and investments in mental health are substantial, especially from the perspective of *the Heckman Equation*¹:

- + **Invest** – investment in educational and developmental resources for children with youth mental health problems provides equal access to successful early human development
- + **Develop** – nurture early development of cognitive and social skills in children from birth to age five
- + **Sustain** – sustain early development with effective education through to adulthood
- = **Gain** – gain a more capable, productive and valuable workforce that pays tax and dividends for generations to come.

Hence, early identification and treatment reduce societal costs and optimise the *Economics of Youth Mental Health*, boosting children’s human capital development.

1.2 Youth Policy Evaluation in Observational Settings

The (former) gold standard for drawing inferences about the effect of interventions in youth mental health is conducting a Randomized Controlled Trial (RCT). Participants in RCTs are randomly assigned to a treatment or control group. That is, treatment allocation is unconditional on (un)observed heterogeneity, enabling differences in efficacy found by statistical comparisons to be attributed to the intervention. Besides the argument that the unobservability of effort compromises external validity of treatment effects in RCTs (Chassang *et al.*, 2012), in many cases RCTs in youth mental health remain difficult or impossible to implement, for financial, political, or ethical reasons, or because the population of interest is too small. For example, it would be unethical to prevent pupils from attending school in order to study the causal effect of education on child development, and politically unthinkable to study the effect of youth care by randomly withholding financial budgets to selected municipalities. Moreover, governmental policies affecting youth mental health are often implemented nationally, challenging the identification of the so-called *counterfactual* – the course of the outcome variable of interest in the treatment group if the policy had not been implemented. Consequently, a large share of research about interventions and policy questions relies on observational data – data where treatment allocation was determined in a way other than through controlled random assignment.

¹ Founded in 2007, the Heckman Equation project is a strategic communications programme designed to support dissemination and amplification of Professor James Heckman’s research in early childhood development. The project offers online resources to understand the great gains to be had by investing in the early and equal development of human potential. Learn more about the Heckman Equation at heckmanequation.org.

In economics, researchers use a wide variety of strategies for attempting to draw inference from observational data (Athey & Imbens, 2017). These strategies are often referred to as identification strategies or empirical strategies (Angrist & Krueger, 1999). The goal is to use statistical techniques to create a *natural experiment* that resembles an experiment in an observational setting. For example, Ravesteijn *et al.* (2017) examined the effect of higher patient cost sharing on mental healthcare use and downstream effects, a policy question that is unethical to assess in a RCT. They compared changes in mental healthcare use by adults, who experienced an increase in cost sharing, with changes in youths, who did not experience the increase and thus formed the control group. The findings suggested that though the cost-sharing reform created overall net savings, for seriously ill patients it may lead to substantial downstream costs in involuntary admissions and acute mental healthcare.

In this thesis we used different strategies to draw inference from both experimental and observational data. Specifically, in **Chapter 2** we used registry data to estimate the persistence in youth mental health trajectories. In **Chapter 3** we employed experimental data to examine the effect of the leading forms of ADHD treatment on adolescent's juvenile delinquency. In **Chapter 4** administrative data were used to assess the effect of a compulsory schooling reform on pupil's test scores, and, finally, in **Chapter 5** we used survey data to disentangle the effect of government expenditures in special needs education on children's mental health.

Chapter |2|

2. The Persistence of Child and Adolescent Mental Healthcare: Results from Registry Data¹

Abstract

Previous studies on the persistence of child and adolescent mental healthcare do not consider the role of time-invariant individual characteristics. Estimating persistence of healthcare using standard linear models yields biased estimates due to unobserved heterogeneity and the autoregressive structure of the model. This study provides estimates of the persistence of child and adolescent mental healthcare taking these statistical issues into account. We use registry data of more than 80,000 Dutch children and adolescents between 2000 and 2012 from the Psychiatric Case Registry Northern Netherlands (PCR-NN). In order to account for autocorrelation due to the presence of a lagged dependent variable and to distinguish between persistence caused by time-invariant individual characteristics and a direct care effect we use GMM-IV estimation. All estimation results for the direct care effect (true state-dependence) showed a positive coefficient smaller than unity with a main effect of 0.215 ($p < 0.01$), which indicates that the process is stable. Persistence of care is found to be 0.065 ($p < 0.05$) higher for females. The results indicated that a substantial part of persistence is due to time-invariant individual characteristics. Additionally, we found that in the absence of further shocks a sudden increase of 10 care contacts in the present year is associated with approximately 3 additional care contacts at some point in the future. This result provides essential information about the necessity of budget increases for future years in the case of exogenous increases in healthcare use.

Keywords: Mental Healthcare, Register Data, Panel Data Models

JEL Codes: I11, C23, C26

2.1 Background

Around 20% of the working age population in OECD countries are currently suffering from a mental disorder and the lifetime prevalence is even twice as high (OECD, 2012). These disorders often originate from childhood (Knapp *et al.*, 2016; Kessler *et al.*, 2007) and have

¹ Joint work with Rob Alessie, Hermien Dijk, and Jochen Mierau. The data were retrieved from the Psychiatric Case Registry-Northern Netherlands (PCR-NN), which is maintained by the Rob Giel Onderzoekscentrum (RGOc). I thank Ellen Visser for managing and providing the data.

This study benefitted from comments received at the Boston 2017 Congress of the international Health Economics Association (iHEA), the 2017 European Network for Mental Health Service Evaluation (ENMESH), the 20th Annual European Congress of the International Society For Pharmacoeconomics and Outcomes Research (ISPOR), and at the CPB Netherlands Bureau for Economic Policy Analysis. I want to thank everyone present for their comments and feedback. I thank Erik Buskens, Maarten Postma, Minke Remmerswaal, Talitha Veenstra and Tom Wansbeek for helpful comments.

long-lasting effects throughout the lifespan due to worse health and educational outcomes (Currie & Stabile, 2006; Currie, 2009; Johnston *et al.*, 2014).

Since mental health problems appear to be highly persistent (Knapp *et al.*, 2016), it is important to understand whether child and adolescent mental healthcare is also persistent. If, in a certain year, there is an increase in the amount of mental healthcare required, knowledge on the persistence of that care provides information about the necessity of budget increases for future years. Consequently, understanding the persistence of care is also an important component of cost-effectiveness research, as it allows for a more accurate prediction of child and adolescent mental healthcare costs.

In addition, knowledge on the nature of the persistence of care in children and adolescents provides insights about the effectiveness of budget increases to reduce future healthcare use. If the persistence of care is largely the result of children's time-invariant underlying characteristics (*spurious state-dependence*), such as genetic predisposition (Lee *et al.*, 2013), children currently in care are likely to receive care for many years to come, which, assuming the reception of care is strongly related to mental health states, suggests that care is mostly targeted at alleviating and managing symptoms but that it does not lead to full remission (Heckman, 1991). In that case, broad budget increases in mental healthcare are unlikely to yield future reductions in required care, unless they alter the nature of the care provided. If the role of individual time-invariant characteristics is small, either mental health problems in themselves dissipate over time, care appears to have long-term effects, or the mechanism at work consists of a combination of both. We will refer to persistence that is not caused by time-invariant individual characteristics as the direct care effect of persistence (*true state-dependence*).

Only few studies have focused on the persistence of child and adolescent mental healthcare. Farmer *et al.* (1999) and Shenkman *et al.* (2007) find presence of persistence in child (mental) healthcare in the US, but do not consider the role of time-invariant individual characteristics in this persistence. Several studies on the persistence of child and adolescent mental health problems found that most of the persistence in childhood and adolescence is likely to be due to time-invariant individual characteristics (Contoyannis & Li, 2007; Roy & Schurer, 2013; Wichstrøm *et al.*, 2017).

One might assume that this time-invariant persistence in mental health translates to time-invariant persistence of mental healthcare. However, not all individuals with mental health problems will automatically receive mental healthcare (Kieling *et al.*, 2011). Additionally, studies on the persistence of all healthcare expenditures of elderly US citizens generally find that for these individuals, time-invariant individual characteristics appear to

play a relatively small role in overall persistence of care (Feenberg & Skinner, 1994; French & Jones, 2004). Hence, the mechanism underlying the persistence of child and adolescent mental healthcare remains unclear.

Therefore, this study investigates the nature of the persistence of child and adolescent mental healthcare by distinguishing between persistence due to time-invariant individual characteristics and the direct care effect. We do so using Dutch registry data of psychiatric care of more than 80,000 children and adolescents in the Northern Netherlands, who received care between 2000 and 2012. The use of such a unique registry dataset results in a large representative sample of individuals in care in the Northern Netherlands. Furthermore, it circumvents reporting bias that might be present in survey self-reports of healthcare use (Drapeau *et al.*, 2011). Hence, this allows us to obtain estimates of persistence in daily practice, which enhances the generalizability of the results. Additionally, during the period of observation, three major reforms took place of which we analyse the effects.

2.2 Data

We used a unique registry dataset from the Psychiatric Case Registry Northern Netherlands (PCR-NN), which is a large longitudinal record of care contacts at the largest psychiatric institutions in the Northern Netherlands between 2000 and 2012. The PCR-NN contains year of birth, sex and diagnoses of the individuals in care, as well as entries denoting each care contact an individual received, which contained information on the date of the care contacts and the type of care.

As soon as individuals had their first appointment, or received their first diagnosis, at one of the institutions they entered the PCR-NN. Each separate appointment or diagnosis is a new entry in the dataset. An individual might not be observed in the original sample at a particular point of time for several reasons: (1) the individual did not receive secondary psychiatric care; (2) the individual did receive secondary psychiatric care, but not at a reporting institution; (3) the individual is deceased. This third possibility can be ruled out if at a later point that individual reappears in the set. Additionally, mortality in the Netherlands for the age group 5-25 was continuously below 0.03% for all years 2000-2012 (CBS, 2017). While the mortality rates for the individuals in our sample may be higher than those of the general population, they are unlikely to be so to a problematic degree as the direct mortality for mental illnesses is generally low (OECD, 2014; WHO, 2018). Furthermore, as previously mentioned, the institutions in the dataset accounted for most of the secondary psychiatric care provided in the Northern Netherlands. Consequently, it was assumed that individuals receive no secondary psychiatric care when they are not observed. With these assumptions we transformed the PCR-NN into a panel dataset with time intervals of one year.

The original sample of individuals aged 4 to 23 contains 5,975,096 observations of care contacts and diagnoses corresponding to 106,523 individuals. This sample was restricted to 5,083,812 care contacts and diagnoses from 93,786 individuals for Ordinary Least Squares (OLS) estimation, as a few of these care contacts were logged before January 2000 and estimation of persistence automatically excludes individuals with only one available time period. This data was transformed so that observations represented care contacts per year, leading to 485,072 observations from 93,786 individuals. Furthermore, identification of the direct care effect required the availability of at least three consecutive time periods per individual. Consequently, the final sample for estimation of the direct care effect contained a total of 391,286 care contacts per year from 81,525 individuals.

Descriptive statistics of both samples are provided in **Table 1**. T-tests on the mean differences revealed that the two samples do not differ on the included baseline characteristics ($p > 0.05$), which indicates that the results in this study are not biased due to choices made regarding the sample selection.

Table 1 Summary statistics

	Mean	Standard Deviation	Minimum	Maximum
OLS sample ($N = 93,786$)				
Year of birth	1992.30	5.39	1978	2007
Age	15.74	4.75	5	23
Female	0.41			
Care contacts per year per individual	8.36	36.57	0	764
Selected sample ($N = 81,525$)				
Year of birth	1992.27	5.08	1979	2006
Age	16.14	4.50	6	23
Female	0.40			
Care contacts per year per individual	7.09	34.55	0	764

2.3 Empirical Strategy

We assumed that the persistence of care can be described as

$$\text{Care}_{i,t} = \beta \cdot \text{Care}_{i,t-1} + \boldsymbol{\Omega} \cdot \mathbf{X}_{i,t} + c_i + \varepsilon_{i,t}, \quad (1)$$

where $\text{Care}_{i,t}$ and $\text{Care}_{i,t-1}$ denote the number of care contacts individual i receives in year t and $t - 1$, respectively, c_i captures time-invariant individual characteristics and $\mathbf{X}_{i,t}$ is a vector of strictly exogenous control variables containing age and year dummies, $\varepsilon_{i,t}$ denotes the error term and β is the parameter of interest, aimed to capture the direct care effect.

Equation (1) could be estimated using OLS if time-invariant individual characteristics, c_i , are left out of the model. However, this estimation would yield inconsistent estimates β and $\boldsymbol{\Omega}$ because $\text{Care}_{i,t-1}$ is correlated with c_i . To account for these time-invariant individual characteristic, we could estimate equation (1) using first differencing:

$$\Delta \text{Care}_{i,t} = \beta_1 \cdot \Delta \text{Care}_{i,t-1} + \boldsymbol{\beta}_2 \cdot \Delta \mathbf{X}_{i,t} + \Delta \varepsilon_{i,t}, \quad (2)$$

where $\Delta \text{Care}_{i,t} = \text{Care}_{i,t} - \text{Care}_{i,t-1}$, $\Delta \text{Care}_{i,t-1} = \text{Care}_{i,t-1} - \text{Care}_{i,t-2}$, $\Delta \mathbf{X}_{i,t} = \mathbf{X}_{i,t} - \mathbf{X}_{i,t-1}$, and $\Delta \varepsilon_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$.

Note that the right-hand side variable $\Delta \text{Care}_{i,t-1}$ is correlated with the error term $\Delta \varepsilon_{i,t}$ so that OLS estimation of equation (2) will yield inconsistent estimates. Additionally, first differencing introduces another source of autocorrelation since $\Delta \varepsilon_{i,t}$ and $\Delta \varepsilon_{i,t-1}$ both depend on $\varepsilon_{i,t-1}$ (Nickell, 1981). To address these problems, we followed the suggestion of Arellano and Bond (1991) and estimation equation (2) with Generalized Method of Moments with Instrumental Variables (GMM-IV) using past levels of care as instruments for $\Delta \text{Care}_{i,t-1}$. As excluded instrument, we used the first available lag of $\text{Care}_{i,t}$ that does not cause $\Delta \varepsilon_{i,t}$ to be correlated with $\varepsilon_{i,t}$ at a 10-percent significance level.² We only used a single lag to prevent problems due to too many, or weak, instruments (Roodman, 2009b).

Since prevalence rates for certain disorders can differ strongly by sex (Merikangas *et al.*, 2009), persistence might also differ by sex. To test this, we performed the estimation

² $\text{Care}_{i,t-2}$ will be a valid instrument as long as the error term $\varepsilon_{i,t}$ (cf. equation (1)) is serially uncorrelated. It then holds that $E[\text{Care}_{i,t-2} | \Delta \varepsilon_{i,t}] = 0$. Note also that $\Delta \varepsilon_{i,t}$ follows an AR(1) process ($\text{cov}(\Delta \varepsilon_{i,t}, \Delta \varepsilon_{i,t-1}) < 0$ and $\text{cov}(\Delta \varepsilon_{i,t}, \Delta \varepsilon_{i,t-k}) < 0, k \geq 2$) if $\varepsilon_{i,t}$ is serially uncorrelated. We initially ran a difference GMM-IV regression with $\text{Care}_{i,t-2}$ as excluded instrument for $\Delta \text{Care}_{i,t-1}$. We then carried out a Cumby-Huizinga test (Cumby and Huizinga, 1992) to check for the validity of the following hypothesis: $\Delta \varepsilon_{i,t}$ follows an AR(1) process. If the test results indicated that there was a higher order process AR(k), we used $\text{Care}_{i,t-(k+1)}$ as excluded instrument for $\Delta \text{Care}_{i,t-1}$. We again performed the Cumby-Huizinga test to assess whether $\Delta \varepsilon_{i,t}$ follows an AR(k) process. In case the Cumby-Huizinga test indicated another autocorrelation process AR(l), $l \neq k$, at a 10-percent significance level, the excluded instrument for $\Delta \text{Care}_{i,t-1}$ was updated following the same procedure.

separately for males and females. Additionally, we performed a number of sensitivity and robustness analyses. Firstly, we analysed how three different healthcare reforms might have changed the persistence of care over the period of observation. Specifically, we looked at differences after the Dutch healthcare reform in 2006, the introduction of Diagnostic Treatment Combinations (DTCs) in 2008, and the introduction of co-payments for individuals aged 18 plus in 2012.

We also tested whether our estimations are robust to different definitions of care. First, we re-estimated the model using cost estimates of care instead care contacts, after which we did the same using number of days per year an individual received care instead of care contacts. Additionally, we varied the time unit of measurement. Lastly, we performed separate estimations for individuals with a diagnosis of Attention-Deficit/Hyperactivity Disorders (ADHD), Pervasive Developmental Disorders (PDD), anxiety, and Episodic Mood Disorders (EMD) and any of their subtypes. All estimations were performed using STATA/SE 15.0 (STATA; <https://www.stata.com/>), the difference GMM-IV estimations were performed using the command `xtabond2` (Roodman, 2009a).

2.4 Results

Table 2 shows the estimates for β of equation (1).³ The difference GMM-IV estimate only captured the direct care effect and has a value of 0.215, which is smaller than unity, indicating a stable process. Hence, if children or adolescents experience a sudden increase in mental healthcare above a certain individual-specific base level of care in a certain year, they will receive an increased number of care contacts for the following years, but this effect will weaken over time so that eventually they will receive a base level of care again, as long as there are no further shocks. Hence, in the absence of further shocks, a sudden increase of 10 care contacts in the present year is associated with approximately a total of less than 3 additional care contacts in the future above an individual's long-term base-level.

According to the Cumby-Huizinga autocorrelation test (1992), the model in equation (2) does not suffer from a $MA(k)$ process ($p < 0.01$), however we found a second-order autocorrelation process ($p > 0.10$) (i.e., $\Delta\epsilon_{i,t}$ is correlated to $\Delta\epsilon_{i,t-1}$ and $\Delta\epsilon_{i,t-2}$), suggesting that the model from equation (1) suffers from a first-order autocorrelation process ($p < 0.01$). In other words, $\epsilon_{i,t}$ appeared to be correlated with $\epsilon_{i,t-1}$, but not with $\epsilon_{i,t-2}$. This autocorrelation process is likely the result of the inclusion of a lagged dependent variable. As a result, $Care_{i,t-3}$ is the first valid instrument.

³ The FE estimate, β_1 in equation (2), functioned as a first check, as the difference GMM-IV estimate should lie between the OLS and FE estimate (Bond, 2002). The results demonstrate that this is the case and, consequently, that the difference GMM-IV estimate is likely consistent.

Table 2 Estimates of persistence of care

Care contacts	OLS	FE	d.GMM-IV
Care contacts (-1)	0.539*** (0.0064)	0.189*** (0.0016)	0.215*** (0.0156)
Age dummies	YES	YES	YES
Year dummies	YES	YES	YES
Observations	485,072	485,072	391,286
R-squared	0.268	0.211	
Number of ID	93,786	93,786	81,525

d.GMM-IV = difference Generalized Methods of Moments with Instrumental Variables; Robust standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Cumby-Huizinga autocorrelation test results yielded p -values of 0.000 {AR(1)}; 0.000 {AR(2)}; 0.215 {AR(3)}; 0.977 {AR(4)}.

To prevent large reductions in sample size due to required availability of lags of $Care_{i,t-1}$, missing values for $Care_{i,t-3}$ in the first stage equation were replaced by the zeros. This will not bias the results (Holtz-Eakin *et al.*, 1988). Since weak instruments might become a problem, we performed an F-test to determine the joint significance of the instruments for $\Delta Care_{i,t-1}$. We found an F-statistic of 1,746.95 ($p > 0.10$) using cluster robust standard errors which indicates that $Care_{i,t-3}$ is a relevant instrument for $\Delta Care_{i,t-1}$.⁴

In addition, the OLS estimate of equation (1) of 0.539 differs substantially from the difference GMM-IV estimate, suggesting that the majority of observed persistence is associated with time-invariant individuals characteristics. In other words, to a large extent, children currently in care appear to receive care for years.⁵ If we assume that the reception of care is strongly related to children's mental health states, this finding of the large role of time-invariant individual characteristics suggests that a substantial amount of care might be targeted at alleviating and managing symptoms, but not lead to full remission.

Results of the difference GMM-IV estimation on sex differences can be found in **Table 3**. For 23 individuals, gender was unknown, hence these individuals were excluded from the estimation. Females have a higher persistence of care than males (0.247 and 0.181, respectively). Both the interaction between the gender dummy and the lagged dependent

⁴ We also extended the set of instruments by including interactions between year dummies and $Care_{i,t-1}$. The estimation results are barely affected by the inclusion of those extra instruments. Additionally, when we also include $Care_{i,t-4}$ up to $Care_{i,t-10}$ as excluded instruments, results do not change substantially: the direct care effect ranges between 0.218 ($p < 0.01$) and 0.230 ($p < 0.01$), depending on the number of lags as excluded instruments.

⁵ Since OLS estimation requires less available lags of $Care_{i,t}$, the sample differs slightly from the sample used for difference GMM-IV estimation. Consequently, we have also performed the same OLS estimation using the sample used for difference GMM-IV. This estimation resulted in a very similar coefficient of 0.522 ($p < 0.01$).

variable and the F-test for the joint significance of all other interactions with the gender dummies are statistically significant ($p < 0.05$). This suggests that both the level and persistence of care is statistically significantly higher for females than for males. This difference in persistence might be the result of difference prevalence rates across different diagnoses between males and females (Kerig *et al.*, 2012).

Table 3 GMM-IV estimation on sex differences

Care contacts	Full sample	Males	Females
Care contacts (-1)	0.181*** (0.0207)	0.181*** (0.0207)	0.247*** (0.0236)
Care contacts (-1) $\times$ female	0.065** (0.0314)		
Age dummies	YES	YES	YES
Year dummies	YES	YES	YES
Interaction terms females	YES	NO	NO
Observations	391,177	235,835	155,342
Number of ID	81,502	46,149	35,353
Robust standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.			

2.5 Supplementary Analyses

In the supplementary analyses we demonstrate that the identification strategy is robust to policy reforms and different definitions of care. Furthermore, we show that although results differ across the distribution, conclusions remain similar. Finally, we focus on decomposition by diagnosis.

2.5.1 Policy Reforms

We first assessed the effects on the persistence of care of several healthcare reforms that took place in the period 2000-2012, using structural breaks. We found that the Dutch healthcare reform of 2006 did not statistically significantly affect persistence of care ($p > 0.10$), whereas the introduction of Diagnosis Treatment Combinations (DTCs) in 2008 appeared to be associated with a weakly statistically significant increase in the persistence of care ($p > 0.10$).⁶ This increase in the persistence of care might be the result of the upcoding of DTCs: by placing patients in higher DTCs than medically required providers can obtain higher

⁶ A combined F-test for both structural breaks rejected the null hypothesis of no structural breaks with $p < 0.01$.

reimbursements (van Herwaarden *et al.*, 2020). The introduction of co-payments for individuals aged 18 plus in 2012 does not affect our results: when we performed the estimation with and without the observations from 2012, the estimates for the direct care effect did not differ ($p > 0.10$). Results can be found in **Table A4** in the Appendix.

2.5.2 Definitions of Care

Second, we tested whether our estimations are robust to different definitions of care. First, we re-estimated the model using cost estimates of care instead care contacts, after which we did the same using number of days per year an individual received care instead of care contacts. Both results were very similar to our initial estimate, indicating that our initial results are robust to different definitions of care. We also varied the time unit of measurement by re-estimating the model again with number of care contacts per quarter - instead of number of care contacts per year - as our variable of interest. The results of this estimation showed a substantially larger coefficient for the direct care effect of persistence ($p < 0.01$). This difference might arise when the reception of care is largely episodic within a year (i.e., individuals receive care for a short number of weeks, or months, and no care outside this period), but individuals receive repeated ‘care episodes’ over multiple years. Results can be found in **Table A5** in the Appendix.

2.5.3 Decomposition by Diagnoses and Distribution

We also performed separate estimations for individuals with a diagnosis of Attention-Deficit/Hyperactivity Disorder (ADHD), Pervasive Developmental Disorders (PDD), anxiety, and Episodic Mood Disorders (EMD) and any of their subtypes. We found no statistically significant differences in the direct care effect between the different diagnosis groups. When we estimated the direct care effect of persistence for the highest care users in 2000 ($p < 0.01$), we found a significantly higher estimate of the direct care effect for these individuals. This suggests that the persistence of care might depend on individuals’ initial place in the distribution of care use. Results can be found in **Table A6** in the Appendix.

2.6 Concluding Remarks

In this chapter we estimated a coefficient of the year-to-year direct care effect of persistence of Dutch psychiatric care of 0.215. In the different sensitivity analyses, this coefficient varied depending on gender, the introduction of DTCs, the duration over which care was measured. Comparison of the OLS and difference GMM-IV results indicated that a substantial part of persistence is due to time-invariant individuals characteristics. These results seem to be in line with previous studies on the persistence of child and adolescent mental health problems (Contoyannis & Li, 2007; Roy & Schurer, 2013; Wichstrøm *et al.*, 2017). For example,

Wichstrøm *et al.* (2017) found coefficients of 2-year homotypic persistence that, depending on the disorder, lie between 24% and 56% of estimates of persistence that also include persistence due to time-invariant characteristics. This study is the first that considers the distinction between persistence of mental healthcare due to time-invariant characteristics and the direct care effect, which provides important information about the nature of care for policy makers and future research.

Nevertheless, this study has some limitations, which we will discuss here. First, the PCR-NN tracks individuals across institutions in the Northern Netherlands. However, not all institutions are included in the set, and individuals might obtain care at institutions outside the Northern Netherlands or in primary care. Consequently, at some point individuals in the set might have received secondary psychiatric care at institutions outside the set. Since we assume that individuals that are not observed receive no care, the persistence of care might be underestimated. However, as previously mentioned, the PCR-NN covers most psychiatric care in the Northern-Netherlands. Consequently, this bias is likely to be small.

Additionally, while the PCR-NN contains observations on a large number of individuals between 2000-2012, it lacks information on individual characteristics aside from gender, age and diagnoses. As such, the current study is unable to investigate which time-invariant characteristics in particular are responsible for the persistence of care not explained by the direct care effect. Hence, this is a topic for further research. Literature showing a strong correlation between socioeconomic status and certain mental health problems (Goodman *et al.*, 2003), as well as the probability of receiving mental healthcare (Daley, 2004; Mandell *et al.*, 2009), suggests that there might be a link between socioeconomic status and time-invariant persistence of care.

Furthermore, in this study we performed a number of robustness and sensitivity analyses. It should be noted that the multiplicity problem might arise: the more analyses there are performed, the higher the probability that one or more of the results are generated by random chance. Additionally, our estimates on the persistence of care should not be conflated with the necessity for care. There might be large groups of individuals with mental health problems who have never been in care and thus not represented in our sample (Kieling *et al.*, 2011). Hence, budgeting decisions based on our estimates should take factors in the accessibility of care into account, especially since individuals who might require care but are somehow unable to access it might be among the most vulnerable among society.

Concerning the identification strategy, instead of using the levels in the first-stage and first differences in the second-stage equation, we could have used lagged changes in $Care_{i,t}$ as additional moment conditions in the first-stage equation and levels in the second stage to

increase efficiency, as suggested by Arellano and Bover (1995) and Blundell and Bond (1998). This estimator is only consistent if persistence of care follows a random walk, i.e. if individuals are at, or close to a steady state (Roodman, 2009a). However, since we focus on young individuals it seems unlikely that this stationarity assumption holds, since the median age of onset of mental disorders is around the age of 14 (Kessler *et al.*, 2007). Hence, most individuals in our sample will not have reached a steady state yet. As a result, system GMM-IV would not be sensible as one of its main assumptions is violated in our sample.

Furthermore, Hsiao *et al.*, (2002) introduced a Maximum Likelihood (ML) approach which appears to dominate the GMM-IV approach in terms of efficiency of the estimator (Hsiao *et al.*, 2002; Moral-Benito *et al.*, 2019). Wichstrøm *et al.* (2017) were the first to apply this estimator in child and adolescent psychiatry. However, this ML estimator is only more efficient if there are few time-series observations per individual (i.e. small T), the cross-section dimension of the data is small (i.e. small N) or when time-invariant explanatory variables should be included in the second stage (Moral-Benito *et al.*, 2019). Hence, applying current ML estimators would only result in very minor efficiency gains due to our large dataset and statistically significant results.

The contribution of the results with respect to the purchasing process in Dutch psychiatric care is twofold. First, we show that an sudden increase in youth mental healthcare leads to a diminishing increase in care for the subsequent years. This information leads to more accurate budget allocation decisions in mental healthcare purchasing. Second, persistence in youth mental healthcare trajectories is partially unalterable with broad budget increases, as a substantial part of the persistence is associated with individual time-invariant characteristics. Hence, this finding suggests that personalised investments in youth mental health are required to ensure maximum care efficiency.

Appendix

A2.1 Figures & Tables

Table A4 The 2006, 2008 and 2012 healthcare reforms

Care contacts	2006	2008	2012
Care contacts (-1)	0.183*** (0.0280)	0.186*** (0.0215)	0.201*** (0.0170)
Care contacts (-1) × reform	0.047 (0.0366)	0.064* (0.0374)	
Age dummies	YES	YES	YES
Year dummies	YES	YES	YES
Structural breaks	YES	YES	NO
Observations	391,286	391,286	332,907
Number of ID	81,525	81,525	74,259

Robust standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A5 Different definitions of care

Care	Number of care days	Cost analysis	Care contacts per quarter
Care (-1)	0.224*** (0.0147)	0.231*** (0.0180)	0.627*** (0.006)
Age dummies	YES	YES	YES
Year dummies	YES	YES	YES
Observations	391,286	391,286	2,009,510
Number of ID	81,525	81,525	100,515

Robust standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A6 Diagnosis groups and highest care-users

Care	ADHD	Anxiety	EMD	PDD	Upper tail
Care (-1)	0.181*** (0.0282)	0.183*** (0.0334)	0.220*** (0.0438)	0.182*** (0.0255)	0.364*** (0.0966)
Age dummies	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES
Observations	100,609	43,919	17,951	82,783	2,113
Number of ID	19,666	10,175	4,311	14,870	354
Robust standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.					

A2.2 List of Abbreviations

AR	=	Autoregressive
ADHD	=	Attention-Deficit/Hyperactivity Disorder
DTC	=	Diagnosis Treatment Combination
EMD	=	Episodic Mood Disorders
GMM-IV	=	Generalized Method of Moments with Instrumental Variables
PCR-NN	=	Psychiatric Case Registry Northern Netherlands
PDD	=	Pervasive Developmental Disorders
OLS	=	Ordinary Least Squares

Chapter |3|

3. Cost-Effectiveness of Treatments in Children with Attention-Deficit/Hyperactivity Disorder: A Continuous-Time Markov Modelling Approach¹

Abstract

This study aimed to assess the cost-effectiveness of treatments for Attention-Deficit/Hyperactivity Disorder (ADHD) in children through prevention of serious delinquent behaviour. Cost-effectiveness was assessed in Net-Monetary Benefit (NMB). To evaluate the three major forms of ADHD treatment (medication management, behavioural treatment, and the combination thereof) relative to community-delivered treatment (control condition), we used data from 448 children, aged 7 to 10, who participated in the NIMH's Multimodal Treatment Study of Children with ADHD (MTA study). We developed a three-state continuous-time Markov model (no delinquency, minor to moderate delinquency, serious delinquency) to extrapolate the results ten years beyond the 14-month trial period at a 3% discount rate. Serious delinquency was considered an absorbing state to enable assessment in Life-Years (LYs) of serious delinquent behaviour prevented. The Willingness-To-Pay (WTP) threshold was set equal to the annual cost associated with serious delinquency in children with ADHD of \$12,370. The economic evaluation revealed a NMB of \$95,449, €88,553, \$90,536 and \$98,660 for medication management, behavioural treatment, combined treatment, and routine community care, respectively. Estimates remained stable after linearly increasing the WTP threshold between \$0 and \$50,000 in the deterministic sensitivity analyses. Treatment evaluation in broader societal outcomes is essential for policy makers, as the controlled treatments turned out to be inferior in cost-effectiveness to the control condition.

Keywords: Mental Healthcare, Statistical Simulation, Cost-Benefit Analysis

JEL Codes: I11, C15, D61

3.1 Background

Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder characterised by developmentally inappropriate and impairing inattention, hyperactivity, and impulsivity, mostly seen and diagnosed in children and adolescents (Faraone *et al.*, 2015; Feldman & Reiff, 2014). According to the meta-analysis of Polanczyk *et al.* (2015) of 41

¹ Joint work with Jochen Mierau, Jurjen van der Schans, Annabeth Groenman, Pieter Hoekstra, Maarten Postma, Erik Buskens and Qi Cao. Data was retrieved from the Multimodal Treatment of Attention-Deficit/Hyperactivity Disorder (MTA) study, regulated by the National Institute of Mental Health (NIMH) in the United States.

This chapter is based on Freriks, R.D., Mierau, J.O., van der Schans, J., Groenman, A.P., Hoekstra, P.J., Postma, M.J., Buskens, E., & Cao, Q. (2019). Cost-Effectiveness of Treatments in Children with Attention-Deficit/Hyperactivity Disorder: A Continuous-Time Markov Modelling Approach. *MDM Policy & Practice*, 4(2), 2381468319867629. doi:10.1177/2381468319867629

studies in 27 countries from every world region, the prevalence of ADHD in children and adolescents is estimated around 3.4% (Polanczyk *et al.*, 2015). Academic failure, poor self-esteem, troublesome peer and family relationships, substance abuse and delinquent behaviour are associated with ADHD and patients are often diagnosed with one or more co-occurring psychiatric disorders (Faraone *et al.*, 2015; Feldman & Reiff, 2014; Currie & Stabile, 2006; Currie *et al.*, 2010). The majority of children and adolescents diagnosed continue to have impairing symptoms into adulthood (Faraone *et al.*, 2015; Feldman & Reiff, 2014; Currie *et al.*, 2010). The negative impact of ADHD within and beyond the health system during childhood and the long-lasting impact into adulthood result in significant long-term personal and societal costs (Matza *et al.*, 2005; Hakkaart-van Roijen *et al.*, 2007; Doshi *et al.*, 2012; Le *et al.*, 2014).

The major treatments to mitigate the related (economic) burden of ADHD are medication management and behavioural treatment, alone or in combination (MTA Cooperative Group, 1999). Jensen *et al.* (2005) evaluated the cost-effectiveness of these treatments using data from the NIMH's Multimodal Treatment Study of Children with ADHD (MTA study) (Tambour *et al.*, 1998), in which 579 children with ADHD were assigned to 14 months of controlled medication management, behavioural treatment, the combination of medication management and behavioural treatment (also referred to as combined treatment), or routine community care (control condition) (MTA Cooperative Group, 1999, 2004). The cost-effectiveness of medication management proved superior to behavioural treatment at the end of 14-month trial (Jensen *et al.*, 2005). The combined treatment is less superior than medication management due to the considerable increase in costs associated with behavioural treatment (Jensen *et al.*, 2005). We build on this study by focusing on the cost-effectiveness beyond the trial period of the MTA study. Hence, a decision model was developed to evaluate the ten-year cost-effectiveness of the treatments of the MTA study.

Contrary to previous Markov models for ADHD treatment evaluation (Faber *et al.*, 2008; Wu *et al.*, 2012; Schawo *et al.*, 2015; van der Schans *et al.*, 2015), we proposed a different model structure. First, ADHD is a chronic condition which makes a full remission state unlikely. Second, the level of ADHD symptoms tends to be highly persistent over time (Asherson *et al.*, 2016; van Lieshout *et al.*, 2016). Thus, decision models with diseases states based on ADHD symptoms end up with extremely low or high transition probabilities, which limits the applicability of the model. Third, treatments for ADHD are mostly evaluated in terms of symptomatic outcomes, while the (economic) burden of ADHD often extends to society at large (Matza *et al.*, 2005; Hakkaart-van Roijen *et al.*, 2007; Doshi *et al.*, 2012; Le *et al.*, 2016). For example, anti-social behaviours and delinquency associated with ADHD result in significant costs for society (Swensen *et al.*, 2003, 2004; Jones & Foster, 2009).

Specifically, D’Amico *et al.* (2014) demonstrated that conduct disorders in childhood are associated with a two- to threefold increase in early adulthood costs, mainly driven by criminal acts and judicial contacts (D’Amico *et al.*, 2014). Therefore, we defined delinquency states for our decision model. Importantly, delinquency is a distinct indicator for children’s behaviour and participation in society. Also, robust correlations between delinquency and the level of ADHD symptoms are found in the literature (Barkley *et al.*, 2004; Jensen *et al.*, 2007; Satterfield *et al.*, 2007). Fourth, we followed common practice and based the extrapolation on data within the trial period of the MTA study. Subsequently, contrarily to previous studies, we used follow-up data to assess the accuracy of the model’s long-term prediction to ensure reliability of modelling estimates in future economic evaluation. Finally, previous Markov models for ADHD treatment evaluation consider discrete time periods (Faber *et al.*, 2008; Wu *et al.*, 2012; Schawo *et al.*, 2015; van der Schans *et al.*, 2015). Consequently, changes in states occur only at the beginning or end of predefined time intervals (van Rosmalen *et al.*, 2013). We have relaxed this assumption to build our model in continuous time. This relaxation was previously shown to result in more accurate estimates (Soares *et al.*, 2013).

3.2 Data

3.2.1 MTA Study

For this study we used data from the MTA study, a multi-site randomised controlled trial that was conducted in the United States (US) and was designed to evaluate the major forms of ADHD treatment (MTA Cooperative Group, 1999, 2004). Children had been randomly assigned to one of the three active treatments – medication management, behavioural treatment or the combination thereof (hereafter combined treatment) – or routine community care. Routine community care is the control condition and reflects the nature of less intensive (and less costly) community-delivered treatment. The MTA study involved 14-months of controlled treatment in 579 children with ADHD, aged 7 to 10 years, with naturalistic follow-ups for up to 16 years after the end of the trial period. Follow-up assessments were carried out during childhood (2 and 3 years after baseline), (late-)adolescence (6, 8, and 10 years after baseline), and adulthood (12, 14, and 16 years after baseline). We used the follow-up data of childhood and late-adolescence periods, since our modelled outcome variable (delinquency) was not assessed in adulthood. Summary statistics of the baseline characteristics age, gender, comorbidity, intelligence, ethnic background and occupation-based socioeconomic family status are presented in **Table 7**.

Table 7 Summary statistics

	Mean	Standard Deviation
Dropped sample (<i>N</i> = 131)		
Age	7.90	0.84
Female	0.168	
Comorbidity	0.260	
Intelligence, WISC-III IQ	102.06	13.08
Non-Caucasian background	0.374	
High occupation-based socioeconomic family status	0.298	
Selected sample (<i>N</i> = 448)		
Age	7.74	0.80
Female	0.205	
Comorbidity	0.252	
Intelligence, WISC-III IQ	100.16	15.22
Non-Caucasian background	0.392	
High occupation-based socioeconomic family status	0.317	
WISC = Wechsler Intelligence Scale for Children; IQ = Intelligence Quotient.		

Comorbidity is a dummy variable that equals 1 for the presence of anxiety and/or depression. Intelligence is the child's total Intelligence Quotient (IQ) measured with the Wechsler Intelligence Scale for Children-III (WISC-III). Ethnic background is a dummy variable that equals 1 for children from a non-Caucasian background. Finally, occupation-based socioeconomic family status is a dummy variable that equals 1 for children from a high-socioeconomic family status. Further details on the four treatment modes of the MTA study and other baseline characteristics are available in previous publications (MTA Cooperative Group, 1999, 2004; Molina *et al.*, 2007, 2009). All study procedures had been approved by institutional review boards and were carried out in accordance with the Declaration of Helsinki. Participants and parents were informed of the procedures and provided written informed consent (MTA Cooperative Group, 1999).

3.2.2 Delinquency Outcome

In this study, a six-point scale on delinquency was used as primary outcome variable (Molina *et al.*, 2009), coded ordinally from two parent-report measures, the DISC-IV-CD Module and the Parent DSM-IV Aggression and Conduct Disorder Rating Scale (American Psychiatric Association, 1994), and two self-report measures. Specifically, the Self-Reported Antisocial Behavior questionnaire (SRA) (Loeber *et al.*, 1989) at the 2-year assessment and the Self-Reported Delinquency questionnaire (SRD) (Elliott *et al.*, 1985) at the 3-year assessment. By using all available procedures participants were assigned (retrospectively) a delinquency classification code at each assessment point (Wolfgang *et al.*, 1985; Lee *et al.*, 2004; Loeber *et al.*, 1991, 1998). The coding scheme of the Pittsburgh Youth Study was used to contribute items to each code (Loeber *et al.*, 1991, 1998).

The delinquency scale was then categorised as: 0 = no delinquency; 1 = minor delinquency only at home (e.g., theft of less than \$5 or vandalism); 2 = minor delinquency outside of the home (e.g., vandalism, cheating someone, shoplifting less than \$5); 3 = moderately serious delinquency (e.g., vandalism, theft of \$5 or more, carrying a weapon); 4 = serious delinquency (e.g., breaking and entering, drug selling, attacking someone with the intent to seriously hurt or kill, rape); 5 = engagement in two or more different level 4 offences. This variable was assessed at baseline, after the 14-month trial period, and at the follow-up assessments after 2, 3, 6, and 8 years.

3.3 Empirical Strategy

3.3.1 Simulation Model

To predict the trajectories of delinquent behaviour during adolescence in relation to the four treatment modes of the MTA study, we developed a continuous-time Markov model based on three delinquency states (**Figure 3**): no delinquency (state 1); minor to moderate delinquency (state 2); and serious delinquency (state 3). The states were discerned based on the delinquency scale mentioned above, in which a 0 score was considered no delinquency, 1 to 3 scores minor to moderate delinquency, and scores 4 and 5 serious delinquency.

The specification, parameter estimation, and evaluation of this model were conducted using vertical modelling formulation (Nicolaie *et al.*, 2010; Cao *et al.*, 2016). Briefly, this includes the specification of the Markov process by means of two main parameters: (i) sojourn time distributions and (ii) the probabilities of the next state visited (hereafter future state probabilities). The treatment indicator was incorporated in the corresponding parametric survival and multinomial regression models. Finally, the modelled outcomes were evaluated using Monte-Carlo simulation of the whole procedure.

Figure 3 Schematic representation of the Markov model

To assess the cost-effectiveness of the treatments through prevention of serious delinquent behaviour, we considered serious delinquency to be an absorbing state. This assumption does not allow participants to either make transition out of this state (i.e., decrease the delinquency level after entering the serious delinquency state) or start in this state (i.e., enrolling in this study with delinquency levels 4 or 5). Consequently, within the total sample size of 579 children, our Markov model was built based on the delinquency data from 448 children. The reason for exclusion of these 131 children was that they either already had a 4 or 5 delinquency score when enrolling in the study or follow-up data was missing.

Exponential survival models were subsequently used to estimate the cumulative distribution functions of the sojourn time in both states 1 and 2 ($F_i(u)$, $i = 1, 2$). The future state probabilities (π_{ij} , $i = 1, 2$; $j = 1, 2, 3$; $i \neq j$) were estimated by fitting logistic regression models to the event indicator of entering the serious delinquency state from the no delinquency state π_{13} and from the minor to moderate delinquency state π_{23} . The transition probabilities between the no delinquency state and minor to moderate delinquency state can then be calculated using one minus the two previously mentioned two probabilities. The treatment indicator was incorporated as the covariate in both exponential survival and logistic regression models.

We developed our model based on the individual trajectory of delinquency seriousness within the trial period, as we use the ten-year follow-up data to validate accuracy of long-term prediction. The outcome of our model is the average time of not reaching the serious delinquency state, defined as the Life-Years (LYs) of serious delinquent behaviour prevented. Model performance was internally validated by comparing the predicted probability (average across 100,000 simulation runs) of serious delinquent behaviour prevented (based on a Kaplan-Meier estimate) to the observed empirical survival curves with the same outcome variable. The simulated results reflected a ten-year time horizon. The available follow-up data in the MTA study enabled the unique opportunity to internally validate the modelling prediction at ten-year follow-up for the four different treatments.

3.3.2 Net-Monetary Benefit Framework

We included treatment costs in which the following three components were taken into account: medication cost, visit cost for teachers and aides, and cost of psychiatrist, psychologist and paediatrician. Per treatment group, we converted the longitudinal costs into daily costs. The respective resulting daily costs were \$0.52 for routine community care, \$0.62 for medication management, \$3.18 for behavioural treatment and \$3.53 for the combined treatment (Jensen *et al.*, 2005). Total treatment costs were calculated by multiplying the LYs of serious delinquent behaviour prevented with the daily costs. The annual rate of discounting was set at 3% for both cost and effectiveness outcomes (Gold *et al.*, 1996).

We compared the cost-effectiveness outcomes among the treatments in terms of the Net-Monetary Benefit (NMB) framework (Tambour *et al.*, 1998). The Willingness-To-Pay (WTP) threshold was set equal to the annual cost associated with serious juvenile delinquency in children with ADHD retrieved from previous research in the US (Matza *et al.*, 2005). They assessed the individual's criminal history through self-report, including crimes, juvenile detention, probation, and jail based on information from the Bureau of Justice Statistics, the Federal Bureau of Investigation, and the Criminal Justice Institute. The mean total criminal costs were \$12,868 and \$498 for children with and without ADHD, respectively. Hence, we incorporated the adjusted difference of \$12,370 as WTP threshold – the cost avoided in the serious delinquency state. The cost-effectiveness can then easily be calculated within this framework as

$$\text{NMB} = \text{mean prevented LYs} \times \text{WTP} - \text{mean total treatment cost.} \quad (3)$$

3.3.3 Robustness Checks

First, the WTP threshold is a key parameter, hence determining the conclusions drawn from the NMB analysis. Therefore, we re-evaluated equation (3) with linearly increasing WTP

thresholds between \$0 and \$50,000. Second, we used logistic regression models to control for sample selection and the likelihood of absorbance. In the first model we controlled for the effect of sample selection by estimating Odds Ratios (ORs) of model exclusion conditional on the relevant covariates age, gender, comorbidity, intelligence, ethnic background and occupation-based socioeconomic family status. We established whether the covariates mentioned above were independent predictors for model exclusion at a five-percent significance level. Third, we focused on the likelihood of the absorbance in the serious delinquency state. We compared the adjusted OR of absorbance in the serious delinquency state with moving out of this state at a five-percent significance level for both the included and excluded children. All of above statistical analyses were performed with R 3.2.4 (R Foundation; <https://www.r-project.org/foundation/>) and STATA/SE 15.0 (STATA; <https://www.stata.com/>).

3.4 Results

3.4.1 Model Validation

The model parameters estimation results are presented in **Table 8**. As is shown in **Figure 4**, the predicted survival curves obtained from our Markov model closely resembled the observed survival curves. The only exception was the curve for children who were assigned to medication management, in which a clear overestimation of the probability of preventing serious delinquent behaviour was detected. Our model provided excellent predictions for children assigned to routine community care and the combined strategy of medication management and behavioural treatment. Specifically, the mean difference in percentage of LYs of serious delinquent behaviour prevented between the modelled and observed trajectories is 8.5% for medication management, 7.8% for behavioural treatment, 5.8% for the combined treatment and 5.7% for routine community care.

3.4.2 Economic Evaluation

32 of the 448 children (7%), who started at baseline in the no delinquency or mild to moderate delinquency state, reached the serious delinquency state within ten years' time. For policy makers this is a substantial percentage, looking at the description of the delinquency levels associated with this state and taking into account the annual cost associated with serious juvenile delinquency in children with ADHD of \$12,370 (Matza *et al.*, 2005). **Table 9** presents the cost-effectiveness results stratified by the four treatment modes considered in the MTA study. We performed both undiscounted and discounted analysis.

Table 8 Specification and parameter estimation of the Markov model

Markov model	α	$\beta_T T$
$F_1(u)$		
T1	6.89 (0.22)	
T2	6.89 (0.22)	0.08 (0.32)
T3	6.89 (0.22)	0.13 (0.32)
T4	6.89 (0.22)	-0.10 (0.30)
$F_2(u)$		
T1	7.02 (0.17)	
T2	7.02 (0.17)	-0.16 (0.22)
T3	7.02 (0.17)	-0.02 (0.24)
T4	7.02 (0.17)	-0.15 (0.22)
π_{13}		
T1	-3.13 (1.10)	
T2	-3.13 (1.10)	0.61 (1.25)
T3	-3.13 (1.10)	0.64 (1.25)
T4	-3.13 (1.10)	0.07 (1.44)
π_{23}		
T1	-3.13 (0.67)	
T2	-3.13 (0.67)	0.20 (0.49)
T3	-3.13 (0.67)	0.23 (0.50)
T4	-3.13 (0.67)	-0.33 (0.54)
$\begin{cases} F_1(u; \lambda) = F_2(u; \lambda) = 1 - \exp(-\lambda u) \\ \lambda = \exp(-\alpha - \beta_T T) \end{cases}$		$\begin{cases} \theta(\pi_{13}) = \alpha + \beta_T T + \beta_G G \\ \theta(\pi_{23}) = \alpha + \beta_T T + \beta_G G \end{cases}$
<p>Standard errors are presented in parentheses; T2, T3, T4 respectively represent routine community care, medication management, behavioral treatment and combined treatment, with routine community care as the reference category for $\beta_T T$; $\theta(\pi) = \log(\pi(1 - \pi))$ is the logit function.</p>

Figure 4 Results of the model validation for the four treatment modes: (a) routine community care, (b) medication management, (c) behavioural treatment, (d) combined treatment; straight (dashed) lines represent observed data (model prediction).

Table 9 Results economic evaluation stratified by treatment mode

Treatment mode	Meant total cost (\$)	Mean LYs prevented (years)	Mean NMB (\$)
Undiscounted			
a. Routine community care	1,769	9.32	113,519
b. Medication management	2,043	9.03	109,658
c. Behavioural treatment	10,539	9.08	101,781
d. Combined treatment	12,111	9.40	104,167
Discounted			
a. Routine community care	1,537	8.10	98,660
b. Medication management	1,779	7.86	95,449
c. Behavioural treatment	9,170	7.90	88,553
d. Combined treatment	10,527	8.17	90,536

LY = Life-Year; NMB = Net-Monetary Benefit.

In ten years' time, the discounted average LYs of serious delinquent behaviour prevented were 7.86 for medication management, 7.90 for behavioural treatment, 8.17 for the combined treatment, and 8.10 for routine community care. Although the combined treatment had the highest LYs prevented, routine community care turned out to be the optimal strategy in terms of cost-effectiveness with the highest mean NMB due to the substantial difference in treatment cost.

3.5 Supplementary Analyses

In the supplementary analyses we demonstrate that the modelling strategy is robust to the choice of WTP threshold. Furthermore, we show that the sample selection is not associated with the observed individual characteristics. Finally, we focused on the likelihood of absorbance in the serious delinquency state.

3.5.1 WTP Threshold

Figure 5 illustrates that the difference in NMB between routine community care and the two active treatments medication management and behavioural treatment increases as the WTP increases, while the difference in NMB between routine community care and the combined treatment reduces. The latter is due to the fact that the difference in treatments cost becomes, relatively, less determinative for the NMB results.

Figure 5 NMB plot with linearly increasing WTP thresholds between \$ 0 and \$50,000

3.5.2 Sampling Assumptions

Table A10 demonstrates that the results are not driven by sample selection, as none of the ORs of model exclusion conditional on the relevant covariates age, gender, comorbidity, intelligence, ethnic background or occupation-based socioeconomic family status were significant at a five-percent level.

Furthermore, we determined the likelihood of absorbance in the serious delinquency state once a child entered this state. We found for the 448 included children an adjusted OR of continuation in the serious delinquency state of 1.47 ($p < 0.01$), against an adjusted OR of moving out of -1.42 ($p < 0.01$). Similarly, we found for the 131 excluded children an adjusted OR of continuation in the serious delinquency state of 0.927 ($p < 0.01$), against an adjusted OR of moving out of -0.953 ($p < 0.01$). Hence, the serious delinquency state is more likely absorbent than transient for both groups.

3.6 Concluding Remarks

In this study we assessed the long-term cost-effectiveness of the three major forms of ADHD treatment and routine community care beyond the 14-month trial period of the MTA study (MTA Cooperative Group, 2004). For this we developed a Markov model with an innovative model structure. We considered the high-cost serious delinquency state to be absorbing to predict the LYs of serious delinquent behaviour prevented over a time period of ten years. The availability of long-term follow-up data in the MTA study enabled to assess accuracy of modelling prediction. Modelled and observed outcomes matched closely with a mean

difference of 6.9%. Hence, the model delivered reliable estimates when used in future economic evaluations. By setting the WTP equal to the cost avoided in the serious delinquency state, we calculated cost-effectiveness in terms of NMB (Tambour *et al.*, 1998). Results of the economic evaluation revealed that the combined treatment was the only active treatment mode that further decreases serious delinquent behaviour compared with routine community care. However, the substantial difference in treatment cost rendered the routine community care to be the optimal treatment strategy in terms of NMB. Moreover, results were robust to manipulating the WTP threshold in deterministic sensitivity analyses.

These findings are in line with previous research with the MTA study (Molina *et al.*, 2007, 2009). Molina *et al.* (2007, 2009) demonstrated that while after three years the active ADHD treatments had improved symptomatic symptoms, no differences were found with respect to delinquent behaviour (Molina *et al.*, 2007, 2009). Although the MTA study was not originally designed to examine delinquency, we argue that it is interesting for policy makers to see whether the pursued medical effects of the treatments translate in positive effects with respect to behaviour of these children in society. Clearly, the economic impact of ADHD often extends to society beyond the health system (Matza *et al.*, 2005; Hakkaart-van Roijen *et al.*, 2007; Doshi *et al.*, 2012; Le *et al.*, 2016) and treatment effects on ADHD symptoms and broader societal aspects of the disorder differ substantially (Molina *et al.*, 2007, 2009). Additionally, Schawo *et al.* (2015) plead for model-based studies in ADHD using empirical data for model validation and broader societal outcomes as the modelled outcome. As such, the evaluation form in this study has its potential to be extended to a broader perspective.

One of the innovative structures of the Markov model in this study is the use of a more or less transient state in real-life as absorbing state in the model. As such, it enables model validation according to the guidelines for simulating a continuous-time Markov model (van Rosmalen *et al.*, 2013). Without including an absorbing state, the simulation along with our vertical modelling approach reaches an endless process. The only reason for a process to end is when the processing time is beyond our study time horizon of ten years. Furthermore, with this model structure we were able to evaluate the treatments of the MTA study through prevention of serious delinquent behaviour. The latter is highly relevant for policy makers as serious delinquency is associated with substantial societal cost (Swensen *et al.*, 2003, 2004; Jones & Foster, 2009). A limitation of this assumption is that we had to delete 22.6% of the original sample. However, we demonstrated that this assumption did not affect the modelled outcomes, as model exclusion was not predicted by the relevant child and family characteristics age, gender, comorbidity, intelligence, ethnic background and occupation-based socioeconomic family status. Moreover, results revealed that the serious delinquency state is more likely absorbent than transient for both included and excluded children.

Appendix

A3.1 Figures & Tables

Table A10 Logistic regression results on model exclusion

Model exclusion indicator	Odds ratios	Marginal effects
Age	-0.232 (0.1237)	-0.041 (0.0214)
Female	0.266 (0.2607)	0.046 (0.0454)
Comorbidity	-0.083 (0.2301)	-0.014 (0.0401)
Intelligence, WISC-III IQ	-0.007 (0.0076)	0.0012 (0.0013)
Non-Caucasian background	0.090 (0.2109)	0.016 (0.0368)
High occupation-based socioeconomic family status	0.094 (0.2224)	0.016 (0.0388)
WISC = Wechsler Intelligence Scale for Children; IQ = Intelligence Quotient; Clustered standard errors in parentheses. Inference: */** indicate significance level of 5%/1%.		

A3.2 List of Abbreviations

ADHD	=	Attention/Deficit-Hyperactivity Disorder
IQ	=	Intelligence Quotient
LY	=	Life-Year
MTA study	=	Multimodal Treatment Study of Children with ADHD
NMB	=	Net-Monetary Benefit
OR	=	Odds Ratio
WISC	=	Wechsler Intelligence Scale for Children
WTP	=	Willingness-To-Pay

Chapter |4|

4. The Effect of Monitoring Compliance with Compulsory Education on Test Scores: A Non-Parametric Bounds Analysis¹

Abstract

This study investigated the effect of a law reform that tightened the monitoring compliance with the compulsory education act in Dutch regular primary schools. Schools are from 1 January 2012 directly responsible for pupil's relative absenteeism and receive a penalty of €100,000 if they do not comply. We used longitudinal data from 41,049 pupils in the first and last class of primary school between 2007 and 2014 to analyse differences in test scores cross cohorts. We used a non-parametric bounds analysis to estimate the effect of the law reform on pupil's math and language test scores in the presence of non-random sample selection due to selective attrition. The lower and upper bound correspond to extreme assumptions about the missing information that are consistent with the observed data. The bounds suggested that the law reform improved children's test scores in both classes on average with 0.85 and 0.73 standard deviation, respectively. Estimates were robust to tightening the bounds on selected covariates using ℓ_1 -regularized regression. Heterogeneity analyses revealed that care pupils and children from an ethnic minority or low socio-economic environment benefited more from the policy.

Keywords: Achievement Gap, Education Policy, Non-Parametric Selection

JEL Codes: C14, C52, I24

4.1 Background

4.1.1 Compulsory Education & Child Development

Previous studies have shown that education is an important determinant of child development, as education relates to a variety of outcomes in adulthood. The literature demonstrates that more education leads to a higher probability of work (Oreopoulos, 2006a, 2006b, 2007), a higher income (Bhuller *et al.*, 2017; Devereux & Hart, 2010; Leigh & Ryan, 2008), better health (Brunello *et al.*, 2016; Galama *et al.*, 2018; Grimard & Parent, 2007; Jürges *et al.*, 2011), a smaller chance of becoming ill (Kemptner *et al.*, 2011), fewer teenage pregnancies (Cygan-Rehm & Maeder, 2013; Silles, 2011) and less crime (Cullen *et al.*, 2006;

¹ The data were retrieved from the Cohort Onderzoek OnderwijsLoopbanen (COOL⁵⁻¹⁸), publicly available at <https://easy.dans.knaw.nl>.

This study benefitted from comments received at the RES Annual Conference 2020 (digital edition due to the COVID-19 outbreak). I am grateful to RES for selecting this study for a young economists' travel grant for the 2021 edition. Furthermore, I thank Jos Lubberman, Jochen Mierau and Sándor Sóvágó for helpful comments.

Deming, 2011; Amin *et al.*, 2016). Indeed, differences in child educational performance have persistent impact throughout the life-cycle. Consequently, policies to reduce educational disparities are on the top of the political agenda worldwide.

One of the most discussed educational policy measures in the economic literature has been the impact of years of schooling in the form of compulsory education. Previous research conclude that there is a positive consistent relationship between compulsory schooling and labour market outcomes (Angrist & Krueger, 1991; Harmon & Walker, 1995; Oreopoulos, 2006a, 2006b, 2007), health (Kemptner *et al.*, 2011; Fischer *et al.*, 2013; Crespo *et al.*, 2014; Avandano *et al.*, 2020), and criminal behaviour (Lochner & Moretti, 2004). However, these studies focused mainly on the impact of additional years of compulsory schooling without taking into account the compliance after implementation in terms of absenteeism.

In this study we estimated the effect of a law reform that tightened the monitoring compliance with the compulsory education act in Dutch regular primary schools. From 1 January 2012 onward schools were directly responsible for pupil's absenteeism and received a fine of €100,000 if they did not comply with the absence rules. We employed longitudinal data from 41,049 pupils in the first and last class of primary school between 2007 and 2014 to analyse differences in test scores across cohorts. Using non-parametric bounds we showed that the law reform improved the test scores in both classes on average with 0.85 and 0.73 Standard Deviation (SD), respectively. In the heterogeneity analyses we demonstrated that the average effect is 0.18, 0.09 and 0.05 SD higher for pupils with special needs, children from an ethnic minority, and a low socio-economic environment, respectively.

4.1.2 Dutch Education System

A flowchart of the Dutch education system is presented in **Figure 6**. Most children start primary school at the age of four, although they are not required by law to attend school until the age of five. After eight years primary education pupils receive an advice which type of secondary education would be the most appropriate. The advice is based on the judgment of the primary school teachers and on objective test scores. About 95% of the schools use the national test provided by Cito, the Dutch Testing and Assessment Company.² Based on this combined advice parents and pupils can apply for the secondary school they prefer and that provides the appropriate educational programme.

² We refer the reader to <https://www.cito.com/> for more information about Cito.

Figure 6 Flowchart of the Dutch education system

Secondary education consists of pre-vocational secondary education (Dutch: Voorbereidend Middelbaar BeroepsOnderwijs (VMBO)), general secondary education (Dutch: Hoger Algemeen Voortgezet Onderwijs (HAVO)) and pre-university education (Dutch: Voorbereidend Wetenschappelijk Onderwijs (VWO)). Schools are free in the composition of the classroom (homo- or heterogeneous) and in the manner they design education towards the attainment standards. Some schools in secondary education are combined schools, which makes it possible to make the first year of secondary education as common as possible in so called bridge classes. This makes it easier to switch from one type of secondary education. Combined schools also give better opportunities for switches in the upper levels of secondary education. Students continue though largely at the starting level of their lower class.

VMBO (four years) is meant as a preparation for senior secondary vocational education (Dutch: Middelbaar BeroepsOnderwijs (MBO)). HAVO is designed to prepare students for higher professional education (Dutch: Hoger BeroepsOnderwijs (HBO)). In practice, however, HAVO-leavers also go on to the upper years of VWO or MBO. Finally, VWO(six years) is designed to prepare students for university (Dutch: Wetenschappelijk Onderwijs (WO)) and in most cases pre-university education certificate-holders go on to university. Some of them though opt for HBO.

For children with disabilities, or such special needs that it is difficult to attend regular education, special education is provided. There exist special primary schools for children of primary education age (Dutch: Speciaal Onderwijs (SO)) and special secondary schools for children from the age of twelve onwards (Dutch: Voorgezet Speciaal Onderwijs (VSO)). For students in special secondary schools the so-called on the job training schools are developed. These students are stimulated to combine school with work from age sixteen on. This provision is aimed for an adequate labour market perspective of these students.

4.1.3 Monitoring Compulsory Education

School absenteeism occurs when a pupil is not present at school when presence is obligatory. Absence may be permitted after the school has given permission. If that is not the case, this is called truancy. Legally speaking, truancy at the age of compulsory education is a violation of the Dutch compulsory education act (Wettenbank, 2012).

The law distinguishes between absolute and relative absenteeism. In the case of absolute absenteeism, a child is not enrolled in a school at the age of compulsory education. Relative absenteeism means that a pupil is enrolled in a school but is not present during class or practice time. In the case of relative absenteeism, a distinction is made between 'luxury' absenteeism and 'signal' absenteeism. Luxury absenteeism means that a student goes on holiday without permission during the school period. Signal absenteeism is a result of problems in the environment of the pupil. In practice, if a pupil is absent for more than 16 hours for a period of four weeks without a valid reason, the school must report this directly to the Education Implementation Agency (Dutch: Dienst Uitvoering Onderwijs (DUO)) of the Ministry of Education, Culture and Science. DUO then reports the truant student to the municipality of the pupil's place of residence.

As of 1 January 2012 primary schools are directly responsible for relative absenteeism, as the monitoring of compliance with the compulsory education act was legally transferred from the municipality to the Dutch Inspectorate of Education (Wettenbank, 2012). Until 2012, the municipality's compulsory education officer was responsible for monitoring compliance by the school board. The pre-reform situation encountered obstacles, such as a lack of monitoring capacity and disruption of ambiguous relationships between compulsory education officers and school boards, with a potentially opposing effect on absenteeism. The law reform should lead to better compliance with the compulsory education act by the school boards. If school boards do not comply, the law reform might cause the school to incur a fine by the Inspectorate of Education up to €100,000. As a result, the law reform ostensibly caused a drop in relative absenteeism, as illustrated by **Figure 7**.

Figure 7 Relative absenteeism between 2005 and 2016 (Source: Statistics Netherlands)

4.2 Data

4.2.1 COOL⁵⁻¹⁸

This chapter used data from the Cohort Onderzoek OnderwijsLoopbanen (COOL⁵⁻¹⁸)³, a cohort study in which pupils from 5 to 18 years were followed in their cognitive, social and emotional development. Additionally, the cohort was linked on individual level to administrative data of Statistics Netherlands to better understand the pupil's socio-economic and demographic living situation. The study consisted of three waves of data corresponding to the academic years 2007/2008, 2010/2011 and 2013/2014. The pupils in wave 3 were automatically allocated to the treatment group, as the law reform has been introduced on 1 January 2012. Consequently, the pupils in waves 1 and 2 were allocated to the control group. We used data of 41,049 pupils in the first and last class in primary school.

4.2.2 Demographic Characteristics

We controlled for observed heterogeneity in demographic characteristics using information on pupil's age, gender, ethnicity, family situation, parental educational attainment, care level, urbanicity and school performance. We used a dummy variable for ethnicity with non-Caucasian background as reference category for at least one foreign-born parent. To define

³ The COOL⁵⁻¹⁸ is publicly available at <https://easy.dans.knaw.nl> and subsidised by Nationaal Regieorgaan Onderwijsonderzoek/Programmaraad voor het Onderwijsonderzoek (NRO/PROO).

the family situation, we distinguished between two- and one-parent families. Educational attainment of the parents was categorised according to three levels: score 1 - low (primary education, lower tracks of secondary education); score 2 - middle (higher tracks of secondary education, senior secondary vocational education); and score 3 - high (higher professional education, university). Using a dummy variable, we distinguished between regular and special needs pupils. Furthermore, we used a categorical variable (1-5) for urbanicity to separate children living in lower and higher urban areas. Finally, we used the official scores of the education inspectorate's annual reports to control for the overall school performance (1-7). These reports were based on the quality of education and teaching, financial administration, quality assurance and the school board ambitions.

4.2.3 Sample Selection

Children's cognitive development was measured with test scores on official Cito exams in language and arithmetic. To control for non-random sample selection in the empirical strategy, we distinguished between pupils with missing and registered test scores. We observed attrition rates of 51.0 and 26.3 percent on average for the control and treatment group, respectively. Corresponding summary statistics of pupils in the first and last class of primary school are provided in **Table 11** and **Table 12**, respectively.

Table 11 Summary statistics of control and treatment group in the first class

	Control		Treatment	
	Missing test scores <i>N</i> = 16,315	Registered test scores <i>N</i> = 12,641	Missing test scores <i>N</i> = 690	Registered test scores <i>N</i> = 8,305
<i>Pupil</i>				
Age (5-8)	6.25	6.26	6.21	6.16***
Females	0.48	0.48	0.47	0.48
Non-Caucasian background	0.48	0.38***	0.51	0.35***
Care pupil	0.24	0.23	0.28	0.23***
<i>Family background</i>				
One-parent family	0.35	0.14***	0.24	0.10***
Educational attainment mother (1-3)	1.72	1.69***	1.94	1.89
Educational attainment father (1-3)	1.80	1.81	2.05	2.05
<i>Demographic</i>				
Urbanicity (1-5)	2.85	2.83	2.76	2.68
School performance (1-7)	2.56	2.58	2.88	2.68***
Inference: */**/** indicate significant differences at the 10%/5%/1% level based on the mean differences of the dropped and selected samples, assessed with a t-test.				

The samples with registered test scores consisted of less children from an ethnic minority, less special needs students and children grow up more frequently in a two-parent family in both classes compared to the samples with missing test scores. Furthermore, we observed in the first class that the maternal educational attainment is lower in the control group of the sample with registered test scores and the school score is higher in the treatment group. Additionally, in the last class we found mixed differences in parental educational attainment between the both sample. Similar pattern for school scores. Finally, children in the sample with registered test scores of the last class consisted of more children living in higher urban areas compared to the samples with missings.

Table 12 Summary statistics of control and treatment group in the last class

	Control		Treatment	
	Missing test scores <i>N</i> = 11,850	Registered test scores <i>N</i> = 14,374	Missing test scores <i>N</i> = 4,330	Registered test scores <i>N</i> = 5,729
<i>Pupil</i>				
Age (10-14)	12.48	12.48	12.36	12.35
Females	0.50	0.50	0.50	0.51
Non-Caucasian background	0.45	0.34***	0.35	0.31***
Care pupil	0.27	0.24***	0.28	0.25***
<i>Family background</i>				
One-parent family	0.40	0.14***	0.11	0.09***
Educational attainment mother (1-3)	1.76	1.47***	1.39	1.55***
Educational attainment father (1-3)	1.86	1.58***	1.51	1.63***
<i>Demographic</i>				
Urbanicity (1-5)	2.82	2.97***	2.52	3.03***
School performance (1-7)	2.39	2.62***	2.26	2.23
Inference: */**/** indicate significant differences at the 10%/5%/1% level based on the mean differences of the dropped and selected samples, assessed with a t-test.				

4.3 Empirical Strategy

4.3.1 Ordinary Least Squares

The linear approximation can be described as

$$Y_{i,c} = \alpha + \beta \cdot L_i + \boldsymbol{\Omega} \cdot \mathbf{X}_i + \varepsilon_{i,c}, \quad (4)$$

where $Y_{i,c}$ is the test score of each individual i in class c , L_i is the binary treatment indicator for the law reform in 2012, $\mathbf{X}_i$ are the control variables and $\varepsilon_{i,c}$ denotes the error term. The

coefficient β represents the effect of the law reform on test scores. We used equation (4) to estimate two coefficients, i.e. the effect of the law reform in 2012 on average test scores in both the first and last class of primary school. All estimations were performed using STATA/SE 15.0 (STATA; <https://www.stata.com/>).

4.3.2 Non-Parametric Bounds

We observed an average difference in attrition of 24.7 percent before and after the law reform in 2012, i.e. between control and treatment group (51.0 vs 26.3 percent, respectively). We assumed that this difference in attrition is associated with the policy indicator, L_i , as **Figure 7** illustrates a decline in absenteeism after the law reform in 2012, which leads to more registered test scores and hence lower attrition. Previous research has shown that absenteeism clusters in low-ability pupils (Arulampalam *et al.*, 2012; Aucejo & Romano, 2016). Consequently, we expected that the non-random selection caused downward biased estimates of β in equation (4), $E(L_i \epsilon_{i,c}) < 0$, as the composition of the treatment group consists of more low-ability pupils compared to the control group.

To control for selective attrition, we used a bounds estimator that estimates an interval for the true value of the treatment effect in the presence of non-random sample selection (Lee, 2009). The estimator rested on two assumptions: random assignment of treatment and monotonicity. Monotonicity implies that assignment to the treatment group affects attrition in one direction, which corresponds to the observed pattern of lower absenteeism, and thus more registered test scores, after the law reform (**Figure 7**). As such, observations for which the outcome variable is observed irrespective of the assigned treatment status were included. Additionally, the actual estimation sample included either observations where the outcome is observed because of receiving the treatment or observations where the outcome is observed because of not receiving the treatment, but not both simultaneously. The bounds estimator trimmed either the treated or non-treated observations so that the share of observations with the observed outcome was equal for both groups. Trimming is either from above or from below. This corresponds to two extreme assumptions about missing information that were consistent with the observed data and a one-sided selection mechanism. That is, in the group that suffers less from attrition, the treatment group, either the largest or the smallest values of the outcome were excluded. As such, the bounds represented the effect of the law reform on test scores in the presence of non-random sample selection due to selective attrition (Lee, 2009).

Estimating treatment-effect bounds requires calculation of a raw group mean and two trimmed group means of the test scores, $Y_{i,c}$. Furthermore, L_i is the binary treatment indicator, S_i the binary selection indicator with $S_i = 1$ indicating attriters for which $Y_{i,c}$ is

observed. The shares of observations with $Y_{i,c}$ in the treatment and control group, q_T and q_C , can then be written as

$$q_T = \frac{\sum_i 1(L_i = 1, S_i = 1)}{\sum_i 1(L_i = 1)} \quad (5)$$

$$q_C = \frac{\sum_i 1(L_i = 0, S_i = 1)}{\sum_i 1(L_i = 0)} \quad (6)$$

where $1(\cdot)$ denotes the indicator function. The notation can be simplified, as the treatment group suffered less from attrition, $q_T > q_C$, resulting in

$$q = \frac{q_T - q_C}{q_T} \quad (7)$$

and $1 - q$ determines then the quantiles at which the distribution of Y in the treatment group are trimmed to exclude extreme values from the analysis, which can be written as

$$y_q^T = G_{Y|T=1, S=1}^{-1}(q) \quad (8)$$

$$y_{1-q}^T = G_{Y|T=1, S=1}^{-1}(1 - q) \quad (9)$$

determine the marginal values y_q^T and y_{1-q}^T of the outcome that enter the trimmed means, with G_Y^{-1} denoting the inverse empirical distribution of $Y_{i,c}$. Using this notation, we calculated estimates for the lower and upper bounds as

$$\hat{\theta}_{\text{lower}} = \frac{\sum_i 1(T_i = 1, S_i = 1, Y_{i,c} \leq y_{1-q}^T) Y_{i,c}}{\sum_i 1(T_i = 1, S_i = 1, Y_{i,c} \leq y_{1-q}^T)} - \frac{\sum_i 1(T_i = 0, S_i = 1) Y_{i,c}}{\sum_i 1(T_i = 0, S_i = 1)} \quad (10)$$

$$\hat{\theta}_{\text{upper}} = \frac{\sum_i 1(T_i = 1, S_i = 1, Y_{i,c} \geq y_q^T) Y_{i,c}}{\sum_i 1(T_i = 1, S_i = 1, Y_{i,c} \geq y_q^T)} - \frac{\sum_i 1(T_i = 0, S_i = 1) Y_{i,c}}{\sum_i 1(T_i = 0, S_i = 1)}. \quad (11)$$

4.4 Results

Table 13 displays the standardised estimates of β in equation (4) with additional control variables age, gender, ethnicity, care level, family situation, parental educational attainment, urbanicity and school performance, respectively. The results suggest that the law reform increased the average test scores of pupils in the first and last class of primary school with 0.81 and 0.69 SD, respectively.

Table 13 OLS estimates of standardised test scores

	First class	Last class
Reform	0.814*** (0.013)	0.693*** (0.011)
Covariates	YES	YES
R-squared	0.36	0.64
Number of ID	20,946	20,103
Robust standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.		

Table 14 and **Table 15** display the estimates of the lower and upper bounds, $\hat{\theta}_{\text{lower}}$ and $\hat{\theta}_{\text{upper}}$ in equations (7) and (8), of pupils in the first and last class of primary school, respectively. The bounds in **Table 14** (**Table 15**) suggest an average effect, $\bar{\theta}$, of the law reform on pupil's test scores in the first (last) class of primary school of 0.851 (0.734) SD. In both classes is $\bar{\theta}$ higher than the OLS point estimates, which confirms our hypothesis that the law reform decreased in particular the absenteeism of low-ability pupils, causing a downward bias of β in equation (4).

Table 14 Lower and upper bound in the first class

	Lower	Upper
Reform	0.122*** (0.015)	1.58*** (0.017)
Number of ID	20,946	20,946
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.		

Table 15 Lower and upper bound in the last class

	Lower	Upper
Reform	0.287*** (0.017)	1.18*** (0.015)
Number of ID	20,103	20,103
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.		

4.5 Supplementary Analyses

4.5.1 Tightening Bounds

We extended the original estimator by tightening the bounds on selected (interaction between) covariates using ℓ_1 -regularized regression based on an iterative algorithm for estimating the optimal lambda (Belloni *et al.*, 2012). That is, we assumed assignment of treatment under which the differences in conditional and unconditional expectations of Y do not coincide. Covariates that have some explanatory power for attrition were used to split the sample into cells, and bounds are separately calculated for each cell. Finally, a weighted average of cells' bounds was computed. The appropriate weights, the probabilities of cell membership for those that never suffer from attrition, are unknown. However, because of random assignment of treatment and monotonicity, they can consistently be estimated by

$$F = \frac{\sum_i 1(J_i = 1, S_i = 1, T_i = 0)}{\sum_i 1(S_i = 1, T_i = 0)} \quad (12)$$

for each cell J, where $J_i = 1$ indicates membership in J. As such, we controlled for selective attrition and demographic composition.

The results in **Table A16** and **Table A17** in the Appendix show that tightening the bounds yields similar conclusions with an average increase in test scores for both classes of 0.84 and 0.73, respectively. Hence, this suggests that the main estimates are robust to the observed heterogeneity between control and treatment group.

4.5.2 Effect Heterogeneity

In this section we determined the effect heterogeneity, as the main results only provide the bounds for the entire sample, and previous research has shown that school attendance clusters in societal subgroups (Sacerdote, 2002; Banerjee & Dulfo, 2006). Furthermore, pupils from a more advantaged background capitalize on education policies to a larger extent (Chetty *et al.*, 2014). **Chapter 5** below confirmed this for the Dutch setting, as pupils from a more vulnerable background benefitted substantially less from special needs funding compared to their more privileged peers. We therefore estimate the lower and upper bounds stratified by pupil's care level, ethnicity, and socio-economic environment. Socio-economic environment is classified as low Socio-Economic Status (SES) for low parental educational attainment and high SES for middle and high parental educational attainment.

Results are provided in **Table A18 – Table A23** in the Appendix. The bounds suggest that the average effect is 0.18, 0.09 and 0.05 SD higher for pupils with special needs, children from an ethnic minority, and from a low socio-economic environment, respectively.

4.6 Concluding Remarks

This study contributes to the discussion on the impact of compulsory education. We show besides duration of compulsory school a law reform that tightens the monitoring compliance with the compulsory education act enhances pupil's test scores substantially. Specifically, after the Dutch government introduced a penalty if primary schools did not comply to the law on relative absenteeism, pupil's test scores in the first and last class of primary school increased with 0.85 and 0.73 SD, respectively.

Education is a pivotal element of human capital in child development, as academic performance is positively associated with various labour market outcomes (Bhuller *et al.*, 2017; Devereux & Hart, 2010; Leigh & Ryan, 2008; Oreopoulos, 2006a, 2006b, 2007), late-life health (Brunello *et al.*, 2016; Galama *et al.*, 2018; Grimard & Parent, 2007; Jürges *et al.*, 2011; Kemptner *et al.*, 2011), and criminal behaviour (Cullen *et al.*, 2006; Deming, 2011; Amin *et al.*, 2016). Hence, monitoring compliance to relative absenteeism in primary schools has long-lasting impact on the development of children.

Previous research has shown that absenteeism clusters in low-ability pupils (Arulampalam *et al.*, 2012; Aucejo & Romano, 2016). Consequently, a compulsory education act that attempts to reduce absenteeism will likely lead to more registered test scores of low-ability pupils in the treatment group, hence causing bias estimates. In this study we tackled the potential issue of selective attrition by taking non-parametric bounds corresponding to extreme assumptions about the missing information that are consistent with the observed data. Moreover, we showed that the estimates are robust to observed heterogeneity by tightening the bounds on selected covariates using ℓ_1 -regularized regression (Belloni *et al.*, 2012). Specifically, we showed that the bounds are not affected by differences in age, gender, ethnicity, care level, family situation, parental educational attainment, urbanicity and school performance between the control and treatment group.

Education can for disadvantaged pupils be an excellent opportunity to rise above their position. Previous research in the Netherlands demonstrates, however, that investing additional money in care pupils affects pupils from a more advantaged background to a larger extent than their disadvantaged peers (**Chapter 5**). The findings of this study are contradictory, as the average effect is higher for care pupils and children from an ethnic minority or a low socio-economic environment. These findings suggest that a law reform on pupil's absenteeism better reaches the more vulnerable groups compared to increasing school resources. However, further research is needed to find out if effect heterogeneity across different type of policy instruments is spurious or follows a consistent pattern.

Appendix

A4.1 Figures & Tables

Table A16 Tightened bounds in the first class

	Lower	Upper
Reform	0.271*** (0.051)	1.41*** (0.056)
Number of ID	20,946	20,946
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.		

Table A17 Tightened bounds in the last class

	Lower	Upper
Reform	0.348*** (0.053)	1.12*** (0.045)
Number of ID	20,103	20,103
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.		

Table A18 Bounds in the first class stratified by care level

	Regular pupils		Special needs pupils	
	Lower	Upper	Lower	Upper
Reform	0.169*** (0.017)	1.60*** (0.020)	0.141*** (0.029)	1.59*** (0.039)
Number of ID	16,128	16,128	4,818	4,818
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.				

Table A19 Bounds in the first class stratified by care level

	Regular students		Special needs students	
	Lower	Upper	Lower	Upper
Reform	0.287*** (0.018)	1.23*** (0.015)	0.103*** (0.034)	1.26*** (0.033)
Number of ID	15,278	15,278	4,825	4,825
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.				

Table A20 Bounds in the first class stratified by ethnic background

	Caucasian		Non-Caucasian	
	Lower	Upper	Lower	Upper
Reform	0.192*** (0.018)	1.44*** (0.020)	0.052** (0.025)	1.68*** (0.029)
Number of ID	13,221	13,221	7,725	7,725
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.				

Table A21 Bounds in the last class stratified by ethnic background

	Caucasian		Non-Caucasian	
	Lower	Upper	Lower	Upper
Reform	0.315*** (0.021)	1.07*** (0.018)	0.280*** (0.031)	1.33*** (0.026)
Number of ID	12,665	12,665	7,348	7,348
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.				

Table A22 Bounds in the first class stratified by socio-economic environment

	Low SES		High SES	
	Lower	Upper	Lower	Upper
Reform	0.264*** (0.026)	1.62*** (0.027)	0.06*** (0.018)	1.62*** (0.022)
Number of ID	7,474	7,474	13,472	13,472
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.				

Table A23 Bounds in the last class stratified by socio-economic environment

	Low SES		High SES	
	Lower	Upper	Lower	Upper
Reform	0.390*** (0.023)	1.08*** (0.021)	0.21*** (0.027)	1.28*** (0.022)
Number of ID	9,655	9,655	10,448	10,448
Bootstrapped standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.				

A4.2 List of Abbreviations

COOL	=	Cohort Onderzoek OnderwijsLoopbanen
OLS	=	Ordinary Least Squares
SD	=	Standard Deviation
SES	=	Socio-Economic Status

Chapter |5|

5. Long-run Returns to Government Expenditures in Special Needs Education: Evidence from the Netherlands¹

Abstract

This study investigated the long-run returns to government expenditures in special needs education. From the 1st of August 2003 the Dutch government provided regular primary and secondary schools with additional funding for the inclusion of special needs students. Schools were responsible in deciding how to allocate this special needs funding (~ 14,400 euros per eligible student). We used a unique cohort study tracking 1,072 regular and 146 special needs students from the end of primary school to adulthood between 2000 and 2012. We used a differences-in-differences design to estimate the causal effect of special needs funding on the mental health of eligible students. The results suggest that special needs funding improved their mental health by 0.41 standard deviation. Moreover, we demonstrated a cost/benefit-ratio of 0.69 when we extrapolate the found mental health effect to life-cycle returns in mental healthcare use, educational attainment and labour market participation. Heterogeneity analyses present a cautionary note, as girls and children from an ethnic minority or a low socio-economic environment benefited less from the policy.

Keywords: Government Expenditures, Health and Inequality, Returns to Education

JEL Codes: H52, I14, I21, I26

5.1 Background

5.1.1 Government Expenditures & Education

Students disadvantaged by low-ability or worse health have a lower opportunity profile through educational setbacks early in life (Galama *et al.*, 2018; Grossman, 2015). Topic of discussion on the worldwide political agenda have been the impact of government expenditures in education on disparities in student outcomes. Previous research conclude that there hardly is a consistent relationship between additional funding and student achievement (Hanushek, 2003, 2006). Ignoring endogeneity problems could explain the lack

¹ Joint work with Jochen Mierau. The data were retrieved from the TRacking Adolescents' Individual Lives Survey (TRAILS). I thank the TRAILS Data Group for managing and providing the data.

This study benefitted from comments received at the Tinbergen Institute Seminar 2019, the Royal Economic Society (RES) Annual Conference 2019 and the International Association of Applied Econometrics (IAAE) Annual Conference 2019. I am grateful to RES and IAAE for selecting this study for a young economists' travel grant. Furthermore, I thank Gabriella Conti, Owen O'Donnell, Beatriz Eugster, Martin Salm, Sarah See, Hans van Kippersluis and Bram Wouterse for helpful comments.

of a consistent relationship, as variation to evaluate policy measures is commonly correlated with (un)observed school and student characteristics. There are several recent studies that use a quasi-experimental design to tackle the problem of endogeneity. These studies obtain contradictory findings of the causal impact of additional funding on student outcomes. Leuven *et al.* (2007} and van der Klauw (2008) find negative effects on students' achievement, while Papke (2005), Holmlund *et al.* (2010), Machin *et al.* (2010), de Haan (2015) and Gibbons *et al.* (2017) all find positive effects. Bénabou *et al.* (2009) find ambiguous effects of additional funding. Moreover, there is a lack of studies investigating the cost/benefit-ratio of government expenditures in education.

These studies focus mainly on policy measures that are implemented nationwide for each student and are evaluated in achievement outcomes only. By contrast, this study investigated the impact of government expenditures in special needs education on students' mental health. We exploited a natural experiment in special needs funding that provided additional funding to regular primary and secondary schools for the inclusion of special needs students from 1st of August 2003 onwards (Wettenbank, 2003). Unlike most studied resource policies, special needs funding was implemented conditionally on eligibility terms set by the Dutch Ministry of Education, Culture and Science. This exogenous variation allowed us to identify the counterfactual change over time in mental health of special needs student if the policy had not been implemented. To do so, we used a rich cohort study tracking both regular and special needs students from the end of primary school to adulthood between 2000 and 2012.

We show that additional government expenditures in special needs education increases the mental health of the eligible students by 0.41 Standard Deviation (SD). The special needs funding scheme has been abolished on 1st of August 2014, with the purpose to reduce the budget by 300 million euros per year (Wettenbank, 2014). We feel this simplistic budget reduction may been taken without proper evaluation and shed new light on this decision. We demonstrate a cost/benefit-ratio of 0.69 within working age by extrapolating the found mental health effect to life-cycle returns in mental healthcare use, educational attainment and labour market participation, turning the decision in 'penny wise, pound foolish'.

5.1.2 Special Needs Funding

In the Netherlands Regional Expertise Centra (REC) are aimed at the indication of special needs students. Each school for special education participates in one of these 34 RECs. Until the academic year 2003/2004 special needs students were largely assigned to special schools. To stimulate the development of special needs students, the Dutch Ministry of Education, Culture and Science introduced a policy measure on the 1st of August 2003 that provided

additional funding to regular primary and secondary schools for inclusion of special needs students in regular education (Wettenbank, 2003). This special needs funding was implemented conditionally on eligibility terms set by the Dutch Ministry of Education, Culture and Science. The Committee of Indication of Assessment became responsible for the assignment and assessment of referred students. Once a student was assigned to the policy measure a budget was allocated to the corresponding school. Schools were responsible in deciding how to allocate the additional money of 14,400 euro per eligible student.² The money was mainly used for additional school materials or teaching assistance to support the assigned students to catch up with their regular peers. **Figure 8** illustrates the expected pattern with a significant inflow of special needs students in regular education after the policy introduction in 2003.

Figure 8 Special needs students in regular education between 2000 and 2018

The policy for special needs students was heavily criticised, because the additional budget required to support the funding was exploding and credible estimates on the effect were lacking. The special needs funding scheme was therefore abolished on 1st of August 2014. From this date appointed regional partnerships, consisting of delegates from regular and special schools and the corresponding municipality, became responsible for the

² The direct educational costs differ substantially between students. Special needs students cost on average 18,400 euros per year if the student is assigned to special needs funding and about 8,900 euros per year if the student is still in special education. The costs of a regular student are on average 4,000 euros per year (Minne *et al.*, 2009).

indication of special needs students. As such, the government aimed at a more effective and efficient allocation of money. In practice it was no more than a blunt budget cut in special needs funding of 300 million euros per year with an outflow of special needs students from regular to special schools as a result (**Figure 8**).

5.2 Data

5.2.1 TRAILS

We used data of the TRacking Adolescents' Individual Lives Survey (TRAILS). TRAILS sampled participants born between 1 October 1989 and 30 September 1991, who lived in one of the three provinces in the North of the Netherlands at the time of the baseline assessment in 2000 (Huisman *et al.*, 2008). Children from both urban and rural areas were included to construct a representative sample. In total 2,230 children were included, corresponding to a response rate of 76%. The mean age was 11.1 and 51% of the sample were girls (Huisman *et al.*, 2008). Subsequent data collection took place every two to three years with 1,584 individuals of the original cohort participating in all five waves (Oldehinkel *et al.*, 2015). The final sample comprised 1,218 individuals due to missings in mental health scores. We discuss the impact of the observed attrition on treatment allocation in **Section 5.5.1**.

5.2.2 Mental Health

Scores on the Youth Self Report (YSR) for the years 2000, 2003 and 2006 and on the Adult Self Report (ASR) for the years 2009 and 2012 were used to measure mental health. YSR and ASR are questionnaires of 113 items, validated by the Achenbach System of Empirically Based Assessment (ASEBA) (Achenbach & Rescorla, 2013). All items were scored by the participant with a 0 (not applicable at all), 1 (a little or sometimes applicable) or 2 (clear or often applicable). We convert the scores into average scores between 0 and 100 in which (0) 100 represents (bad) good mental health. The questions about behaviour consist of the problem scales depressed, anxious, social problems, attention problems, norm-deviant behaviour and aggressive behaviour. The first three problem scales together form the broad-band syndrome “internalizing problems” and the latter three the broad-band syndrome “externalizing problems”. All problem scales together form the scale “total problems” (Achenbach & Rescorla, 2013), which we used as mental health outcome in the main analysis. We split the broad-band syndromes in **Section 5.5.3**.

5.2.3 Demographic Characteristics

We controlled for observed heterogeneity in demographic characteristics. We used a dummy variable for ethnicity with non-Caucasian background as reference category for at least one foreign-born parent. To define the family composition, we use the number of children in the family. Educational attainment of the parents and individual was categorised according to three levels: score 1 - low (primary education, lower tracks of secondary education); score 2 - middle (higher tracks of secondary education, senior secondary vocational education); and score 3 - high (higher professional education, university). Furthermore, family income is a categorical variable with nine monthly household income classes, ranging from less than €680.67 (score 1), between €680.67 and €1,134.45 (score 2), ..., between €3,403.35 and €3,857.13 (score 8), to more than €3,857.13 per month (score 9). Finally, we used a dummy variable for urbanicity to separate children living in a larger city (Groningen, Leeuwarden, Assen) or in a village (Winschoten, Dantumadeel, Grijpskerk).

5.2.4 Indication Process

Eligibility for special needs funding was assessed by the Committee of Indication of Assessment of the Dutch Ministry of Education, Culture and Science. We identified special needs students accordingly. Baseline mental health scores were discerned into normal-borderline-clinical cutoffs based on the validated Dutch norms of the ASEBA (Achenbach & Rescorla, 2013).³ Students with a score above the clinical cutoff were assumed eligible, resulting in 146 students assigned to special needs funding (12%), which is representative for national averages (Minne *et al.*, 2009). The other 1,072 students were categorised as regular students, non-eligible for special needs funding.

Baseline characteristics for both groups are provided in **Table 24**. The sample of special needs students consists of less boys and children live more frequently in a high-urban area ($p < 0.05$). Furthermore, we observe that both groups are from a comparable in age, ethnic background, family composition, parental educational attainment and family income.

³ We multiplied the original baseline scores (0-2) with 119. Subsequently, we followed the validated cut-offs for boys younger than twelve years old of scores below 39 (normal), between 39 and 48 (borderline) and above 48 (clinical). The validated cut-offs for boys of twelve years or older are scores below 40 (normal), between 40 and 52 (borderline) and above 52 (clinical). Similarly, we followed the validated cut-offs for girls younger than twelve years old of scores below 36 (normal), between 36 and 48 (borderline) and above 48 (clinical). Finally, the validated cut-offs for girls of twelve years or older are scores below 36 (normal), between 36 and 44 (borderline) and above 44 (clinical).

Table 24 Summary statistics of regular and special needs students

	Regular students <i>N</i> = 1,072	Special needs students <i>N</i> = 146
<i>Student</i>		
Mental health (0-100)	88.41	77.95***
Age (10-12)	11.08	11.09
Males	0.48	0.39***
Non-Caucasian background	0.07	0.06
<i>Family background</i>		
Number of children (0-8)	2.53	2.61
Educational attainment mother (1-3)	1.82	1.82
Educational attainment father (1-3)	1.93	1.93
Family income (1-9)	4.85	4.80
<i>Demographic</i>		
High urbanicity	0.82	0.87**
Inference: */**/** indicate significant differences at the 10%/5%/1% level based on the mean differences of the two samples, assessed with a t-test.		

5.3 Empirical Strategy

We used a Differences-In-Differences (DID) design to control for time-invariant unobserved heterogeneity, enabling direct identification of special needs funding on the mental health of special needs students. The official time cut-off was 1 August 2003 (Wettenbank, 2003). Hence, we defined waves 1 and 2 of TRAILS as the pre-policy period ($t = 0$) and waves 3 to 5 as the follow-up period ($t = 1$). Furthermore, only special needs students were eligible for indication to special needs funding. Hence, special needs students were allocated to the treated group ($Z_i = 0$) and regular students to the control group ($Z_i = 1$).

Furthermore, we assumed absence of policy changes in the baseline for either group ($D_{i,t=0} = 0, Z_i = 0$) and the policy to have a positive effect on the mental health of students in the treated group ($D_{i,t=1} = 1, Z_i = 1$). For the outcome variable ($Y_{i,t}$) the population DID effect is then given by the mean difference in mental health between special needs students and regular students before and after the policy. The corresponding DID setting is given by

$$\begin{aligned} \text{DID} = \{ & E(Y_{i,t=1} | D_{i,t=1} = 1, Z_i = 1) - E(Y_{i,t=1} | D_{i,t=1} = 0, Z_i = 0) \} \\ & - \{ E(Y_{i,t=0} | D_{i,t=0} = 1, Z_i = 1) - E(Y_{i,t=0} | D_{i,t=0} = 0, Z_i = 0) \} \end{aligned} \quad (13)$$

In the DID extended model we added the individual-level control variables (X_i) age, gender, ethnicity, family composition, parental educational attainment, family income, and urbanicity.⁴ The corresponding extended DID setting is given by

$$\begin{aligned} \text{DID} = \{ & E(Y_{i,t=1} | D_{i,t=1} = 1, Z_i = 1, X_i) - E(Y_{i,t=1} | D_{i,t=1} = 0, Z_i = 0, X_i) \} \\ & - \{ E(Y_{i,t=0} | D_{i,t=0} = 1, Z_i = 1, X_i) - E(Y_{i,t=0} | D_{i,t=0} = 0, Z_i = 0, X_i) \} \end{aligned} \quad (14)$$

The linear approximation of the population DID in (14) is given by

$$Y_{i,t} = \alpha + \beta \cdot F_{i,t} + \gamma_i + \delta_t + \boldsymbol{\Omega} \cdot \mathbf{X}_i + \varepsilon_{i,t}, \quad (15)$$

where $Y_{i,t}$ is the mental health for each individual i in wave t , $F_{i,t}$ is the indicator for allocation to special needs funding, γ_i captures the baseline difference in mental health, δ_t captures the time trend, $\mathbf{X}_i$ are the individual-level control variables and $\varepsilon_{i,t}$ denotes the error term. The coefficient β represents the DID estimand of interest – the effect of special needs funding on the mental health of special needs students. All estimations were performed using STATA/SE 15.0 (STATA; <https://www.stata.com/>).

5.4 Results

5.4.1 Mental Health

Figure 9 shows the average mental health of regular students and special needs students with the introduction of special needs funding illustrated by the vertical dotted line. We observe both parallel paths before the policy introduction in 2003 and from 2006 onwards ($p > 0.10$). Moreover, previous research demonstrated that regular students were not affected by the inclusion of special needs students (Ruijs & Peetsma, 2009; Ruijs, 2017). Hence, we were able to identify the counterfactual change in mental health over time for special needs students as if the policy had not been implemented.

Table 25 displays the standardised estimates of β in equation (15) with and without additional control variables, respectively. The results suggest that special needs funding increased the mental health of special needs students with 0.41 SD (~ 38.0 percent).

⁴ Note that adding individual-level control variables only reduces the within-group variance. The within-group variation does not affect the identification of the policy effect but may reduce standard errors. Between-group variance becomes important when observed heterogeneity may confound the identification strategy. Given the features of our DID setting, observed heterogeneity may confound the identification of the policy effect. Therefore, in **Section 5.5.2** of the supplementary analyses we added non-experimental methods to the original DID setting to reduce the between-group variance. Specifically, we followed Heckman *et al.* (1997, 1998) and matched the special needs students with their regular peers at baseline on the propensity of eligibility for the policy conditional on the control variables in (15), (X_i). In the second stage we added the kernel weights obtained in the first stage to the DID setting to estimate the exempt policy effect.

Figure 9 Average mental health between 2000 and 2012

Table 25 DID estimates for mental health

	(1)	(2)
DID	0.414*** (0.041)	0.414*** (0.028)
Covariates	NO	YES
R-squared	1,218	1,218
Number of ID	0.20	0.22
Clustered standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.		

In **Figure 10** we show the standardised estimates of β in equation (15) across the distribution of students' baseline mental health with the main result of **Table 25** as horizontal dashed line. The corresponding ten-percent quantile DID estimates are presented in **Table 26**. The policy yields an effect at the median of 0.42 SD. Furthermore, the policy was the most (least) effective at the sixth (third) quantile with an effect of 0.44 (0.32) SD.

Figure 10 Ten-percent quantile DID estimates

Table 26 Quantile DID estimates for mental health

Quantile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
DID	0.418* (0.087)	0.327* (0.072)	0.316* (0.066)	0.335* (0.063)	0.420* (0.060)	0.443* (0.062)	0.440* (0.055)	0.438* (0.056)	0.340* (0.047)
Covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES
Number of ID	1,218	1,218	1,218	1,218	1,218	1,218	1,218	1,218	1,218
R-squared	0.16	0.16	0.16	0.15	0.13	0.12	0.10	0.06	0.06
Clustered standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.									

5.4.2 School Indicators

We used two indicators to measure the standardised difference in school performance between 2003 (*before policy*) and 2006 (*after policy*). First, we measured the performance in language and math with a validated questionnaire, which consisted of three items: 1. work pace; 2. effort; and 3. test scores. The teacher assessed all items with a score ranging from 1 ('strongly disagree') to 5 ('strongly agree'). We used the average score between 0 and 5 with higher values for better academic performance. Second, we used apprenticeship scores, which are combined scores of a student's relative level and progression in education (Bosker *et al.*, 1985). **Table 27** suggest that the educational gap between regular and special needs students reduced after the introduction of special needs funding.

Table 27 School indicators of regular and special needs students

	Difference (SD)
<i>Before policy</i>	
Language performance	0.20***
Math performance	0.20***
Apprenticeship score	0.16***
<i>After policy</i>	
Language performance	0.09
Math performance	0.01
Apprenticeship score	0.11*
Inference: */**/** indicate significant differences at the 10%/5%/1% level based on the mean differences of the two samples, assessed with a t-test.	

Finally, we show in **Figure 11** for both groups the distribution of educational attainment nine years later in 2015. The Wilcoxon rank-sum test revealed that the median of special needs students only differs statistically significantly from their regular peers at a ten-percent level. We were not able to identify the counterfactual distribution if the policy was not implemented, but the descriptives in **Table 27** suggest a partial shift of special needs students towards lower educational attainment.

5.4.3 Cost/Benefit-ratio

Mental health is a pivotal element of human capital formation (WHO, 2013). The Dutch Ministry decided to abolish special needs funding in 2014 solely based on seemingly high expenditures, while the potential benefits had not been discussed (Wettenbank, 2014).

Figure 11 Educational attainment at labour market entry

In this section we estimate the cost/benefit-ratio of special needs funding by extrapolating the mental health impact estimated to benefits in mental healthcare use, educational attainment and labour market participation. Four assumptions were key for our cost/benefit-analysis. First, the average annual cost of special needs funding equals €14,400 (Minne *et al.*, 2009). Second, 35.3 percent of the adolescents with special needs in our sample are registered in mental healthcare (Jörg *et al.*, 2016). The average annual cost of patients in mental healthcare equals €5,909, as in 2016 total mental healthcare cost counted €6.5 billion for 1.1 million registered patients (CBS, 2017a; Vektis Intelligence, 2017). Hence, a 38 percent improvement in mental health would be associated with annual benefits in mental healthcare use of $0.38 \times 0.353 \times €5,909 = €793$. Third, a 0.41 SD increase in mental health would be associated with a 12.3 percent increased probability in labour market participation (Frijters *et al.*, 2014). The average annual difference in labour income between employed and non-employed individuals between 2012 and 2018 amounts to €20,583 (CBS, 2017b). Hence, the policy associated with increased in labour market participation would result in $0.123 \times €20,583 = €2,532$ annually. Fourth, the difference in apprenticeship scores between regular and special needs students decreased by 31.3 percent after the policy. We link this with **Figure 11** by assuming a counterfactual shift of 31.3 percent in special needs students' distribution from high to middle educational attainment. The annual labour income of individuals with middle and high educational attainment is estimated at €32,400 and

€53,300, respectively (CBS, 2017c). Hence, the policy effect would have resulted in benefits in labour income of $0.313 \times 0.442 \times €20,900 = €2,891$ annually.

Subsequently, we assumed that (1) special needs students received funding from the age of twelve to eighteen, (2) benefits in mental healthcare use sustained from the age of eighteen onward, (3) benefits in labour market participation occur from the age of twenty-one, (4) indirect benefits on labour income start from the age of twenty-four, and (5) the discount rate r in education equals three percent. These assumptions enabled us to estimate the breakeven point in time t of the government's investment by solving equation (16).

$$\sum_{t=1}^6 \frac{€14,400}{(1 + 0.03)^t} = \sum_{t=6}^t \frac{€793}{(1 + 0.03)^t} + \sum_{t=9}^t \frac{€2,532}{(1 + 0.03)^t} + \sum_{t=12}^t \frac{€2,891}{(1 + 0.03)^t} \quad (16)$$

Equalizing returns to investment yields a reimbursement time t^* of 33 years and 275 days from the starting point of investment. When we extrapolated benefits beyond t^* until retirement age T , we found a cost/benefit-ratio of 0.69 – horizontal dashed line in **Figure 12**. However, this ratio will be largely determined by how much the government values future benefits (CPB, 2017). Therefore, we performed a sensitivity analysis for the discount rate r . Hence, the benefits of special needs funding would outweigh the cost up to a discount rate of 4.7 percent (**Figure 12**).

Figure 12 Cost/benefit-ratios for discount rates between 1 and 5 percent

5.5 Supplementary Analyses

In the supplementary analyses we demonstrate that the identification strategy is robust to attrition and propensity reweighting. Furthermore, we show that although results differ cross broad-band syndromes, conclusions remain similar. Finally, we focused on the effect heterogeneity for different classes in society.

5.5.1 Attrition Bias

Consistent estimation of the DID estimate requires availability of mental health scores in all five waves. As a result, we dropped 366 individuals of the original sample of which 350 (16) would have been allocated to the control (treatment) group (**Table 28**).

Table 28 Attrition matrix

	Treated	0	1
Selected			
0		350	16
1		1,072	146

To control for the robustness to attrition we regressed the selection indicator on the treatment indicator in a logit model given by

$$p(Z_i = 1) = \beta_0 + \beta_1 \cdot S_i + \varepsilon_i, \quad (17)$$

where S_i is a binary variable taking the value 0 for the 366 dropped individuals and 1 for the 1,218 included individuals, and ε_i denotes the error term. Estimation of equation (17) yields an odds ratio β of -0.222 ($p > 0.10$). Hence, the treatment allocation is not affected by attrition in the sample selection.

5.5.2 Propensity Reweighting

In the third specification we followed Heckman *et al.* (1997, 1998) to reduce the between-group variance. We used the observed covariates X_i of (15) to estimate the propensity score (the likelihood of indication to special needs funding) and to calculate kernel weights. Instead of only accounting for the within-group variance, this method matches the special needs students and regular students according to their propensity score. First, we obtained the propensity score (p_i) for both groups using probit estimation,

$$p_i = E(Z_i = 1|X_i). \quad (18)$$

Then according to Heckman *et al.* (1997), the kernel matching is given by the propensity score conditional on the covariates $\mathbf{X}_i$, which leads to the calculation of the kernel weights,

$$w_i = \frac{K\left(\frac{p_i - p_k}{h_n}\right)}{\sum K\left(\frac{p_i - p_k}{h_n}\right)} \quad (19)$$

in which $K \cdot$ is the kernel function and h_n is the selected bandwidth, set equal to 0.05. Then we introduced the kernel weights into (13) to obtain a kernel propensity-score matching DID effect. The corresponding DID setting is given by

$$\begin{aligned} \text{DID} = & \{E(Y_{i,t=1} | D_{i,t=1} = 1, Z_i = 1) - w_i \cdot E(Y_{i,t=1} | D_{i,t=1} = 0, Z_i = 0)\} \\ & - \{E(Y_{i,t=0} | D_{i,t=0} = 1, Z_i = 1) - w_i \cdot E(Y_{i,t=0} | D_{i,t=0} = 0, Z_i = 0)\} \end{aligned} \quad (20)$$

To increase the internal validity of the DID estimand, we restricted (20) to the common support of the propensity score for special needs students and regular students, following Rosenbaum & Rubin (1985). The common support is the overlap region of the propensity for both groups. This sample of i individuals is then restricted to the region defined as

$$(i: p_i \in [\max\{\min(p_i | Z_i = 1), \min(p_i | Z_i = 0)\}, \min\{\max(p_i | Z_i = 1), \max(p_i | Z_i = 0)\}]). \quad (21)$$

Then, we showed that in absence of the policy, the outcome variable is orthogonal to the policy indicator given the set of covariates $\mathbf{X}_i$. The balancing property at baseline is given by

$$Y_{i,t=0} \perp Z_i | \mathbf{X}_i. \quad (22)$$

The kernel propensity-score matching DID estimate is given in column (3) in **Table A29** in the Appendix. According to the matching strategy, special needs funding increased the mental health of special needs students with 0.39 SD. Although this estimate is slightly lower than the estimate of our main result of 0.41 SD, estimates are not statistically significantly different ($p > 0.10$). In **Table A30** in the Appendix we show that in absence of the policy, the outcome variable is, indeed, orthogonal to the policy indicator, as on all included covariates the difference between the treated and control group is statistically insignificant ($p > 0.10$).

5.5.3 Broad-band Syndromes

The scores for mental health can be split in the broad-band syndromes internalizing and externalizing socio-emotional skills (Achenbach & Rescorla, 2013; Attanasio *et al.*, 2020). The former (latter) is more present in mental health problems such as mood disorders (conduct disorders). Results in **Table A31** in the Appendix suggest a larger policy effect for special needs students' internalizing socio-emotional skills with a difference of 0.15 SD with their externalizing socio-emotional skills.

5.5.4 Effect Heterogeneity

In this section we determined the effect heterogeneity, as the main result in **Table 25** only provides the mean DID effect for the entire sample. Results are provided in **Table A32 – Table A34** in the Appendix for stratification by gender, socio-economic class and ethnicity, respectively. We show DID estimates for both nested models and decomposed samples (Abadie *et al.*, 2018).

The estimates in **Table A32** suggest that males benefited more from the policy with a relative difference with females of 0.076 SD. In **Table A33** we use a binary variable to define the student's Socio-Economic Status (SES). The SES variable is derived with a factor analysis on the control variables educational attainment mother, occupation mother, educational attainment father, occupation father, and family income. The DID estimates demonstrate that students from a high SES class capitalized to a greater extent from the policy than their low SES peers. Specifically, the relative difference between high and low SES students is 0.007 SD. Finally, the estimates in **Table A34** suggest that students with at least one foreign-born parent benefited less from the policy with a relative difference of 0.096 SD with their peers with both parents born in the Netherlands.

5.6 Concluding Remarks

This study contributes to the discussion on the impact of government expenditures in education. We estimated the causal effect of special needs funding on the mental health of special needs students. The results reveal that the policy measure improved their mental health by 0.41 SD with a potential cost/benefit ratio of 0.69. Importantly, this indicates that the decision to abolish the policy may have been premature and ill-informed

This study distinguishes from other policy evaluations in education by focusing on students' mental health instead of test scores. However, individual longitudinal data on child and adolescent mental health on a large scale is scarce. Therefore, we used data from a unique cohort following children from the end of primary school to adulthood with an excellent retention rate of 71 percent over a time span of twelve years. A balanced panel for estimation required availability of mental health scores in all five waves and, therefore, we dropped 366 of the 1,584 individuals. However, this did not bias the identification, as the attrition appeared not to be associated with the probability of indication to funding.

Within the sample 146 of the 1,218 students were selected into the treatment group based on eligibility terms. However, to a certain extent this selection could be associated with student's background characteristics. For example, Chetty *et al.* (2016) found that student's socio-economic environment affects college attendance and school performance. To test for

selection bias, we extended our DID specification by estimating the propensity of indication to special needs funding based on background characteristics in the first stage. We showed that in absence of the policy, the outcome variable mental health is orthogonal in the second stage to the policy indicator given the student's socio-economic environment. The extended specification yields a reduction compared to the original DID estimate of only 0.02 SD, which indicates that the main analysis is robust to selection bias.

Chetty *et al.* (2014) suggest that advantaged students capitalize on the effect of education policies to a larger extent relative to their disadvantaged peers. Our heterogeneity analyses confirm this, as we found a significant difference between the lowest and highest SES class, as well as between students with Dutch-born parents with peers with a foreign-born parent. Future research could focus on the mechanism behind reaching the more vulnerable students within the treatment group.

The Dutch Ministry of Education, Culture and Science in 2014 decided to cut expenditures on special needs funding by 300 million euros. Notably, the potential benefits of the policy were ignored. We shed a new light on this decision by taking into account the long-run returns to mental health. By linking the found increase in mental health to benefits in mental healthcare use, educational attainment and labour market participation, we showed that the investment required to sustain special needs funding would be earned back within 34 years. Moreover, extrapolating the benefits to retirement age yields a cost/benefit-ratio of 0.69. We controlled for how much the government values future benefits and show that the benefits of special needs funding would outweigh the cost up to a discount rate of 4.7 percent. Hence, the budget cut in special needs funding has been a 'penny wise, pound foolish' decision.

Appendix

A5.1 Figures & Tables

Table A29 Kernel propensity DID estimates

	(1)	(2)	(3)
DID	0.414*** (0.041)	0.414*** (0.028)	0.394*** (0.043)
Covariates	NO	YES	YES
Kernel weights	NO	NO	YES
Number of ID	1,218	1,218	1,218
R-squared	0.20	0.22	0.22
Standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.			

Table A30 Results balancing test with weighted covariates at baseline

Weighted variable	$P_r(T > t)$	Mean control	Mean treated	Diff	t
Mental health	0.0000***	88.778	78/369	-10.409	30.20
Age	0.8042	10.582	10.572	-0.010	0.25
Males	0.6041	0.406	0.391	-0.015	0.52
Non-Caucasian background	0.8270	0.062	0.059	-0.003	0.22
Number of children	0.8467	2.597	2.608	0.012	0.19
Educational attainment mother	0.6914	1.823	1.819	-0.004	0.40
Educational attainment father	0.9134	1.929	1.934	0.005	0.11
Family income	0.7934	4.854	4.802	-0.052	0.20
High urbanicity	0.6721	0.821	0.873	-0.052	0.42
Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.					

Table A31 DID estimates for broad-band syndromes

	Internalizing	Externalizing
DID	0.405*** (0.047)	0.253*** (0.048)
Covariates	YES	YES
R-squared	1,218	1,218
Number of ID	0.22	0.12
Clustered standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.		

Table A32 DID estimates stratified by gender

	Full	Females	Males
DID	0.391*** (0.060)	0.391*** (0.060)	0.466*** (0.069)
DID × Males	0.075*** (0.011)		
Covariates	YES	YES	YES
Number of ID	1,218	693	525
R-squared	0.22	0.22	0.19
Clustered standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.			

Table A33 DID estimates stratified by socio-economic class

	Full	Low SES	High SES
DID	0.410*** (0.104)	0.410*** (0.104)	0.417*** (0.045)
DID × High SES	0.007*** (0.003)		
Covariates	YES	YES	YES
Number of ID	1,218	202	1,016
R-squared	0.21	0.21	0.20
Clustered standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.			

Table A34 DID estimates stratified by ethnicity

	Full	Caucasian	Non-Caucasian
DID	0.419*** (0.047)	0.419*** (0.047)	0.323* (0.182)
DID × Non-Caucasian	-0.096* (0.072)		
Covariates	YES	YES	YES
Number of ID	1,218	1,142	76
R-squared	0.22	0.21	0.31
Clustered standard errors in parentheses. Inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.			

A5.2 List of Abbreviations

ASEBA	=	Achenbach System of Empirically Based Assessment
ASR	=	Adult Self Report
DID	=	Differences-In-Differences
OLS	=	Ordinary Least Squares
REC	=	Regional Expertise Centra
SD	=	Standard Deviation
SES	=	Socio-Economic Status
TRAILS	=	TRacking Adolescents' Individual Lives Survey
YSR	=	Youth Self Report

A5.3 TRAILS Data Statement

This research is part of the TRacking Adolescents' Individual Lives Survey (TRAILS). Participating centers of TRAILS include the University Medical Center and University of Groningen, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group, all in the Netherlands. TRAILS has been financially supported by various grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMw Risk Behavior and Dependence grants 60-60600-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013 and 481-11-001; NWO Vici 016.130.002 and 453-16-007/2735; NWO Gravitation 024.001.003), the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), the European Research Council (ERC-2017-STG-757364 en ERC-CoG-2015-681466), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the Gratama foundation, the Jan Dekker foundation, the participating universities, and Accare Centre for Child and Adolescent Psychiatry. The data that support the findings of this study are available, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Chapter |6|

6. Policy Notes

We know that mental health problems originate from youth, but what is the persistence of mental health issues and healthcare during adolescence after youth enters psychiatric care? Are mental health trajectories largely determined by underlying individual characteristics? And how do these trajectories change or respond to health policy reforms? In **Chapter 2** we used registry data of more than 80,000 Dutch children and adolescents to answer these questions. Previous studies on the persistence of child and adolescent mental healthcare do not consider the role of underlying individual characteristics. Moreover, estimating persistence of healthcare using standard linear models yields biased estimates due to unobserved heterogeneity and the autoregressive structure of the model. We provide estimates of the persistence of care while taking these statistical issues into account. All estimation results showed a positive coefficient smaller than unity, which indicates that young patients return to their base level of mental healthcare after an exogenous shock. Additionally, persistence of care is found not to be affected by large healthcare reforms in the last decade. Indeed, a substantial part of persistence is due to individual characteristics and thus partially unavoidable.

Policy Note I – Tracking mental healthcare trajectories using observational data enables policy makers to make more informative budget allocation decisions.

After early identification of mental health trajectories, effective treatments could mitigate the norm-deviated pattern of mental health problems in youth. In **Chapter 3** we used experimental data of the largest RCT in youth mental health, the NIMH's Multimodal Treatment Study of Children with ADHD (MTA study). The MTA study has been designed to evaluate the clinical effectiveness of the leading treatments for ADHD (medication, behavioural therapy, combination thereof) relative to community-delivered treatment (control condition) and has defined guidelines for ADHD treatment globally. We re-used the experimental data to understand the cost-effectiveness of these treatments beyond the clinical setting, and therefore used juvenile delinquency as societal outcome. To do so, we developed a three-state continuous-time Markov model (no delinquency, minor to moderate

delinquency, serious delinquency) to extrapolate the data ten years beyond the trial period. This innovative model structure in youth mental health enabled to calculate the number of years that children do not enter the worst societal state, in our case serious delinquent behaviour, stratified by treatment mode. The economic evaluation revealed that the controlled treatments turned out to be inferior in terms of cost-effectiveness to the control condition, as the extra costs did not outweigh the impact on juvenile delinquency.

Policy Note II – It is crucial to take broader societal outcomes into account evaluating youth mental health policies, as the burden often extends beyond the health system.

Education is one the most critical predictors of youth development and contributes to the cumulative human capital of society, hence increasing its international position. Previous research has shown that increasing compulsory schooling benefits youth development. Yet we do not know how to monitor the implementation of these compulsory schooling reforms with regard to their effects on the academic performance of youth. Using administrative data we exploited a Dutch law reform in **Chapter 4** to determine the effect of tightening the monitoring of compulsory education on pupil's test scores. The law reform legally transferred monitoring from the municipality to the Dutch Inspectorate of Education with the purpose to get more pupils to attend class. As a result, we observe more registered test scores of low-ability pupils in the treatment group, causing selective attrition. We corrected for this non-random sample selection in our study and conclude that such a monitoring reform significantly increases the test scores of children, both in the first and last class of primary school. Moreover, the heterogeneity analyses reveal that care pupils and children from an ethnic minority or low socio-economic environment benefited more from the policy.

Policy Note III – Understanding the effect of youth policies for different subgroups is crucial, as it informs policy makers on the accessibility of the policy.

We know from **Chapters 2 and 4** that mental health is persistent during adolescence, and that education is a crucial aspect of youth development. In **Chapter 5** we combined these aspects and investigated the long-run returns to government expenditures in special needs education. The Dutch government provided regular primary and secondary schools from the 1st of August 2003 with additional funding for the inclusion of special needs students. In the

first part of the study we focused on the impact of these government expenditures on special needs student's mental health. The regular students form the reference group, and so we used their mental health trajectory as counterfactual for the course of mental health of special needs students if the policy had not been implemented. From this part we concluded that the additional funding substantially improved the mental health of special needs students. Heterogeneity analyses present a cautionary note, as girls and children from an ethnic minority or a low socio-economic environment benefited less from the policy. In the second part we responded to the commonly asked policy question in youth mental health: "Did the investment pay off?". The answer is "Yes", as we demonstrate a cost/benefit-ratio of 0.69 when extrapolating the found mental health effect to life-cycle returns in mental healthcare use, educational attainment and labour market participation.

Policy Note IV – Investments in youth mental health should be seen from a life-cycle perspective in order not to underestimate the benefits.

Dutch Summary

Mentale gezondheid is een essentieel onderdeel van het algemeen welzijn en functioneren binnen de maatschappij en daarmee bepalend voor de samenleving. Mentale gezondheidsproblemen nemen een steeds groter aandeel in van de ziektelast en -kosten wereldwijd (GBD, 2017). In Nederland zijn in totaal 1,1 miljoen mensen geregistreerd als patiënt in de geestelijke gezondheidszorg (GGZ) (Vektis Intelligence, 2017). De ziektekosten geassocieerd met de GGZ zijn gelijk aan 1,2% van het bruto binnenlands product en 6,8% van de totale ziektekosten in Nederland (CBS, 2019, 2020). Dit is echter nog maar een onderschatting van de daadwerkelijke kosten, aangezien de daadwerkelijke impact op de economie verborgen zit in de indirecte kosten. Mentale gezondheidsproblemen ontstaan veelal in de kindertijd en adolescentie (Kessler et al., 2007) en leiden in veel gevallen tot een afwijkende levensloop. Enkele voorbeelden zijn slechtere schoolresultaten (Currie, 2006; De Zeeuw *et al.*, 2017), lagere arbeidsparticipatie (Frijters *et al.*, 2014) en een lager inkomen (OECD, 2012). Mentale gezondheidsproblemen raakt niet alleen het individu, maar ook de omgeving en maatschappij, zoals een toenemende druk op mantelzorg van omringende verzorgers en jeugdhulp (OECD, 2014).

We weten dat veel psychische problemen hun oorsprong vinden in de jeugd. We weten echter veel minder in hoeverre zorgtrajecten en problemen persisteren tijdens de adolescentie nadat de kinderen en jongeren zijn verwezen naar de GGZ? Wordt de persistentie in de jeugd GGZ grotendeels beïnvloed door onderliggende individuele patiëntkenmerken? En wat is de gevoeligheid van deze zorgtrajecten met betrekking tot hervormingen van het Nederlandse gezondheidsstelsel? In Hoofdstuk 2 gebruikten we routinematig verzamelde gegevens van meer dan 80.000 Nederlandse kinderen en jongeren uit het Psychiatrisch Casus Register Noord-Nederland (PCR-NN) om deze vragen te beantwoorden. Eerdere onderzoeken naar de persistentie van jeugd GGZ hielden geen rekening met de rol van onderliggende individuele patiëntkenmerken. Het schatten van de persistentie van zorg met behulp van standaard lineaire modellen leverde vertekende resultaten op vanwege de niet-waargenomen heterogeniteit en de dynamische structuur van het onderliggende model. Wij hebben de persistentie van jeugd GGZ geschat, rekening houdend met deze statistische aspecten. Alle resultaten lieten een positieve coëfficiënt zien tussen 0 en 1, wat aangeeft dat jongeren na een plotselinge toename van zorg zouden terugkeren naar hun basisniveau van GGZ-zorggebruik. Bovendien bleek de persistentie van jeugd GGZ niet te worden beïnvloed door grote hervormingen van het gezondheidsstelsel in het afgelopen decennium en is een substantieel deel van de persistentie te wijten aan individuele patiëntkenmerken en is dus gedeeltelijk onvermijdelijk.

Na vroege identificatie van jeugd-GGZ-trajecten zouden effectieve behandelingen het normafwijkende patroon van psychische problemen bij jongeren kunnen verzachten. In Hoofdstuk 3 gebruikten we experimentele gegevens van een vooraanstaande klinische trial in mentale gezondheid bij jeugd, de Multimodal Treatment Study of Children with ADHD (MTA studie). De MTA studie is gebruikt om de klinische effectiviteit van reguliere behandelingen voor ADHD (medicatie, gedragstherapie, combinatie daarvan) te evalueren ten opzichte van door de gemeenschap geleverde behandeling (controleconditie) en heeft de wereldwijde richtlijnen voor ADHD-behandeling beïnvloed. We hergebruikten de experimentele gegevens om de kosteneffectiviteit van deze behandelingen buiten de klinische setting te onderzoeken met delinquentie als maatschappelijke uitkomst. Hiervoor hebben we een transitie-model met drie maatschappelijke toestanden ontwikkeld (geen delinquentie, lichte tot matige delinquentie, ernstige delinquentie) om voorbij het tijdsinterval van de klinische trial te extrapoleren. Deze innovatieve modelstructuur maakte het mogelijk het aantal jaren te berekenen dat kinderen niet in de slechtste maatschappelijke toestand terechtkomen, in ons geval ernstige delinquentie, na verschillende behandelingen. Uit de economische evaluatie bleek dat de reguliere behandelingen ondergeschikt zijn aan de controleconditie, aangezien de hogere kosten niet opwegen tegen de impact op jeugdcriminaliteit. Het is van cruciaal belang rekening te houden met bredere maatschappelijke uitkomsten bij effectmetingen in de jeugd GGZ, aangezien de invloed verder reikt dan alleen de klinische setting.

Onderwijsprestatie is een van de meest kritische voorspellers van de ontwikkeling van jongeren en draagt het bij aan het cumulatieve menselijke kapitaal van de samenleving. Eerder onderzoek heeft aangetoond dat het verlengen van de leerplicht de ontwikkeling van jongeren ten goede komt. Toch weten we niet hoe het verscherpt monitoren van de Leerplichtwet de onderwijsprestaties van jongeren beïnvloedt. In Hoofdstuk 4 maakten we gebruik van een Nederlandse wijziging in het monitoringssysteem van de Leerplichtwet om met behulp van administratieve gegevens de associatie te bepalen tussen de verwachte afname van het relatieve verzuim van leerlingen op hun testcores. Het beleid streeft naar meer leerlingen in de klas door een lager verzuim, en dus meer geregistreerde testcores, na invoering van de wetwijziging op 1 augustus 2007. Deze onnatuurlijke selectie zorgt voor een vertekend beeld, aangezien uit eerder onderzoek is gebleken dat verzuim vaker voorkomt bij leerlingen met lagere schoolprestaties. We hielden rekening met deze statistische kwestie in onze studie en concludeerden dat een aanscherping van het toezicht op de Leerplichtwet de testcores van kinderen aanzienlijk heeft verhoogd, zowel in de eerste als in de laatste klas van de basisschool. Bovendien lieten de aanvullende analyses zien dat zorgleerlingen en kinderen afkomstig van een etnische minderheid of een lager sociaaleconomische omgeving meer baat hadden bij de hervorming.

We weten uit Hoofdstukken 2 en 4 dat mentale gezondheidsproblemen persistent zijn tijdens adolescentie en dat onderwijs een cruciaal aspect is van de ontwikkeling van jongeren. In Hoofdstuk 5 combineerden we deze aspecten en onderzochten we de lange-termijn rendementen op overheidsuitgaven in het speciaal onderwijs. De Nederlandse regering heeft vanaf 1 augustus 2003 basis- en middelbare scholen voorzien van aanvullende financiering voor de opname van zorgleerlingen uit het speciaal onderwijs, bekend als het Rugzakbeleid. In het eerste deel van dit hoofdstuk richtten we ons op de impact van het Rugzakbeleid op de mentale gezondheid van zorgleerlingen. De reguliere leerlingen vormden de controlegroep en daarom gebruikten we hun ontwikkeling van mentale gezondheid voor het verloop van de mentale gezondheid van zorgleerlingen als het Rugzakbeleid niet was geïmplementeerd. Uit dit deel concludeerden we dat het Rugzakbeleid de mentale gezondheid van zorgleerlingen flink heeft verbeterd. Heterogeniteitsanalyses lieten wel zien dat meisjes en kinderen afkomstig van een etnische minderheid of een lagere sociaaleconomische omgeving minder baat hadden bij het beleid. In het tweede deel gingen we in op de veel voorkomende beleidsvraag in het jeugddomein: “Wat is de monetaire waarde van de investering?”. We hebben aangetoond dat het Rugzakbeleid een kosten/baten-ratio van 0,69 heeft als we het gevonden effect in mentale gezondheid extrapoleerden naar de te verwachte rendementen in GGZ-zorggebruik, opleidingsniveau en arbeidsmarktparticipatie over de levensloop.

Bibliography

A

Abadie, A., Chingos, M.M., & West, M.R. (2018). Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, **100**(4), 567-580.
doi:10.1162/rest_a_00732

Achenbach, T., & Rescorla, L. (2013). Achenbach system of empirically based assessment. In *Encyclopedia of autism spectrum disorders* (pp. 31-39). Springer, New York.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV)*. American Psychiatric Association, Washington.

Amin, V., Flores, C.A., Flores-Lagunes, A., & Parisian, D.J. (2016). The effect of degree attainment on arrests: Evidence from a randomized social experiment. *Economics of Education Review*, **54**(1), 259-273. doi:10.1016/j.econedurev.2016.02.006

Angrist, J.D., & Keueger, A.B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, **106**(4), 979-1014.
doi: 10.2307/2937954

Angrist, J.D., & Krueger, A.B. (1999.) Empirical Strategies in Labor Economics. In *Handbook of Labor Economics* (1277-1366) North Holland, Amsterdam.

Arellano, M., & Bond, S.R. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, **58**(2), 277-297 (1991). doi:10.2307/2297968

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, **68**(1), 29-51. doi:10.1016/0304-4076(94)01642-D

Asherson, P., Buitelaar, J.K., Faraone, S.V., & Rohde, L.A. (2016). Adult attention-deficit hyperactivity disorder: key conceptual issues. *The Lancet Psychiatry*, **3**(6), 568-578.
doi:10.1016/S2215-0366(16)30032-3

Athey, S. & Imbens, G. (2017). State of applied econometrics: Causality & policy evaluation. *Journal of Economic Perspectives*, **31**(2), 3-32. doi:10.1257/jep.31.2.3

Attanasio, O., Blundell, R., Conti, G., & Mason, G. (2020). Inequality in socio-emotional skills: A cross-cohort comparison. *Journal of Public Economics*, forthcoming.
doi:10.1016/j.jpubeco.2020.104171

Aydin, S., Crone, M.R., Siebelink, B.M., Vermeiren, R.R., Numans, M.E., & Westenberg, P.M. (2020). Recognition of anxiety disorders in children: a cross-sectional vignette-based survey among general practitioners. *BMJ Open*, **10**(4), e035799. doi:10.1136/bmjopen-2019-035799

B

Banerjee, A., & Duflo, E. (2006). Addressing absence. *Journal of Economic perspectives*, **20**(1), 117-132. doi:10.1257/089533006776526139

Barkley, R.A., Fischer, M., Smallish, L., & Fletcher, K. (2004). Young adult follow-up of hyperactive children: antisocial activities and drug use. *The Journal of Child Psychology and Psychiatry*, **45**(2), 195-211. doi: /10.1111/j.1469-7610.2004.00214.x

Bayer, J.K., Ukoumunne, O.C., Lucas, N., Wake, M., Scalzo, K., & Nicholson, J.M. (2011). Risk factors for childhood mental health symptoms: national longitudinal study of Australian children. *Pediatrics*, **128**(4), e865-e879. doi:10.1542/peds.2011-0491

Becker, G.S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press, Chicago.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, **80**(6), 2369-2429. doi:10.3982/ECTA9626

Bénabou, R., Kramarz, F., & Prost, C. (2009). The French zones d'éducation prioritaire: Much ado about nothing? *Economics of Education Review*, **28**(3), 345-356. doi:10.1016/j.econedurev.2008.04.005

Bhuller, M., Mogstad, M., & Salvanes, K.G. (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, **35**(4), 993-1030. doi: 10.1086/692509

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, **87**(1), 115-143. doi:10.1016/S0304-4076(98)00009-8

Bond, S.R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, **1**(2), 141-162. doi:10.1007/s10258-002-0009-9

Bosker, R.J., Van der Velden, R.K.W., & Hofman, W.H.A. (1985). Een generatie geselecteerd. deel i: de schoolloopbanen.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis Part III: multivariate data analysis—choosing a model and assessing its adequacy and fit. *British Journal of Cancer*, **89**(4), 605-611. doi:10.1038/sj.bjc.6601120

Brunello, G., Fort, M., Schneeweis, N., & Winter-Ebmer, R. (2016). The causal effect of education on health: What is the role of health behaviors? *Health Economics*, **25**(3), 314-336. doi:10.1002/hec.3141

C

Cao, Q., Buskens, E., Feenstra, T., Jaarsma, T., Hillege, H., & Postmus, D. (2016). Continuous-time semi-Markov models in health economic decision making: an illustrative example in heart failure disease management. *Medical Decision Making*, **36**(1), 59-71. doi:10.1177/0272989X15593080

CBS; Statistics Netherlands (2017). Deceased; sex, age and marital Status. Retrieved from <http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=37168> (Accessed September 21, 2017)

CBS; Statistics Netherlands. (2017a). Employed and non-employed population; average income and employment position. Retrieved from <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83687NED/table?ts=1572421308240> (Accessed November 12, 2019).

CBS; Statistics Netherlands. (2017b). Health, lifestyle, use and provision of care, causes of death; key figures. Retrieved from <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81628NED/table?fromstatweb> (Accessed November 8, 2019)

CBS; Statistics Netherlands. (2017c). Working population; average income; key figures 2017. Retrieved from <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83686NED/table?fromstatweb> (Accessed November 12, 2019)

CBS; Statistics Netherlands (2019). Healthcare expenditures; healthcare providers and financing. Retrieved from <https://opendata.cbs.nl/statline/?ts=1528891083261#/CBS/nl/dataset/84053NED/table> (Accessed April 2, 2020).

CBS; Statistics Netherlands (2020). Structure domestic product (GDP); national accounts. Retrieved from <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84087NED/table?ts=1587544468304> (Accessed April 2, 2020).

Chassang, S., Miquel, P.I., & Snowberg, E. (2012). Selective trials: A principal-agent approach to randomized controlled experiments. *American Economic Review*, **102**(4), 1279-1309. doi:10.1257/aer.102.4.1279

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, **104**(9), 2593-2632. doi:10.1257/aer.104.9.2593

Chetty, R., Hendren, N., & Katz, L.F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review*, **106**(4), 855-902. doi:10.1257/aer.20150572

Contoyannis, P., & Li, J. (2017). The dynamics of adolescent depression: an instrumental variable quantile regression with fixed effects approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **180**(3), 907-922. doi:10.1111/rssa.12225

CPB; Netherlands Bureau for Economic Policy Analysis (2017). Discontovoet onderwijs. Retrieved from <https://www.cpb.nl/publicatie/discontovoet-onderwijs> (Accessed November 16, 2019).

Crespo, L., López-Noval, B., & Mira, P. (2014). Compulsory schooling, education, depression and memory: New evidence from SHARELIFE. *Economics of Education Review*, **43**(1), 36-46. doi:10.1016/j.econedurev.2014.09.003

Cullen, J.B., Jacob, B.A., & Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, **74**(5), 1191-1230. doi:10.1111/j.1468-0262.2006.00702.x

Cumby, R., Huizinga, J. (1992). Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica*, **60**(1), 185-195. doi:10.2307/2951684

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, **47**(1), 87-122. doi:10.1257/jel.47.1.87

Currie, J., & Stabile, M. (2006). Child mental health and human capital accumulation: the case of ADHD. *Journal of Health Economics*, **25**(6), 1094-1118. doi:10.1016/j.jhealeco.2006.03.001

Currie, J., Stabile, M., Manivong, P., & Roos, L.L. (2010). Child health and young adult outcomes. *Journal of Human Resources*, **45**(3), 517-548. doi:10.3368/jhr.45.3.517

Cygan-Rehm, K., & Maeder, M. (2013). The effect of education on fertility: Evidence from a compulsory schooling reform. *Labour Economics*, **25**(3), 35-48. doi:10.1016/j.labeco.2013.04.015

D

D'Agostino, R.B. (2007). Estimating treatment effects using observational data. *JAMA*, **297**(3), 314-316. doi:10.1001/jama.297.3.314

Daley, T.C. (2004). From symptom recognition to diagnosis: children with autism in urban India. *Social Science & Medicine*, **58**(7), 1323-1335. doi:10.1016/S0277-9536(03)00330-7

D'Amico, F., Knapp, M., Beecham, J., Sandberg, S., Taylor, E., & Sayal, K. (2014). Use of services and associated costs for young adults with childhood hyperactivity/conduct problems: 20-year follow-up. *The British Journal of Psychiatry*, **204**(6), 441-447. doi:10.1192/bjp.bp.113.131367

De Haan, M. (2017). The Effect of Additional Funds for Low-ability Pupils: A Non-parametric Bounds Analysis. *The Economic Journal*, **127**(599), 177-198. doi: 10.1111/ecoj.12221

Deming, D.J. (2011). Better schools, less crime?. *The Quarterly Journal of Economics*, **126**(4), 2063-2115. doi:10.1093/qje/qjr036

Devereux, P.J., & Hart, R.A. (2010). Forced to be rich? Returns to compulsory schooling in Britain. *The Economic Journal*, **120**(549), 1345-1364. doi: 10.1111/j.1468-0297.2010.02365.x

De Zeeuw, E.L., van Beijsterveldt, C.E., Ehli, E.A., de Geus, E.J., & Boomsma, D.I. (2017). Attention deficit hyperactivity disorder symptoms and low educational achievement: Evidence supporting a causal hypothesis. *Behavior Genetics*, **47**(3), 278-289. doi: 10.1007/s10519-017-9836-4

Doshi, J.A., Hodgkins, P., Kahle, J., Sikirica, V., Cangelosi, M.J., Setyawan, J., Erder, M.H., & Neumann, P.J. (2012). Economic impact of childhood and adult attention-deficit/hyperactivity disorder in the United States. *Journal of the American Academy of Child & Adolescent Psychiatry*, **51**(10), 990-1002. doi:10.1016/j.jaac.2012.07.008

Drapeau, A., Boyer, R., & Diallo, F.B. (2011). Discrepancies between survey and administrative data on the use of mental health services in the general population: findings from a study conducted in Quebec. *BMC Public Health*, **11**(837), 1-10. doi:10.1186/1471-2458-11-837

Ducat, L., Philipson, L.H., & Anderson, B.J. (2014). The mental health comorbidities of diabetes. *JAMA*, **312**(7), 691-692. doi:10.1001/jama.2014.8040

E

Elliott, D., Huizinga, D., & Ageton, S. (1985). Explaining delinquency and drug use. Sage, Beverly Hills.

F

Faber, A., Van Agthoven, M., Kalverdijk, L.J., Tobi, H., De Jong-Van den Berg, L.T.W., Annemans, L., & Postma, M.J. (2008). Long-acting methylphenidate-OROS in youths with attention-deficit/hyperactivity disorder suboptimally controlled with immediate-release methylphenidate: a study of cost effectiveness in The Netherlands. *CNS Drugs*, **22**(2), 157-170. doi:10.2165/00023210-200822020-00006

Farmer, E.M., Stangl, D.K., Burns, B.J., Costello, E.J., Angold, A. (1999). Use, persistence, and intensity: Patterns of care for children's mental health across one year. *Community Mental Health Journal*, **35**(1), 31-46. doi:10.1023/A:1018743908617

Faraone, S.V., Asherson, P., & Banaschewski, T. *et al.* (2015). Attention-Deficit/Hyperactivity disorder. *Nature Reviews Disease Primers*, **1**(15020), 1-18. doi:10.1038/nrdp.2015.20

Feenberg, D., & Skinner, J. (1994). The risk and duration of catastrophic health care expenditures. *The Review of Economics and Statistics*, **76**(4), 633-647. doi:10.2307/2109767

Feldman, H.D., & Reiff, M.I. (2014). Attention-Deficit/Hyperactivity Disorder in children and adolescents. *New England Journal of Medicine*, **370**(9), 838-846. doi:10.1056/NEJMcip1307215

Fenton, W. S., & Stover, E. S. (2006). Mood disorders: cardiovascular and diabetes comorbidity. *Current Opinion in Psychiatry*, **19**(4), 421-427. doi: 10.1097/01.yco.0000228765.33356.9f

Fischer, M., Karlsson, M., & Nilsson, T. (2013). Effects of compulsory schooling on mortality: evidence from Sweden. *International Journal of Environmental Research and Public Health*, **10**(8), 3596-3618. doi:10.3390/ijerph10083596

French, E., & Jones, J.B. (2004). On the distribution and dynamics of health care costs. *Journal of Applied Econometrics*, **19**(6), 705-721. doi:10.1002/jae.790

Frijters, P., Johnston, D.W., & Shields, M.A. (2014). The effect of mental health on employment: Evidence from Australian panel data. *Health Economics*, **23**(9), 1058-1071. doi:10.1002/hec.3083

G

Galama, T.J., Lleras-Muney, A., & van Kippersluis, H. (2018). The Effect of education on health and mortality: A review of experimental and quasi-experimental evidence (No. w24225). National Bureau of Economic Research. doi:10.3386/w24225

Gibbons, S., McNally, S., & Viarengo, M. (2018). Does additional spending help urban schools? An evaluation using boundary discontinuities. *Journal of the European Economic Association*, **16**(5), 1618-1668. doi:10.1093/jeea/jvx038

Gold, M.R., Siegel, J.E., Russell, L.B., & Weinstein, M.C. (1996). Cost-effectiveness in health and medicine. Oxford University Press, New York.

Goodman, E., Slap, G.B., & Huang, B. (2003). The public health impact of socioeconomic status on adolescent depression and obesity. *American Journal of Public Health*, **93**(11), 1844-1850. doi:10.2105/AJPH.93.11.1844

Grimard, F., & Parent, D. (2007). Education and smoking: Were Vietnam war draft avoiders also more likely to avoid smoking? *Journal of Health Economics*, **26**(5), 896-926. doi:10.1016/j.jhealeco.2007.03.004

Grossman, M. (2015). The relationship between health and schooling: What's new? (No. w21609). National Bureau of Economic Research.

H

Hakkaart-van Roijen, L., Zwirs, B.W.C., Bouwmans, C., Tan, S.S., Schulpen, T.W.J., Vlasveld, L., & Buitelaar, J.K. (2007). Societal costs and quality of life of children suffering from attention deficient hyperactivity disorder (ADHD). *European Child & Adolescent Psychiatry*, **16**(5), 316-326. doi:10.1007/s00787-007-0603-6

Hanushek, E.A. (2003). The failure of input-based schooling policies. *The Economic Journal*, **113**(485), F64-F98. doi:10.1111/1468-0297.00099

Hanushek, E.A. (2006). School resources. In *Handbook of the Economics of Education* (pp. 865-908). North Holland, Amsterdam. doi:10.1016/S1574-0692(06)02014-9

Harmon, C., & Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *The American Economic Review*, **85**(5), 1278-1286.

Heckman, J.J. (1991). Identifying the hand of past: Distinguishing state dependence from heterogeneity. *American Economic Review*, **81**(2), 75-79.

Heckman, J.J., Ichimura, H., & Todd, P.E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, **64**(4), 605-654. doi:10.2307/2971733

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, **65**(2), 261-294. doi:10.1111/1467-937X.00044

Holmlund, H., McNally, S., & Viarengo, M. (2010). Does money matter for schools? *Economics of Education Review*, **29**(6), 1154-1164. doi:10.1016/j.econedurev.2010.06.008

Holtz-Eakin, D., Newey, W., & Rosen, H.S. (1988). Estimating vector autoregressions with panel data. *Econometrica*, **56**(6), 1371-1395. doi:10.2307/1913103

Hsiao, C., Pesaran, M.H., & Tahmiscioglu, A.K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics*, **109**(1), 107-150. doi:10.1016/S0304-4076(01)00143-9

Huisman, M., Oldehinkel, A.J., de Winter, A., Minderaa, R.B., de Bildt, A., Huizink, A.C., Verhulst, F.C., & Ormel, J. (2008). Cohort profile: The Dutch 'TRacking adolescents' individual lives' survey'; TRAILS. *International Journal of Epidemiology*, **37**(6), 1227-1235. doi:10.1093/ije/dym273

I

Institute for Health Metrics and Evaluation (2017). Findings from the Global Burden of Disease Study 2017. IHME, Seattle.

J

Jensen, P. S., Garcia, J. A., Glied, S., Crowe, M., Foster, M., Schlander, M., Hechtman, L. (2005). Cost-effectiveness of ADHD treatments: Findings from the multimodal treatment study of children with ADHD. *American Journal of Psychiatry*, **162**(9), 1628-36. doi:10.1176/appi.ajp.162.9.1628

Jensen, P.S., Arnold, L.E., Swanson, J.M. *et al.* (2007). 3-Year follow-up of the NIMH MTA study. *Journal of the American Academy of Child and Adolescent Psychiatry*, **46**(8), 989-1002. doi:10.1097/CHI.0b013e3180686d48

Jones, D.E., & Foster, E.M. (2009). Service use patterns for adolescents with ADHD and comorbid conduct disorder. *Journal of Behavioral Health Services & Research*, **36**(4): 436-449. doi:10.1007/s11414-008-9133-3

Johnston, D., Propper, C., Pudney, S., & Shields, M. (2014). Child mental health and educational attainment: multiple observers and the measurement error problem. *Journal of Applied Econometrics*, **29**(6), 880-900. doi:10.1002/jae.2359

Jörg, F., Visser, E., Ormel, J., Reijneveld, S. A., Hartman, C.A., & Oldehinkel, A.J. (2016). Mental health care use in adolescents with and without mental disorders. *European Child & Adolescent Psychiatry*, **25**(5), 501-508. doi:10.1007/s00787-015-0754-9

Jürges, H., Reinhold, S., & Salm, M. (2011). Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. *Economics of Education Review*, **30**(5), 862-872. doi: 10.1016/j.econedurev.2011.04.002

K

Kemptoner, D., Jürges, H., & Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics*, **30**(2), 340-354. doi:10.1016/j.jhealeco.2011.01.004

Kerig, P.K., Ludlow, A., & Wenar, C. (2012). *Developmental Psychopathology*. McGraw-Hill Education, New York.

Kessler, R.C., Amminger, G.P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustun, T.B. (2007). Age of onset of mental disorders: a review of recent literature. *Current Opinion in Psychiatry*, **20**(4), 359-364. doi:10.1097/YCO.ob013e32816ebc8c

Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., Rohde, L.A., Srinath, S., Ulkuer, N., & Rahman, A. (2011). Child and adolescent mental health worldwide: evidence for action. *The Lancet*, **378**(9801), 1515-1525. doi:10.1016/S0140-6736(11)60827-1

Knapp, M., Ardino, V., Brimblecombe, N., Evans-Lacko, S., Lemmi, V., King, D., & Snell, T. (2016). *Youth mental health: new economic evidence*. Personal social services research unit, London.

L

Le, H.H., Hodgkins, P., Postma, M.J., Kahle, J., Sikirica, V., Setyawan, J., Erder, M.H., & Doshi, J.A. (2014). Economic impact of childhood/adolescent ADHD in a European setting: the Netherlands as a reference case. *European Child and Adolescent Psychiatry*, **23**(7), 587-598. doi:10.1007/s00787-013-0477-8

Lee, D.S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, **76**(3), 1071-1102. doi:10.1111/j.1467-937X.2009.00536.x

Lee, S.S., & Hinshaw, S.P. (2004). Severity of adolescent delinquency among boys with and without attention deficit hyperactivity disorder: predictions from early antisocial behavior and peer status. *Journal of Clinical Child & Adolescent Psychology*, **33**(4), 705-716. doi:10.1207/s15374424jccp3304_6

Lee, S.H., Ripke, S., Neale, B.M. *et al.* (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, **45**(9), 984-994. doi:10.1038/ng.2711

Leigh, A., & Ryan, C. (2008). Estimating returns to education using different natural experiment techniques. *Economics of Education Review*, **27**(2), 149-160. doi:10.1016/j.econedurev.2006.09.004

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economics and Statistics*, **89**(4), 721-736. doi:10.1162/rest.89.4.721

Lochner, L., & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review*, **94**(1), 155-189. doi:10.1257/000282804322970751

Loeber, R., Stouthamer-Loeber, M., Van Kammen, W.B., & Farrington, D.P. (1989). Development of a new measure of self-reported antisocial behavior for young children: prevalence and reliability. In: *Cross-national Research in Self-reported Crime and Delinquency* (pp. 203-225). Springer, Dordrecht.

Loeber, R., Stouthamer-Loeber, M., Van Kammen, W., & Farrington, D.P. (1991). Initiation, escalation and desistance in juvenile offending and their correlates. *Journal of Criminal Law and Criminology*, **82**(1), 36-82. doi:0091-4169/91/8201-36

Loeber, R., Farrington, D.P., Stouthamer-Loeber, M., & Van Kammen, W.B. (1998). Antisocial behavior and mental health problems. Explanatory factors in childhood and adolescence. Erlbaum, Mahwah.

M

Machin, S., McNally, S., & Meghir, C. (2010). Resources and standards in urban schools. *Journal of Human Capital*, **4**(4), 365-393. doi:10.1086/658634

Mandell, D.S., Wiggins, L.D., Carpenter, L.A., *et al.* (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health*, **99**(3), 493-498. doi:10.2105/AJPH.2007.131243

Matza, L.S., Paramore, C., & Prasad, M. (2005). A review of the economic burden of ADHD. *Cost Effectiveness and Resource Allocation*, **3**(1), 5-13. doi:10.1186/1478-7547-3-5

Merikangas, K.R., Nakamura, E.F., & Kessler, R.C. (2009). Epidemiology of mental disorders in children and adolescents. *Dialogues in Clinical Neuroscience*, **11**(1), 7-20.

Minne, B., Webbink, H.D., & Wiel, H.P. (2009). Zorg om zorgleerlingen: een blik op beleid, aantal en kosten van jonge zorgleerlingen. CPB; Netherlands Bureau for Economic Policy Analysis, The Hague.

Molina, B.S., Flory, K., Hinshaw, S.P., *et al.* (2007). Delinquent behavior and emerging substance use in the MTA at 36 months: prevalence, course, and treatment effects. *Journal of the American Academy of Child and Adolescent Psychiatry*, **46**(8), 1028-1040. doi:10.1097/chi.0b013e3180686d96

Molina, B.S.G., Hinshaw, S.P., Swanson, J.M., *et al.* (2009). The MTA at 8 years: prospective follow-up of children treated for combined-type ADHD in a multisite study. *Journal of the American Academy of Child and Adolescent Psychiatry*, **48**(5), 484-500. doi:10.1097/CHI.0b013e31819c23d0

Moral-Benito, E., Allison, P., & Williams, R. (2019). Dynamic panel data modelling using maximum likelihood: an alternative to Arellano-Bond. *Applied Economics*, **51**(20), 2221-2232. doi:10.1080/00036846.2018.1540854

MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/Hyperactivity disorder. *Archives of General Psychiatry*, **56**(12), 1073-1086. doi:10.1001/archpsyc.56.12.1073

MTA Cooperative Group. (2004). National Institute of Mental Health Multimodal Treatment Study of ADHD follow-up: changes in effectiveness and growth after the end of treatment. *Pediatrics*, **113**(4), 762-769. doi:10.1542/peds.113.4.754

N

Nakash, O., Levav, I., Aguilar-Gaxiola, S. *et al.* (2014). Comorbidity of common mental disorders with cancer and their treatment gap: findings from the World Mental Health Surveys. *Psycho-Oncology*, **23**(1), 40-51. doi:10.1002/pon.3372

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, **49**(6), 1417-1426. doi:10.2307/1911408

Nicolaie, M.A., Van Houwelingen, H.C., & Putter, H. (2010). Vertical modelling: A pattern mixture approach for competing risks modelling. *Statistics in Medicine*, **29**(11), 1190-1205. doi:10.1002/sim.3844

O

OECD (2012). *Sick on the job? Myths and realities about mental health and work*. OECD Publishing, Paris. doi:10.1787/9789264124523-en.

OECD (2014). *Making mental health count: The social and economic costs of neglecting mental health care*. OECD Publishing, Paris. doi:10.1787/9789264208445-en.

Oldehinkel, A.J., Rosmalen, J.G., Buitelaar, J.K., Hoek, H. W., Ormel, J., Raven, D., Verhulst, F.C., Volleberg, W.A.M., & Hartman, C.A. (2015). Cohort profile update: the tracking adolescents' individual lives survey (TRAILS). *International Journal of Epidemiology*, **44**(1), 76-76n. doi:10.1093/ije/dyu225

Oreopoulos, P. (2006a). The compelling effects of compulsory schooling: Evidence from Canada. *Canadian Journal of Economics*, **39**(1), 22-52. doi:10.1111/j.0008-4085.2006.00337.x

Oreopoulos, P. (2006b). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, **96**(1), 152-175. doi:10.1257/000282806776157641

Oreopoulos, P. (2007). Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. *Journal of Public Economics*, **91**(11), 2213-2229. doi:10.1016/j.jpubeco.2007.02.002

P

Papke, L.E. (2005). The effects of spending on test pass rates: evidence from Michigan. *Journal of Public Economics*, **89**(5), 821-839. doi:10.1016/j.jpubeco.2004.05.008

Polanczyk, G.V., Salum, G.A., Sugaya, L.S., Caye, A., & Rohde, L.A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, **56**(3), 345-365. doi:10.1111/jcpp.12381

R

Ravesteijn, B., Schachar, E.B., Beekman, A.T., Janssen, R.T., & Jeurissen, P.P. (2017). Association of cost sharing with mental health care use, involuntary commitment, and acute care. *JAMA Psychiatry*, **74**(9), 932-939. doi:10.1001/jamapsychiatry.2017.1847

Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, **9**(1), 86-136. doi:10.1177/1536867X0900900106

Roodman, D. (2009b). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, **71**(1), 135-158. doi:10.1111/j.1468-0084.2008.00542.x

Rosenbaum, P.R., & Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, **39**(1), 33-38.

Roy, J., & Schurer, S. (2013). Getting stuck in the blues: persistence of mental health problems in Australia. *Health Economics*, **22**(9), 1139-1157. doi:10.1002/hec.2967

Ruijs, N. (2017). The impact of special needs students on classmate performance. *Economics of Education Review*, **58**(2), 15-31. doi:10.1016/j.econedurev.2017.03.002

Ruijs, N.M., & Peetsma, T.T. (2009). Effects of inclusion on students with and without special educational needs reviewed. *Educational Research Review*, **4**(2), 67-79. doi: 10.1016/j.edurev.2009.02.002

S

Sacerdote, B. (2002). The nature and nurture of economic outcomes. *American Economic Review*, **92**(2), 344-348. doi:10.1257/000282802320191589

Satterfield, J.H., Faller, K.J., Crinella, F.M., Schell, A.M., Swanson, J.M., & Homer, L.D. (2007). A 30-year prospective follow-up study of hyperactive boys with conduct problems: adult criminality. *Journal of the American Academy of Child & Adolescent Psychiatry*, **46**(5), 601-610. doi:10.1097/chi.obo13e318033ff59

Schawo, S., Van der Kolk, A., Bouwmans, C., Annemans, L., Postma, M.J., Buitelaar, J., Van Agthoven, M., & Hakkaart-van Roijen, L. (2015). Probabilistic Markov model estimating cost-effectiveness of Methylphenidate Osmotic-Release Oral System versus Immediate-Release Methylphenidate in children and adolescents: which information is needed? *PharmacoEconomics*, **33**(5), 489-509. doi:10.1007/s40273-015-0259-x

Shenkman, E., Knapp, C., Sappington, D., Vogel, B., & Schatz, D. (2007). Persistence of high health care expenditures among children in Medicaid. *Medical Care Research and Review*, **64**(3), 304-330. doi:10.1177/1077558707299864

Silles, M.A. (2011). The effect of schooling on teenage childbearing: evidence using changes in compulsory education laws. *Journal of Population Economics*, **24**(2), 761-777. doi:10.1007/s00148-010-0334-8

Soares, M.O., Canto, E., & Castro, L. (2012). Continuous time simulation and discretized models for cost-effectiveness analysis. *Pharmacoeconomics*, **30**(12), 1101-1117. doi:10.2165/11599380-000000000-00000

Swensen, A.R., Birnbaum, H.G., Secnik, K., Marynchenko, M., Greenberg, P., & Claxton, A.M.I. (2003). Attention-deficit/hyperactivity disorder: increased costs for patients and their families. *Journal of the American Academy of Child & Adolescent Psychiatry*, **42**(12), 1415-1423. doi:10.1097/00004583-200312000-00008

Swensen, A.R., Birnbaum, H.G., Ben Hamadi, R., Greenberg, P., & Cremieux, P.Y. (2004). Incidence and costs of accidents among attention-deficit/hyperactivity disorder patients. *Journal of Adolescent Health*, **35**(4), e1-e9. doi:10.1016/j.jadohealth.2003.12.003

T

Tambour, M., Zethraeus, N., & Johannesson, M. (1998). A note on confidence intervals in cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care*, **14**(3), 467-471. doi:10.1017/S0266462300011442

V

Van der Klaauw, W. (2008). Breaking the link between poverty and low student achievement: An evaluation of Title I. *Journal of Econometrics*, **142**(2), 731-756. doi:10.1016/j.jeconom.2007.05.007

Van der Schans, J., Kotsopoulos, N., Hoekstra, P.J., Hak, E., & Postma, M.J. (2015). Cost-effectiveness of extended-release methylphenidate in children and adolescents with attention-deficit/hyperactivity disorder sub-optimally treated with immediate release methylphenidate. *PLoS One*, **10**(5), e0127237. doi:10.1371/journal.pone.0127237

Van Herwaarden, S., Wallenburg, I., Messelink, J., & Bal, R. (2020). Opening the black box of diagnosis-related groups (DRGs): unpacking the technical remuneration structure of the Dutch DRG system. *Health Economics, Policy and Law*, **15**(2), 196-209. doi:10.1017/S1744133118000324

Van Lieshout, M., Luman, M., Twisk, J.W. *et al.* (2016). A 6-year follow-up of a large European cohort of children with attention-deficit/hyperactivity disorder-combined subtype: outcomes in late adolescence and young adulthood. *European Child and Adolescent Psychiatry*, **25**(9), 1007-1017. doi:10.1007/s00787-016-0820-y

Van Rosmalen, J., Toy, M., & O'Mahony, J.F. (2013). A mathematical approach for evaluating Markov models in continuous time without discrete-event simulation. *Medical Decision Making*, **33**(6), 767-779. doi:10.1177/0272989X13487947

Vektis Intelligence (2017). How many patients are registered in mental healthcare? Retrieved from <https://www.zorgprismapubliek.nl/producten/geestelijke-gezondheidszorg/geestelijke-gezondheidszorg/row-1/hoeveel-mensen-worden-behandeld-in-deggz/?tab=aantallen&jaar=2017&meetwaarde=ap> (Accessed November 8, 2019).

W

Wettenbank. (2003). Voorlichting over leerlinggebonden financiering (lgf). Retrieved from <https://wetten.overheid.nl/BWBR0015058/2003-05-28/o> (Accessed October 10, 2018).

Wettenbank (2012). Leerplichtwet 1969. Retrieved from <https://wetten.overheid.nl/BWBR0002628/2020-07-01> (Accessed November 16, 2019).

Wettenbank. (2014). Regeling stimulering invoering passend onderwijs in het primair onderwijs, voortgezet onderwijs en cluster 1 en 2, 2013{2014. Retrieved from <https://wetten.overheid.nl/BWBR0033584/2013-08-01> (Accessed October 10, 2018).

WHO. (2013). Investing in mental health: evidence for action. WHO, Geneva.

WHO. (2018). Global health observatory data repository. Causes of child death in the Netherlands. Retrieved from <http://apps.who.int/gho/data/view.main-wpro.ghe3002015-NLD?lang=en> (Accessed April 11, 2018).

WHO. (2019). Mental health. Retrieved from <https://www.who.int/news-room/facts-in-pictures/detail/mental-health> (Accessed April 2, 2020).

Wichstrøm, L., Belsky, J., & Steinsbekk, S. (2017). Homotypic and heterotypic continuity of symptoms of psychiatric disorders from age 4 to 10 years: a dynamic panel model. *Journal of Child Psychology and Psychiatry*, **58**(11), 1239-1247. doi:10.1111/jcpp.12754

Wolfgang, M., Figlio, R., Tracy, P., & Singer, S. (1985). The national survey of crime severity. Government Printing Office, Washington.

Wu, E.Q., Hodgkins, P., Ben-Hamadi, R., Setyawan, J., Xie, J., Sikirica, V., Du, E.X., Yan, S.Y., & Erder, M.H. (2012). Cost-effectiveness of pharmacotherapies for attention-deficit hyperactivity disorder: a systematic literature review. *CNS Drugs*, **26**(7), 581-600. doi:10.2165/11633900-000000000-00000