Direct and indirect effects of public policies and interventions

Direct and Indirect Effects of Public Policies and Interventions

ANNELORE MARIE-CLAIRE VERHAGEN

ISBN 978-90-832301-2-2

Cover and printing: Proefschriftenprinten.nl

Direct and Indirect Effects of Public Policies and Interventions

DISSERTATION

to obtain the degree of Doctor at Maastricht University
on the authority of the Rector Magnificus, Prof. Dr. Pamela Habibović
in accordance with the decision of the Board of Deans,
to be defended in public on Wednesday 4 May 2022 at 13.00 hours

by

ANNELORE MARIE-CLAIRE VERHAGEN

SUPERVISORS

Prof. Dr. Lex Borghans

Prof. Dr. Thomas Dohmen

Prof. Dr. Bart Golsteyn

ASSESSMENT COMMITTEE

Prof. Dr. Trudie Schils (chair)

Prof. Dr. Michèle Belot (Cornell University)

Prof. Dr. Mark Levels

Prof. Dr. Maarten Lindeboom (Vrije Universiteit Amsterdam)

Acknowledgements

Many people had direct and indirect effects on the content and completion of this dissertation. I would like to dedicate these pages to thanking them. Unless stated otherwise, I will do so in alphabetical order.

First, I would like to express my sincere gratitude to my Three Wise Men: Bart, Lex and Thomas. Bart, thank you for the great co-operation and your everlasting enthusiasm and ambition for our papers. I truly enjoy our discussions, whether they are about my dissertation, housing, children, gossip, or tank stations: they are as inspiring as they are entertaining. Lex, thank you for all of your guidance and support, and for encouraging me to continue working on my papers until I did not only have an answer to my research questions, but until I felt I did everything I could. Thomas, thank you for giving such detailed and timely feedback whenever I needed it and for helping me put my research questions in a broader perspective.

I would also like to thank the members of the assessment committee, Trudie Schils (chair), Michèle Belot, Mark Levels and Maarten Lindeboom, for taking the time to read this dissertation and for their valuable comments.

Thanks also to Dave and Maria. Working together and getting your public health perspective on physical activity and school performance has added great value to this dissertation.

This dissertation is not only a milestone for my research career but also for my time at Maastricht University. I would like to thank all members of the Economics of Education group at Maastricht University and all the office mates and colleagues I had over the years and the people I met who made my PhD experience so much more pleasant and interesting. Special thanks go out to Olga, Sergio and Stan for our friendship throughout our PhD life, for understanding society with me in all the meanings of the words, and for letting me make fun of them and all their witty comebacks in return. I would also like to thank my colleagues and friends at the OECD for keeping me motivated to finish my PhD.

Further thanks go to my former colleagues at ROA, because that is where this all began. Not only would I probably not have started a PhD in economics if it were not for you; you have been the witnesses, enablers and sometimes even the direct cause of what turned out to be some of my major life events. Thank you for hiring me as a researcher a decade ago, for showing me that economics is not as scary as I thought it was, for always making me feel welcome and part of the family even after I left, for all the eauverleggen and after-work drinks, and for introducing me to my husband.

Speaking of whom: Dear Eric, thank you so much for being my travel companion from the start of this dissertation all the way through to its end and beyond. I know the ride was not always easy for you either. Thank you for your sincerity and sense of humor, for staying calm when I freak out, for convincing me with valid arguments why you believe in me, and for helping me express my thoughts and intuitions in economic terms.

Lieve Maxime, delen van dit boek heb ik geschreven met jou in mijn buik. Ik weet zeker dat het daardoor extra goed is geworden. Als bedankje zal ik je beloven dat dit een van de weinige boeken op de plank zal zijn waar ik je niet ongevraagd uit zal voorlezen.

Tot slot wil ik graag mijn ouders bedanken voor alle positieve vibes, en voor het aanmoedigen van het meisje dat zo vastberaden was de handstand te kunnen doen.

Summary

All actions have consequences. Some consequences are intentional while others are unintentional, and some have beneficial consequences for individuals or society whereas do not. The same holds for public policies and social interventions: although the intention is to improve society, sometimes these actions have unintended consequences as well. The intended effects are called the direct effects, while other outcomes are called indirect effects. Not only direct effects, but also indirect effects of public policies and interventions should be carefully monitored and evaluated, even when previous research or common sense suggest that direct or indirect effects are desirable. That is the main take-away from this doctoral dissertation, which evaluates the direct effects and specific indirect effects of a quasi field experiment that encourages children to become more physically active in their everyday life (i.e., the *Active Living Program*), and a public policy that incentivizes people to become an organ donor after death (i.e., *opt-out consent for organ donation*).

It is important to emphasize that both policies are successful in reaching most of their direct effects: the Active Living Program successfully increases children's time spent on physical activity during school time, and opt-out consent for organ donation is related to significantly more kidney, liver and heart transplantations from deceased donors. However, both policies also appear to have undesirable indirect effects.

The Active Living Program unintentionally decreases school performance, particularly among the worst-performing students and among boys (Chapter 1). Moreover, the Program significantly increases the prevalence of Attention-Deficit/Hyperactivity Disorder (ADHD)-like symptoms in boys (Chapter 2). The results of these two chapters highlight that policymakers and researchers need to be very careful generalizing previously found desirable (side) effects to different contexts and policy designs: the studies in Chapters 1 and 2 are among the first that identify the causal indirect effects of encouraging physical activity in everyday life among children.

Chapter 3 analyzes whether an intended indirect policy effect is met, by studying whether organ-patient mortality rates are indeed lower in opt-out consent systems for organ donation. This is an indirect effect, because opt-out systems cannot literally save lives: mortality rates can only be affected through other channels, such as deceased-donor transplantation rates. However, organ-patient mortality rates appear to barely differ between consent systems. Moreover, for kidneys and livers, one more deceased-donor transplantation does not imply that one organ-patient life is saved. This indicates the existence of factors related to opt-out consent systems that are positively related to kidney and liver-patient mortality. The chapter demonstrates that even seemingly obvious relationships merit explicit evaluation.

Table of Contents

Introduction

MOTIVATION

Over the past 30 years, evidence-based policymaking has become the gold standard for public policy decisions. Rather than being led by ideology, political opportunity or common sense, it is expected more and more that policymakers base their decisions on rigorous and robust scientific evidence that policy instruments achieve their intended outcomes (see, e.g., Davies, Nutley & Smith, 2000; Pawson, 2006). Although the idea is not new, it is receiving increasing attention since the late 20th century. In the United Kingdom, for instance, Tony Blair's New Labour government was elected in 1997 with the philosophy that '*what matters is what works*,' in 2001, the European Commission published a white paper stating that scientific and other experts play an increasingly significant role in preparing and monitoring decisions (European Commission, 2001, p15) and more recently, the Foundations for Evidence-Based Policymaking Act of 2018 was signed into law in the United States.

As stated above, the aim of evidence-based policymaking is to find rigorous and robust scientific evidence that policies reach their intended outcomes, i.e. that they reach their direct effects. Although this is very commendable, and a necessary condition for evidence-based policymaking, it may not be sufficient. After all, what about policies that reach their intended outcomes, but that also have other (undesirable) effects? For instance, concerning the current COVID-19 pandemic, Haug et al. (2020) find that the closure of educational institutions is one of the most effective policies for decreasing the spread of the virus (i.e., decreasing the reproduction number by -.15 to -.21). Yet, Engel, Frey and Verhagen (2021) show that primary school closures in the Netherlands resulted in a learning loss of about 0.08 standard deviations (equivalent to one-fifth of a school year: the same period that schools in the Netherlands remained closed). This example indicates that ignoring general equilibrium effects by focusing on the direct effects only can have severe consequences for society. Moreover, carefully monitoring *intended* indirect effects is important, too. For example, Liu et al. (2018) show that featuring healthy foods on a store endcap increases healthy-food sales, but that it does not significantly decrease indulgent-food sales (while featuring indulgent food does decrease healthy-food sales). Angelucci and De Giorgi (2009) provide an example showing that positive indirect effects of successful policies are important as well. They argue that previously shown benefits of PROGRESA – a Mexican policy that provides sizeable cash transfers to poor households with the aim to improve their education, health, and nutrition levels – are underestimated, because they do not take into account that the policy not only increases consumption of eligible households, but also of ineligible households living in the same villages. These examples highlight that making truly informed policy decisions requires not only having access to convincing evidence about direct effects of policies and interventions, but also about their indirect effects, thereby making policy decisions multidimensional. Indirect effects are here defined as effects on other variables than the intended outcomes and effects that are only reached through other channels.

Next to having access to convincing evidence about direct and indirect effects, it is important that this evidence is interpreted correctly. For instance, since previous studies have often found positive effects of physical activity on health, well-being, educational and labor market outcomes (e.g., Barron, Ewing, & Waddell, 2000; Cabane, Hille, & Lechner, 2016; Dills, Morgan, & Rotthoff, 2011; Felfe, Lechner, & Steinmayr, 2016; Lipscomb, 2007; Pfeifer & Cornelissen, 2010; Rees & Sabia, 2010; Stevenson, 2010), it is commonly believed that any kind of physical activity will generally be 'good for you.' For instance, the World Health Organization (2020) published a long list of benefits of physical activity without being specific about which types of physical activity may result in these positive outcomes, and without mentioning any potential negative side effects. The belief that physical activity will generally be good for you is further strengthened by various public awareness campaigns encouraging people to 'stay active' or 'get moving.'[1] However, it is not certain whether this belief is correct, since the existing evidence does not cover the effects of just *any* type of physical activity: it focuses on 'formal physical activity,' such as moderate to vigorous physical exercise, doing sports in clubs, physical education, or participating in high school athletics. Indeed, the results in Chapters 1 and 2 indicate that it may be wrong to assume that the findings from formal physical activity policies are generalizable to informal physical activity, such as active transportation or active play.

Chapter 3 is an illustration of another type of misconception based on existing scientific literature. Previous studies have consistently found that countries with opt-out consent systems for organ donation conduct significantly more transplantations from deceased donors (e.g., Abadie & Gay, 2006; Johnson & Goldstein, 2003; Johnson & Goldstein, 2004; Rithalia et al., 2009; Shepherd, O'Carroll & Ferguson, 2014; Ugur, 2015). Common sense tells us that every transplantation performed is an organ-patient's life saved, and hence, that every deceased-donor transplantation decreases the number of organ-patients that dies by one patient. Although reducing organ-patient mortality has become one of the primary aims on the policy agenda regarding organ donation, consent systems do not directly affect organ-patient mortality. After all, consent systems cannot literally save lives: they can only affect organ-patient mortality rates through other channels, such as deceased-donor transplantations. The assumption that every deceased-donor transplantation performed reduces organ-patient mortality by one patient, completely discards the existence of any desirable or undesirable indirect behavioral effects of opt-out consent systems that are related to organ-patient mortality. Chapter 3 argues that opt-out consent systems do not only affect deceased-donor transplantation rates, but that they also change incentives, which may, in turn, affect organ-patient mortality. For instance, the chapter shows empirically that, by increasing the number of deceased-donor transplantations, opt-out systems may decrease the incentive for people to become a living donor. Since this decreases the difference in the total number of transplantations performed between consent systems, it explains why organ-patient mortality rates in opt-out systems are not as low as expected based on deceased-donor transplantation rates alone.

[1] Examples of policies and campaigns that encourage physical activity are the *Move it AUS* campaign in Australia, the *Move your way* campaign in the US, *MOOVIN'* in Bergamo (Italy) or the *#BeActive Workplace Initiative* in Malta.

RESEARCH QUESTIONS

This dissertation aims to investigate the direct and specific indirect effects of two public policies and interventions that directly or indirectly aim to improve public health: the *Active Living Program* and *opt-out consent for organ donation*. The main research questions and sub-questions of this thesis are as follows:

Part I

Direct and Indirect Effects of Encouraging Physical Activity in Everyday Life

- What is the direct effect of encouraging physical activity in everyday life on children's time spent on physical activity? (Chapter 1)
- What is the indirect effect of encouraging physical activity in everyday life on children's school performance? (Chapter 1)
- What is the indirect effect of encouraging physical activity in everyday life on the prevalence of ADHD-like symptoms among children? (Chapter 2)

Part II

Direct and Indirect Effects of Opt-Out Consent for Organ Donation

- What is the direct effect of opt-out consent for organ donation on deceased-donor transplantation rates? (Chapter 3)
- What is the indirect effect of opt-out consent for organ donation on organ-patient mortality rates? (Chapter 3)
- Does the effect of opt-out consent for organ donation on deceased-donor transplantation rates translate into a similar effect (though with opposite sign) on organ-patient mortality rates? (Chapter 3)

OUTLINE

Chapter 1

This chapter investigates whether encouraging children to become more physically active in their everyday life affects their primary school performance. It uses data from a quasi field experiment called the Active Living Program, which aimed to increase active modes of transportation to school and active play among 8- to 12-year-olds living in low socioeconomic status (SES) areas in the Netherlands. Difference-in-differences estimates reveal that the program increases time spent on physical activity during school hours, but negatively affects school performance, especially among the worst-performing students. The results suggest that the commonly found positive effects of exercising or participating in sports on educational outcomes may not be generalizable to physical activity in everyday life. Therefore, policymakers and educators who seek to increase physical activity in everyday life need to weigh the health and well-being benefits against the probability of increasing inequality in school performance.

Chapter 2

Using data from the Active Living Program, this chapter explores the effects of stimulating physical activity in everyday life on the prevalence of ADHD-like symptoms in 12-year-old boys and girls. Difference-in-differences estimates across cohorts of sixth graders show that, due to the program, boys more often report symptoms of inattentiveness and hyperactivity-impulsivity; effects that remain robust across a variety of specifications. No significant treatment effects on girls' ADHD-like symptoms are found. The desirable direct effects of the Active Living Program on children's time spent on physical activity notwithstanding, this chapter highlights that certain physical activity interventions can have undesirable side effects. Therefore, when implementing physical activity interventions, potential side effects need to be monitored carefully.

Chapter 3

Previous research shows that countries with opt-out consent systems for organ donation conduct significantly more deceased-donor organ transplantations than those with opt-in systems. This chapter investigates whether the higher transplantation rates in opt-out systems translate into equally lower death rates among organ patients registered on a waiting list (i.e., organ-patient mortality rates). It shows that the difference between consent systems regarding kidney- and liver-patient mortality rates is significantly smaller than the difference in deceased-donor transplantation rates. This is likely due to different incentives between the consent systems. Empirical evidence that opt-out systems reduce incentives for living donations is found, which explains the findings for kidneys. The results imply that focusing on deceased-donor transplantation rates alone paints an incomplete picture of opt-out systems' benefits, and that there are important differences between organs in this respect.

SCIENTIFIC AND POLICY RELEVANCE

This dissertation provides an important contribution to the scientific literature as well as to the policy debate regarding children's health behavior (particularly: physical activity), inequality in school performance, and consent systems for organ donation.

First, Chapter 1 highlights that policymakers, educators, and researchers around the world cannot assume that every type of physical activity intervention will only have positive indirect effects if it is effective in increasing time spent on physical activity. Previous studies have shown that formal physical activity can have desirable indirect effects on educational outcomes and ADHD-like symptoms, but these results may not be generalizable to informal physical activity such as active transportation or active play.

An important takeaway of Chapters 1 and 2 combined is that the costs and benefits of physical activity may not be the same for all children. Although the effect of the Active Living Program on time spent on physical activity is similar across groups of children, the worst-performing students, who arguably need their study- and instruction-time the most, may also suffer the most if they or other children (as Chapter 2 indicates: mostly boys) are increasingly inattentive, hyperactive or impulsive during instruction time. Before resources are allocated to increase time spent on physical activity among children and adolescents, we first need to improve our understanding of the causal effects of such interventions on short-term and long-term educational and behavioral outcomes. Future research needs to focus on which forms of physical activity do not have undesirable indirect effects.

The main implication of Chapter 3 is that it is not evident that opt-out consent systems reduce kidney and liver-patient mortality rates. For kidneys and livers—by far the most commonly transplanted organs—the higher deceased-donor transplantation rates in opt-out systems give a false impression that presumed consent is related to lower organ-patient mortality rates. Although lower organ-patient mortality rates are one of the main indirect aims of opt-out consent systems, this relationship has never been analyzed before in the literature.

In sum, this dissertation highlights the importance of feedback loops in policymaking (also known as iterative policymaking or policy learning), so that policies can be adjusted as more evidence is found about the direct and indirect effects on more outcomes.

Part I

Direct and Indirect Effects of Encouraging Physical Activity in Everyday Life

Chapter 1. Active Bodies, Inactive Minds?

The Effect of Encouraging Active Living on Physical Activity and School Performance

1.1. INTRODUCTION[2]

Due to the global increase of childhood obesity and sedentary behavior, governments and institutions advocate increasing the amount of time primary school children spend on physical activity.[3] Although the benefits of physical activity for physical and mental health are well-documented, the current literature is still unsettled regarding the causal effects on educational outcomes. Physical activity is expected to positively affect cognition and thereby school performance from a physiological viewpoint,[4] but the effect of increasing physical activity on school performance is theoretically ambiguous. Even if physical activity has positive effects on children's cognitive abilities, increasing sports participation may crowd out time investments in other potentially beneficial activities such as studying or active play.[5] Physical activity may, however, also crowd out potentially harmful activities such as smoking or watching television, leading to a zero or positive effect on school performance. The net effect on school performance thus depends on the relative gain (or harm) from the activity that was crowded out compared to the gains from physical activity.

This chapter investigates with a quasi field experiment whether the encouragement of informal physical activity, such as active play, affects primary school performance. The quasi-experiment, called the Active Living Program, is organized in low-SES areas of the Netherlands. We analyze whether the interventions affect physical activity and school performance.

The Active Living Program aims to increase physical activity and decrease sedentary behavior among 8- to 12-year-olds, through encouraging active modes of transportation to school and active play at school and during leisure time. The interventions, which were implemented between April

---

[2] This chapter is part of a paper that was published in the *Journal for Policy Analysis and Management* (Golsteyn, Jansen, Van Kann & Verhagen, 2020). It has benefited from valuable comments by Eric Bonsang, Lex Borghans, Stefano DellaVigna, Thomas Dohmen, Ilyana Kuziemko, Michael Lechner, and Bas Ter Weel, as well as participants at the 21st IZA Summer School in Labour Economics, the Maastricht University Workshop in Economics, the Applied Economics Lunch Seminar at the Paris School of Economics, the Learning and Work seminar at Maastricht University, the Netherlands Economists Day 2017, the 29th annual conference of the European Association of Labour Economists, the 31st Annual Conference of the European Society for Population Economics, and the Human Enhancement and Learning conference. Special thanks go out to all schools and children participating in this study and to all Maastricht University colleagues involved in the data collection. This study was partly funded by the Netherlands Organization for Scientific Research (VIDI grant 452-16-006) and the Netherlands Organization for Health Research and Development (ZonMW), Project Number 200130003.

[3] See, e.g., Centers for Disease Control and Prevention (2011); European Commission (2014); World Health Organization (2010).

[4] The medical literature shows that physical activity can positively affect cognitive functioning through increased blood and oxygen flow to the brain and increased growth factors that help create new nerve cells and support synaptic plasticity, decreased stress, and improved mood due to higher levels of norepinephrine and endorphins (Singh et al., 2012). Erickson, Hillman, and Kramer (2015), for example, find that more-active children show greater hippocampal and basal ganglia volume, greater white matter integrity, elevated and more efficient patterns of brain activity, superior cognitive performance, and higher scholastic achievement.

[5] As Lizandra et al. (2016) point out, sedentary activities could either be academic (e.g., doing homework, reading, or studying), technology-based (e.g., playing video games or watching television) or social-based (e.g., sitting with friends or chatting via social networks); each of these behaviors has a different expected effect on academic performance.

2013 and June 2014, did not affect instruction time or the structure or content of classes. Pre- and post-treatment data regarding school performance and time spent on physical activity were collected in 10 treatment schools and 11 control schools. School performance is measured through nationally standardized language and math/calculating tests that use a grading scheme common to all schools. Physical activity data were collected both pre- and post-treatment by means of accelerometers that children wore for five consecutive days. We use a difference-in-differences technique to estimate the effects of the Active Living Program on school performance and time spent on physical activity.

The results indicate that the Active Living Program causes a significant decrease in school performance. Our most conservative estimate of the average treatment effect is -5.9 percent of a standard deviation. The negative effect on school performance is strongest among the worst-performing students and among boys. Event study analyses including three pre-treatment years and several robustness checks provide strong indications that the effects on school performance are indeed due to the interventions in the treatment schools, which allows for a causal interpretation of our estimates. We explore which mechanisms may account for our main findings. First, we find that the Active Living Program causes a significant increase in time spent on physical activity during school hours (0.34 standard deviations or 9.30 minutes per day), but no significant effects on leisure-time activity. Although the results must be interpreted with caution, they imply that an additional half-hour spent on physical activity during school hours, decreases school performance by 0.19 standard deviations. Second, we investigate whether the Active Living Program crowds out participation in activities that may be cognitively stimulating (i.e., playing music, chess or doing arts and crafts), but we find no significant effects.

This chapter contributes to the literature that studies causal effects of physical activity interventions on educational outcomes of children and adolescents. Earlier papers focus on the effect of participating in physical education classes, club sports, or other forms of formal or professional sports (e.g., Barron, Ewing, & Waddell, 2000; Cabane, Hille, & Lechner, 2016; Dills, Morgan, & Rotthoff, 2011; Felfe, Lechner, & Steinmayr, 2016; Lipscomb, 2007; Pfeifer & Cornelissen, 2010; Rees & Sabia, 2010; Stevenson, 2010). Our study contributes to this literature by being the first to analyze the effect of the encouragement of informal activities such as active transportation and active play. Moreover, due to the quasi-experimental design, our identifying assumptions are less restrictive than in prior studies.

Our results suggest that the commonly found positive effects of exercising or participating in sports on educational performance may not be generalizable to physical activity in everyday life. This has important implications for policymakers, educators, and researchers in other high- or middle-income countries, who aim to increase physical activity among children and adolescents within or outside of low-SES regions. Our findings highlight that: (1) policymakers must remain alert when implementing interventions and assure that all the (assumed) benefits and potential costs are carefully monitored; (2) policymakers must consult research evidence to make informed decisions that weigh the health and well-being benefits against the potential increased inequality in school performance; and (3) schools may need targeted interventions that increase low-performing students' physical activity without further decreasing their educational performance.

This chapter proceeds as follows. Section 1.2 describes our contribution to the literature. Section 1.3 presents the empirical strategy and discusses the quasi field experiment. Section 1.4 describes the data. Section 1.5 provides the main results and robustness checks, and investigates the heterogeneity of the treatment effects. Section 1.6 analyzes the effects of the Active Living Program on time spent on physical activity and investigates the external validity of our findings. Section 1.7 analyzes potential mechanisms. Section 1.8 concludes.

1.2. CONTRIBUTION TO THE LITERATURE

Earlier papers on the consequences of physical exercise mostly have shown positive effects on health (see, e.g., Cawley, Frisvold, & Meyerhoefer, 2013; Cawley, Meyerhoefer, & Newhouse, 2007; Strong et al., 2005) and on cognitive functioning, particularly in older adults (see, e.g., Colcolmbe & Kramer, 2003; Hillman, Erickson, & Kramer, 2008). A limited set of papers analyzes the causal effect of physical activity (PA) on educational outcomes, such as school performance, for younger individuals.

When analyzing the effect of PA on school performance, there is a risk of reverse causality or selection based on unobserved factors. For example, unobserved health deterioration may cause a simultaneous decrease in school performance and time spent on PA. For example, children with decreasing school performance may be encouraged to spend more time doing homework: a sedentary activity. Previous attempts to overcome such problems include adopting an individual-fixed-effects approach (Lipscomb, 2007; Rees & Sabia, 2010), and selection-on-observables or matching (Cabane, Hille, & Lechner, 2016; Felfe, Lechner, & Steinmayr, 2016; Pfeifer & Cornelissen, 2010). Other studies adopt an instrumental variable approach in which students' sports participation is instrumented by their height (Barron, Ewing, & Waddell, 2000; Pfeifer & Cornelissen, 2010; Rees & Sabia, 2010), by certain parental characteristics such as income, or by school characteristics such as school size and the books-per-student ratio (Barron, Ewin, & Waddell, 2000). Stevenson (2010) and Dills, Morgan, and Rotthoff (2011) use changes in state-level policies to instrument for changes in sports participation at school.[6]

Our focus on informal PA makes several important contributions to the literature. First, being physically active entails more than just dedicated physical exercise. During a day, the amount of time spent on informal PA is (or can be) substantially larger than the amount of time spent on dedicated physical exercise. For several reasons, it is unclear whether we should expect the effects of the encouragement of informal PA to be similar to the effects usually found for formal PA. First, given that informal PA is usually less intense than participation in sports, our study provides more insight into the dose-response relationship between PA and school performance. Previous

[6] Stevenson (2010) finds a positive effect of increased high school athletic participation on education and labor market outcomes for women. Dills, Morgan, and Rotthoff (2011) do not find any significant effects of increased recess or physical education time on school performance.

research has shown that the positive relation between PA on health and educational outcomes may be driven mostly by moderate-intensity and vigorous PA (see, e.g., Felez-Nobrega et al., 2017; Singh et al., 2012; Van den Berg et al., 2016). Second, formal and informal PA may foster the development of different skills. Sports teach players competitiveness and perseverance—skills that could be beneficial for educational outcomes, but which are likely less pronounced (if present at all) in active play. Additionally, athletes may receive increased attention and encouragement from their teachers or parents, leading to increased self-esteem and increased peer pressure to succeed (Stevenson, 2010). Developing these "sports-related skills" implies that potential crowding-out effects will be less pronounced. In other words, if children reallocate study hours to PA, the effect of engaging in sports on school performance might have been negative had it not been for the athlete's development of these additional sports-related skills.

Our research also contributes to the literature in the sense that in our quasi-experiment, unlike other PA intervention studies, none of the interventions affected school time or the curriculum. Moreover, our PA data are collected by accelerometers that children wore throughout the day, which allows us to capture any substitution or complementarity effects between in- and out-of-school PA, and between formal and informal forms of PA. This is an important benefit relative to other studies, which focus on (self-reported) measures of formal PA.

1.3. EMPIRICAL STRATEGY

In order to stimulate PA, we organized the Active Living Program: a quasi field experiment that took place between 2012 and 2014 in 21 primary schools located in low-SES areas in the Southern-Limburg region of the Netherlands[7] (see Figure A1.5 in Appendix 1.4 for a map of the research area). The target group consists of children who were enrolled in grades 4 or 5 during the 2012/2013 school year and transferred to grades 5 or 6 during the 2013/2014 school year (i.e., 8- to 12-year-olds).

We focus on 8- to 12-year-old children because children in this age category experience more independent mobility due to fewer parental restrictions than younger children. For instance, they can independently use transportation or visit a playground in their neighborhood. Independent mobility has been shown to be an important predictor of children's PA (Carver et al., 2013; Hume et al., 2009; Pont et al., 2009). Moreover, independent mobility implies that interventions can be directly targeted at children rather than indirectly influencing their behavior through their parents.

Ten primary schools agreed to participate as treatment schools.[8] We matched treatment schools to control schools based on the level of neighborhood deprivation and the level of urbanization (urban vs. rural). One treatment school was matched to an additional control school

[7] The Limburg region has about 609,000 inhabitants and a population density of 922 inhabitants per square kilometer. This density is almost twice as high as the average population density in the Netherlands, which is 496 inhabitants per square kilometer. The region consists of 18 municipalities. The average disposable household income in the region is about €31,500 per year, which is somewhat lower than the national average of €34,200 per year (Statistics Netherlands, 2013).

[8] See Textbox A1.1 and Figure A1.6 in Appendix 1.4 for a more detailed description of how treatment and control schools were selected.

because the school to which it was initially matched relocated during the 2012/2013 school year. After the pre-treatment PA measurement, no further changes were observed in this control school, making both control schools eligible for study inclusion. We exclude one treatment school from the data because it organized a sports day during the pre-treatment PA measurement, resulting in exceptionally high levels of PA that were not representative of children's average school days.

Each treatment school had a working group that consisted of representatives of the school, parents, municipal authorities and other stakeholders, and a public health services (PHS) employee who served as chair. The researchers trained the PHS employee in "physical and social environmental thinking" and evidence-based PA intervention opportunities. The working groups were responsible for choosing, designing, and implementing intervention elements, but the research team had final approval over the final plan costs. Although the overall scope was the same, the intervention packages differed in magnitude and design across treatment schools, depending on local needs.

All working groups received an intervention budget of €2,000 at the start of the project. However, costs for environmental adaptations in public spaces (i.e., school surroundings) were covered by municipalities, and if intervention plans exceeded the budget, the working group could apply for additional funding from the research team and municipal authorities (which had a representative as a member of the working group). Additionally, the province of Limburg provided funding for a sports day at all treatment schools, and several treatment schools organized charity events such as a charity run in order to raise money for new playground equipment. Treatment schools could implement the interventions as soon as the pre-treatment PA measurement was conducted at their school (see Section 1.4 below for more information on the PA data collection). All interventions were implemented before the second PA measurement in the spring of 2014. Table 1.1 presents an overview of all interventions per treatment school. Table A1.7 in Appendix 1.4 provides a more detailed description of the implemented interventions per treatment school.

Each intervention was intended to decrease sedentary behavior or increase PA at school or during leisure time by nudging children to become more active. These nudges could be explicit, e.g., by organizing additional sports days, but also could be implicit, e.g., by creating safer traffic environments around the treatment schools.[9] It is important to note that none of the interventions affected school time or the structure or content of the curriculum because this may have otherwise affected school performance. For the same reason, interventions requiring structural changes in the school yard (e.g., installing new fixed equipment, a ball backstop, or sound equipment in the school yard) were implemented before or after school hours.

---

[9] The only implemented interventions that were specifically aimed at increasing time spent on group PA were the "Active Living Games" and the establishment of a school soccer team. All other interventions are not specifically aimed at either individual or group PA. For example, children could choose whether they wanted to play with the new playground equipment with others or by themselves.

Table 1.1. Overview of implemented interventions per treatment school

	T1	T2	T3	T4	T5	T6	T7	T8	T9
Panel A. Active transportation to school									
Develop safe route to school	−	−	+	+	+	−	−	−	+
Mobilize crossing guards	−	−	+	+	−	−	−	−	−
Adapt unsafe intersection in school environment	−	−	+	−	−	−	−	−	−
Make bicycle racks available	−	+	−	−	−	+	−	−	−
School pedestrian crossing indicators	+	+	−	−	−	+	−	−	−
Create safer parking situation around school	−	−	+	−	−	−	−	−	−
Create traffic circle in schoolyard environment	+	+	−	−	−	−	−	−	−
Sticker competition for active transport	+	+	+	+	−	−	+	−	−
School stimulation documentation on safe active transport	−	−	+	+	−	−	−	−	−
Introduce ”Walk/Bike-to-school-day”	−	−	−	−	−	+	−	−	−
Speed-check action performed by children	−	−	+	+	−	−	−	−	−
Lessons to improve bicycle skills	−	−	−	−	+	−	−	−	+
Panel B. PA in school									
New fixed equipment in schoolyard	−	+	−	+	+	+	−	+	+
New loose equipment in schoolyard	−	+	+	−	+	+	+	−	+
Playground markings	−	+	−	−	−	+	+	−	−
Establish ball game area	+	+	−	−	−	+	−	+	−
Put a ball backstop beside railway	+	−	−	−	−	−	−	−	−
Sound equipment in schoolyard environment	−	+	−	−	−	−	−	−	−
Additional sports day in schoolyard	+	+	+	+	−	−	−	+	−
Sports clinics at recess	+	+	+	+	+	+	−	−	+
Use of schoolyard games	−	+	+	−	−	−	+	+	−
Contest for best PA encouragement idea	+	−	−	−	−	−	−	+	−
Panel C. PA in leisure time									
Establish training circuit	−	−	+	−	−	−	−	−	−
Active Living Games	+	+	+	+	+	+	+	−	+
Establish out-of-school PA program	+	+	+	+	+	−	−	+	+
Establish school soccer team	−	−	+	−	−	−	−	−	−
Establish PA activities by children for local residents	−	−	+	−	−	−	−	−	−

Notes: Tn = Treatment school (number); + = intervention was implemented; − = intervention was not implemented. One of the initial 10 treatment schools is excluded from the analyses because it organized a sports day during the pre-treatment physical activity measurement week. Control schools did not implement any interventions. See Table A1.7 in Appendix 1.4 for a detailed explanation of each intervention.

The intervention budget of 2,000 euros was not conditional on actual implementation of interventions. However, the research team visited the treatment schools regularly pre- and post-treatment, which allowed us to confirm that changes had been made to the physical school environment. Picture A2.1 in Appendix 2.4 provides examples of the interventions. Children were not obliged to engage in additional PA. All implemented interventions were intended to nudge children toward increased PA time (i.e., intention to treat), but there was no guarantee that children increased their time spent on PA.

To decrease the probability of contamination of the control group, all interventions were implemented within an 800-meter radius around the treatment schools. In the Netherlands, the majority of the primary school children live within 800-meters of the school (Statistics Netherlands, 2014). Control schools are located outside of the 800-meter radius around treatment schools, and they did not implement any Active Living interventions. When conducting the pre- and post-treatment PA measurements, the research team confirmed that control schools did not make any changes to the physical school environment. However, we cannot completely rule out the possibility that control schools changed something between the pre- and post-treatment year that had an effect on the children's time spent on PA. We therefore conduct robustness checks in which we analyze whether there was an unexpected treatment effect in the control schools (see Section 1.5.1 below).

1.3.1. Identifying the Treatment Effect on School Performance

After the data collection (see the next section for more information), we verified whether treatment and control schools were similar with respect to pre-treatment observables. These analyses revealed that before any interventions were implemented, treatment and control schools did differ in various domains: the most important differences were that children in treatment schools performed significantly better on math tests and spent an average of 15 minutes per day less on PA during school hours (see Table A1.8 in Appendix 1.4). This means that matching appears not to have been successful. Therefore, we could not merely compare post-treatment outcomes between children in treatment and control schools because post-treatment differences may be due to self-selection.

When there are pre-treatment differences between the treatment and control groups, a difference-in-differences technique can be used as an alternative strategy for identifying treatment effects. Underlying the technique is the identifying assumption that the initial differences between the treatment and control groups would have remained similar if the Active Living Program had not been implemented. Common trends in the pre-treatment period suggest that this assumption holds (see the next subsection).

In order to identify the effect of the Active Living Program on school performance, we estimate the following model:

$$Score_{ist} = \alpha_0 + \alpha_1(Treat_{is} * Post_t) + \alpha_2 Post_t + \alpha_3 age_{ist} + \alpha_4 age^2{}_{ist} + \gamma_{1is} + \varepsilon_{1ist} \quad (1.1)$$

In equation (1.1), *i* indexes the individual pupil, *s* indexes the school, and *t* indexes time. In the regressions, *Treat* is an indicator variable with a value of one if a pupil is in a school that belongs to the treatment group and zero if he or she is enrolled in one of the control schools.[10] *Post* is an indicator variable with a value of one for tests taken during the post-treatment year of 2013/2014 and zero for tests taken during the pre-treatment year of 2012/2013.[11] We are interested in the difference-in-differences estimator α_1, i.e., the interaction of *Treat* and *Post*. In our main specification, we control for age and child fixed effects γ.[12]

[10] The sample did not include any children who changed schools between the pre- and post-treatment year.

[11] Note that the interventions could be implemented throughout the 2013/2014 school year (March through June 2014), which implies that some tests taken during this school year may be wrongly defined as post-treatment tests. However, as this would attenuate results, we consider it a minor issue.

[12] It is possible to include both age and year fixed effects because in our sample, age is measured in months when the tests are taken, and test dates vary by year. However, note that including child fixed effects implies that we cannot also include *Treat* because this variable does not vary across time.

1.3.2. Common Trends: Event Study Analysis

To analyze whether pre-treatment trends are indeed parallel, we conduct an event study analysis including three pre-treatment years by estimating the following model:

$$Score_{ist} = \alpha_5 + \alpha_6(Treat_{is} * Post_t) + \alpha_7(Treat_{is} * Pre2_t) + \alpha_8(Treat_{is} * Pre3_t) + \delta_t + \alpha_9 age_{ist} + \alpha_{10} age^2_{ist} + \gamma_{2is} + \varepsilon_{2ist} \quad (1.2)$$

The variables *Post*, *Pre2*, and *Pre3* are indicator variables for tests taken during the school years 2013/2014, 2011/2012, or 2010/2011, respectively (we use the 2012/2013 school year, i.e., *Pre1*, as the reference group). δ_t stands for year fixed effects, and similar to equation (1.1), we control for age and child fixed effects. The treatment effect in the treatment year is given by the interaction between *Treat* and *Post* (α_6), which should be almost identical to α_1 in equation (1.1). The parallel trends assumption is supported if α_7 and α_8, i.e., the estimated treatment effects in the pre-treatment years, are both not significantly different from zero. We formally test whether these two estimators are individually and jointly significantly different from zero.

1.3.3. Robust Standard Errors

We use robust clustered standard errors because school performance may be correlated among pupils within the same school. However, only 21 schools participated in the Active Living Program, which creates a risk that our robust standard errors will be substantially downward biased (see, e.g., Angrist & Pischke, 2008; Bell & MacCaffrey, 2002; Bertrand, Duflo, & Mullainathan, 2004; Cameron, Gelbach, & Miller, 2008; Cameron & Miller, 2015; Donald & Lang, 2007).

To overcome this problem, we use the method proposed by Donald and Lang (2007), which has been shown to work reasonably well in difference-in-differences settings. They suggest that inferences be based on a *t*-distribution with degrees of freedom equal to the number of clusters minus the number of variables that are constant within clusters, rather than the standard normal distribution. By doing so, one essentially recognizes that the fundamental unit of observation is a cluster and not an individual within a cluster. We verified that inference based on this approach is indeed the most conservative compared to no clustering, clustering without adjusted critical *t*-values, or clustering with corrected standard errors based on the Moulton factor (as proposed by Angrist & Pischke, 2008).[13] Furthermore, in order to avoid serial correlation, we compare the pre-treatment year (2012/2013) to the post-treatment year instead of using test scores in all pre-treatment years as the baseline measure (Bertrand, Duflo, & Mullainathan, 2004).

[13] In some cases, inference based on one of the other strategies yields slightly more conservative results, but in these cases, the interpretation of the results does not change, i.e., the rejection of the null hypothesis stands.

1.4. DATA

Our analyses are based on two data sets. The first contains data on individual test score histories from children in nearly all schools in the Southern Limburg region (school performance data). The second contains data on time spent on PA as measured by accelerometers (PA data) worn by children in the 21 schools that participated in the Active Living Program (Active Living schools). We discuss these data sets in turn. See Figure 1.1 for a timeline of the data collection process.

1.4.1. School Performance Data

The data on school performance are collected from over 200 schools in a cooperative project between elementary schools, school boards, municipalities, and Maastricht University, called the *Onderwijs Monitor Limburg* (OML). The 21 Active Living schools are included in these 200 schools. We use data on the school performance of children in the schools that did not participate in the Active Living Program as a treatment or control group (non-Active Living schools) in the robustness checks.

The school performance data include a wide variety of standardized tests that children took between Kindergarten and grade 6 (i.e., ages 4 to 12) in various domains such as reading, spelling, vocabulary, calculating, and math. The tests are graded by the teachers who use a grading scheme that is standardized across all schools in the Netherlands. Each subject has its own range of scores. For each subject, there are two tests per grade with similar content. The teacher can decide whether a child takes both tests and whether the child takes the same test more than once.

For the purpose of this study, we limit our sample to the target group of the Active Living Program, i.e., children who were enrolled in grades 4 or 5 in 2012/2013 and in grades 5 or 6 in 2013/2014.[14] We exclude tests that were taken during the PA measurements to prevent reactivity effects.[15] Consequently, we exclude all tests that were taken between April 3 and August 11, 2013, and between March 26 and August 24, 2014.[16]

Because children can take multiple tests of the same subject within a school year, we first aggregate children's subject-scores to the average within each school year.[17] To create a comparable scale across subjects (as each has their own range of scores), we standardize children's average subject-scores across schools (Active Living schools as well as non-Active Living schools) and school years to a mean of zero and a standard deviation of one.

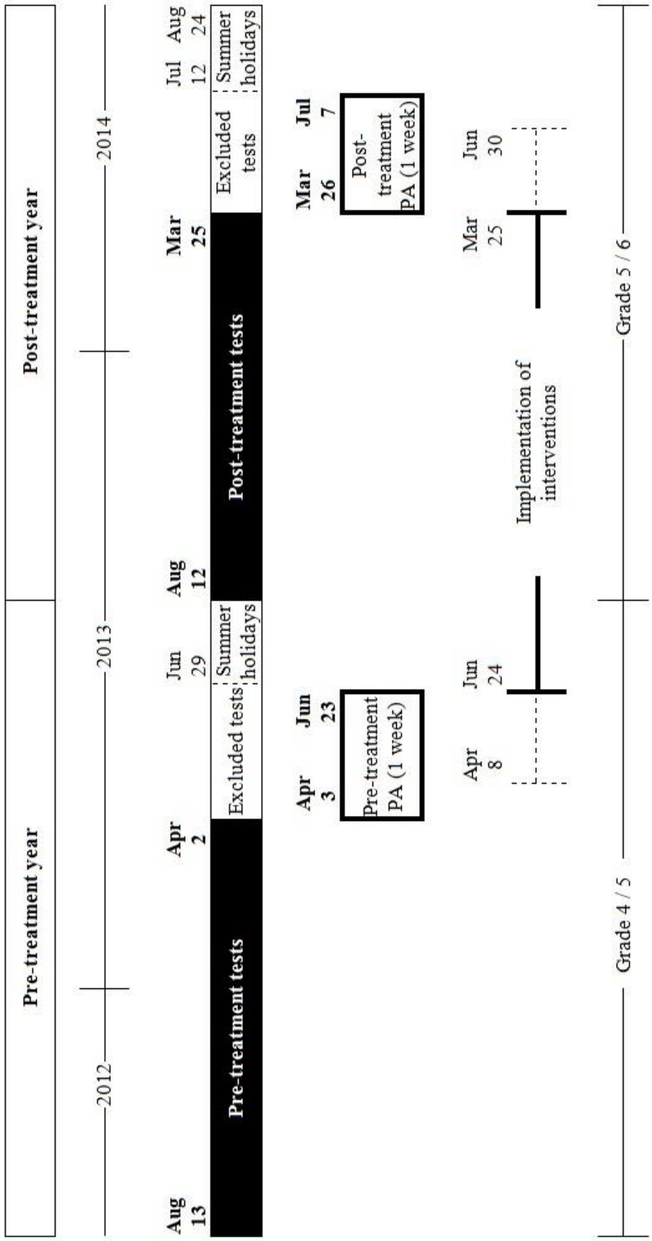
[14] In our sample, no children repeated grade 4 in 2013/2014. We include in the analyses children who repeated grade 5 in 2013/2014, as well as those who skipped grade 5 and went directly from grade 4 in 2012/2013 to grade 6 in 2013/2014.

[15] We also exclude the few tests that were taken between the last PA measurement and the summer holidays. No tests are taken during the summer holidays.

[16] The results from analyses including these tests remain qualitatively robust.

[17] Around 80 percent of the tests are taken in January and February (school years run from August until June), both in the pre- and post-treatment years. Therefore, we cannot estimate the treatment effect non-parametrically in the post-treatment period, e.g., by month, due to a lack of observations in all other months.

Figure 1.1. Timeline of the data collection



Notes: In each school, PA data were collected during one week between April and June 2013 and again during one week between March and July 2014. Treatment schools could implement their interventions between the last day of the pre-treatment PA measurement week in 2013 and the first day of the post-treatment PA measurement week in 2014. Implementation periods can therefore differ by school as indicated by the dashed lines. In some analyses, pre-treatment test scores taken before August 13, 2012 are also included.

We can then construct an overall school performance variable by calculating children's average score across their standardized subject-scores within a school year. For interpretation purposes, we rescale this variable on a scale from zero to 100. This is the dependent variable in our main analyses on school performance. We repeat this strategy for language and math tests separately.

1.4.2. PA Data

The school performance data can be merged at the individual level with the PA data. Time spent on PA was measured during one week between April 3 and June 24, 2013 and again during one week 12 months later (March 26 to July 7, 2014). Prior to the pre-treatment PA measurement, 61.6 percent of the children in the target group obtained written parental consent to wear an accelerometer. We verified that these children were not significantly different with respect to observable characteristics compared to those for whom we did not receive parental consent. In the pre-treatment year, 791 children were fitted with an accelerometer, which they were asked to wear on their waist for at least five consecutive days, including at least one weekend day.[18] To correct for potential seasonal effects (see, e.g., Rich, Griffiths, & Dezateux, 2012), we assessed every treatment-control school pair on the same dates.

PA data were collected using accelerometers (ActiGraph GT3X+; 30 Hz, 10-second epochs). The accelerometers' sensors measure acceleration in units of gravity on three axes: vertical, horizontal, and perpendicular. The accelerometers run on a battery and are not turned on or off by the participants. They register acceleration continuously. We use Evenson's cutoffs (Evenson et al., 2008) to determine PA intensity levels: accelerations of more than 100 counts per minute are recorded as time spent on PA and accelerations of 100 counts per minute or less are recorded as time spent on sedentary behavior.[19] We distinguish between sedentary behavior and a child's not-wearing the accelerometer based on Choi's classification criteria (Choi et al., 2011). We use the ActiLife software (version 6.10.4, ActiGraph, Pensacola, U.S.A.) to transform the raw data into data per hour.

"School-time PA" is defined according to regular school times of Dutch primary schools, i.e., Mondays, Tuesdays, Thursdays, and Fridays from 9:00 a.m. until 3:00 p.m. and Wednesdays from 9:00 a.m. until noon. Leisure time is defined as the hours before and after school time and the weekends. PA data collected between 11:00 p.m. and 6:00 a.m. are excluded from the analyses,

[18] Parental consent was requested of parents of 1,322 children and obtained for 815 children (61.6 percent). Out of the 815 children for whom we obtained parental consent to wear an accelerometer, 791 were fitted with an accelerometer in the pre-treatment year (97 percent); 24 children did not wear their accelerometer.

[19] Counts per minute is a unit of activity that is commonly used in health sciences. Light intensity PA (e.g., walking, biking slowly, playing catch) is defined as 101 to 2,295 counts per minute, 2,296 to 4,011 counts per minute is defined as moderate intensity PA (e.g., brisk walking, jumping on a trampoline, recreational swimming), and 4,012 or more counts per minute is defined as vigorous intensity PA (e.g., running, jumping rope, swimming laps). Fewer than 100 counts per minute is defined as sedentary behavior, e.g., watching TV, gaming, or doing homework (Evenson et al., 2008).

because these data are defined as sleeping hours.[20] The PA data are collected per hour, per day. In our main analyses, we sum the PA data per day per part (school time/leisure time) for each child.

We set a minimum amount of accelerometer wearing time, as otherwise, we would have to assume, for example, that if a child wore the accelerometer for only five minutes during an entire day, those five minutes are representative of the child's activity levels throughout that day. We define minimum wear time as at least half the available daily amount of school/leisure time.[21] This means that, regarding school-time PA, minimum wear time is defined as at least 180 minutes (three hours) between 9:00 a.m. and 3:00 p.m. on Mondays, Tuesdays, Thursdays, and Fridays, and at least 90 minutes (1.5 hours) between 9:00 a.m. and noon on Wednesdays.[22] Minimum wear time for leisure-time PA is defined as at least 330 minutes (5.5 hours) on Mondays, Tuesdays, Thursdays, and Fridays, at least 420 minutes (seven hours) on Wednesdays, and at least 480 minutes (eight hours) on Saturdays and Sundays.[23]

We aggregate the children's daily time spent on PA (school-time/leisure-time) to the average amount of time spent on PA per measurement week. This means that in our analyses, PA is measured in average number of minutes per day (per measurement week).

1.4.3. Summary Statistics

In our estimation samples, we have school performance data from 1,014 children, school-time PA data from 536 children, and leisure-time PA data from 509 children. Around half of the sample is enrolled in a treatment school. On average, children took five tests per year, with an average score of 57.28 in the pre-treatment year and 65.16 in the post-treatment year.

Children wore their accelerometer for four days on average during each PA measurement week. During these days, they wore their accelerometer for approximately five hours per day during school hours (309 minutes) in the pre-treatment year, out of which they spent 1.5 hours (93 minutes) on PA. Figure 1.2 shows that most of the school-time PA (black bars) takes place between noon and 1:00 p.m., which is also the time when most schools have recess. During leisure time in the pre-treatment year, children wore their accelerometer for approximately nine hours per day (548 minutes), out of which they spent 3.5 hours (204 minutes) on PA. Figure 1.2 shows that time spent on PA during leisure time does not differ substantially between school days (grey bars) and weekend days (dotted line). However, the figure does highlight the "incarceration effect" of school on PA; children are much more active between 9:00 a.m. and 3:00 p.m. on weekend days than during school days.

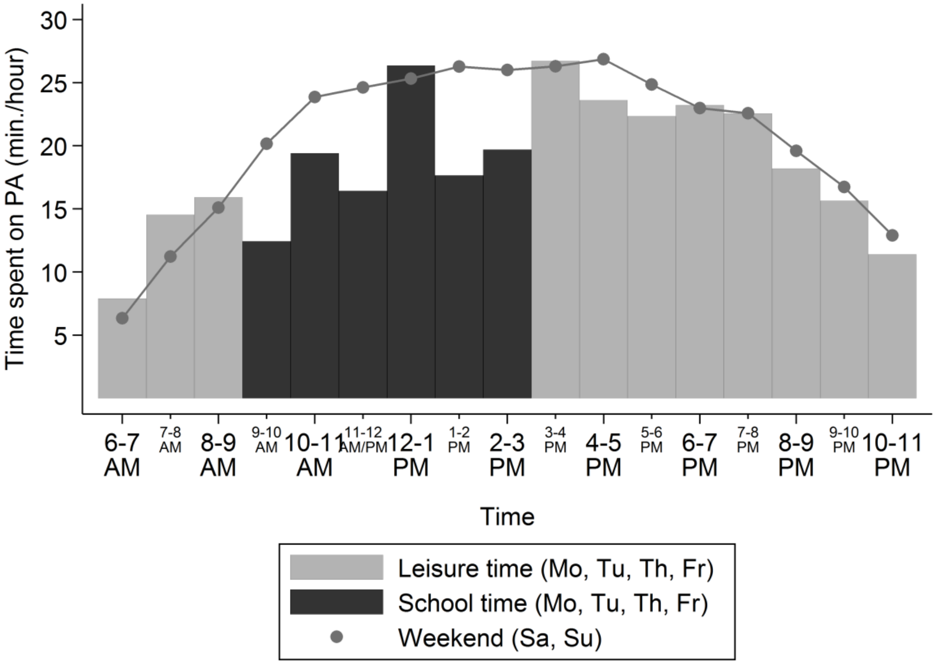
[20] Usually, these data are already excluded because they are recognized as non-wearing time (Choi et al., 2011). Excluding the 11:00 p.m. to 6:00 a.m. data reduces the number of child-hour observations by 1.91 percent.

[21] In our analyses per hour, we define minimum wear time as at least 30 minutes per hour.

[22] This reduces the number of child-day-school-time observations by 1.89 percent. The results are qualitatively robust to this exclusion.

[23] This reduces the number of child-day-leisure-time observations by 11.58 percent. The results are qualitatively robust to this exclusion.

Figure 1.2. Pre-treatment time spent on PA (excluding Wednesdays)



Source: PA data.

Between the pre- and post-treatment year, average time spent on PA during school time did not change significantly, but average leisure-time PA decreased by 21 minutes per day.[24] Appendix 1.1 describes the control variables we use in the analyses and provides summary statistics of all variables.

1.5. RESULTS

We first show the "naïve" OLS estimates of the relationship between time spent on PA and school performance before the Active Living interventions were implemented, i.e., in 2012/2013. The results in Table 1.2 show that, in our sample, every additional 10 minutes spent on PA during school time is related to a 0.52 points lower test score (5.6 percent of a standard deviation in the full sample). This negative correlation between school-time PA and school performance is significantly different from zero at the 1 percent level. Pre-treatment leisure-time PA is not significantly related to pre-treatment school performance.

[24] Note that this trend is controlled for in the difference-in-differences estimations.

1.5.1. Main results

The results in Table 1.2 could be biased due to reverse causality or omitted variables. It is possible that children who spend less time on PA at school use that time to invest in other cognitively stimulating activities such as studying, reading, or playing music. To account for this, we will now turn to our difference-in-differences estimations based on the Active Living Program (see Table 1.3).

The difference-in-differences estimations reveal that the Active Living Program has a negative effect on school performance, which is significantly different from zero at the 5 percent level. Due to the interventions, the increase in school performance of children in the treatment schools is 1.17 points smaller (0.13 standard deviations) compared to the increase of those in the control schools. This result is robust across all specifications, i.e., without control variables, including child fixed effects, controlling for age as well as child fixed effects, and when we adjust *p*-values for the small number of clusters. Additional analyses reveal that the effect of the Active Living Program is similar for school performance on language tests (-0.12 standard deviations) and math tests (-0.15 standard deviations; see Table A1.8 in Appendix 1.4).

To analyze whether the treatment effects persist over time, we estimate the treatment effect one additional year post-treatment, i.e., in 2014/2015. An important point is that we can only estimate this long-term treatment effect for half our sample, namely for those who were enrolled in grade 4 in 2012/2013 and were (therefore) enrolled in grade 6 in 2014/2015. Those who were enrolled in grade 5 in the pre-treatment year of 2012/2013 finished primary school in 2013/2014 and were enrolled in secondary education in 2014/2015. The results from these analyses reveal that the point estimate remains similar in the longer run (2014/2015). However, the estimates in both 2013/2014 and in 2014/2015 are not statistically significant, which may be due to the smaller sample size (see Figure A1.7 in Appendix 1.4).

Table 1.2. Pre-treatment correlation between time spent on physical activity and school performance

Dependent variable: overall school performance	(1)	(2)	(3)	(4)	(5)	(6)
	PA during school time			PA during leisure time		
	Base	Incl. controls from school perf. data	Incl. controls from PA data	Base	Incl. controls from school perf. data	Incl. controls from PA data
Time spent on PA (min./day)	-0.033	-0.031	-0.052	-0.008	-0.007	-0.009
	(0.013)	(0.011)	(0.012)	(0.005)	(0.005)	(0.005)
Adjusted p-value	[0.018]	[0.012]	[0.004]	[0.142]	[0.214]	[0.116]
Age when tests were taken (months)		0.966	0.969		1.112	0.938
		(0.783)	(0.704)		(0.778)	(0.595)
Age squared		-0.005	-0.005		-0.005	-0.004
		(0.003)	(0.003)		(0.003)	(0.002)
Cohort (1 = 5th graders, 0 = 4th graders)		10.183	10.096		10.080	9.657
		(0.819)	(0.879)		(0.957)	(1.032)
Gender (1 = boy)		0.089	0.326		0.193	0.255
		(0.981)	(0.962)		(1.100)	(1.027)
Wearing time of accelerometer (min./day)			0.013			0.129
			(0.101)			(0.050)
Wearing time squared			-0.000			-0.000
			(0.000)			(0.000)
Outside temperature during PA measurement week (Celsius*10)			-0.054			-0.052
			(0.031)			(0.040)
Temperature squared			0.000			0.000
			(0.000)			(0.000)
Amount of sunshine during PA measurement week (hours/day)			1.125			0.591
			(0.530)			(0.659)
Sunshine squared			-0.158			-0.081
			(0.085)			(0.101)
Rain during PA measurement week (1 = yes)			-0.562			-0.507
			(0.924)			(1.011)
Constant	60.654	9.015	9.258	59.427	-2.588	-26.163
	(1.393)	(47.984)	(49.228)	(1.046)	(48.199)	(43.406)
Observations	549	549	512	533	533	500
R-squared	0.012	0.270	0.288	0.003	0.263	0.281
Mean school performance in estimation sample	57.665	57.665	57.668	57.743	57.743	57.767
Standard deviation	(7.795)	(7.795)	(7.687)	(7.827)	(7.827)	(7.681)
Mean school performance in full sample	57.803	57.803	57.803	57.803	57.803	57.803
Standard deviation	(8.304)	(8.304)	(8.304)	(8.304)	(8.304)	(8.304)

Notes: The table shows the results of six ordinary least squares regressions. The dependent variable in each column is overall school performance in 2012/2013. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: Merged school performance and PA data.

Table 1.3. Difference-in-differences estimations of the effect of the Active Living Program on school performance

Dependent variable: overall school performance	(1) OLS	(2) FE	(3) FE
Treatment * Post	-1.289	-1.090	-1.170
	(0.567)	(0.543)	(0.544)
Adjusted p-value	[0.037]	[0.061]	[0.048]
Treatment (1 = Treatment group)	0.809		
	(0.846)		
Post (1 = 2013/2014, 0 = 2012/2013)	8.589	8.448	2.950
	(0.471)	(0.444)	(3.798)
Age when tests were taken (months)			0.972
			(0.400)
Age squared			-0.002
			(0.001)
Child fixed effects	No	Yes	Yes
Constant	56.860		
	(0.531)		
Observations	2,064	2,028	2,028
R-squared	0.199	0.809	0.812
Number of children in the estimation sample	1,050	1,014	1,014
Mean school performance in estimation sample	61.312	61.222	61.222
Standard deviation	(8.910)	(8.890)	(8.910)
Mean school performance in full sample	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)

Notes: The table shows the results of three ordinary least squares regressions. The dependent variable in each column is overall school performance. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

1.5.2. Robustness Checks

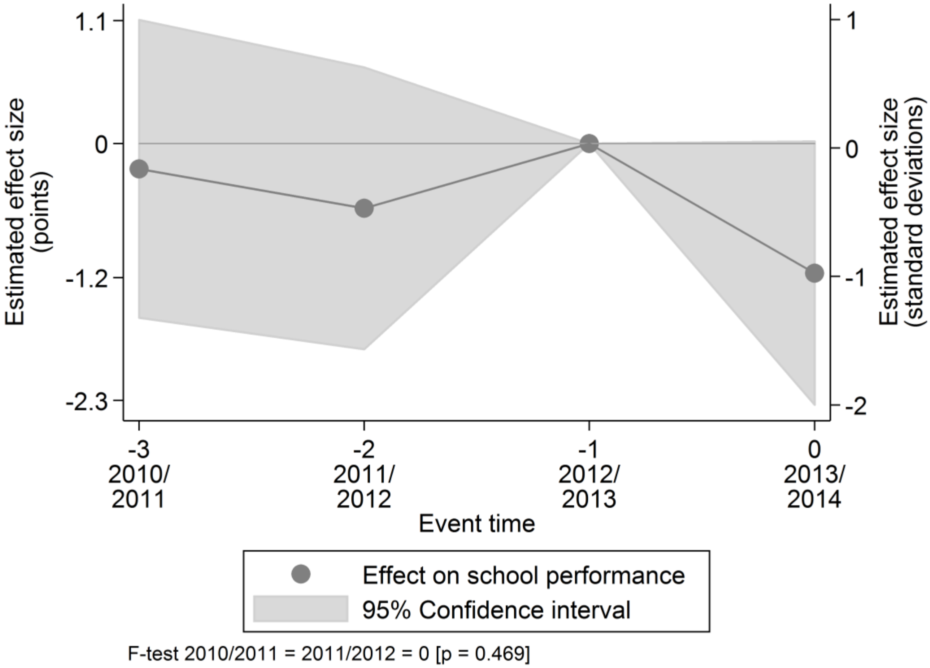
Event Study Analysis

In this section, we investigate whether it is plausible that the negative effects on school performance are indeed due to the Active Living interventions. We therefore conduct an event study analysis in which we repeat our difference-in-differences analysis using interactions of the treatment indicator with three pre-treatment years and the post-treatment year (leaving the 2012/2013 school year as an omitted category).

The results from this analysis show that the difference between the treatment and control group does not change significantly in the pre-treatment years (see Figure 1.3). This indicates that the pre-treatment trends of the treatment and control groups are relatively parallel. Second, the figure shows that the difference between the treatment and control groups becomes significantly negative in the post-treatment year. This is in line with our previous findings as presented in Table 1.3. Moreover, there is no sign of a downward trend in the treatment effects in the pre-treatment years, which makes it less likely that the difference between the treatment and control group would also have become negative in the absence of the Active Living Program. Taken together, we conclude that Figure 1.3 provides strong indications that the parallel trends assumption holds, allowing for a causal interpretation of our estimates of the effects of the Active Living Program.

Repeating the event study analyses for our different specifications of the treatment effect on school performance indicates that the pre-treatment trends in school performance on language tests appear to be parallel as well. However, the estimates of the treatment effect on math performance should be interpreted with caution, given that we only find weak support for the parallel trends assumption (see Figure A1.8 in Appendix 1.4).

Figure 1.3. Event study analysis of the effect of the Active Living Program on school performance



Notes: The figure shows the results of an ordinary least squares regressions. The dependent variable is overall school performance. The estimated effect sizes based on a difference-in-differences estimation are plotted on the y-axis. Independent variables are year fixed effects, age, age squared, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: School performance data.

Checking for Outliers

In Appendix 1.2, we show the results of several robustness checks. In the first section of Appendix 1.2, we investigate whether outliers drive our estimates. Due to the relatively small number of schools included in this study and the heterogeneity in the implemented interventions across treatment schools, there is a risk that one particular treatment or control school drives our results.

However, the results remain robust across regressions in which we estimate the treatment effects while excluding one different school at a time. We therefore conclude that outlier treatment or control schools do not drive the estimates.

Configuration of Design Choices

In the second section of Appendix 1.2, we discuss which configuration of design choices (i.e., intervention package) is most strongly associated with the observed negative effect on school performance. Although running the difference-in-differences regressions while *including* only one treatment school at a time (and all control schools) reveals that none of the design choices lead to a positive effect on school performance; the results do not shed more light on the question. The configuration of design choices is so heterogeneous across schools that it is not possible to identify a particular intervention (or combination of interventions) that may drive the results.

Comparison to Non-Active Living Schools

In the third section of Appendix 1.2, we analyze whether the negative treatment effect we find is (partially) due to an unexpected increase in school performance in the control group rather than a decrease in school performance in the treatment group. However, difference-in-differences analyses in which we compare control schools to non-Active Living schools lead us to conclude that, even if control schools changed their attitudes or policies toward PA, it does not appear to have had a significant effect on school performance, either in statistical or economic terms. It is therefore unlikely that unintentional and unobserved changes within control schools bias our estimates of the treatment effect on school performance.

Running the difference-in-differences analyses in which we compare the treatment schools to non-Active Living schools suggests that the estimated effects as presented in Table 1.3 might be slightly biased upward, and that the more conservative estimate of the average effect of the Active Living Program on school performance is 5.9 percent of a standard deviation.

1.5.3. Heterogeneous Treatment Effects

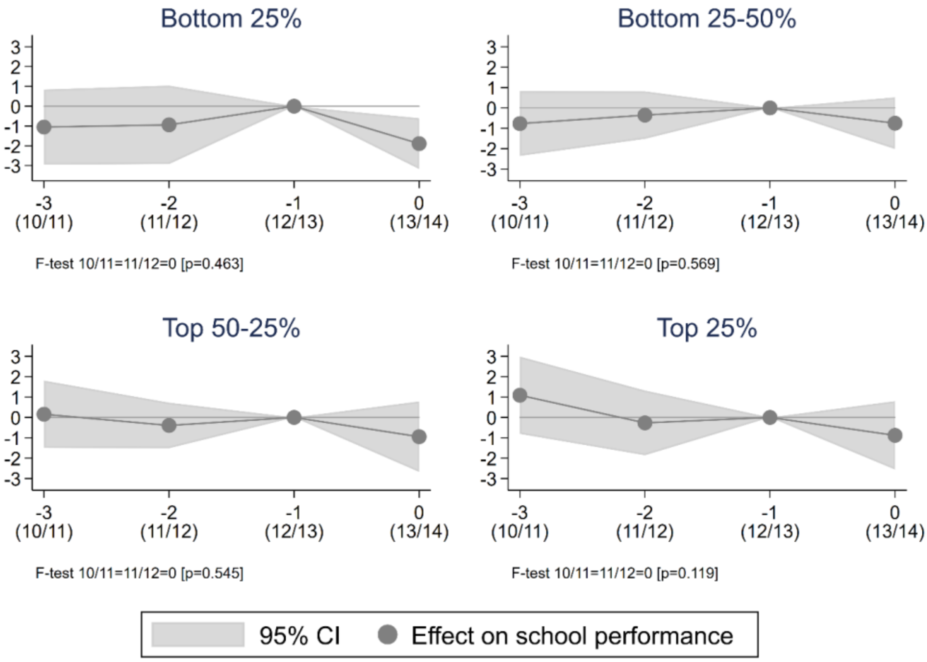
Although the average treatment effect is convincingly negative, this effect may be driven by certain groups of children, while other groups did experience an increase in school performance due to the Active Living Program. In this section, we therefore investigate whether the effects of the Active Living Program on school performance differ by subgroups of children.

Pre-Treatment School Performance

First, we analyze whether the effect of the Active Living Program on school performance is different for the worst-performing students compared to the best-performing students. The results in Figure 1.4 indicate that the Active Living Program negatively affects school performance across the distribution. However, the negative effect is strongest and statistically significant among the 25% of students who were performing the worst at school during the pre-treatment year of

2012/2013. This leads us to conclude that the Active Living Program increases inequality in school performance.

Figure 1.4. Event study analyses of the effect of the Active Living Program on school performance, across the pre-treatment school performance distribution



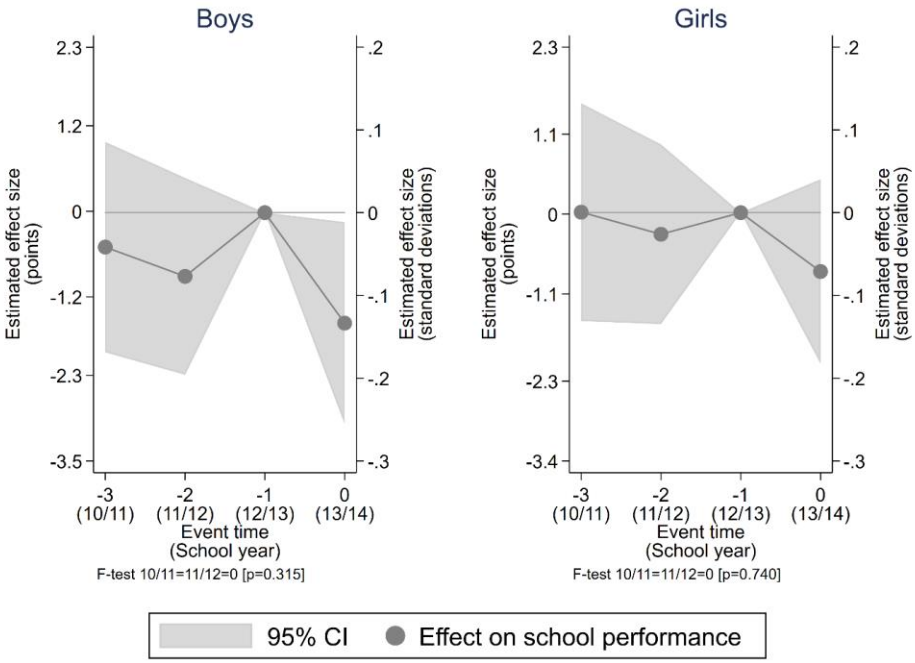
Notes: The figure shows the results of four ordinary least squares regressions, by quartile of the pre-treatment school performance distribution in 2012/2013. The dependent variable in each regression is overall school performance. The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axes. Independent variables are year fixed effects, age, age squared, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: School performance data.

When estimating these heterogeneous effects nonlinearly, using a moving window of 200 individuals across the school performance distribution in the pre-treatment year of 2012/2013, we find that the negative effect is statistically significant in the bottom 18 to 38% of the distribution (see Figure A1.9 in Appendix 1.4).

Gender

Next, we investigate whether the effect of the Active Living Program on school performance is different for boys and girls. Figure 1.5 shows that, although the treatment effect is negative for both genders, it is larger and only statistically significant for boys. Due to the Active Living Program, the school performance of boys in treatment schools decreased by 13.4 percent of a standard deviation compared to boys in control schools.

Figure 1.5. Event study analyses of the effect of the Active Living Program on boys' and girls' school performance



Notes: The figure shows the results of two ordinary least squares regressions, gender. The dependent variable in each regression is overall school performance. The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axes. Independent variables are year fixed effects, age, age squared, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: School performance data.

1.6. THE TREATMENT EFFECT ON PA AND EXTERNAL VALIDITY

The robustness checks give strong indications that the negative effect of the Active Living Program on school performance is due to the implemented interventions in the treatment schools. Considering that the Active Living interventions aim to decrease sedentary behavior, we hypothesize that time spent on PA is an important driver of this effect. This also means that in the setting we study, it is still possible to conclude that PA has a positive effect on school performance if the Active Living Program (unexpectedly) has a negative effect on time spent on PA. In order to analyze this, we use the accelerometer data.[25]

We identify the effect of the Active Living Program on time spent on PA by estimating the following model:

[25] Additional analyses indicate that wearing an accelerometer does not in itself have any statistically or economically significant effects on school performance.

$$PA_{ist} = \beta_0 + \beta_1(Treat_{is} * Post_t) + \beta_2 Post_t + \boldsymbol{X'_{ist}}\beta + \gamma_{3is} + \varepsilon_{3ist} \quad (1.3)$$

The *Treat* and *Post* variables, as well as child fixed effects γ, are identical to those in equation (1.1). ***X'*** is a vector of variables, including the time children wore the accelerometer per day and several variables indicating the weather conditions during each PA measurement week (i.e., temperature, hours of sunshine, and rainfall). To allow for nonlinear relationships between these variables and time spent on PA, we also control for accelerometer wearing time squared, temperature squared, and hours of sunshine squared.[26] Considering the Active Living interventions implemented, we expect to find that the difference-in-differences estimator β_1 is positive and significantly different from zero.

Table 1.4 reports the results of the difference-in-differences analyses on time spent on PA during school time and leisure time. Based on these results, we conclude that the Active Living Program significantly increases time spent on PA during school hours by 9.30 minutes per school day. This effect is equal to 0.34 standard deviations. There appears to be a small crowding-out effect of leisure-time PA, as the Active Living Program decreases time spent on PA during leisure time by 3.8 minutes per day (0.08 standard deviations), but this effect is not significantly different from zero.

Additional analyses indicate that the Active Living Program has a positive effect on time spent on PA during school hours of approximately 10 minutes per day across the pre-treatment school performance distribution. The Active Living Program does not have a significant effect on leisure-time PA in almost all parts of the school performance distribution (see Figure A1.10 in Appendix 1.4). Moreover, for girls compared to boys, the Active Living Program has a stronger positive effect on time spent on PA during school time, but it also has a stronger negative effect on time spent on PA during leisure time. In total, boys' time spent on PA increases by approximately 6 minutes per day, due to the Active Living Program, while girls' time spent on PA decreases by approximately 2.5 minutes (see Tables A1.10 and A1.11 in Appendix 1.4).

[26] Because the study design for the Active Living Program required that the pre- and post-treatment PA measurements were scheduled around the same time for each school, age is approximately constant in the PA data once we control for year and child fixed effects.

Table 1.4. Difference-in-differences estimations of the effect of the Active Living Program on time spent on physical activity during school time and leisure time

Dependent variable: time spent on PA (min./day)	(1)	(2)	(3)	(4)	(5)	(6)
		School time			Leisure time	
	OLS	FE	FE	OLS	FE	FE
Treatment * Post	25.337	22.544	9.298	-2.247	-5.638	-3.814
	(8.552)	(8.611)	(3.283)	(7.836)	(7.126)	(3.853)
Adjusted p-value	[0.009]	[0.017]	[0.016]	[0.778]	[0.439]	[0.344]
Treatment (1 = Treatment group)	-14.990			1.414		
	(5.243)			(5.293)		
Post (1 = 2013/2014, 0=2012/2013)	-15.175	-13.062	-13.963	-18.765	-19.374	-24.136
	(5.872)	(6.179)	(3.182)	(4.426)	(4.494)	(5.104)
Wearing time of accelerometer (min./day)			0.790			0.412
			(0.305)			(0.191)
Wearing time squared			-0.001			-0.000
			(0.000)			(0.000)
Outside temperature during PA measurement week (Celsius*10)			-0.269			0.184
			(0.069)			(0.146)
Temperature squared			0.001			-0.000
			(0.000)			(0.001)
Amount of sunshine during PA measurement week (hours/day)			8.483			10.339
			(2.258)			(4.660)
Sunshine squared			-0.771			-0.853
			(0.268)			(0.520)
Rain during PA measurement week (1 = yes)			1.441			-2.597
			(1.820)			(4.654)
Child fixed effects	No	Yes	Yes	No	Yes	Yes
Constant	100.280			200.506		
	(4.111)			(3.049)		
Observations	1,355	1,204	1,072	1,301	1,120	1,018
R-squared	0.058	0.111	0.503	0.041	0.137	0.435
Number of children in the estimation sample	753	602	536	741	560	509
Mean time spent on PA in estimation sample	91.510	91.973	93.500	191.819	192.564	192.978
Standard deviation	(26.989)	(26.756)	(25.183)	(49.545)	(48.696)	(46.767)
Mean time spent on PA in full sample	91.510	91.510	91.510	191.819	191.819	191.819
Standard deviation	(26.989)	(26.989)	(26.989)	(49.545)	(49.545)	(49.545)

Notes: The table shows the results of six ordinary least squares regressions. The dependent variable in columns (1) through (3) is time spent on PA during school time (min./day). The dependent variable in columns (4) through (6) is time spent on PA during leisure time (min./day). Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: PA data.

1.6.1. Instrumental Variables Estimation

The estimated effects of the Active Living Program on time spent on PA can help interpret the size of the effect of PA on school performance. The causal effect of time spent on PA on school performance can be calculated using the Wald estimator, by dividing the estimated treatment effect on school performance (i.e., the reduced form, captured by α_1) by the estimated treatment effect on time spent on PA (i.e., the first stage, captured by β_1). Under the assumption that the Active Living Program only affects school performance through school-time PA, we would conclude that, on average, an extra 30 minutes spent on PA during school hours reduces school performance by 1.76 score points (18.9 percent of a standard deviation).[27] This result needs to be interpreted with caution; although the effect of the Active Living Program on school-time PA is statistically significant (*t*-statistic of 2.89), it does not meet the criterion of a strong instrument.

Note that the assumption of instrument exogeneity appears to be strong as it is conceivable that the Active Living Program also affected other potential confounders (e.g., increased leisure-time PA and improved health and well-being from adjusted playgrounds or safer roads around school). However, as these examples indicate, it is implausible that such confounders are *negatively* related to school performance. Therefore, even if the exclusion restriction would be violated by these confounders, they cannot explain the *negative* effect on school performance; controlling for such variables would make the negative effect on school performance even stronger. Considering the robust negative effects of the Active Living Program on school performance, we therefore conclude that time spent on PA during school hours appears to have a negative effect on school performance in the context of this study.

1.6.2. External Validity

Appendix 1.3 discusses the external validity of our findings. We use data from the 2015 Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) to show that exercising or practicing a sport before going to school is negatively correlated with 15-year-old students' performance within 60 countries and economies across the world. Moreover, in all countries, this correlation is even more negative in the full sample than within the sample of children of low-educated parents. Evidently, these analyses are not causal in nature and should be interpreted with caution. However, the similarity between our "naïve" pre-treatment OLS estimates and the Wald estimator indicates that the bias of the OLS might be small, and that, if anything, the OLS estimate is slightly biased toward zero. The results in Appendix 1.3 therefore seem to indicate that it is possible that increasing PA in the mornings of school days can also have negative effects on school performance in other high-income countries such as the United States, Sweden, or Japan, as well as in middle-income countries such as Brazil, Colombia, or Tunisia. Additionally, they suggest that our main results are generalizable outside of low-SES regions.

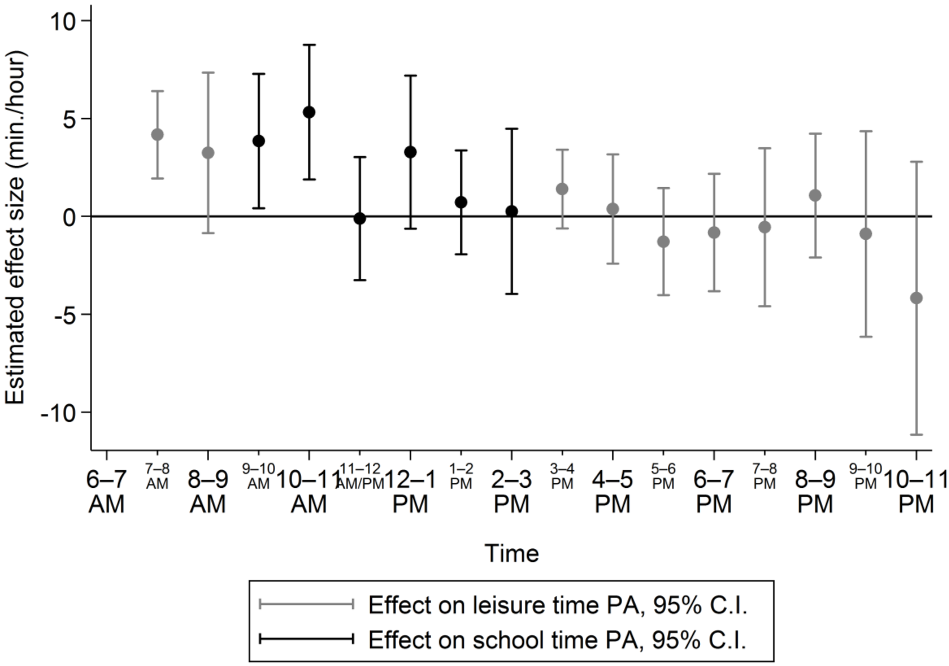
1.7. MECHANISMS

There may be several mechanisms through which time spent on PA during school hours may affect school performance. Before analyzing potential mechanisms, we first analyze in more detail

[27] When using results from Table A1.6 (column 6) in Appendix 1.2 and Table 2.4 (column 3) in the main text, a back-of-the-envelope calculation of the Wald estimator is -0.059 (-0.546 score points / 9.298 minutes), implying that for each minute increase in school-time PA, overall school performance decreases by 0.059 points (0.63 percent of a standard deviation). For boys, the back-of-the-envelope calculation of the Wald estimator is -0.229 points per minute during school time (1.99 percent of a standard deviation), and for girls, it is -0.081 points per minute during school time (0.71 percent of a standard deviation).

at what time of the day the effect on school-time PA is strongest. Figure 1.6 shows the estimated treatment effects on time spent on PA by hour on school days.[28] When interpreting this figure, it is important to note that children typically commute to school between 8:00 a.m. and 9:00 a.m., and that recess time typically takes place between 10:00 a.m. and 11:00 a.m. (around 15 minutes) and between 12:00 p.m. to 1:00 p.m. (around one hour), meaning that most instruction time takes place between 9:00 a.m. and 10:00 a.m., 11:00 a.m. and 12:00 p.m., and 1:00 p.m. and 3:00 p.m.

Figure 1.6. The effect of the Active Living Program on time spent on PA, by hour on school days (excluding Wednesdays)



Notes: The figure shows the results of 16 ordinary least squares regressions. The dependent variable in each regression is time spent on PA on school days (excluding Wednesdays). The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axis. Independent variables are a year dummy, wearing time of the accelerometer, wearing time squared, weather variables, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The effect of the Active Living Program on time spent on PA between 6:00 a.m. and 7:00 a.m. could not be estimated because the number of observations is too small.
Source: PA data.

It appears that the Active Living Program affects time spent on PA between 7:00 a.m. and 1:00 p.m., but not in the afternoons or evenings of school days. The figure highlights that the Active Living Program does not appear to crowd out leisure-time PA on school days; if anything,

[28] We exclude Wednesdays because on these days, children have a different time schedule than on the other weekdays.

it may have increased active transportation to school (positive effect between 8:00 a.m. and 9:00 a.m.). Second, the figure highlights that some of the additional time spent on PA during school hours has taken place during instruction time (positive effect between 9:00 a.m. and 10:00 a.m.).

Figure 1.6 suggests that PA increased in the mornings of schooldays. This insight can help us pinpoint which mechanisms may underlie the negative effects of the Active Living Program on school performance. The effects on school performance may be negative because increased time spent on PA crowds out other beneficial activities in the mornings of school days. The effect could also be negative due to behavioral changes *after* school. Because time spent on PA after school is not affected, the negative effect on school performance may be due to a substitution of sedentary activities, meaning that children substitute more cognitively stimulating activities for less cognitively stimulating ones.

1.7.1. Crowding Out of Cognitively Stimulating Morning Activities

Lateness Due to Active Transportation

The most important (sedentary) activity that affects school performance and takes place in the mornings of schooldays is instruction time. Although the interventions do not affect the curriculum and instruction time, one could argue that increased active transportation to school (positive effect on PA between 8:00 a.m. and 9:00 a.m.) may increase lateness or attendance, which would decrease instruction time received by the child. Although we cannot exclude the possibility that children in treatment schools arrived later at school than usual, this is unlikely because schools are usually very strict about the time students need to be at school. Moreover, it is unlikely that active transportation instead of inactive transportation delayed children by more than five minutes because the average distance from home to primary school is approximately 700 meters (Statistics Netherlands, 2017).

Crowding Out of Other Cognitively Stimulating Morning Activities

Children’s cognitively stimulating morning activities that take place outside of instruction time (i.e., before going to school and in recess time) could include doing homework or talking with parents and friends. However, PA can be a socializing activity as well, making it unlikely to be less cognitively stimulating than talking with parents or friends. Although it is unlikely that children do homework during recess, it remains possible that active transportation to school crowded out doing homework before going to school. Due to data limitations, we unfortunately cannot test this hypothesis.

1.7.2. Substitution of Cognitively Stimulating Afternoon Activities

The negative effect on school performance could be explained by another mechanism: increased time spent on PA during school hours increases fatigue after school hours, leading children to substitute sedentary cognitively stimulating activities in leisure time (e.g., playing music, reading, or doing homework) for other sedentary but less cognitively stimulating activities (e.g., watching television).[29]

Moreover, we can investigate this potential mechanism in more detail across cohorts of sixth graders. As part of the Onderwijs Monitor Limburg (OML) project, every year since 2008, sixth graders have been asked to complete a questionnaire. The following questions are related to sedentary activities that may be cognitively stimulating: "*Which of the following activities do you do in a club or a group? Music lessons, Chess, Arts and crafts.*"[30] These questions were asked to sixth graders from 2012/2013 through 2014/2015.[31] Children who play music and who do not do arts and crafts, have significantly higher school performance in the absence of treatment (see Table A1.12 in Appendix 1.4). Playing chess in a group or club is not significantly related to school performance in the absence of treatment, which makes this variable a less plausible mechanism for the treatment effects on school performance.

Given that these questions are only asked of sixth graders, we do not have pre- and post-treatment measurements for the same individuals. We can, however, run difference-in-differences analyses by comparing pre- and post-treatment cohorts within treatment and control schools. The key assumption for causal identification with this strategy is that the difference between the characteristics in the 2012/2013, 2013/2014, and 2014/2015 cohorts in treatment schools would have been similar to the difference between these cohorts in control schools if the Active Living Program had not taken place. We estimate the following model:

$$Arts_{isc} = \theta_0 + \theta_1(Treat_{is} * Post'_c) + \theta_2 Treat_{is} + \theta_3 Post'_c + \varepsilon_{4isc} \qquad (1.4)$$

In this equation, *i* indexes individual pupil, *s* indexes school, and *c* indexes cohorts of sixth graders. *Arts* stands for any of the three cognitively stimulating sedentary activities that can be identified in the OML questionnaire. The *Post'* variable is a cohort indicator that takes the value of one for children who were enrolled in grade 6 in 2013/2014. The difference-in-differences estimator then takes on the value of one for sixth graders in treatment schools in 2013/2014 (i.e., the treatment year of the Active Living Program), and it takes on the value of zero for sixth graders in treatment schools in 2012/2013, as well as for sixth graders in control schools in any given school year. The *Treat* variable is identical to those in equations (1.1) through (1.3). Considering the negative treatment effect on school performance, we hypothesize that the Active Living Program may have decreased the probability of taking music lessons, playing chess or doing arts

[29] Because we do not find that the Active Living Program increased time spent on PA during leisure time, we assume that children substituted one sedentary activity for another during leisure time.

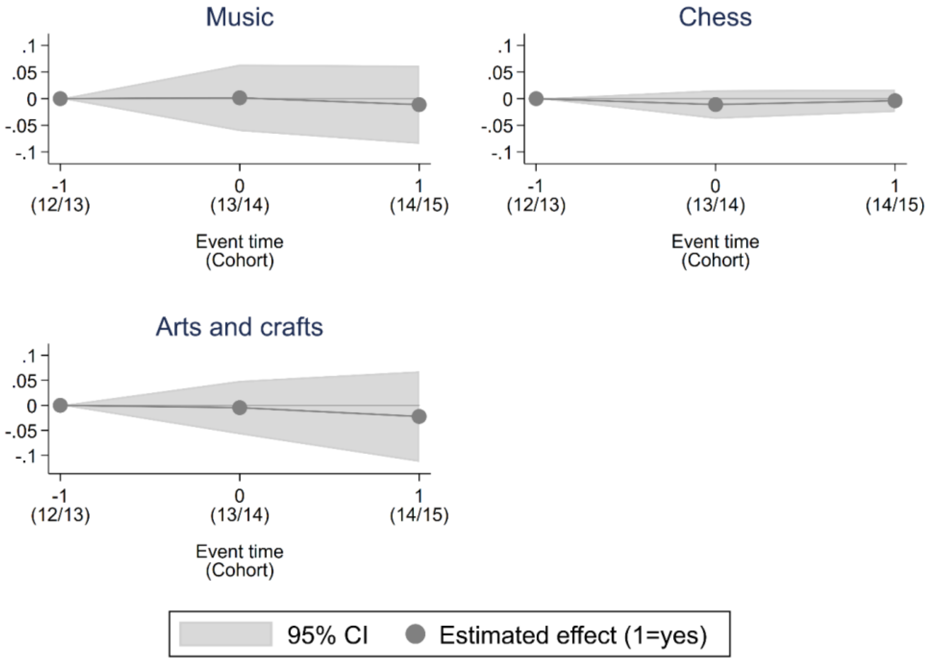
[30] Other activities that could be ticked are: football, tennis, hockey, scouts, dancing/ballet, swimming, and horseback riding. We do not find evidence that the sports activities in this list are significantly affected by the Active Living Program, which is not surprising considering that we do not find effects on PA in the afternoons of schooldays.

[31] The variables provide the lower bound of the number of children who play a musical instruments, play chess or do arts and crafts, because some children may do these activities without being member of a group or club.

and crafts (assuming that these activities are substituted by less cognitively stimulating ones in leisure time).

Figure 1.7 shows that we cannot conclude that the Active Living Program has any effect on the probability of taking music lessons, playing chess or doing arts and crafts in a club or a group; the estimated treatment effects are not significant, either in statistical or economic terms. Due to data limitations, we also cannot rule out the possibility that the Active Living Program had a (negative) effect on time spent doing homework, reading, or participating in other cognitively stimulating post-school activities.

Figure 1.7. The Effect of the Active Living Program on Taking Music Lessons, Playing Chess and Doing Arts and Crafts



Notes: The figures show the results of three ordinary least squares regressions. The dependent variables are whether the child takes music lessons in a club or a group (1 = yes), whether the child plays chess in a club or a group (1=yes) and whether the child does arts and crafts in a club or a group (1=yes). The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axes. Independent variables are cohort fixed effects, a treatment dummy, parental education, age and gender. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: Authors' calculations based on Onderwijs Monitor Limburg questionnaires 2013-2015.

1.7.3. Other Potential Mechanisms

Although every additional minute spent on PA must come at the cost of a decrease in time spent on something else, based this section we conclude that it seems implausible that, in the case of the Active Living Program, time spent on PA crowded out cognitively stimulating activities in the mornings of schooldays. Given the results in Figure 1.7, it also seems implausible that a decrease in playing music, chess or arts and crafts can explain our findings. Due to data limitations, other potential mechanisms that may account for the negative treatment effect on school performance remain uncertain. For instance, PA may have crowded out investments in potentially beneficial activities during leisure time, such as reading, studying, or social activities. Moreover, it remains possible that PA also crowded out potentially harmful activities such as smoking or watching television, which would attenuate the negative effect on school performance.

1.8. CONCLUSIONS

This chapter investigates whether encouraging children to become more physically active in their everyday life affects their primary school performance. Our estimations are based on a quasi field experiment called the Active Living Program, which aimed to increase physical activity and decrease sedentary behavior among 8- to 12-year-olds living in low-SES areas in the Netherlands. The interventions focus on active transportation and active play at school and during leisure time. Results from difference-in-differences analyses reveal that the Active Living Program significantly decreases school performance by at least 5.9 percent of a standard deviation in the treatment year. This result is robust across a variety of specifications, but the negative effect is strongest among the worst-performing students. The results of event study analyses and several robustness checks provide strong indications that the negative effect is indeed due to the Active Living interventions causing a significant decrease in the treatment group's school performance.

Besides its effects on school performance, the Active Living Program causes a significant increase in time spent on physical activity during school time of 0.34 standard deviations (9.30 minutes per school day). The effect remains robust across the pre-treatment school performance distribution. The Active Living Program does not significantly affect leisure-time physical activity. We conclude that time spent on physical activity during school hours appears to have a negative (short-term) effect on children's school performance in the context of this study.

We can put our estimates of the treatment effects into perspective by comparing them to the estimated effect sizes on teenagers' academic achievement of other well-known interventions. Anderson (2008) reviews the treatment effects of Abercedarian (ABC), the Perry Preschool, and Early Training Projects (ETP) in the United States.[32] Although these interventions may not be directly comparable to the Active Living Program due to different target groups and contexts, they do provide potential benchmarks for our estimates. If we simply compute an average of the

[32] Beginning as early as 1962, these programs targeted disadvantaged African-Americans in North Carolina, Michigan, and Tennessee. Through random assignment, treated children in each experiment received several years of preschool education (with intensity differing across programs). Interventions continued until the children began regular schooling (Anderson, 2008).

treatment effects weighted by sample size,[33] the overall impact of the three programs on teen outcomes is about 0.28 standard deviations. This is almost five times as large as our estimates of the effect of the Active Living Program on school performance.

Our results shed new light on the effects of the encouragement of physical activity among children and adolescents. To our knowledge, this study is the first to analyze the causal effect of encouraging informal physical activity on educational outcomes. Previous studies have shown a positive relationship between formal physical activity (i.e., participating in sports or physical education) and school performance, but based on our findings, we conclude that these results appear not to be generalizable to informal physical activity such as active transportation or active play. One reason for this may be that formal physical activity is more effective for fostering skills that are beneficial for school performance, such as confidence, discipline, or competitiveness, than informal physical activity. Moreover, because exercising is generally more intense than being active in everyday life, formal physical activity may yield higher returns in terms of health and well-being. If children usually reallocate study hours (or other cognitively stimulating tasks such as paying attention during instruction hours) to physical activity, the combination of the absence of the development of sports-related skills and the potentially lower health benefits could explain why the encouragement of informal physical activity has a negative effect on school performance.

Additionally, our results indicate that the encouragement of informal physical activity may increase inequality in school performance. The Active Living Program focuses on children living in low-SES areas, where (pre-treatment) school performance is lower. This means that the Active Living Program negatively affects school performance in the worse-performing regions of the Netherlands, and that within these regions, the negative effect is strongest among the worst-performing students.

Although our results may not be generalizable to interventions that encourage formal physical activity or that focus on adults, we do find suggestive evidence of the possibility that increasing adolescents' physical activity in the mornings of school days can also have negative effects on school performance in other high- or middle-income countries and outside of low-SES regions. This means that this study has important implications for policymakers, educators, and researchers around the world who aim to increase physical activity among children and adolescents. First, this study highlights that they cannot assume that every type of physical activity intervention will only have positive externalities if it is effective in increasing time spent on physical activity. To make an informed decision about whether to increase the time children spend on physical activity, policymakers and educators must weigh the benefits for health and well-being of the child relative to the potential cost of decreased educational performance. In this process, they implicitly or explicitly make a choice as to what extent education outcomes and health outcomes are important for children. This means that despite the existing literature that points toward a positive relationship between physical activity and educational outcomes, policymakers must remain alert when implementing interventions, and carefully monitor all the (assumed) benefits and potential

[33] This calculation is identical to Deming (2009) and is based on Table 3 of Anderson (2008).

costs. Another important takeaway is that the costs and benefits of physical activity may not be the same for all children. The worst-performing students, who need their study and instruction time the most, may also suffer the most if other cognitively stimulating activities are crowded out or if the quality of their instruction time decreases because they are tired or distracted from being active. To prevent increased educational inequalities, targeted interventions may be needed that increase low-performing students' physical activity without further decreasing their educational performance. Finally, before resources are allocated to increase time spent on physical activity among children and adolescents, we first need to improve our understanding of the causal effects of such interventions on short-term and long-term educational outcomes. Future research needs to focus on which forms of physical activity do not crowd out desirable activities.

Chapter 2. Hyper-Active Living

The Effect of Encouraging Active Living on ADHD-like Symptoms Among Boys and Girls

2.1. INTRODUCTION[34]

Attention-Deficit/Hyperactivity Disorder (ADHD) is a chronic behavioral and mental health disorder that negatively affects cognitive development and functioning in daily life. According to the American Psychiatric Association (2013), it is characterized by symptoms of inattentiveness, and hyperactivity or impulsiveness at levels that are inconsistent with age or developmental level, and that occur in multiple contexts (e.g., both at home and at school or work). It is one of the most common neurodevelopmental disorders in childhood, particularly for boys (Skogli et al., 2013; Swanson et al., 1998; WHO, 2019). ADHD is typically first diagnosed in childhood, where it is associated with decreased school performance and social functioning, and an increased probability of school dropout and demonstrating risky behaviors such as smoking and lower rates of contraceptive use (Barkley, 2002; Currie & Stabile, 2004; Fletcher & Wolfe, 2008; Kollins, McClernon & Fuemmeler, 2005; Mannuzza & Klein, 2000). When symptoms persist in adult years, ADHD is associated with impairments of social, academic and occupational functioning, such as increased risk of substance abuse, criminality, traffic accidents and poorer labor market outcomes (American Psychiatric Association, 2013; Fletcher & Wolfe, 2008; Freeman, 1996; Hoy et al., 1978; Kessler et al., 2005; Mannuzza & Klein, 2000; Satterfield, Hoppe & Schell, 1982; Vaa, 2014). Early and effective management of the disorder is therefore crucial.

This chapter investigates whether encouraging children to spend more time on physical activity in their everyday life affects the prevalence of ADHD-like symptoms among 12-year-old boys and girls. Data on ADHD-like symptoms are taken from the *Onderwijs Monitor Limburg:* a yearly survey among sixth graders, which includes six questions that are similar to the internationally recognized DSM-5 criteria for ADHD. In order to identify the causal effect of stimulating physical activity on ADHD-like symptoms, we use a quasi field experiment called the Active Living Program, which has been shown to successfully increase time spent on physical activity.[35] By conducting difference-in-differences analyses across treated and non-treated cohorts of sixth graders in treatment and control schools, we can identify the causal effect of stimulating physical activity in everyday life on the prevalence of inattentiveness and hyperactivity-impulsivity among adolescents.

We conclude that boys' self-assessed ability to do exactly what the teacher asks them to do (one of three indicators of inattentiveness) decreases significantly two years post-treatment; the probability that they fail themselves for this 'subject' increases by 2 percentage points, which is an increase of 100% compared to boys' pre-treatment probability to fail at doing exactly what the teacher asks them to do. Boys also become significantly worse at being calm and quiet when the teacher wants them to (one of three indicators of hyperactivity-impulsivity); the probability that

[34] This chapter has benefited from valuable comments by Eric Bonsang, Lex Borghans, Thomas Dohmen and Bart Golsteyn. Special thanks go out to all schools and children participating in this study and to all Maastricht University colleagues involved in the data collection. This study was partly funded by the Netherlands Organization for Health Research and Development (ZonMW), Project Number 200130003.

[35] See Chapter 1 for a detailed description about the Active Living Program, as well as its treatment effects on time spent on physical activity, which amount to 0.34 standard deviations or 9.30 minutes per day during school-time.

treated boys fail at being calm and quiet increases by 3 to 4 percentage points two years post-treatment, which is an increase of 45 to 87% compared to boys' pre-treatment probability. We find no significant treatment effects on ADHD-like symptoms for girls, which may explain the non-significant treatment effects in the full sample. Considering that boys are more likely to display ADHD-like symptoms and be diagnosed with ADHD than girls, our results imply that the Active Living Program increases gender differences in the prevalence of ADHD-like symptoms.

This chapter contributes to the literature that studies the prevalence and treatment of ADHD and ADHD-like symptoms in children and adolescents. Previous studies that found that physical activity can decrease ADHD-like symptoms, focus on formal physical activity such as doing moderate to vigorous intensity cardio or aerobic exercise (Den Heijer et al., 2016; Gapin, Labban and Etnier, 2011; Ng et al., 2017; Pan et al., 2016; Smith et al, 2013). Our study shows that these effects may not be generalizable to informal or light intensity physical activity, such as active transportation or active play. Second, this chapter analyzes the overall effect of stimulating physical activity on ADHD-like symptoms for representative groups of children, rather than focusing on the symptom-reducing effects for diagnosed or symptomatic individuals alone. By focusing on diagnosed or symptomatic individuals, one may overlook potential side- or spill-over effects for non-symptomatic or non-diagnosed individuals. By estimating the overall effect, our estimates incorporate both the treatment effect on ADHD-like symptoms of (previously) symptomatic children, as well as side effects on ADHD-like symptoms of non-symptomatic children, and spill-over effects when the changed ADHD-like symptoms of some children affect the ADHD-like symptoms of others. Finally, since the Active Living Program was not introduced with the aim to affect ADHD-like symptoms in children, and was successful in reaching its direct effect of increased time spent on physical activity, this chapter shows that focusing on policies' and interventions' direct effects alone may leave unforeseen side effects and spill-over effects underexposed.

This chapter proceeds as follows. Section 2.2 briefly explains the concept of ADHD. Section 2.3 discusses the data and Section 2.4 presents the empirical strategy. Section 2.5 provides the main results and robustness checks. Section 2.6 concludes.

2.2. WHAT IS ADHD?

ADHD is a neurodevelopmental disorder that is characterized by symptoms of inattentiveness, hyperactivity and/or impulsiveness. Examples of behavioral manifestations of ADHD in children are the inability to stay on task, appearing not to listen, losing materials, or having difficulty sustaining focus (i.e., inattentiveness), fidgeting excessively, inability to stay seated, inability to wait, or intruding into other people's activities (i.e., hyperactivity-impulsivity). The American Psychiatric Association's DSM-5 diagnostic criteria for ADHD, which are well established and extensively used by clinicians across the world to identify ADHD in children, adolescents and adults, are listed in Appendix 2.1. The essential features of ADHD are: i) a persistent pattern of inattention, hyperactivity or impulsivity, ii) that impairs social, academic or occupational functioning, iii) which is not solely a manifestation of oppositional behavior, a failure to understand tasks or instructions, or another mental disorder (American Psychiatric Association, 2013).

While some individuals with ADHD have symptoms of inattentiveness as well as hyperactivity-impulsivity, others are predominantly inattentive or hyperactive-impulsive. Moreover, since symptoms must be a persistent pattern, they must be present in two or more settings, e.g., both at home and at school. Another additional requirement is that several symptoms were present before the age of 12, which means that, by definition, ADHD begins in childhood.

However, ADHD is not necessarily recognized as such in children. While Polanczyk et al. (2007) estimate that in most cultures around the world around five percent of children have ADHD, the prevalence of children with ADHD-like symptoms is much higher. For instance, in the Netherlands around 2.9 percent of children under the age of 18 are diagnosed with ADHD, whereas 3.5 percent of parents of 8-to-12 year-olds think their child has ADHD, and 25 percent of primary school children report having ADHD-like symptoms (RIVM, 2021). Although this might mean that the severity of some children's ADHD-like symptoms does not meet the threshold for ADHD diagnosis, it might also mean that many children are under-diagnosed for ADHD, and are therefore under-treated.

2.3. DATA

As part of the *Onderwijs Monitor Limburg* (OML) project, sixth graders in the Southern Limburg region of the Netherlands are asked to complete a questionnaire containing questions about their personality, behavior, preferences and out-of-school activities. This leads to a repeated cross-sectional dataset of sixth graders from 2008 onwards. Although the questionnaire is not designed to identify children with ADHD, it includes several questions that are very similar to one or more of the DSM-5 criteria for diagnosing ADHD. Table A2.1 in Appendix 2.1 lists which OML questions are related to DSM-5 criteria for ADHD, and in which school years these questions were asked to sixth graders.

For the purpose of this study, we can only include questions that were asked to at least one pre-treatment cohort (i.e., 2012/2013 or earlier) and at least one treated cohort (i.e., 2013/2014 or 2014/2015). We therefore include the following OML variables as indicators of inattentiveness: "*Give yourself a report card score. You can give yourself the following scores: 5 = insufficient, 6 = sufficient, 7 = more than sufficient, 8 = good. How good are you at:*"

- "*Finishing and handing in my work neatly and tidily*" (2011/2012 – 2014/2015)
- *"Focusing on something"* (2011/2012 – 2014/2015)
- *"Doing exactly what the teacher asks me to do"* (2011/2012 – 2014/2015)

The following OML variables are included in this study as indicators of hyperactivity-impulsivity: "*Give yourself a report card score. You can give yourself the following scores: 5 = insufficient, 6 = sufficient, 7 = more than sufficient, 8 = good. How good are you at:*"

- "*Remaining seated when the teacher wants me to*" (2011/2012 – 2013/2014)
- "*Being calm and quiet when the teacher wants me to*" (2011/2012 – 2014/2015)
- "*Being studious and not being a pest*" (2011/2012 – 2013/2014)

Since the DSM-5 criteria for ADHD are unspecified qualitative assessments (i.e., "often" displaying certain behavior), they are not directly comparable to one specific answer category of the OML variables. In this study, we therefore define 'ADHD-like symptoms' as a scale where lower scores on the OML variables represent stronger ADHD-like symptoms. This applies to each of the six indicators.

Note that children who have symptoms that are similar to the DSM-5 criteria for ADHD according to the OML questionnaire, would not necessarily be diagnosed with ADHD by a healthcare professional. For instance, we cannot measure whether the symptoms manifest themselves in multiple contexts or whether they impair children's development or functioning, which are essential for ADHD diagnosis. That is why we refer to these OML variables as 'ADHD-*like* symptoms,' so as to separate them from ADHD symptoms that could lead to diagnosis of ADHD.

2.3.1. Summary statistics

Table 2.1 shows the summary statistics for the selected OML variables of ADHD-like symptoms in the pre-treatment year 2012/2013 (see Section 2.4 for more information). Although most sixth graders give themselves a report card score of 7 ('more than sufficient') on each of the variables, there are some differences between children in treatment and control schools. Compared to children in control schools, children in treatment schools are significantly more likely to give themselves an 8 ('good') rather than a 7 ('more than sufficient') for 'focusing on something.' For each of the other five indicators, children in treatment schools are more likely to give themselves a 7 rather than an 8 compared to those in control schools, although these differences are only statistically significant for 'doing exactly what the teacher asks me to do,' and 'being calm and quiet when the teacher wants me to.' There are no statistically significant differences between children in treatment and control schools with respect to the share that gives themselves a 5 ('insufficient') on an ADHD-like symptom in the pre-treatment year 2012/2013.

Previous studies show that ADHD-like symptoms are more common in males than in females (see, e.g., Swanson et al., 1998; World Health Organization, 2019). Boys are also more likely to be diagnosed with ADHD, although it remains unclear to which extent this is related to gender-bias in diagnosing ADHD in children (Skogli et al., 2013). In our data, we also find gender differences in the prevalence of ADHD-like symptoms. For most indicators, boys are more likely than girls to give themselves a 5 ('insufficient') or a 6 ('sufficient'), and less likely to give themselves an 8 ('good') for the indicators of ADHD-like symptoms. The gender differences are largest for 'finishing and handing in my work neatly and tidily,' 'doing exactly what the teacher asks me to do,' and 'remaining seated when the teacher wants me to.'

Table 2.1. Prevalence of ADHD-like symptoms among sixth graders in the pre-treatment year 2012/2013

How good are you at :	Total	Treatment schools	Control schools	Treat – Control	*p*-value	Boys	Girls	Boys - Girls	*p*-value
Inattention									
Doing exactly what the teacher asks me to do									
5 (insufficient)	1.5%	1.5%	1.6%	-0.1%	0.858	1.8%	1.3%	0.5%	0.641
6 (sufficient)	25.0%	21.3%	28.9%	-7.6%	0.504	29.8%	21.5%	8.3%	0.031
7 (more than sufficient)	52.5%	57.1%	47.4%	9.7%	0.001	54.6%	51.0%	3.6%	0.419
8 (good)	21.0%	20.1%	22.1%	-2.0%	0.001	13.8%	26.2%	-12.4%	0.001
Focusing on something									
5 (insufficient)	3.9%	5.2%	2.4%	2.8%	0.516	4.0%	4.0%	-0.3%	0.872
6 (sufficient)	18.3%	17.6%	18.9%	-1.3%	0.272	19.5%	17.3%	2.2%	0.525
7 (more than sufficient)	46.0%	44.6%	47.4%	-2.8%	0.028	42.8%	48.3%	-5.5%	0.214
8 (good)	31.8%	32.6%	31.3%	1.3%	0.003	34.0%	30.3%	3.6%	0.385
Finishing and handing in my work neatly and tidily									
5 (insufficient)	6.0%	3.7%	8.3%	-4.6%	0.239	10.0%	3.3%	6.4%	0.002
6 (sufficient)	25.1%	24.4%	25.8%	-1.3%	0.906	32.7%	19.7%	13.0%	0.001
7 (more than sufficient)	44.3%	48.5%	39.7%	8.8%	0.192	44.2%	44.4%	-0.2%	0.969
8 (good)	24.6%	23.3%	26.2%	-2.9%	0.369	13.4%	32.6%	-19.2%	0.000
Hyperactivity-impulsivity									
Being calm and quiet when the teacher wants me to									
5 (insufficient)	5.8%	5.6%	6.0%	-0.4%	0.671	6.0%	5.3%	1.1%	0.588
6 (sufficient)	24.8%	24.3%	25.4%	-1.1%	0.162	28.0%	22.5%	5.5%	0.155
7 (more than sufficient)	41.5%	42.9%	40.1%	2.8%	0.004	41.7%	41.4%	0.4%	0.936
8 (good)	27.9%	27.2%	28.6%	-1.3%	0	23.9%	30.8%	-6.9%	0.082
Remaining seated when the teacher wants me to									
5 (insufficient)	2.3%	1.9%	2.8%	-0.9%	0.424	3.0%	2.0%	0.8%	0.567
6 (sufficient)	15.2%	15.7%	14.7%	1.0%	0.118	17.4%	13.6%	3.9%	0.228
7 (more than sufficient)	44.4%	46.3%	42.5%	3.8%	0.731	48.6%	41.4%	7.2%	0.102
8 (good)	38.1%	36.2%	40.1%	-3.9%	0.184	31.2%	43.0%	-11.9%	0.006
Being studious and not being a pest									
5 (insufficient)	6.2%	5.2%	7.2%	-1.9%	0.327	3.0%	5.6%	-2.9%	0.556
6 (sufficient)	24.3%	23.6%	25.1%	-1.5%	0.22	29.5%	20.6%	8.9%	0.019
7 (more than sufficient)	46.5%	49.8%	43.0%	6.8%	0.441	42.4%	49.5%	-7.1%	0.11
8 (good)	23.0%	21.3%	24.7%	-3.4%	0.094	21.2%	24.3%	-3.1%	0.416

Notes: For each variable, the descriptive statistics are based on the estimation sample for the main results on that variable (Tables 2.2 and 2.3).
Source: Author's calculations based on OML questionnaire 2013.
.

Boys' tendency to display stronger ADHD-like symptoms is also reflected in their time spent on physical activity (PA). On average, boys spend 97.25 minutes on PA during school-time in the pre-treatment year 2012/2013, whereas girls spend 90.95 minutes; a difference of 6.3 minutes (or 7%) that is significant at the 1% level (see Table A2.2 in Appendix 2.2).[36] Boys and girls spend similar amounts of time on PA during leisure time.[37]

2.4. EMPIRICAL STRATEGY

Section 1.3 in Chapter 1 provides a detailed explanation of the Active Living Program and the interventions it includes. To summarize, the Active Living Program is a quasi field experiment that took place between 2012 and 2014 in 21 primary schools located in low-SES areas in the Southern-Limburg region of the Netherlands. Interventions focus on encouraging PA in everyday life (i.e., both at school and in leisure time), such as active transportation to school and active play during recess and in leisure time. Since the interventions ideally have a longer lasting effect on children's active lifestyle in everyday life, as well as the fact that several interventions are permanent changes to the school environment (e.g., new fixed equipment on the school's playground), one could expect longer lasting treatment effects that go beyond the treatment year 2013/2014. The target group consists of children who were enrolled in grades 4 or 5 during the 2012/2013 school year (the 'baseline year') and transferred to grades 5 or 6 during the 2013/2014 school year (the 'treatment year'). This means that children who received treatment were in grade 6 during the treatment year 2013/2014, and one year post-treatment (the 2014/2015 school year).

As in Chapter 1, we use a difference-in-differences technique for identifying treatment effects of the Active Living Program. However, contrary to Chapter 1, where we have access to panel data on children's school performance, we now compare data on ADHD-like symptoms across cohorts of sixth graders in treatment and control schools. The key assumption for causal identification with this strategy is that the difference between the characteristics in the pre-treatment cohorts (i.e., sixth graders in 2012/2013 or before) and the treated cohorts (i.e., sixth graders in 2013/2014 and 2014/2015) in treatment schools would have been similar to the difference between these cohorts in control schools if the Active Living Program had not taken place.[38] Since the dependent variables are ordinal in nature, we estimate the following ordered probit model:

[36] Note that these statistics are based on third and fourth graders in 2012/2013, whereas the statistics in Table 2.3 are based on sixth graders in 2012/2013.
[37] See Section 1.4 in Chapter 1 for more detailed summary statistics of the PA data.
[38] Keep in mind that the key assumption for causal inference is not that there are no pre-treatment differences between children in treatment and control schools, but that, if the Active Living Program had not taken place, the pre-treatment differences would have remained similar.

$$\boldsymbol{ADHD}^*_{isc} = \beta_{21}(AL_{is} * Pre2_c) + \beta_{22}(AL_{is} * Post1_c) + \beta_{23}(AL * Post2_c) + \beta_{24}AL_{is} + \delta_c + \boldsymbol{X}'_{isc}\boldsymbol{\beta_{25}} + \varepsilon_{21isc} \quad (2.1)$$

$$ADHD_{isc} = 5\ [-\infty < ADHD^*_{isc} < \gamma_{21}],$$
$$ADHD_{isc} = 6\ [\gamma_{21} \leq ADHD^*_{isc} < \gamma_{22}],$$
$$ADHD_{isc} = 7\ [\gamma_{22} \leq ADHD^*_{isc} < \gamma_{23}],$$
$$ADHD_{isc} = 8\ [ADHD^*_{isc} \geq \gamma_{23}].$$

In this equation, *i* indexes individual pupil, *s* indexes school, and *c* indexes cohorts of sixth graders. ***ADHD**** stands for the latent variables for the six indicators of ADHD-like symptoms: γ are the threshold parameters to be estimated. Note that, as a robustness check, we also estimate the treatment effect linearly, as ordinary least squares regressions.

AL is an indicator variable with a value of one if pupils are enrolled in a school that belongs to the treatment group of the Active Living Program and zero if they are enrolled in one of the control schools.[39] The variables *Pre2, Post1* and *Post2* are indicator variables that take the value of one for children who were enrolled in grade six in 2011/2012, 2013/2014 or 2014/2015, respectively (we use the 2012/2013 school year, i.e., *Pre1*, as the reference group). We also control for year fixed effects δ as well as for ***X'***; a vector of children's characteristics, i.e., gender, age and parental education.[40] We are interested in the difference-in-differences estimators β_{22} and β_{23}, i.e., the interactions of *AL* and *Post1* and *Post2*. Considering that treatment took place among fifth and sixth graders in the 2013/2014 school year, β_{22} is the treatment effect in the treatment year, and β_{23} is the treatment effect one year post-treatment. The parallel trends assumption is supported if β_{21}, i.e., the treatment effect two years before treatment, is not significantly different from zero, because it indicates that the difference between children in control and treatment schools is not significantly different between the two pre-treatment years.[41]

We estimate equation 2.1 for the full sample, as well as for boys and girls separately. The reason for this is that boys have stronger baseline levels of ADHD-like symptoms and spend more time on PA during school-time in the pre-treatment year. Moreover, the Active Living Program has a stronger effect on their time spent on PA (see Chapter 1). It is therefore likely that the Program's effect on ADHD-like symptoms also differs by gender.

2.4.1. Robust standard errors

It is plausible that the prevalence of ADHD-like symptoms, as well as other characteristics of children that may affect these symptoms, are correlated within schools. However, since only 21 schools participated in the Active Living Program, clustering at the school-level creates a risk that clustered standard errors are downward biased (see, e.g., Angrist & Pischke, 2008; Bell & MacCaffrey, 2002; Bertrand, Duflo, & Mullainathan, 2004; Cameron, Gelbach, & Miller, 2008;

[39] Children who changed schools are excluded from the sample.
[40] All regressions also include dummy variables for missing values on the independent variables.
[41] Ideally, we would include more than two pre-treatment years to formally test the parallel trends assumption. However, this is not possible due to data limitations.

Cameron & Miller, 2015; Donald & Lang, 2007). A common way to overcome this problem is to bootstrap the standard errors. Since re-sampling should ideally occur for entire schools rather than for individual children, the best solution would be to block-bootstrap the standard errors, which generally works well with small numbers of groups (Cameron, Gelbach & Miller, 2008; Angrist & Pischke, 2008). Another method, which has been shown to work reasonably well in difference-in-differences settings, is to base inference on a *t*-distribution with degrees of freedom equal to the number of clusters minus the number of variables that are constant within clusters, rather than the standard normal distribution (Donald & Lang, 2007). By doing so, one essentially recognizes that the fundamental unit of observation is a cluster and not an individual within a cluster. In the remainder of this chapter, the correction proposed by Donald and Lang (2007) will be referred to as the "DL-correction."

For all analyses, we verify which method results in the largest standard errors (i.e., not clustering the standard errors, clustering the standard errors, bootstrapping clustered standard errors with 5,000 replications, block-bootstrapping clustered standard errors with 5,000 replications, and applying the DL-correction), and make note when inference based on a certain approach leads to different conclusions.

2.5. RESULTS

2.5.1. Main results

Table 2.2 provides the main results for the three indicators of inattentiveness. The non-significant results in column 1 imply that the common trends assumption holds: differences between children in treatment and control schools with respect to how they grade themselves on these three indicators of ADHD do not differ significantly between the two pre-treatment cohorts of sixth graders. This, in turn, means that the results in columns 2 and 3 can be interpreted as the causal effects of the Active Living Program on indicators of inattentiveness. However, we do not find any statistically significant effects on the three indicators. Also with respect to the three indicators of hyperactivity-impulsivity, we find no statistically significant effects of the Active Living Program (see Table 2.3). [42]

[42] Note that the standard errors in Tables 2.2 and 2.3 are not clustered at the school-level, because inference based on clustered standard errors is less conservative than without clustered standard errors, despite the applied corrections for small cluster-numbers. See Tables A2.3-A2.8 in Appendix 2.3.

Table 2.2. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit)

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.013	Ref.	0.002	0.001
	(0.014)		(0.015)	(0.014)
6 (sufficient)	0.030		0.004	0.001
	(0.032)		(0.033)	(0.032)
7 (more than sufficient)	-0.004		-0.001	-0.000
	(0.004)		(0.004)	(0.004)
8 (good)	-0.039		-0.005	-0.002
	(0.042)		(0.043)	(0.042)
Observations	2,251		2,251	2,251
Focusing on something				
5 (insufficient)	-0.002	Ref.	0.010	0.008
	(0.009)		(0.009)	(0.009)
6 (sufficient)	-0.005		0.025	0.021
	(0.022)		(0.024)	(0.023)
7 (more than sufficient)	-0.001		0.006	0.005
	(0.005)		(0.006)	(0.005)
8 (good)	0.009		-0.040	-0.034
	(0.036)		(0.039)	(0.038)
Observations	3,740		3,740	3,740
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.008	Ref.	0.004	0.008
	(0.005)		(0.005)	(0.005)
6 (sufficient)	0.059		0.030	0.061
	(0.038)		(0.039)	(0.038)
7 (more than sufficient)	-0.007		-0.003	-0.007
	(0.005)		(0.005)	(0.005)
8 (good)	-0.060		-0.030	-0.062
	(0.038)		(0.040)	(0.039)
Observations	2,253		2,253	2,253

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Standard errors are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table 2.3. Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity-impulsivity (marginal effects after ordered probit)

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	0.000	Ref.	-0.001	-
	(0.007)		(0.007)	
6 (sufficient)	0.002		-0.006	-
	(0.030)		(0.031)	
7 (more than sufficient)	0.001		-0.003	-
	(0.013)		(0.014)	
8 (good)	-0.003		0.009	-
	(0.051)		(0.053)	
Observations	1,670		1,670	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	0.004	Ref.	0.017	0.017
	(0.012)		(0.012)	(0.012)
6 (sufficient)	0.010		0.048	0.047
	(0.033)		(0.034)	(0.033)
7 (more than sufficient)	-0.000		-0.001	-0.001
	(0.001)		(0.002)	(0.002)
8 (good)	-0.014		-0.064	-0.063
	(0.044)		(0.046)	(0.045)
Observations	2,255		2,255	2,255
Being studious and not being a pest				
5 (insufficient)	0.014	Ref.	-0.001	-
	(0.014)		(0.014)	
6 (sufficient)	0.033		-0.003	-
	(0.031)		(0.033)	
7 (more than sufficient)	-0.008		0.001	-
	(0.007)		(0.007)	
8 (good)	-0.039		0.003	-
	(0.038)		(0.039)	
Observations	1,671		1,671	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Standard errors are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Results by gender

Splitting the sample by gender reveals important differences compared to the main results. Table 2.1 showed that, pre-treatment, boys generally show stronger symptoms of inattentiveness and hyperactivity-impulsivity than do girls. The results in Table 2.4 reveal that, due to the Active Living Program, boys become even more likely to give themselves a 5 (+2.2 percentage points) or a 6 (+16 percentage points) rather than a 7 (-6.1 percentage points) or an 8 (-12.1 percentage points) for 'doing exactly what the teacher asks me to do' (i.e., increased inattentiveness among boys). This effect becomes stronger and significant at the 5% level two years post-treatment.[43] Again, the statistically insignificant effects in column 1 indicate that these effects can be interpreted as the causal effects of the Active Living Program. We find no statistically significant treatment effects on indicators of inattentiveness for girls (see Tables A2.15-A2.18 in Appendix 2.7).

Table 2.4. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit), boys only

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.016	Ref.	0.044	0.017
	(0.047)		(0.031)	(0.036)
6 (sufficient)	0.020		0.053	0.020
	(0.056)		(0.036)	(0.042)
7 (more than sufficient)	-0.014		-0.039	-0.015
	(0.041)		(0.028)	(0.031)
8 (good)	-0.022		-0.058	-0.022
	(0.061)		(0.039)	(0.047)
Observations	1,061		1,061	1,061
Focusing on something				
5 (insufficient)	-0.003	Ref.	0.010	0.015
	(0.014)		(0.013)	(0.011)
6 (sufficient)	-0.008		0.028	0.041
	(0.039)		(0.037)	(0.031)
7 (more than sufficient)	-0.001		0.004	0.006
	(0.006)		(0.005)	(0.004)
8 (good)	0.012		-0.042	-0.061
	(0.059)		(0.056)	(0.045)
Observations	1,801		1,801	1,801
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.016	Ref.	0.015*	0.022**
	(0.013)		(0.009)	(0.008)
6 (sufficient)	0.118		0.112	0.160**
	(0.093)		(0.065)	(0.060)
7 (more than sufficient)	-0.045		-0.043*	-0.061**
	(0.037)		(0.024)	(0.022)
8 (good)	-0.089		-0.085	-0.121**
	(0.070)		(0.050)	(0.048)
Observations	1,063		1,063	1,063

Notes: The table shows the marginal effects after three ordered probit regressions when selecting only boys. All regressions include a treatment dummy, year fixed effects, and controls for parental education and age. Robust standard errors, clustered at the school-level and corrected with the Donald&Lang-correction, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

[43] Table 2.4 shows the results when the standard errors are clustered at the school-level and corrected with the Donald&Lang-correction, because this leads to the most conservative results (see Tables A2.9-A2.11 in Appendix 2.5).

Table 2.5. Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity and impulsivity (marginal effects after ordered probit), boys only

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	-0.004	Ref.	0.007	-
	(0.018)		(0.014)	
6 (sufficient)	-0.016		0.024	-
	(0.064)		(0.050)	
7 (more than sufficient)	-0.003		0.005	-
	(0.013)		(0.010)	
8 (good)	0.023		-0.036	-
	(0.096)		(0.074)	
Observations	785		785	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	0.027	Ref.	0.043**	0.047*
	(0.028)		(0.019)	(0.025)
6 (sufficient)	0.061		0.096*	0.105*
	(0.065)		(0.046)	(0.055)
7 (more than sufficient)	-0.014		-0.022*	-0.024*
	(0.016)		(0.012)	(0.013)
8 (good)	-0.075		-0.117**	-0.128*
	(0.077)		(0.054)	(0.067)
Observations	1,066		1,066	1,066
Being studious and not being a pest				
5 (insufficient)	0.016	Ref.	0.016	-
	(0.029)		(0.022)	
6 (sufficient)	0.037		0.036	-
	(0.066)		(0.052)	
7 (more than sufficient)	-0.017		-0.017	-
	(0.032)		(0.025)	
8 (good)	-0.035		-0.035	-
	(0.063)		(0.049)	
Observations	785		785	-

Notes: The table shows the marginal effects after three ordered probit regressions when selecting only boys. All regressions include a treatment dummy, year fixed effects, and controls for parental education, and age. Robust standard errors, clustered at the school-level and corrected with the Donald&Lang-correction, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Due to the Active Living Program, boys also score significantly lower on 'being calm and quiet when the teacher wants me to' (i.e., increased hyperactivity-impulsivity: see Table 2.5). They become around 4 percentage points more likely to give themselves a 5, and 10 percentage points more likely to give themselves a 6 on this indicator, rather than a 7 (-2 percentage points) or an 8 (-12 percentage points). This effect is stable across the two post-treatment years, although it depends on the specification of the standard errors whether this effect is statistically significant at the 5% level in the first treatment year 2013/2014, or in the second treatment year 2014/2015.[44] Again, we find no statistically significant treatment effects for girls (see Tables A2.19-A2.22 in Appendix 2.7).

2.5.2. Robustness checks

In this section, we verify whether the statistically significant effects of the Active Living Program on boys' inattentiveness (i.e., a decreased ability to do exactly what the teacher asks them to do) and hyperactivity-impulsivity (i.e., a decreased ability to be calm and quiet when the teacher wants them to) are robust for different specifications.

Heterogeneous treatment effects across treatment schools

As Table 1.1 in Chapter 1 indicated, there is considerable heterogeneity in the implemented interventions across treatment schools. Having such heterogeneity in implementation increases the possibility that one outlier treatment school is driving the results. We verify whether this is the case, by running the difference-in-differences analyses on boys' ability to do exactly what the teacher asks them to do and to be calm and quiet when the teacher wants them to, while excluding the treatment schools one-by-one.[45]

The results in Table A2.23 in Appendix 2.8 show that the effects of the Active Living Program on 'doing exactly what the teacher asks me to do' are robust for one-by-one exclusion of treatment schools: the results are similar to those in Table 2.4. Moreover, there is not one school that, when excluded, leads to the finding that boys' self-assessed ability to do exactly what the teacher asks them to do *increases*. This implies that there is not one particular treatment school that is driving the inattentiveness-increasing effect of the Active Living Program for boys. Nevertheless, it appears that the treatment effect on boys' inattentiveness may be weaker in treatment school number 7: excluding this school from the analyses leads to slightly stronger negative point estimates, particularly among the first treated cohort (i.e., sixth graders in 2013/2014).

Table A2.24 in Appendix 2.8 shows the results on boys' ability to be calm and quiet when the teacher wants them to, while excluding treatment schools one-by-one. Although one-by-one exclusion of treatment schools sometimes leads to statistically insignificant results (which may be related to relatively small sample sizes), the point estimates remain similar to those in Table 2.5. Moreover, there is again not one treatment school that, when excluded, leads to the finding that boys' self-assessed ability to be calm and quiet when the teacher wants them to *increases*, leading us to conclude that there is not one particular treatment school that drives the hyperactivity-impulsivity-inducing effect of the Active Living Program for boys. Nevertheless, treatment school

[44] Table 2.5 shows the results when the standard errors are clustered at the school-level and corrected with the DL-correction, which leads to the most conservative results (see Tables A2.12-A2.14 in Appendix 2.6).
[45] An alternative would be to *include* treatment schools one by one (while selecting on boys). Unfortunately, however, those analyses are not based on enough observations to lead to reliable results.

7 again leads to slightly stronger negative results when excluded, implying that the treatment effect on boys' hyperactivity/impulsivity is weaker in this school.

Table A2.25 in Appendix 2.8 shows the interventions that were implemented by treatment school 7 (see also Table 1.1 in Chapter 1). It is possible that this school's unique configuration of design choices prevented the effects of stimulating PA in everyday life on boys' ADHD-like symptoms to become too negative.

Unforeseen treatment effects in control schools

Next, we analyze whether the negative treatment effects we find for boys are (partially) due to an improvement in boys' inattentiveness and hyperactivity-impulsivity in control schools, rather than a decline among boys in treatment schools. Difference-in-differences analyses in which we compare control schools to schools in the Limburg region of the Netherlands that did *not* participate to the Active Living Program as a treatment or control school (i.e., "non-Active Living schools"), indeed leads to positive effects among boys in control schools. However, these effects are very small and at best only significant at the 10% level (see columns 1-3 in Table A2.26 in Appendix 2.8).

Running the difference-in-differences analyses in which we compare the treatment schools to Non-Active Living schools (see columns 4-6 in Table A2.26 in Appendix 2.8), suggests that the very small positive treatment effects in the control schools may indeed have led to a small upward bias of the estimated effects on boys' inattentiveness and hyperactivity-impulsivity, as presented in Tables 2.4 and 2.5. The more conservative estimate of the effect of the Active Living Program on boys' ability to do exactly what the teacher asks them to do, is that it increases their probability to score a 5 or a 6 by 10 percentage points (rather than 18 percentage points, as reported in Table 2.4). The more conservative estimate of the effect of the Active Living Program on boys' ability to be calm and quiet when the teacher wants them to is that it increases their probability to score a 5 or a 6 by 11 percentage points (rather than 14 percentage points, as reported in Table 2.5).

Estimating the treatment effects linearly

Although ordered probit regressions are appropriate for ordinal data, for which it is uncertain whether the distance between outcome values is uniform, there may be some doubt as to whether our dependent variables are truly ordinal in nature. One might argue that, since children are asked to give themselves a report card score, the dependent variables are measured on a ratio scale, even if children can only fill in a round number ranging from 5 to 8. Therefore, as a robustness check, we estimate the effects of the Active Living Program on boys' ability to do exactly what the teacher asks them to do, and to be calm and quiet when the teacher wants them to, linearly as ordinary least squares regressions. Since the results in Table A2.26 in Appendix 2.8 show that the size of the estimates depends on whether the treatment group is compared to the control group or to Non-Active Living schools, we run this robustness check with both comparison groups separately.

Table 2.6. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness and hyperactivity/impulsivity (ordinary least squares analyses), boys only

	(1)	(3)	(2)	(4)
How good are you at:	Doing exactly what the teacher asks me to do		Being calm mand quiet when the teacher wants me to	
	Treatment vs. Control	Treatment vs. Non-Active Living	Treatment vs. Control	Treatment vs. Non-Active Living
Treat * Pre2	-0.241	-0.060	-0.188	-0.004
	(0.190)	(0.162)	(0.196)	(0.157)
Treat * Pre1	Ref.			
Treat * Post1	-0.228*	-0.119	-0.300**	-0.175
	(0.131)	(0.091)	(0.142)	(0.120)
Treat * Post2	-0.330**	-0.197**	-0.325*	-0.242**
	(0.123)	(0.086)	(0.171)	(0.115)
Post2 (2014/2015 = 1)	0.109	-0.021	0.062	-0.016
	(0.093)	(0.031)	(0.125)	(0.034)
Post1 (2013/2014 = 1)	0.114	0.001	0.150**	0.029
	(0.093)	(0.029)	(0.068)	(0.030)
Pre2 (2011/2012 = 1)	0.156	-0.028	0.172	-0.010
	(0.103)	(0.030)	(0.119)	(0.033)
Treatment (1=Treatment group)	0.113	0.012	0.097	0.011
	(0.100)	(0.085)	(0.106)	(0.085)
Parental education				
Lower secondary or below (≤ ISCED 2)	Ref.			
Upper secondary / lower vocational (ISCED 3/4)	0.016	0.003	0.024	0.037
	(0.085)	(0.039)	(0.075)	(0.042)
Higher vocational / university (≥ ISCED 5)	0.069	0.033	0.130	0.118***
	(0.087)	(0.036)	(0.092)	(0.041)
Age (in months)	-0.106	0.053*	-0.028	-0.019
	(0.136)	(0.027)	(0.154)	(0.040)
Age²	0.000	-0.000*	0.000	0.000
	(0.000)	(0.000)	(0.001)	(0.000)
Constant	14.909	2.907	9.232	8.584***
	(10.266)	(2.029)	(11.587)	(2.952)
Observations	1,063	8,064	1,066	8,064
R-squared	0.015	0.004	0.020	0.009
Mean in estimation sample	6.802	6.832	6.804	6.832
[standard deviation]	[0.723]	[0.749]	[0.722]	[0.748]

Notes: The table shows the results of four ordinary least squares analyses when selecting only boys. Robust standard errors, clustered at the school-level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

The results in Table 2.6 show that the conclusions remain robust for this specification.[46] First, the parallel trends assumption holds for both variables, irrespective of whether we compare the treatment group to the control group or to Non-Active Living schools (indicated by the non-significant coefficients on the first row). Second, we still find that the Active Living Program significantly decreases boys' ability to do exactly what the teacher asks them to. The size of the

[46] Results based on clustered standard errors, corrected with the DL-correction, leads to the most conservative estimates.

effect is larger when the comparison group consists of control schools rather than Non-Active Living schools. It ranges from -.12 points (-.16 SD; not statistically significant) to -.23 points (-.32 SD; statistically significant at the 10% level) in the first treatment year, and from -.20 points (-.26 SD) to -.33 points (-.46 SD) in the second treatment year (statistically significant at the 5% level). Third, we still find that the Active Living Program significantly decreases boys' ability to be calm and quiet when the teacher wants them to. The size of this effect ranges from -.18 points (-.23 SD; not statistically significant) to -.30 points (-.42 SD; statistically significant at the 5% level) in the first treated cohort, and from -.24 points (-.32 SD; statistically significant at the 5% level) to -.33 points (-.45 SD; statistically significant at the 10% level) in the second treated cohort.

2.6. CONCLUSIONS

This chapter investigates whether encouraging children to become more physically active in their everyday life affects the prevalence of ADHD-like symptoms among 12-year-old boys and girls. Our estimations are based on a quasi field experiment called the Active Living Program, which successfully increases time spent on physical activity by 8- to 12-year-olds living in low-SES areas in the Netherlands. The interventions focus on active transportation and active play at school and during leisure time. Results from difference-in-differences analyses across cohorts of sixth graders reveal that the Active Living Program significantly increases the prevalence of ADHD-like symptoms among boys. This result is robust across a variety of specifications, leading us to conclude that the effects we find are indeed due to the Active Living interventions. We do not find any significant treatment effects on ADHD-like symptoms among girls. Since boys are more likely to show ADHD-like symptoms and be diagnosed with ADHD, the results of this chapter point in the direction of an ADHD-like symptom-inducing effect of the Active Living Program, and an increase of gender differences in ADHD-like symptoms.

Our results shed new light on the effects of the encouragement of physical activity among children and adolescents. To our knowledge, this study is the first to analyze the causal effect of encouraging informal physical activity on the prevalence of ADHD-like symptoms among representative groups of boys and girls. Previous studies have shown that formal physical activity (e.g., doing cardio or aerobics exercises) can have ADHD-symptom-reducing effects for symptomatic children. However, based on our findings, we conclude that these results may not be generalizable to informal physical activity such as active transportation or active play, or that – at the group level – the positive effects for symptomatic children are cancelled out by negative effects among (previously) asymptomatic children.

There may be several potential explanations for the effects we find. Since we found in Chapter 1 that the Active Living Program is slightly more effective in increasing boys' time spent on physical activity, a first potential explanation is that increasing time spent on informal physical activity may reduce participation in other activities that are better for reducing ADHD-like symptoms. For instance, biking to and from school may reduce doing sports in the afternoon (assuming that doing sports is more effective in reducing ADHD-like symptoms in boys than informal physical activity). Second, encouraging children to spend more time on physical activity may have different effects on symptomatic and asymptomatic children. It is possible that, even though physical activity has a symptom-reducing effect for children with ADHD-like symptoms, other students might be distracted by such an intervention. As a result, the previously asymptomatic students may now become more restless in class, leading to an average increase in ADHD-like symptoms among pupils. These potential spill-over effects have received little to no attention in the current literature on the relationship between physical activity and ADHD. Third, physical activity may be fun and distracting for all children, and lower-intensity or informal physical activity (e.g., active transportation and active play) may not be as effective in transforming the distraction into a ADHD-symptom-reducing effect as higher-intensity or formal physical activity. More research is needed to identify the mechanisms through which different types of physical activity can affect ADHD-like symptoms in boys and girls.

The results of this chapter have important implications for policymakers, researchers, healthcare professionals, educators, and parents alike, who aim to increase children’s time spent on physical activity and/or reduce their ADHD-like symptoms. Although physical activity is an attractive potential treatment of ADHD-like symptoms since it is less invasive than medication, this study shows that the effectiveness of using physical activity to reduce ADHD-like symptoms depends greatly on the type of interventions that are implemented. Certain physical activity interventions may accidentally cause ADHD-like symptoms to increase for certain children. This study also highlights the importance of not only studying direct effects of policies and interventions (in our case, the effects of encouraging physical activity on time spent on physical activity), but to also investigate potential side- or spill-over effects, even if previous literature has shown desirable relationships with those variables. In order to make informed decisions about whether or not to implement certain policies or interventions, we need as much information as possible about other aspects they may affect.

Part II

Direct and Indirect Effects
of Opt-Out Consent for Organ Donation

Chapter 3. Deceased by Default

Consent Systems and Organ-Patient Mortality

3.1. INTRODUCTION[47]

Organ failure is a widely occurring medical condition with lethal consequences if left untreated. The most important treatment for end-stage organ failure is the transplantation of a donor organ. However, due to a shortage of donor organs, many patients die while waiting for an organ transplant (see, e.g., Howard, 2007; Kessler & Roth, 2014). Decreasing the mortality rate among organ patients is one of the primary goals of the organ donation policy agenda.

Almost every country has a system to formalize the default consent for deceased-donor organ donation. Systems in which the default is to become a donor after death are called "opt-out" or "presumed consent" systems. Systems in which the default is not to become a donor after death are "opt-in" or "informed consent" systems. Earlier research has shown that the number of transplantations from deceased donors is substantially higher in countries with opt-out systems than those with opt-in systems (Abadie & Gay, 2006; Johnson & Goldstein, 2003; Johnson & Goldstein, 2004; Rithalia et al., 2009; Shepherd, O'Carroll & Ferguson, 2014; Ugur, 2015). A positive correlation between opt-out systems and organ donation rates has also been found using hypothetical questions in laboratory experiments (Dalen & Henkens, 2014; Davidai, Gilovich & Ross, 2012; Li, Hawley & Schnier, 2013) and using survey questions on the willingness to donate an organ (Mossialos, Costa-Font & Rudisill, 2008). These findings seem to imply that opt-out systems reduce mortality among organ patients (see, e.g., Johnson & Goldstein, 2003). However, the relationship between consent systems and organ-patient mortality has not yet been explicitly investigated yet. This chapter aims to fill that gap in the literature.

The relationship between consent systems for organ donation and organ-patient mortality rates is not as obvious as it may seem, because a substantial difference in deceased-donor transplantation rates between consent systems also changes incentives. For example, a larger supply of organs from deceased donors reduces individuals' incentives to become a living donor (see, e.g., Arshad, Anderson & Sharif, 2019; Fernandez, Howard & Stohr Kroese, 2013; Shepherd, O'Carroll & Ferguson, 2014), or to purchase an organ or to obtain it abroad (i.e., organ trafficking, organ tourism, or legal organ trade within a transplant network). Consent systems may also change the incentives for medical professionals to enter, keep, or remove certain organ patients on/from the waiting list, and change governments' and researchers' incentives to invest in research and technology to increase organ-patient longevity. These incentives may affect behavior, and, as a result, lead to differences in organ-patient mortality between consent systems that are smaller (or larger) to what one would expect based on the differences in deceased-donor transplantation rates alone.

This chapter is the first to analyze whether organ-patient mortality rates are indeed lower in countries with opt-out consent systems for organ donation compared to those with opt-in systems.

---

[47] This chapter was published in *PLoS ONE* (Golsteyn & Verhagen, 2021). It has benefited from valuable comments by Eric Bonsang, Lex Borghans, Thomas Dohmen, Maarten Lindeboom, Arjan Non, Anders Stenberg, and Ulf Zölitz, as well as participants at the 2018 ESPE conference and the OECD Applied Economics Work in Progress Seminar. Special thanks go out to André Hildmann and Tifanny Istamto, who were involved in an early stage of this project.

We study between-country differences, using data from newsletters from the Organización Nacional de Trasplantes on the number of kidney, liver, heart, and lung transplantations performed and the number of patients who died while registered on a waiting list to receive a donation of one of those organs. We gathered data from several other sources to determine whether a country has an opt-out or opt-in system. Combining these data, we create the largest data set available on organ-patient mortality and organ transplantations; our sample consists of 69 countries across a period of 15 years (2001–2015, unbalanced panel). Additionally, we obtained data on several potentially confounding factors from the World Bank, the World Health Organization, and the World Values Survey.

There is a considerable difference between a patient with kidney disease (who may remain on dialysis for many years or may receive a live donor kidney), a patient with heart disease (who can receive a left ventricular assist device or "artificial heart" as a bridge to the transplant), a patient with lung disease (who may choose a ventilator), and a patient with liver disease (for whom live split-liver donation is possible, but artificial devices are not available). We therefore show separate analyses for each organ and do not pool the organs. We estimate two regressions for kidneys, livers, hearts, and lungs separately, while controlling for year fixed-effects as well as for a country's GDP, general mortality rates, health expenditures, and religious denomination. The first regression shows the relationship between countries' consent systems for organ donation and the number of deceased-donor transplantations per million population (hereafter, "deceased-donor transplantation rates"). The second regression shows the relationship between consent systems and patients who died while on a waiting list for a donor organ per million population (hereafter, "organ-patient mortality rates"). By estimating the two equations as seemingly unrelated regressions, we can analyze whether the correlation between opt-out systems and deceased-donor transplantation rates is significantly different from the correlation between opt-out systems and organ-patient mortality rates.

For hearts, we find that opt-out systems have a small but statistically significant advantage in terms of deceased-donor transplantation rates, and that this translates into similarly lower heart-patient mortality rates compared to opt-in systems. For lungs, however, deceased-donor transplantation rates do not differ significantly between consent systems; this is reflected in the mortality rates, which are also not significantly different from each other between consent systems. Although the advantages of opt-out systems are by far the largest for deceased-donor kidney and liver transplantation rates, these are also the organs for which the differences in organ-patient mortality rates are significantly smaller (i.e., closer to zero) than the differences in deceased-donor transplantation rates between consent systems.

We investigate which incentives could explain the surprisingly small differences between consent systems regarding kidney- and liver-patient mortality rates. For kidneys, we find empirical evidence that lower incentives to become a living donor in opt-out systems, reflected by lower rates of living donations, can explain our findings: the difference between consent systems in kidney-patient mortality rates is similar to what could be expected from the total kidney transplantation rates (i.e., from deceased and living kidney donors combined). For livers, differences in organ-patient mortality rates between consent systems remain significantly smaller than the differences in transplantation rates, even when transplantations from living donors are considered. We do not find empirical evidence that opt-out systems are related to different incentives for kidney and liver patients to purchase an organ on the black market (i.e., organ trafficking or organ tourism): excluding countries with reports of organ trade as either a donor or recipient country does not change our main results.

We investigate the robustness of our results in various sensitivity checks. One potential measurement issue is related to the earlier stated hypothesis that incentives to enter, keep, or remove patients from the waiting list may differ between consent systems. More lenient policies regarding waiting list entry not only affect the probability of patients dying from other factors than organ failure (e.g., co-morbid conditions); they also increase the number of patients registered on the waiting lists and thereby, by construction, the number of deaths on the waiting lists. However, the empirical evidence does not support these hypotheses: taking the conservative approach of assuming that if an organ patient was removed from the waiting list because (s)he was too sick to receive a transplant implies that the organ patient died, does not change our results. Our results also remain unchanged when we control for waiting list length, indicating that countries with different consent systems do not have substantially different policies regarding waiting list entry. These two robust results indicate that, although fatal co-morbid conditions affect the number of deaths on the waiting lists, it is unlikely that they substantially affect the difference in organ-patient mortality rates between consent systems. The results also remain robust when we reduce country heterogeneity by only including Europe, North America, Australia, and New Zealand, when we use different methods to deal with zero values (i.e., excluding zero values and estimating Tobit regressions), and when we deal with potential outliers by truncating the data at the 99^{th} percentile.

The findings provide an important contribution to the literature on organ donation and thereby provide new insights for the discussion on the advantages and disadvantages of the two consent systems for organ donation. First, they highlight that the advantages and disadvantages differ substantially by organ. This is an important addition to the current literature on the topic, which often focuses on one single organ (typically either kidneys or livers). Second, we find empirical evidence that opt-out systems decrease the incentives to become a living kidney donor. Although we are not the first to find this result (see, e.g., Arshad, Anderson & Sharif, 2019; Fernandez, Howard & Stohr Kroese, 2013; Shepherd, O'Carroll & Ferguson, 2014), the fact that living kidney transplantation rates are on average lower in opt-out systems remains a (dis)advantage of opt-out systems that receives little to no attention in the debate on the pros and cons of different consent systems. Further research is needed to better understand the channels through which consent systems can affect organ-patient mortality rates, channels such as technological investments regarding organ-patient longevity and organ trade through transplant networks, as well as to what extent and why these channels differ by organ.

This chapter proceeds as follows: Section 3.2 discusses the empirical strategy, Section 3.3 discusses the data, and Section 3.4 gives the results. Section 3.5 concludes.

3.2. EMPIRICAL STRATEGY

3.2.1 The relationship between consent systems and organ-patient mortality rates

The aim of this chapter is to analyze whether organ-patient mortality rates differ between countries with opt-out or opt-in consent systems for organ donation. In our main specifications, we estimate the following model for kidneys, livers, hearts, and lungs separately:

$$mortality_{ct} = \alpha_{30} + \alpha_{31} optout_c + \delta_{31t} + \boldsymbol{X'}_{ct}\gamma + \varepsilon_{31ct} \qquad (3.1)$$

In this equation, *c* indexes the country, and *t* indexes time in years. For simplicity, organ-specific subscripts are suppressed from the equation. *Mortality* is the organ-patient mortality rate, i.e., the number of people who died throughout year *t* while on a waiting list to receive either a donor kidney, liver, heart, or lung, divided by the size of the population (in millions) of that person's country of residence. The variable *optout* is an indicator variable with a value of 1 if a country has an opt-out consent system for organ donation and 0 if that country has an opt-in system. In our main specification, we exclude countries that changed their consent system within our data window (see more about this in Section 3.4). Therefore, the opt-out dummy variable does not vary over time. We also control for year fixed effects δ as well as for ***X'***; a vector of country characteristics, i.e., the country's GDP, health expenditures, age-standardized death rates (all causes), and religious denomination (see Section 3.3 below and Appendix 3.1 for more information). All regressions also include dummy variables for missing values on the independent variables. ε is the normally distributed error term, which is allowed to correlate with the error-term of equation (3.2) (more on this in Section 3.2.3 below). Because organ-patient mortality rates are likely to be correlated over time within countries, we cluster the standard errors at the country level.

Please note that for simplicity, we sometimes use the term "effect" to describe the correlation between consent systems and our dependent variables. However, this chapter does not identify causal effects of consent systems.

3.2.2 The relationship between consent systems and deceased-donor transplantation rates

Besides the analyses on organ-patient mortality, we also analyze the relationship between consent systems and deceased-donor transplantation rates. If deceased-donor transplantation rates are the only mechanism through which consent systems affect organ-patient mortality rates, the relationship between consent systems and organ-patient mortality rates should be similar (although with the opposite sign) to the relationship between consent systems and deceased-donor transplantation rates. We estimate the following model for each organ separately (again, organ-specific subscripts are suppressed for simplicity):

$$DDTX_{ct} = \beta_{30} + \beta_{31} optout_c + \delta_{32t} + \boldsymbol{X'}_{ct}\mu + \varepsilon_{32ct} \qquad (3.2)$$

In this equation, $DDTX_{ct}$ stands for the deceased-donor transplantation rate in country *c* in year *t*, i.e., the number of transplantations from deceased donors divided by the size of the population (in millions). Based on previous findings in the literature, we expect $\hat{\beta}_{31}$ to be larger than zero, i.e., that deceased-donor transplantation rates are higher in countries with opt-out systems. To reduce unobserved heterogeneity and improve the efficiency of estimator $\hat{\beta}_{31}$, we again include year fixed effects δ as well as $\boldsymbol{X'}$, which is a vector of variables that are identical to those in equation (3.1). Again, we cluster the standard errors at the country level.

3.2.3 Seemingly unrelated regressions

If the difference between consent systems regarding organ-patient mortality rates is significantly different from the difference between the systems regarding deceased-donor transplantation rates (i.e., if $\hat{\alpha}_{31}+\hat{\beta}_{31}\neq 0$), consent systems must affect organ-patient mortality rates through channels other than deceased-donor transplantation rates alone. However, even if there would be no one-to-minus-one relationship between the opt-out effect on deceased-donor transplantation rates and the opt-out effect on organ-patient mortality rates, it is unrealistic to assume that these two effects are completely unrelated to each other. This means that it is unrealistic to assume that the covariance between the error terms of regressions (3.1) and (3.2) is equal to zero. By estimating both regressions simultaneously using a seemingly unrelated regression (SUR) technique, the covariance between both regressions can be taken into account when testing whether $\hat{\alpha}_{31}+\hat{\beta}_{31}=0$. Estimating equations (3.1) and (3.2) as SUR rather than as separate regressions does not change the regression results in any way.

3.2.4 Using countries that changed consent systems

Although it is tempting to try to exploit the fact that some countries in our data changed their consent system for organ donation within our data window, this would be inappropriate for several reasons. First, a country's decision to change its consent system is likely to be endogenous, i.e., related to previous trends in organ-patient mortality rates, the population's attitudes toward organ donation, or technological advances in increasing organ-patient longevity, for instance. Therefore, estimates based on within-country analyses do not necessarily reveal causal effects of a consent system change but rather pick up the continuation of a previously existing trend.

Second, even if the decision to change the consent system was exogenous, implementing this change may have lagged effects on transplantation rates and therefore on organ-patient mortality rates: it may take several years before countries have sufficiently increased their health care capacity (e.g., medical equipment, hospital beds, and medical personnel) to use the entire additional supply of organ donors. Therefore, merely adjusting the opt-out dummy from the year of the consent system change onward, thereby ignoring any lagged effects, would likely lead to an estimation of α_{31} and β_{31} that is biased toward zero. However, trying to take into account the lagged effects poses problems as well because it is uncertain how big the lag would be and because the size of the lag may vary from one country/consent system change to another. Moreover, even if the size of the lag would be known, the fact that there is a lag makes it substantially less likely that nothing else changed that may affect transplantation or organ-patient mortality rates since the consent system changed, again leading to biased estimates of α_{31} and β_{31}.

Finally, only a few countries in our sample switched from one consent system to another within our data window. Therefore, the statistical power to study within-country differences is limited, which may lead to biased estimates of α_{31} and β_{31}.

3.3. DATA

3.3.1 Data sources[48]

We obtained the data from a great variety of sources, leading to the largest data set available to study the relationships between consent systems and organ-patient mortality rates and transplantation rates. In this section, we briefly introduce the data. In Appendix 3.1 and Appendix 3.2, we give the sources of the variables and discuss data selection.

Our main data source is newsletters from the Organización Nacional de Trasplantes. From Newsletter Transplant 2002 (reporting data on calendar year 2001) onward, the newsletters include data on the number of kidney, liver, heart, and lung transplantations performed, as well as the number of patients who died while on a waiting list to receive a donor kidney, liver, heart, or lung. For each organ, we divide both variables by the size of the population (in millions) to obtain the rates of organ-patient mortality and transplantations.

Between 2001 and 2015, the number of countries included in our sample increased from 35 to 66. In Table A3.2 in Appendix 3.1, we show which countries are added to the newsletters by year. In additional analyses, we also use data from the Newsletters Transplant on the annual number of kidney and liver transplantations from living donors, the annual number of patients who were on the waiting list to receive an organ on December 31 in each year, and the annual number of patients who entered the waiting list for the first time in that year.

From other sources, we obtained information on whether countries have opt-in or opt-out consent systems (see Table A3.3 in Appendix 3.1). Every year, around 60% of the countries in the sample had an opt-out consent system for organ donation. We control for countries' annual gross domestic product (GDP) per capita, annual health expenditures per capita, annual general mortality rates (all causes and age standardized), and religion (measured as the average share of the population between 2000 and 2014 that indicated they were Roman Catholic). We obtained these data from the World Bank, the global health expenditures database of the WHO, the WHO mortality database, and the World Values Survey, respectively (see Appendix 3.1 for more information). For one of the robustness checks, we also obtained data from Eurotransplant and

[48] The data on the number of deceased-donor and living transplantations, the number of patients who died while on waiting lists, and the length of the waiting lists can be downloaded without restriction at http://www.ont.es/publicaciones/Paginas/Publicaciones.aspx [retrieved on December 9, 2016]. These data are only available in pdf format. Researchers interested in receiving a copy of the spreadsheet and the programs to replicate the analyses can contact the corresponding author. Data on consent systems and removals from waiting lists are taken from a variety of sources. We have added the list with consent systems across countries in Appendix 3.1, including the sources.

Scandiatransplant on the number of patients who were removed from organ transplantation wait lists.

3.3.2 Descriptive statistics

Most deceased-donor transplantations (per million population: pmp) are kidney transplantations, and most patients who died while on the waiting list (pmp) were waiting for a kidney transplant (see Appendix 3.4). This indicates that the organ shortages problem is driven by a kidney shortage. More than half of all deceased-donor transplants (pmp) are kidney transplants (55%), 26% are liver transplants, 9% are heart transplants, and 10% are lung transplants. Of all patients who died while on a waiting list for organ transplantation (pmp), 53% were on a waiting list for kidney transplants, 25% were on the liver waiting list, 12% were on the heart waiting list, and 10% were on the waiting list for a lung transplant.

In 2015, the average deceased-donor transplantation rate in countries with an opt-out consent system was 22.2 pmp for kidneys (17.6 in opt-in systems), 9.3 (7.0) for liver transplantations, 3.6 (2.9) for heart transplantations, and 2.9 (3.9) for lung transplantations. Although there are minor fluctuations and differences between consent systems, every year the organ-patient mortality rate is around 5 pmp among patients who were on a kidney waiting list, 2 pmp among those who were waiting for a liver transplant, 1 pmp among those who were waiting for a heart transplant, and less than 1 patient pmp among those who were on the waiting list for a lung transplant.

Table 3.1 provides additional descriptive statistics by consent system for the estimation sample of the main results for kidneys (for the full set of results for all organs, see Tables A3.4-A3.7 in Appendix 3.1). It shows that countries with different consent systems on average do not significantly differ from each other on these key variables. This holds for the estimation sample of the main results for each organ, except for hearts. In the estimation sample for the main results for hearts (column 3 of Table 3.2), countries with an opt-out system, on average, have significantly smaller populations and spend significantly less money on health per capita than those with an opt-in system.

Table 3.1. Descriptive statistics of key variables by consent system in the estimation sample for kidneys

	(1) Total	(2) Opt-in	(3) Opt-out	(4) Diff.	(5) [*p*-value]
Population size (millions)	35.12	50.80	23.67	-27.13	[0.119]
	(57.46)	(78.68)	(32.10)		
Western countries (%)	53.33	47.37	57.69	10.32	[0.504]
	(50.45)	(51.30)	(50.38)		
Total deaths all causes (pmp, age-standardized)	54.45	51.48	56.12	4.64	[0.373]
	(14.78)	(10.99)	(16.54)		
GDP per capita (current US$/10,000)	21,434.12	20,983.99	21,763.06	779.07	[0.902]
	(20,552.31)	(17,540.11)	(22,839.46)		
Health expenditures per capita (current US$/1,000)	1,940.21	2,177.59	1,785.00	-392.59	[0.541]
	(2,024.45)	(2,189.79)	(1,937.27)		
Religious denomination (%)					
None	19.93	19.29	20.45	1.16	[0.857]
	(16.78)	(17.54)	(16.69)		
Muslim	9.74	11.29	8.47	-2.82	[0.743]
	(22.45)	(20.14)	(24.74)		
Roman Catholic	32.06	25.78	37.16	11.38	[0.338]
	(31.24)	(24.96)	(35.51)		
Orthodox	15.68	12.49	18.27	5.79	[0.624]
	(30.82)	(29.07)	(32.88)		
Protestant	6.43	6.80	6.13	-0.67	[0.892]
	(12.94)	(9.30)	(15.60)		
Number of countries	45	19	26		

Notes: Based on the estimation sample for the main results for kidneys (Table 3.2, column 1). See Tables A3.4-A3.7 in Appendix 3.1 for the full set of descriptive statistics for the other organs.

3.4. RESULTS

3.4.1 Main results

Table 3.2 provides the main results. We first verify whether our data support the finding of other studies that countries with an opt-out consent system for organ donation have higher deceased-donor transplantation rates (i.e., conduct more transplantations from deceased donors pmp) than countries with opt-in systems. The results in the top panel of Table 3.2 show that this is indeed the case for kidneys, livers, and hearts. For lungs, we do not find a significant difference between opt-out and opt-in systems in deceased-donor transplantation rates. Deceased-donor kidney transplantation rates are on average 7.1 pmp higher in opt-out than in opt-in systems, there are 4.4 pmp more deceased-donor liver transplantations, and 1.0 pmp more deceased-donor hearts transplantations in opt-out relative to opt-in systems.

Table 3.2. The relationship between consent systems and the number of transplantations from deceased donors (pmp) and the number of patients who died while on the waiting list (pmp), by organ

		(1) Kidney	(2) Liver	(3) Heart	(4) Lung
DDTX	α_{31}: Opt-out (1 = yes)	7.082***	4.357***	1.031**	-0.106
		(2.331)	(1.258)	(0.430)	(0.678)
	Total deaths all causes (pmp, age-standardized)	0.076	-0.057	-0.002	-0.038
		(0.129)	(0.057)	(0.024)	(0.029)
	GDP pc (current US$/10,000)	-3.276	-0.194	0.097	0.670
		(2.097)	(0.835)	(0.248)	(0.494)
	Health expenditures pc (current US$/1,000)	6.763***	2.562***	0.865***	0.359
		(1.910)	(0.551)	(0.180)	(0.344)
	Religious denomination (% Roman Catholic)	0.054	0.049*	0.013*	0.003
		(0.039)	(0.027)	(0.007)	(0.010)
	Constant	8.165	3.824	1.413	1.972
		(7.804)	(3.563)	(1.426)	(2.191)
Mortality	β_{31}: Opt-out (1 = yes)	-0.811	0.434	-0.171	-0.331*
		(1.423)	(0.494)	(0.280)	(0.197)
	Total deaths all causes (pmp, age-standardized)	0.010	-0.020	0.002	-0.007
		(0.031)	(0.018)	(0.011)	(0.007)
	GDP pc (current US$/10,000)	-1.815***	-0.979***	-0.036	0.252***
		(0.673)	(0.294)	(0.132)	(0.094)
	Health expenditures pc (current US$/1,000)	1.980**	0.981***	0.034	-0.116
		(0.776)	(0.234)	(0.116)	(0.110)
	Religious denomination (% Roman Catholic)	-0.028	-0.008	-0.004	0.000
		(0.028)	(0.013)	(0.006)	(0.004)
	Constant	4.040	3.401**	1.250*	1.088*
		(2.590)	(1.568)	(0.739)	(0.639)
Chi² test: $\alpha_{31}+\beta_{31}=0$ [*p*-value]		[0.047]	[0.003]	[0.192]	[0.564]
Observations		549	514	492	390
Number of countries in the estimation sample		45	42	38	30
Mean DDTX rate in estimation sample		19.849	8.510	3.121	3.134
	Standard deviation	*13.056*	*7.295*	*2.575*	*3.208*
Mean organ-patient mortality rate in estimation sample		4.547	1.966	0.906	0.648
	Standard deviation	*5.859*	*1.699*	*0.780*	*0.660*

Notes: Each column presents the results of a seemingly unrelated OLS regression. The dependent variables in each column are the number of transplantations from deceased donors per million population (DDTX), and the number of organ patients who died while on the waiting list per million population (Mortality). All regressions include year fixed effects and dummy variables for missing values on the independent variables. We select countries that did not change their consent system after 1999, and we only include country-year observations for which both the number of transplantations from deceased donors and the number of deaths while on the waiting lists are non-missing. Robust standard errors clustered at the country level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

The second set of estimates in Table 3.2 shows the relationship between consent systems and organ-patient mortality rates (i.e., the number of patients who died while on a waiting list for organ transplantation pmp). For kidneys, livers and hearts, we find no significant difference regarding the number of patients who died while on the waiting lists (pmp) between consent systems. For lungs, we do find a significant difference: there are 0.3 fewer lung patients (pmp) who died while on the waiting list in opt-out consent systems than in opt-in systems.

However, the more important question is whether the difference between consent systems regarding organ-patient mortality rates is similar to what could be expected based on the difference in deceased-donor transplantation rates, i.e., whether if $\hat{\alpha}_{31}+\hat{\beta}_{31}=0$. The answer to this question is given by the Chi² tests, for which the *p*-values are given in the bottom of the table. The results show that, for kidneys and livers, the difference between opt-out and opt-in systems in terms of patient mortality rates is smaller than the difference in terms of deceased-donor transplantation rates. For hearts and lungs, on the other hand, the differences in terms of patient mortality and deceased-donor transplantation rates are similar, which is what one would expect if there is a one-to-minus-one relationship between deceased-donor transplantations and organ-patient mortality.

3.4.2 Potential mechanisms

The results in Table 3.2 indicate that consent systems do not affect kidney and liver-patient mortality rates through the deceased-donor transplantation rates alone. There must be other factors related to opt-out (opt-in) systems that increase (decrease) the number of deaths pmp on the waiting lists for kidneys and livers. In this section, we discuss in more detail the hypotheses on how differences in incentives between consent systems may explain our findings, and—where possible —we empirically analyze whether these hypotheses are likely to hold.

Transplantations from living donors

One important example of potential differences in incentives between consent systems is that the larger supply of organs from deceased donors in opt-out systems may reduce people's incentives to become a living donor. Considering that transplantations from living donors almost exclusively apply to kidneys and livers because living heart transplantations are not medically possible and living lung transplantations are still extremely rare, this is a particularly likely mechanism for our findings. Moreover, Shepherd, O'Carroll and Ferguson (2014) and Arhad, Anderson and Sharif (2019) already report evidence supporting the hypothesis that consent systems affect people's incentives to become a living donor, as they find that countries with opt-out systems conduct fewer transplants from living donors than those with opt-in systems. Moreover, Fernandez, Howard and Stohr Kroese (2013) show for the United States that an increase in the supply of cadaveric donors causes a decrease in the supply of living donors. This would mean that analyzing deceased-donor transplantation rates alone may overestimate the difference in transplantation rates between consent systems, and that the more accurate measure would be to analyze the *total* transplantation rates (i.e., deceased-donor plus living transplantations pmp). We can indeed analyze this with our data.

Similar to the result in our main analysis, Table 3.3 shows that the total transplantation rates are significantly higher in opt-out relative to opt-in systems. However, the difference between consent systems is smaller for the total kidney transplantation rate than for the deceased-donor kidney transplantation rate alone (i.e., 5.0 pmp compared to 7.1 pmp in Table 3.2). Moreover, now that the estimated opt-out effect on kidney transplantation rates is smaller, the estimated opt-out effect on kidney-patient mortality rates is no longer significantly different from it (p = 0.20 for kidneys). This means that the lower incentive to become a living kidney donor in opt-out systems (reflected by lower living kidney transplantation rates) can explain why kidney-patient mortality rates are not as low in opt-out systems as one would expect based on the deceased-donor kidney transplantation rates alone.

Table 3.3. The relationship between consent systems and the number of transplantations from deceased and living donors (pmp) and the number of patients who died while on the waiting list (pmp), by organ

		(1) Kidney	(2) Liver
Total TX	α_{32}: Opt-out (1 = yes)	5.046**	5.211***
		(2.369)	(1.371)
	Total deaths all causes (pmp, age-standardized)	0.005	-0.111*
		(0.135)	(0.064)
	GDP pc (current US$/10,000)	1.596	-0.740
		(1.846)	(0.957)
	Health expenditures pc (current US$/1,000)	3.785***	2.762***
		(1.445)	(0.638)
	Religious denomination (% Roman Catholic)	-0.044	0.026
		(0.053)	(0.029)
	Constant	20.959***	8.516**
		(7.931)	(4.070)
Mortality	β_{32}: Opt-out (1 = yes)	-0.784	0.485
		(1.405)	(0.495)
	Total deaths all causes (pmp, age-standardized)	-0.001	-0.026
		(0.033)	(0.018)
	GDP pc (current US$/10,000)	-2.276**	-1.056***
		(0.899)	(0.288)
	Health expenditures pc (current US$/1,000)	2.309***	1.010***
		(0.840)	(0.239)
	Religious denomination (% Roman Catholic)	-0.031	-0.009
		(0.029)	(0.012)
	Constant	5.226*	3.803**
		(3.016)	(1.534)
Chi^2 test: $\alpha_{32}+\beta_{32}=0$ [*p*-value]		[0.200]	[0.001]
Observations		534	485
Number of countries in the estimation sample		44	39
Mean total TX rate in estimation sample		28.270	9.632
	Standard deviation	*15.879*	*1.711*
Mean organ-patient mortality rate in estimation sample		4.618	2.026
	Standard deviation	*5.906*	*7.482*

Notes: Each column presents the results of a seemingly unrelated OLS regression. The dependent variables in each column are the number of transplantations from deceased donors per million population plus the number of transplantations from living donors per million population (Total TX), and the number of organ patients who died while on the waiting list per million population (Mortality). All regressions include year fixed effects and dummy variables for missing values on the independent variables. We select countries that did not change their consent system after 1999, and we only include country-year observations for which both the number of transplantations from deceased donors, the number transplantations from living donors, and the number of waiting list deaths are non-missing. Robust standard errors clustered at the country level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

In contrast to the advantage of receiving a kidney from a living donor relative to receiving one from a deceased donor (Nemati et al., 2012), live liver transplants have no mortality advantage over cadaveric liver transplants (Thuluvath & Yoo, 2004). Therefore, the incentive to become a living liver donor is not the same as the incentive to become a living kidney donor. Our analysis indeed shows that the pattern we find for kidneys does not hold for livers. The point estimates, and the difference between the point estimates for the transplantation rates and the liver-patient mortality rates, remain similar to those in Table 3.2. It is therefore unlikely that liver transplantations from living donors are a mechanism for the surprisingly small differences in liver-patient mortality rates between consent systems.

Participation in organ trade

The smaller deceased-donor and total supply of donor organs in opt-in systems creates incentives for organ-patients in countries with opt-in systems to try to speed up the process by purchasing an organ. However, despite the worldwide shortage of organs, the commercial trade of organs is illegal in almost all countries. When there is a black market for organ transplantations, the true number of transplantations in the donor's country may be higher than what is reported in our data. At the same time, organ patients who consider purchasing their organ may not even enter the waiting list. If the purchase was successful in the sense that the recipient (who did not enter the waiting list) does not die, we overestimate the organ-patient mortality rates in the recipient's country. However, if the (potential) recipient who did not enter the waiting list dies either before or after having purchased the organ, we underestimate the organ-patient mortality rates in the (potential) recipient's country. Both mechanisms may lead to biased estimates of the opt-out effects on transplantation rates and organ-patient mortality rates.

Despite the illegality of commercial organ trade, there have been reports of organ sales in the following countries in our data (i.e., "organ-exporting countries"): Bolivia, Brazil, Colombia, Egypt, Iran, Moldova, Peru, and Turkey (Shimazono, 2007). According to the same source, major organ-importing countries (i.e., the recipients' countries of origin) included in our data are: Australia, Canada, Israel, and the United States. Paid kidney donation is legal in Iran, but there is strict regulation of the allocation of organs to non-local citizens, thereby restricting the international organ trade (Shimazono, 2007). According to Shepherd, O'Carroll and Ferguson (2014), there are also high incidences of organ trafficking in Ukraine.

In Table 3.4, we show the results when estimating the same regressions as in Table 3.2, while excluding this list of organ-exporting and organ-importing countries. Looking at the *p*-values in the bottom of the table, the conclusions are identical to those in Table 3.2: the difference between consent systems with regards to heart- and lung-patient mortality rates is similar to what could be expected based on the difference in deceased-donor transplantation rates, but the difference in kidney- and liver-patient mortality rates is significantly smaller (closer to zero) than the difference in deceased-donor transplantation rates. This implies that, although the incentive to purchase an organ on the black market may be higher in opt-in systems, there is no empirical evidence that this substantially affects deceased-donor transplantation rates or organ-patient mortality rates.

Table 3.4. The relationship between consent systems and the number of transplantations from deceased donors (pmp) and the number of patients who died while on the waiting list (pmp) in countries without reports of frequent organ trafficking, by organ

		(1) Kidney	(2) Liver	(3) Heart	(4) Lung
DDTX	α_{33}: Opt-out (1 = yes)	9.333***	5.628***	1.350**	0.843
		(2.642)	(1.476)	(0.625)	(0.780)
	Total deaths all causes (pmp, age-standardized)	0.290***	-0.069	0.001	0.002
		(0.097)	(0.084)	(0.029)	(0.029)
	GDP pc (current US$/10,000)	-7.453***	1.032	0.528	-0.643
		(1.824)	(1.824)	(0.551)	(0.870)
	Health expenditures pc (current US$/1,000)	11.704***	1.359	0.469	1.884**
		(2.031)	(1.751)	(0.553)	(0.911)
	Religious denomination (% Roman Catholic)	0.030	0.063*	0.018**	0.011
		(0.034)	(0.034)	(0.008)	(0.012)
	Constant	-4.276	1.705	0.446	-1.454
		(6.207)	(5.063)	(2.029)	(2.478)
Mortality	β_{33}: Opt-out (1 = yes)	-1.975	0.766	-0.163	-0.200
		(1.855)	(0.493)	(0.375)	(0.152)
	Total deaths all causes (pmp, age-standardized)	-0.021	0.003	0.011	0.003
		(0.043)	(0.016)	(0.010)	(0.004)
	GDP pc (current US$/10,000)	-0.498	-0.756	-0.133	0.142
		(0.893)	(0.644)	(0.267)	(0.118)
	Health expenditures pc (current US$/1,000)	0.469	0.922	0.211	0.062
		(1.033)	(0.644)	(0.269)	(0.113)
	Religious denomination (% Roman Catholic)	-0.003	0.011	0.006	0.005*
		(0.026)	(0.008)	(0.004)	(0.002)
	Constant	4.587	0.452	0.345	0.328
		(3.335)	(1.301)	(0.755)	(0.426)
Chi² test: $\alpha_{33}+\beta_{33}=0$ [*p*-value]		[0.069]	[0.000]	[0.216]	[0.400]
Observations		423	398	382	292
Number of countries in the estimation sample		34	33	29	23
Mean DDTX rate in estimation sample		21.324	8.768	3.163	3.084
	Standard deviation	*13.352*	*1.451*	*2.626*	*3.375*
Mean organ-patient mortality rate in estimation sample		4.444	1.762	0.893	0.602
	Standard deviation	*5.875*	*7.652*	*0.765*	*0.579*

Notes: Each column presents the results of a seemingly unrelated OLS regression. The dependent variables in each column are the number of transplantations from deceased donors per million population (DDTX) and the number of organ patients who died while on the waiting list per million population (Mortality). All regressions include year fixed effects and dummy variables for missing values on the independent variables. We select countries that did not change their consent system after 1999, and we only include country-year observations for which both the number of transplantations from deceased donors and the number of waiting list deaths are non-missing. Additionally, we exclude countries with reports of organ trafficking as either an organ-importing or organ-exporting country (i.e., Australia, Bolivia, Brazil, Canada, Colombia, Egypt, Iran, Israel, Moldova, Peru, Turkey, Ukraine and the United States). Robust standard errors clustered at the country level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

3.4.3 Robustness checks

This section discusses the results of a variety of robustness checks. As we established in Section 3.4.2 that it is inappropriate to focus on deceased-donor transplantation rates alone when comparing transplantation rates between consent systems, we focus on robustness checks of the results presented in Table 3.3, i.e., of the opt-out effect on the *total* transplantation rates and organ-patient mortality rates.

Waiting list removals

Patients who are too sick to receive an organ may be taken off the waiting list (see, e.g., Charpentier & Mavanur, 2008). If this occurs less (more) often in opt-out systems than in opt-in systems, we will be underestimating (overestimating) the difference in the organ-patient mortality rates between the systems. We found information on the number of patients taken off the waiting list from various transplantation organizations for 16 of the countries in our data set. Although health deterioration is the main reason that patients are removed from a waiting list, patients may also be taken off the list for reasons other than being too sick. Examples of other registered reasons for waiting list removals are: "condition improved," "recovered," "not sick enough to receive a transplant," "refused transplant," and "other/unknown." The data therefore provide an upper bound of the relation between consent systems and organ-patient mortality rates. Table A3.8 in Appendix 3.1 shows a list of the countries for which we have data on waiting list removals.

Table A3.10 in Appendix 3.3 shows that for this selection of countries and with this new measure of organ-patient mortality rates, the difference between consent systems with respect to kidney-patient mortality rates remains not significantly different from the difference in the total kidney transplantation rates. In other words: lower incentives to become a living kidney donor in opt-out systems remain a plausible mechanism for the surprisingly small difference in kidney-patient mortality rates between consent systems when waiting list removals are taken into account. For livers, the results remain robust to those shown in Table 3.3: when waiting list removals are taken into account, the estimated opt-out effect on total liver transplantation rates remains significantly different from the opt-out effect on liver-patient mortality rates.

Length of the waiting list

By construction, there are fewer patients who can die while on the waiting list when waiting lists are shorter. Therefore, if waiting lists are generally longer (shorter) in opt-out systems, our estimated opt-out effects on organ-patient mortality rates will be biased. We measure waiting list length in $year_t$ by taking the number of organ patients registered on the waiting list on December 31 in $year_{t-1}$, and adding to this the number of patients who entered the waiting list throughout the current $year_t$. These data are obtained from the Newsletters Transplant.

On average, the number of patients who receive a kidney transplant from either a living or deceased donor is equal to 29.2% of the number of patients registered on a kidney waiting list (see column 1 in Table A3.11 in Appendix 3.3). In opt-out systems, these "relative kidney transplantation rates" are on average 1.3 percentage points higher than in opt-in systems (not statistically significant). For livers, the relative transplantation rates, as well as the differences between consent systems, are much larger, i.e., 45.6% and 12.2 percentage points, respectively (significant at the 5% level). On average, the number of patients who died while on a waiting list for kidney transplantation is equal to 3.6% of the number of patients who were on the kidney waiting list. For liver patients, the relative mortality rate is 9.0%.

The results in Table A3.11 in Appendix 3.3 indicate that the length of the waiting lists does not vary substantially between consent systems because for both kidneys and livers, the conclusions of Table 3.3 remain robust when taking waiting list length into account.

Co-morbid conditions

Co-morbid conditions may affect the number of deaths on the waiting list, particularly for kidneys, since patients with end-stage renal disease can live for years on dialysis. Indeed, while organ patients are waiting for a donor organ, they might die from something else than organ failure, such as co-morbid conditions or even a traffic accident. That is why we control for countries' general mortality rates in our analyses, thereby taking into account that in some countries, people - including organ patients - may be more likely to develop (fatal) co-morbid conditions than in others, and that this may systematically differ between consent systems.

This method works best if organ patients have a similar probability of dying from other factors than organ failure as their fellow countrymen. If organ patients have an increased probability of developing co-morbid conditions, controlling for general mortality rates might not be sufficient. It is important to keep in mind that this can only change our conclusions if the probability of having (fatal) co-morbid conditions is different between countries with opt-out and opt-in systems. The main reasons why one consent system would systematically have more kidney patients who die from co-morbid conditions than the other are (1) longer waiting time, e.g., due to smaller transplantation rates or more inefficient systems (note that, while kidney failure can be acute, the chance to suffer from co-morbid conditions will by construction increase as time goes by), and (2) unhealthier kidney patients are allowed to enter or remain on the waiting list, thereby increasing their probability to have or develop co-morbid conditions that ultimately become fatal. We do not have data to investigate these points, so we need to leave this point for future research.

Dealing with zero values

Both the data on deceased-donor and living transplantations and on organ-patient mortality are heavily right-skewed with many zero values. This may pose two problems in the data. First, although we already set all country-organ observations to missing when only zeros and missing values are reported for that organ across years, it is possible that not all remaining zeros are "true zeros." Some countries might, for instance, have reported a zero when there were no (reliable) data in a specific year. That means that there might be different mechanisms driving the increase from zero to one transplantation or death on the waiting list, than those driving the increase from one to two or infinity. We therefore run the same analyses as in Table 3.3, while excluding all zero values. Table A3.12 in Appendix 3.3 shows that the conclusions remain robust for this specification.

Second, because the data are censored at zero with many zero values, a Tobit regression might better fit our data than OLS. However, Table A3.13 in Appendix 3.3 shows that the conclusions remain robust for Tobit regressions.

Country heterogeneity within consent systems

We also analyze whether the non-significant relationship between consent systems and organ-patient mortality rates is due to the potentially large heterogeneity in organ-patient mortality among countries that have the same consent system. To decrease heterogeneity within consent systems, we run the same analyses as in Table 3.3 while only including European countries, Australia, Canada, New Zealand, and the United States (i.e., excluding Latin American, Asian, and African countries). By doing so, the included countries are likely to be more homogenous with respect to unobservable characteristics that might affect transplantation and mortality rates. Table A3.14 in Appendix 3.3 shows that the patterns shown in Table 3.3 remain robust with this specification.

Truncating the data

Finally, we analyze whether our results are driven by outliers in either the transplantation or the organ-patient mortality data. However, running the analyses after truncating both variables at the 99^{th} percentile leads to the same results (see Table A3.15 in Appendix 3.3).

3.5. CONCLUSIONS

This chapter analyzes the relationship between consent systems and the number of organ patients who died while waiting for a donor organ. Previous research has shown that countries with an opt-out consent system for organ donation conduct more deceased-donor transplantations than countries with an opt-in system. One might therefore expect that organ-patient mortality rates are lower in opt-out systems. We find that organ-patient mortality rates (i.e., the number of patients who died while on a waiting list, divided by the size of the population) among heart patients are indeed as low in countries with opt-out systems as could be expected based on the higher deceased-donor transplantation rates in these countries. For lungs, deceased-donor transplantation rates do not differ significantly between consent systems, and this is reflected in the mortality rates that are also not significantly different from each other. However, the difference between consent systems with respect to kidney- and liver-patient mortality rates is significantly different from the difference in deceased-donor transplantation rates. While deceased-donor kidney and liver transplantations are significantly higher in opt-out systems compared to opt-in systems, the number of patients (per million population) who died while on the waiting list for one of these organs is similar between consent systems. For kidneys, an explanation for this surprising result is that opt-out consent systems create fewer incentives to become living donors. Transplantations from living donors do not explain the findings for livers. These results are robust for various specifications.

Our results indicate that focusing on the deceased-donor transplantation rates alone paints an incomplete picture of the benefits of opt-out systems compared to opt-in systems. For kidneys and livers, consent systems for organ donation apparently do not affect organ-patient mortality rates through deceased-donor transplantation rates alone; there are other factors related to consent systems that affect the number of deaths for people on the kidney and liver waiting lists. We find that the number of living transplantations can explain the findings regarding kidneys, but not for livers. Waiting list removals and the length of the waiting lists appear to be unlikely mechanisms.

Moreover, it is worth noting that the number of kidney patients who died while on the waiting list is similar between consent systems, even though kidney transplantations from living donors are lower in opt-out systems compared to opt-in systems. Previous research shows that graft and patient survival rates are higher among those who received a kidney from a living donor than among those who received one from a deceased donor (Nemati et al., 2012). This indicates that living-donor kidney recipients may be less likely to re-enter the waiting list because their body rejected the donor kidney, and would therefore, by construction, be less likely to die while on a kidney waiting list. However, even if this were true, the impact on the number of deaths of people on the kidney waiting list must be minimal; otherwise, we would have found that the number of deaths among kidney patients (per million population) in opt-out systems is significantly higher than in opt-in systems. This is an important addition to the discussion on the benefits and drawbacks of the different consent systems.

More research is needed to identify determinants of organ-patient mortality rates. One potentially important topic for future research concerns the fact that countries with large organ shortages also have an incentive to participate in legal organ trade within a so-called "transplant network." Examples are Eurotransplant or Scandiatransplant, where organs that cannot be allocated within the donor's country because there is no matching recipient are allocated to a matching recipient from another country within the network. If countries with opt-in systems could acquire more organs (and particularly more kidneys and livers) through such networks than countries with opt-out systems, it might explain why the difference in kidney- and liver-patient mortality rates is so small between consent systems. Unfortunately, we cannot empirically test whether this hypothesis holds, because (1) almost every country in our data participates in a transplant network, either formally or through bilateral agreements, and (2) we do not have data on countries of origin of organ donors or recipients. A second potentially important topic for future research concerns investments in medical technology to decrease organ-patient mortality. The smaller supply of donor organs in countries with opt-in systems provides an incentive for these countries' governments and researchers to invest in research and technology to increase organ-patients' longevity or search for alternative sources of organs such as xenotransplants (i.e., transplanting animal organs into humans) or synthetic organs. If successful (for kidneys and livers), these investments would, in turn, decrease organ-patient mortality rates in opt-in systems, which would explain the surprisingly small differences between consent systems with respect to kidney- and liver-patient mortality rates. Although we do control for countries' overall health expenditures, we unfortunately do not have data on the size of the investments in these specific medical technologies that may decrease organ-patient mortality nor on the share of the overall health expenditures that are allocated to them. It is therefore not possible to provide empirical evidence for the hypothesis that differences between consent systems with respect to the incentives for investing in organ-patient mortality-decreasing medical technology may explain our findings.

The main implication of our analyses is that it is not evident that opt-out consent systems reduce kidney and liver-patient mortality rates. For kidneys and livers—by far the most commonly transplanted organs—the higher deceased-donor transplantation rates in opt-out systems give a false impression that presumed consent is related to lower organ-patient mortality rates.

Conclusions

SUMMARY

This doctoral dissertation investigates the direct effects and certain indirect effects of two public policies and interventions: the *Active Living Program*, a quasi field experiment which aims to increase children's physical activity in everyday life (**Part I**), and *opt-out consent for organ donation*, which aims to decrease organ-patient mortality through increased deceased-donor transplantations (**Part II**). The dissertation consists of three self-contained chapters, each using and combining different data and applying different empirical strategies.

Although the ambition of this dissertation is not to identify all potential indirect effects of the selected policies and interventions, it does highlight the importance of carefully monitoring and evaluating indirect effects, even when previous research or common sense suggest that these effects are desirable. This does not mean that policymakers are expected to find existing evidence on the universe of potential direct and indirect effects before implementing a policy, since *all* evidence cannot be collected, and waiting until a substantial portion of evidence is collected would render policymaking impossible. It also does not mean that policymakers and researchers are expected to monitor and evaluate the universe of effects after implementing a policy, since this would be practically impossible to do. However, this dissertation does call for increased awareness of potential indirect effects, both in policy design and in its evaluation, particularly the effects for which previous research or common sense provide indications of plausible relationships. It also highlights the importance of feedback loops in policymaking (also known as iterative policymaking or policy learning), so that policies can be adjusted as more evidence is found about the direct and indirect effects on more outcomes.

PART I

ANSWERS TO THE RESEARCH QUESTIONS

The first part of this dissertation focuses on the direct and indirect effects of encouraging physical activity in everyday life, and addresses three research questions: (1) *What is the direct effect of encouraging physical activity in everyday life on children's time spent on physical activity?* (2) *What is the indirect effect of encouraging physical activity in everyday life on children's school performance?* and (3) *What is the indirect effect of encouraging physical activity in everyday life on the prevalence of ADHD-like symptoms among children?* Below, these questions are answered in turn. Each answer is based on data from the Active Living Program: a quasi field experiment that was conducted in 10 treatment schools and 11 control schools during the 2012/2013 and the 2013/2014 school years in low-SES regions in the Southern Limburg region in the Netherlands. The interventions focused on active transportation and active play at school and during leisure time.

What is the direct effect of encouraging physical activity in everyday life on children's time spent on physical activity?

For any policy or intervention, it is important to evaluate its direct effects. Using unique data on physical activity, obtained from accelerometers that children wore throughout the day both before and after the implementation the Active Living Program, Chapter 1 applies linear difference-in-differences analyses to estimate the direct effect of encouraging physical activity in everyday life on children's time spent on physical activity. The results show that, on average, encouraging physical activity in everyday life through the Active Living Program increases 5th and 6th graders' time spent on physical activity during school-time by 9.3 minutes per day (equal to 0.34 standard deviations). No significant treatment effects are found on time spent on physical activity during leisure time. These results remain robust across the pre-treatment school performance distribution, but the effect on time spent on physical activity is slightly stronger for boys than for girls, and are strongest in the mornings of school-days.[49] In summary, the conclusion is that the Active Living Program successfully increases time spent on physical activity during school-time, without crowding out time spent on physical activity during leisure time.

What is the indirect effect of encouraging physical activity in everyday life on children's school performance?

The main focus of Chapter 1 lies on answering this question, which is an interesting and important one for two reasons. First, physical activity is expected to positively affect cognition and thereby school performance from a physiological viewpoint, but even if physical activity has positive effects on children's cognitive abilities, increasing sports participation may crowd out time investments in other potentially beneficial activities such as studying or active play. Second, it is uncertain whether previously found positive effects of formal physical activity such as doing moderate to vigorous intensity cardio or aerobic exercise on educational outcomes (e.g., Barron, Ewing, & Waddell, 2000; Cabane, Hille, & Lechner, 2016; Dills, Morgan, & Rotthoff, 2011; Felfe, Lechner, & Steinmayr, 2016; Lipscomb, 2007; Pfeifer & Cornelissen, 2010; Rees & Sabia, 2010; Stevenson, 2010) is generalizable to informal physical activity.

Chapter 1 uses panel data from the *Onderwijs Monitor Limburg* on individual children's performance on nationally standardized language and math tests from grades 1 to 6. This allows for conducting difference-in-differences event study analyses including three pre-treatment years. The results show that encouraging physical activity in everyday life through the Active Living Program decreases 5th and 6th graders' school performance by 5.9 percent of a standard deviation (most conservative estimate of the average treatment effect). The negative effect on school performance is robust for a variety of specifications and is strongest among the worst-performing students and among boys. No empirical evidence could be found that crowding out of other

[49] Although the empirical strategy used in Chapter 1 differs from the one applied by Van Kann et al. (2016), resulting in different sizes of the estimates, the general conclusion of Chapter 1 is similar, i.e. the Active Living Program successfully increases children's time spent on physical activity.

cognitively stimulating activities, such as playing music, chess or doing arts and crafts, is a potential mechanism for the negative treatment effects on school performance. Although the results may not be generalizable to interventions that encourage formal physical activity or that focus on adults, suggestive evidence is found of the possibility that increasing adolescents' physical activity in the mornings of school days can also have negative effects on school performance in other high- or middle-income countries and outside of low-SES regions.

What is the indirect effect of encouraging physical activity in everyday life on the prevalence of ADHD-like symptoms among children?

Attention-deficit/hyperactivity disorder (ADHD) is a chronic behavioral and mental health disorder that negatively affects cognitive development and functioning in daily life. Early and effective management of the disorder is therefore crucial. Previous studies that found that physical activity can decrease ADHD-like symptoms, focus on formal physical activity such as doing moderate to vigorous intensity cardio or aerobic exercise (e.g., Den Heijer et al., 2016; Gapin, Labban and Etnier, 2011; Ng et al., 2017; Pan et al., 2016; Smith et al, 2013). Chapter 2 analyzes the indirect effect of encouraging physical activity in everyday life (i.e. informal physical activity) on the prevalence of ADHD-like symptoms among children.

Repeated cross-sectional data on self-reported ADHD-like symptoms are taken from 6th grader questionnaires of the *Onderwijs Monitor Limburg*. Ordered probit difference-in-differences analyses including two pre-treatment years reveal that encouraging physical activity in everyday life through the Active Living Program significantly decreases male 6th graders' self-assessed ability to do exactly what the teacher asks them to do (one of three indicators of inattentiveness), and to be calm and quiet when the teacher wants them to (one of three indicators of hyperactivity-impulsivity). The probability that boys fail for these 'subjects' increases by 2 and 3- percentage points, respectively, which is equal to 100% and 45 to 87% compared to boys' pre-treatment probabilities. No significant treatment effects on ADHD-like symptoms are found for girls.

INTERPRETATION OF THE FINDINGS

Whereas the Active Living Program successfully increases children's time spent on physical activity during school-time, it unintentionally decreases school performance and increases the prevalence of ADHD-like symptoms among boys. Although the indirect effects found in Chapters 1 and 2 may appear to be contradictory to the existing literature on the topics, this is not necessarily the case. After all, the studies in Chapters 1 and 2 are among the first that identify the causal indirect effects of encouraging *informal* physical activity (i.e., physical activity in everyday life, including active play and active transportation). The results suggest that the commonly found positive effects of exercising or participating in sports (i.e., formal physical activity) on educational outcomes and on ADHD-like symptoms may not be generalizable to informal physical activity.

There are several potential explanations for the negative treatment effects on school performance and ADHD-like symptoms that are found in this dissertation, whereas the existing literature points towards a positive relationship. One explanation may be that formal physical activity is more effective for fostering skills, such as discipline, that are beneficial for school performance and for decreasing ADHD-like symptoms than informal physical activity. Moreover, because exercising is generally more intense than being active in everyday life, formal physical activity may yield higher returns than informal physical activity in terms of health, well-being, and (therefore) ADHD-like symptoms and school performance. If, due to the Active Living Program, children usually substitute formal physical activity for informal physical activity, it would explain the negative effects that are found. Another potential explanation is that physical activity may be fun and distracting, and lower-intensity or informal physical activity may not be as effective in transforming the distraction into an ADHD-symptom-reducing effect as higher-intensity or formal physical activity. Moreover, the distraction of the physical activity interventions may negatively affect school performance, not only of those who are distracted (as Chapter 2 indicates: mostly boys), but – due to negative spill-over effects – also of the distracted students' classmates. The worst-performing students, who arguably need their study- and instruction-time the most, may also suffer the most if they or other children are increasingly inattentive, hyperactive or impulsive during instruction time.

PART II

ANSWERS TO THE RESEARCH QUESTIONS

The second part of this dissertation focuses on the direct and indirect effects of opt-out consent systems for organ donation, and addresses three research questions: (1) *What is the direct effect of opt-out consent for organ donation on deceased-donor transplantation rates?* (2) *What is the indirect effect of opt-out consent for organ donation on organ-patient mortality rates?* and (3) *Does the effect of opt-out consent for organ donation on deceased-donor transplantation rates translate into a similar effect (though with opposite sign) on organ-patient mortality rates?* In order to answer these research questions, the largest dataset available on organ-patient mortality and organ transplantations was created, by combining organ transplantation and organ-patient mortality data from 69 countries across a period of 15 years from newsletters from the Organización Nacional de Trasplantes, with data from 26 different sources to determine whether a country has an opt-out or opt-in system, and data on several potentially confounding factors from the World Bank, the World Health Organization, and the World Values Survey.

What is the direct effect of opt-out consent for organ donation on deceased-donor transplantation rates?

Previous research has shown that countries with an opt-out consent system for organ donation conduct more deceased-donor transplantations than countries with an opt-in system (see, e.g. Abadie & Gay, 2006; Johnson & Goldstein, 2003; Johnson & Goldstein, 2004; Rithalia et al., 2009; Shepherd, O'Carroll & Ferguson, 2014; Ugur, 2015). For kidneys, livers and hearts, this finding is confirmed in Chapter 3: countries with opt-out consent systems for organ donation on average conduct 7.1 more deceased-donor kidney transplantation per million population (pmp), 4.4 pmp more deceased-donor liver transplantations and 1.0 pmp more deceased-donor heart transplantations than countries with opt-in systems. On average, there are no statistically significant differences between consent systems regarding the number of lung transplantations from deceased donors that are performed pmp. The results are based on ordinary least squares regressions, controlling for year fixed-effects as well as for countries' GDP, general mortality rates, health expenditures, and religious denomination.

What is the indirect effect of opt-out consent for organ donation on organ-patient mortality rates?

Since countries with opt-out consent systems conduct significantly more deceased-donor transplantations, one might expect that organ-patient mortality rates are lower in opt-out systems. Chapter 3 is the first study to analyze whether organ-patient mortality rates are indeed lower in countries with opt-out consent systems for organ donation compared to those with opt-in systems.

Although lower organ-patient mortality rates are one of the aims of opt-out consent systems, it is not the direct effect, since opt-out systems cannot literally save lives: mortality rates can only be affected through other channels, such as deceased-donor transplantation rates. However, using the same empirical strategy as above, Chapter 3 shows that organ-patient mortality rates (i.e., the number of patients who died while on a waiting list, divided by the size of the population) rarely differ significantly between consent systems. This conclusion holds for all organs, and across all specifications.

Does the effect of opt-out consent for organ donation on deceased-donor transplantation rates translate into an equal effect (though with opposite sign) on organ-patient mortality rates?

Common sense tells us that every transplantation performed is an organ-patient's life saved, and hence, that every deceased-donor transplantation decreases the number of organ-patients that dies by one patient. Although this conclusion may hold in a partial equilibrium, Chapter 3 is the first study that analyzes whether it also holds in a general equilibrium, in which organ-patient mortality rates can be affected by more than deceased-donor transplantations alone. The results of seemingly unrelated regressions show that, for hearts and lungs, the effect[50] of opt-out consent for organ donation on deceased-donor transplantation rates indeed translate into a similar effect (though with opposite sign) on organ-patient mortality rates. For kidneys and livers, however, this is not the case. For these organs, the effect of opt-out consent for organ donation on deceased-donor transplantation rates is significantly larger (further away from zero) than the effect on organ-patient mortality rates.

[50] Please note that for simplicity, the term "effect" is used to describe the correlation between consent systems and the dependent variables. However, Chapter 3 does not identify causal effects of consent systems.

INTERPRETATION OF THE FINDINGS

The main conclusion of Part II is that it is not evident that opt-out consent systems reduce kidney and liver-patient mortality rates. For kidneys and livers—by far the most commonly transplanted organs—the higher deceased-donor transplantation rates in opt-out systems do not translate into equally lower mortality rates among kidney and liver patients. This indicates the existence of factors related to opt-out consent systems that are positively related to kidney and liver-patient mortality.

For kidneys, Chapter 3 shows that one such factor appears to be a decreased incentive to become a living donor in countries with opt-out systems (plausibly due to the higher deceased-donor transplantation rates). However, transplantations from living donors do not explain the findings for livers. Waiting list removals and the length of the waiting lists appear to be unlikely mechanisms. More plausible explanations for the findings are (1) that countries with opt-in systems acquire more organs (and particularly more kidneys and livers) through transplant networks than countries with opt-out systems and (2) that the smaller supply of donor organs in countries with opt-in systems provides an incentive for these countries' governments and researchers to invest in research and technology to increase organ-patients' longevity or search for alternative sources of organs such as xenotransplants (i.e., transplanting animal organs into humans) or synthetic organs.

GENERAL IMPLICATIONS AND VALORIZATION

The results in this dissertation have important implications for policymakers, researchers and other experts in the fields of health and education. They highlight that these stakeholders would be advised to: (1) remain alert when implementing policies and interventions and assure that all the (assumed) benefits and potential costs are carefully monitored. Since it is practically impossible to monitor and evaluate the universe of potential direct and indirect effects, they could focus on the effects for which previous research or common sense provide indications of plausible relationships; (2) consult research evidence to make informed decisions that weigh the benefits against the costs (both financial and otherwise); (3) develop targeted interventions to ensure that the assumed positive indirect effects are met.

Chapter 1

This chapter highlights that policymakers cannot assume that every type of physical activity intervention will only have positive indirect effects if it is effective in increasing time spent on physical activity. To make an informed decision about whether to increase the time children spend on physical activity, policymakers and educators must weigh the benefits for health and well-being of the child relative to the potential cost of decreased educational performance. In this process, they implicitly or explicitly make a choice as to what extent education outcomes and health outcomes are important for children. This means that, although the existing literature points toward a positive relationship between physical activity and educational outcomes, policymakers must remain alert when implementing interventions, and carefully monitor the (assumed) benefits and potential costs (focusing on – but not limited to – those for which previous research and common sense provide indications of plausible relationships).

Another important implication is that, to prevent increased educational inequalities, targeted interventions may be needed that increase low-performing students' physical activity without further decreasing their educational performance. Finally, before resources are allocated to increase time spent on physical activity among children and adolescents, we first need to improve our understanding of the causal effects of such interventions on short-term and long-term educational outcomes. Future research needs to focus on which forms of physical activity do not crowd out desirable activities.

Chapter 2

The results of this chapter have important implications for policymakers, researchers, healthcare professionals, educators, and parents alike, who aim to increase children's time spent on physical activity and/or reduce their ADHD-like symptoms. Although physical activity is an attractive potential treatment of ADHD-like symptoms since it is less invasive than medication, this study shows that the effectiveness of using physical activity to reduce ADHD-like symptoms depends greatly on the type of interventions that are implemented. Certain physical activity interventions may accidentally cause ADHD-like symptoms to increase for certain children. This study also highlights the importance of not only studying direct effects of policies and interventions (in this case, the effects of encouraging physical activity on time spent on physical activity), but to also investigate potential indirect effects, even if previous literature has shown desirable relationships with those variables. In order to make informed decisions about whether or not to implement certain policies or interventions, we need as much information as possible about other aspects they may affect.

Chapter 3

The main implication of this chapter is that focusing on the deceased-donor transplantation rates alone paints an incomplete picture of the benefits of opt-out systems compared to opt-in systems. This is an important contribution to the policy debate regarding consent systems. Moreover, the fact that living kidney transplantation rates are on average lower in opt-out systems remains a (dis)advantage of opt-out systems that receives little to no attention in the current debate on the pros and cons of different consent systems. Further research is needed to better understand the channels through which consent systems can affect organ-patient mortality rates, channels such as technological investments regarding organ-patient longevity and organ trade through transplant networks, as well as to what extent and why these channels differ by organ.

Finally, this doctoral dissertation has important implications for people's everyday life. First, Chapters 1 and 2 have important implications for parents of children and adolescents. Although it is important to encourage children to be (more) physically active, the results in the first two chapters show that this should not be done at all costs. We need to keep in mind that every minute can only be spent once, and make informed decisions about what to encourage children to spend their time on.

Based on Chapter 3, I hope to inspire people to start (or continue) the conversation with their family and loved ones about whether they would like to become an organ donor or not after they die, and better yet, to record their decision in a donor register or will. Without advocating certain consent systems for organ donation, I do want to emphasize that making a clear choice about organ donation will save your loved ones the trouble of making an important and difficult decision for you during a moment of great distress. Until then, I wish you all a life full of happiness and inspiring conversations.

Bibliography

Abadie A. & Gay S. (2006). The impact of presumed consent legislation on cadaveric organ donation: a cross-country study. *Journal of Health Economics*, 25(4): 599-620.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th Edition*. Electronic version accessed on 29 March 2021, via: https://doi.org/10.1176/appi.books.9780890425596

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American Statistical Association, 103, 1481–1495.

Angelucci, M. & De Giorgi, G. (2009). Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption? *American Economic Review*, 99(1): 486–508.

Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton, NJ: Princeton University Press.

Arshad, A., Anderson, B. & Sharif, A. (2019). Comparison of organ donation and transplantation rates between opt-out and opt-in systems. *Kidney International*. 95(6): 1453-1460.

Barkley R. (2002). Major life activity and health outcomes associated with attention deficit/hyperactivity disorder, *Journal of Clinical Psychiatry*, 63:12, pp 10-15.

Barron, J., Ewing, B., & Waddell, G. (2000). The effects of high school athletic participation on education and labor market outcomes. *Review of Economics and Statistics*, 82: 409–421.

Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. Survey Methodology, 28, 169–181.

Berg, V. van den, Salasi, E., Groot, R. H. M. de, Jolles, J., Chinapaw, M. J. M., & Singh, A. S. (2016). Physical activity in the school setting: Cognitive performance is not affected by three different types of acute exercise. *Frontiers in Psychology*, 7: 1–9.

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

Cabane, C., Hille, A., & Lechner, M. (2016). Mozart or Pelé? The effects of adolescents' participation in music and sports. *Labour Economics*, 41: 90–103.

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

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *The Journal of Human Resources*, 50: 318–372.

Carver, A., Timperio, A., & Crawford, D. (2013). Parental chauffeurs: What drives their transport choice? *Journal of Transport Geography*, 26: 72–77.

Cawley, J., Frisvold, D., & Meyerhoefer, C. (2013). The impact of physical education on obesity among elementary school children. *Journal of Health Economics*, 32: 743–755.

Cawley, J., Meyerhoefer, C., & Newhouse, D. (2007). The impact of state physical education requirements on youth physical activity and overweight. *Health Economics*, 16: 1287–1301.

Centers for Disease Control and Prevention. (2011). *Strategies to prevent obesity and other chronic diseases: The CDC guide to strategies to increase physical activity in the community*. Atlanta, GA: U.S. Department of Health and Human Services.

Charpentier, K., & Mavanur, A. (2008). Removing patients from the liver transplant waiting list: a survey of US liver transplant programs. *Liver Transplantation*, 14: 303-307.

Choi, L., Liu, Z., Matthews, C. W., & Buchowski, M. S. (2011). Validation of accelerometer wear and nonwear time classification algorithm. *Medicine, and Science in Sports, and Exercise*, 43: 357–364.

Colcolmbe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive functioning of older adults: A meta-analytic study. *Psychological Science*, 14: 125–130.

Currie, J. & Stabile, M. (2006). Child mental health and human capital accumulation: The case of ADHD, *Journal of Health Economics*, 25(6): 1094-1118.

Dalen, H. &, Henkens, K. (2014). Comparing the effects of defaults in organ donation systems. *Social Science and Medicine*, 106: 137-142.

Davidai, S., Gilovich, T. & Ross, L. (2012). The meaning of default options for potential organ donors. *Proceedings of the National Academy of Sciences*, 109(38): 15201-15205.

Davies, H., Nutley, S. & Smith, P. (2000). Introducing evidence-based policy and practice in public services. In: Davies, H.T.O., Nutley, S.M. & Smith, P.C. (eds.) *What works? Evidence-based policy and practice in public services.* Bristol: The Policy Press.

Deming, D. (2009). Early childhood intervention and life-cycle skills development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1: 111–134.

Dills, A. K., Morgan, H. N., & Rotthoff, K. W. (2011). Recess, physical education, and elementary school student outcomes. *Economics of Education Review*, 30: 889–900.

Donald, D. G., & Lang, K. (2007). Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics*, 89: 221–223.

Engel, P., Frey, A. & Verhagen, M.D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences,* 118(17): e2022376118.

Erickson, K. I., Hillman, C. H., & Kramer, A. F. (2015). Physical activity, brain, and cognition. *Current Opinion in Behavioral Sciences*, 4: 27–32.

European Commission (2001). *European Governance: A White Pape*r. The European Commission of the European Communities, COM(2001) 428. [Electronic version]. Accessed on 7 June 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52001DC0428

European Commission. (2014). *EU action plan on childhood obesity 2014/2020*.

Evenson, K., Catellier, D., Gill, K., Ondrak, K., & McMurray, R. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences*, 26: 1557–1565.

Felez-Nobrega, M., Hillman, C. H., Cirera, E., & Puig-Ribera, A. (2017). The association of context-specific sitting time and physical activity intensity to working memory capacity and academic achievement in young adults. *The European Journal of Public Health*, 27: 741–746.

Felfe, C., Lechner, M., & Steinmayr, A. (2016). Sports and child development. *PLoS ONE*, 11: 1–23.

Fernandez J, Howard D, Stohr Kroese L. The effect of cadaveric kidney donations on living kidney donations: an instrumental variable approach. Economic Inquiry. 2013;51(3): 1696-1714.

Fletcher, J. & Wolfe, B. (2008). Child mental health and human capital accumulation: The case of ADHD revisited, *Journal of Health Economics*, 27(3): 794-800.

Freeman, R.B. (1996). Why do so many young American men commit crimes and what might we do about it? *Journal of Economic Perspectives*, 10(1): 25-42.

Golsteyn, B.H.H., Jansen, M.W.J., Van Kann, D.H.H. & Verhagen, A.M.C. (2020). Does Stimulating Physical Activity Affect School Performance?, *Journal of Policy Analysis and Management*, 39(1): 64–95.

Golsteyn, B.H.H. & Verhagen, A.M.C. (2021). Deceased by Default: Consent Systems and Organ Patient Mortality. *PLoS ONE*, 16(3): e0247719.

Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., Pinior, B., Thurner, S. & Klimek, P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behavior*, 4: 1303–1312.

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: Exercise effects on brain and cognition. *Nature Reviews Neuroscience*, 9: 58–65.

Hume, C., Timperio, A., Salmon, J., Carver, A., Giles-Corti, B., & Crawford, D. (2009). Walking and cycling to school: Predictors of increases among children and adolescents. *American Journal of Preventive Medicine*, 36: 195–200.

Howard, D. (2007). Producing organ donors. *Journal of Economic Perspectives*, 21(3): 25-36.

Hoy, E., Weiss, G., Minde, K., & Cohen, N. (1978). The hyperactive child at adolescence: Cognitive, emotional, and social functioning, *Journal of Abnormal Child Psychology*, 6: 311-324.

Johnson E. & Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649): 1338-1339.

Johnson E. & Goldstein, D. (2004). Defaults and donation decisions. *Transplantation*, 78(12): 1713-1716.

Kann, D. H. H. van, Jansen, M. W. J., Vries, S. I. de, & Kremers, S. P. J. (2015). Active Living: development and quasi-experimental evaluation of a school-centered physical activity intervention for primary school children. *BMC Public Health*, 15: 1–10.

Kann, D. H. H. van, Kremers, S. P. J., Vries, N. K. de, Vries, S. I. de, & Jansen, M. W. J. (2016). The effect of a school-centered multicomponent intervention on daily physical activity and sedentary behavior in primary school children: The Active Living study. *Preventive Medicine*, 89: 64–69.

Kessler, R.C., Adler, L., Ames, M., Barkley, R.A., Birnbaum, H., Greenberg, P., Johnston, J.A., Spencer, T., & Ustün, T.B. (2005). The Prevalence and Effects of Adult Attention Deficit/Hyperactivity Disorder on Work Performance in a Nationally Representative Sample of Workers, *Journal of Occupational and Environmental Medicine*, 47(6): 565-572.

Kessler, J. & Roth, A. (2014). Getting more organs for transplantation. *American Economic Review: Papers and Proceedings*, 104(5): 425-430.

Kollins, S.H., McClernon, F.J., & Fuemmelerm, B. (2005). Association between smoking and attention-deficit/hyperactivity disorder symptoms in a population-based sample of young adults, *Archives of General Psychiatry*, 62: 1142-1147.

Lipscomb, S. (2007). Secondary school extracurricular involvement and academic achievement: A fixed effects approach. *Economics of Education Review*, 26: 463–472.

Liu, P.J., Dallas, S.K., Harding, M. & Fitzsimons, G.J. (2018). The Sales Impact of Featuring Healthy Foods, Indulgent Foods, or Both: Findings from a Large-Scale Retail Field Study. *Journal of the Association for Consumer Research*, 3(3): 255–259.

Lizandra, J., Devís-Devís, J., Pérez-Gimeno, E., Valencia-Peris, A., & Peiró-Velert, C. (2016). Does sedentary behavior predict academic performance in adolescents or the other way round? A longitudinal path analysis. *PLoS ONE*, 11: e0158254.

Li, D., Hawley, Z. & Schnier, K. (2013). Increasing organ donation via changes in the default choice or allocation rule. *Journal of Health Economics*, 32: 1117-1129.

Mannuzza, S. & Klein, R.G. (2000). Long-term Prognosis in Attention-Deficit/Hyperactivity Disorder, *Child and Adolescent Psychiatric Clinics of North America*, 9(3): 711-726.

Mossialos, E., Costa-Font, J. & Rudisill, C. (2008). Does organ donation legislation affect individuals' willingness to donate their own or their relative's organs? Evidence from European Union survey data. *BMC Health Services Research*, 8(48): 1-10.

Nemati, E., Einollahi, B., Pezeshki, M.L., Porfarziani, V. & Fattahi, M.R. (2012). Does kidney transplantation with deceased or living donor affect graft survival? *Nephro-Urology Monthly*, 6(4): e12182.

Pawson, R. (2006). *Evidence-based Policy: a Realist Perspective*. London: SAGE Publications Ltd.

Pfeifer, C., & Cornelissen, T. (2010). The impact of participation in sports on educational attainment—New evidence from Germany. *Economics of Education Review*, 29: 94–103.

Pont, K., Ziviani, J., Wadley, D., Bennett, S., & Abbott, R. (2009). Environmental correlates of children's active transportation: A systematic literature review. *Health & Place*, 15: 849–862.

Rees, D. I., & Sabia, J. J. (2010). Sports participation and academic performance: Evidence from the National Longitudinal Study of Adolescent Health. *Economics of Education Review*, 29: 751–759.

Rich, C., Griffiths, L., & Dezateux, C. (2012). Seasonal variation in accelerometer-determined sedentary behaviour and physical activity in children: A review. *International Journal of Behavioral Nutrition and Physical Activity*, 9: 1–8.

Rithalia, A., McDaid, C., Suekarran, S., Myers, L. & Sowden, A.(2009). Impact of presumed consent for organ donation on donation rates: a systematic review. *BMJ*, 338: 1-8.

Satterfield, J.H., Hoppe, C.M., & Schell, A.M. (1982). A prospective study of delinquency in 110 adolescent boys with attention-deficit disorder and 88 normal adolescent boys, *American Journal of Psychiatry*, 139: 795-798.

Shepherd, L., O'Carroll, R. & Ferguson, E. (2014). An international comparison of deceased and living organ donation/transplant rates in opt-in and opt-out systems: a panel study. *BMC Medicine*, 12(131): 1-14.

Shimazono, Y. (2007). The state of the international organ trade: a provisional picture based on integration of available information. *Bulletin of the World Health Organization*, 85(12): 901-980.

Singh, A., Uijtdewilligen, L., Twisk, J. W. R., Mechelen, W. van, & Chinapaw, M. J. M. (2012). Physical activity and performance at school: A systematic review of the literature including a methodological quality assessment. *Archives of Pediatrics and Adolescent Medicine*, 166: 49–55.

Skogli, E.W., Teicher, M.H., Andersen, P.N., Hovik, K.T., & Øie, M. (2013). ADHD in girls and boys – gender differences in co-existing symptoms and executive function measures, *BMC Psychiatry*, 13(298): 1-12.

Statistics Netherlands (2013). *Statistics on population density* via statline.cbs.nl.

Statistics Netherlands (2014). *Statline: nabijheid voorzieningen; afstand locatie, wijk- en buurtcijfers 2013*. Voorburg/Heerlen: Statistics Netherlands.

Statistics Netherlands (2017). *Statline*. http://statline.cbs.nl/.

Stevenson, B. (2010). Beyond the classroom: Using Title IX to measure the return to high school sports. *Review of Economics and Statistics*, 92: 284–301.

Strong, W., Malina, R., Blimkie, C., Daniels, S., Dishman, R., & Gutin, B. (2005). Evidence-based physical activity for school-age youth. *The Journal of Pediatrics*, 146: 732–737.

Swanson, J.M., Sergeant, J.A., Taylor, E., Sonuga-Barke, E.J.S., Jensen, P.S., and Cantwell, D.P. (1998). Attention-deficit hyperactivity disorder and hyperkinetic disorder, *The Lancet*, 351(9100): 429-433.

Thuluvath, P.J. & Yoo, H.Y. (2004). Graft and patient survival after adult live donor liver transplantation compared to a matched cohort who received a deceased donor transplantation. *Liver Transplantation*, 10(10): 1263-1268.

Ugur, Z. (2015). Does presumed consent save lives? Evidence from Europe. *Health Economics*, 24(12): 1560-1572.

United States of America (2019). *Foundations for Evidence-Based Policymaking Act of 2018*. 132 STAT. 5529. Public Law 115–435. H.R.4174 [Electronic version]. Accessed on 7 June 2021. https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf

Vaa, T. (2014). ADHD and relative risk of accidents in road traffic: A meta-analysis, *Accident Analysis & Prevention*, 62, pp 415-425.

World Health Organization. (2010). *Global recommendations on physical activity for health*. Geneva, Switzerland: WHO Press.

World Health Organization (2019). *Attention Deficit Hyperactivity Disorder (ADHD) (WHO-EM/MNH/214/E)*. Electronic version downloaded on 29 March 2021, via: https://applications.emro.who.int/docs/EMRPUB_leaflet_2019_mnh_214_en.pdf?ua=1&ua=1.

World Health Organization (2020). *Physical activity*. Accessed on 3 August 2021, via: https://www.who.int/news-room/fact-sheets/detail/physical-activity.

Appendix

Appendix 1
Appendix to Chapter 1

1.1. CONTROL VARIABLES AND SUMMARY STATISTICS

1.1.1. Control variables

In the school performance data, age is registered as the average age (in months) on the day the tests were taken. In the PA data, age is registered as age (in months) during the PA measurement week. Both data sets include an indicator variable with a value of one if the child is a boy. We create an indicator variable with a value of one if the child was enrolled in grade 5 during the pre-treatment year and zero if in that school year, the child was enrolled in grade 4. The accelerometer wearing time is registered in minutes. Our weather indicators are mean temperature, hours of sunshine, and rainfall (millimeters) between 6:00 a.m. and 11:00 p.m. during the PA measurement week. Rainfall is transformed into an indicator variable with a value of one if more than zero millimeters of rain fell. The weather indicators are based on information from the weather station of the Royal Dutch Meteorological Institute (KNMI) located at the Maastricht-Aachen airport and are calculated for every day of the measurement week.

1.1.2. Summary statistics

Table A1.1 lists descriptive statistics of our full sample, i.e., of children in the target group (fourth and fifth graders in 2012/2013) who are enrolled in one of the 21 Active Living schools, as well as of their peers in non-Active Living schools. First, the table shows that the children in Active Living schools account for 17 percent of our full sample. It also highlights that Active Living schools are located in low-SES areas; average school performance is significantly lower in Active Living schools compared to the other (non-Active Living) schools in our data. Children in Active Living schools and non-Active Living schools are comparable with respect to age, gender, and grade in which they were enrolled in 2012/2013.

Table A1.2 lists descriptive statistics for the estimation samples of our main results (i.e., including child fixed effects and a vector of control variables as specified earlier in the section on Empirical Strategy). In our estimation samples, we have school performance data from 1,014 children, school-time PA data from 536 children, and leisure-time PA data from 509 children. When the school performance data are merged with the PA data, school-time PA data remain for 422 children and leisure-time PA data for 398 children.

In the school performance data, 52 percent of the sample is enrolled in a treatment school. This share is slightly larger in the PA data (56 percent), which indicates that there are relatively more children in control schools for whom we do not have (valid) PA data.

Table A1.1. Descriptive statistics for Active Living schools and non-Active Living schools.

	Full sample		Non–Active Living schools		Active Living schools		Diff.	P-value
	mean	(sd)	mean	(sd)	mean	(sd)		
Child enrolled in an Active Living school (1 = yes)	0.17	(0.38)	0.00	(0.00)	1.00	(0.00)	-1.00	0.000
Overall school performance	62.28	(9.30)	62.41	(9.33)	61.65	(9.11)	0.76	0.001
Language performance	61.62	(9.25)	61.74	(9.29)	61.04	(9.01)	0.70	0.002
Math performance	60.96	(8.88)	61.10	(8.88)	60.30	(8.86)	0.80	0.000
Age when tests were taken (months)	132.39	(10.25)	132.37	(10.25)	132.46	(10.26)	-0.09	0.718
Gender (1 = boy)	0.49	(0.50)	0.49	(0.50)	0.48	(0.50)	0.02	0.189
Cohort (1 = 5th graders, 0 = 4th graders)	0.61	(0.49)	0.61	(0.49)	0.59	(0.49)	0.02	0.129
Number of children in the estimation sample	6,776		5,671		1,105			
Number of schools	174		153		21			

Notes: Descriptive statistics are based on the pre-treatment year (2012/2013) and the post-treatment year (2013/2014) combined.
Source: School performance data.

On average, children took five tests per year (i.e., the number of tests on which the school performance measure is based). The average school performance in the pre-treatment year is 57.28, with a standard deviation of 7.90. Because the subject scores are standardized across school years and children's performance generally improves as they get older, scores increase over time, to an average of 65.16 with a standard deviation of 8.04 in the post-treatment year.

The children in the estimation sample were, on average, 10.5 years old (126 months) when they took their pre-treatment tests, and 11.5 years old (138 months) when they took tests in the post-treatment year. Because we exclude tests that were taken during the PA measurement weeks, the age of the children in the school performance data is, on average, two months lower than in the PA data.

In each data set, a little less than half the sample is male. The slight overrepresentation of the 5th-grade cohort in the school performance data shows that there are more fourth graders than fifth graders who did not take any tests.

Table A1.2. Descriptive statistics for the estimation samples.

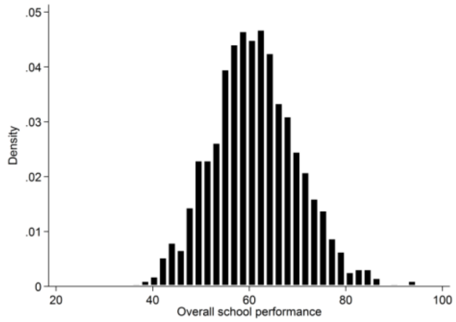
	Pre-treatment year mean	(sd)	Post-treatment year mean	(sd)	Diff.	P-value
Panel A. School performance data						
Treatment (1 = Treatment group)	0.52	(0.50)	0.52	(0.50)	0.00	1.000
Number of tests taken	5.00	(0.93)	4.98	(0.77)	-0.02	0.515
Overall school performance	57.28	(7.90)	65.16	(8.04)	7.88	0.000
Language performance	56.99	(7.89)	64.29	(8.09)	7.31	0.000
Math performance	56.05	(7.99)	63.81	(7.64)	7.76	0.000
Age when tests were taken (months)	126.10	(8.35)	137.91	(8.21)	11.81	0.000
Gender (1 = boy)	0.48	(0.50)	0.48	(0.50)	0.00	1.000
Cohort (1 = 5th graders, 0 = 4th graders)	0.57	(0.49)	0.57	(0.49)	0.00	1.000
Number of children with school performance data	1,014		1,014			
Panel B. PA data						
Treatment (1= Treatment group)	0.56	(0.50)	0.56	(0.50)	0.00	1.000
Number of days the accelerometer was worn	4.10	(1.15)	4.15	(1.46)	0.05	0.547
Time spent on school-time PA	93.36	(24.74)	93.64	(25.64)	0.28	0.854
Time spent on leisure-time PA	203.50	(47.13)	182.46	(43.99)	-21.04	0.000
Wearing time of accelerometer during school (min./day)	309.23	(32.18)	323.73	(31.67)	14.51	0.000
Wearing time of accelerometer during leisure (min./day)	547.89	71.55	527.24	69.71	-20.65	0.000
Age during PA measurement (months)	127.75	(8.22)	139.70	(8.23)	11.96	0.000
Gender (1 = boy)	0.44	(0.50)	0.44	(0.50)	0.00	1.000
Cohort (1 = 5th graders, 0 = 4th graders)	0.45	(0.50)	0.45	(0.50)	0.00	1.000
Outside temperature during PA measurement (Celsius)	13.07	(5.58)	17.04	(3.45)	3.97	0.000
Amount of sunshine during PA measurement week (hours/day)	2.62	(1.59)	5.20	(2.26)	2.58	0.000
Rain during PA measurement week (1=yes)	0.46	(0.50)	0.47	(0.50)	0.01	0.669
Number of children with school-time PA data	536		536			
Number of children with leisure-time PA data	509		509			
Panel C. Merged data						
Treatment (1 = Treatment group)	0.60	(0.49)	0.60	(0.49)	0.00	1.000
Overall school performance	57.67	(7.94)	65.58	(8.14)	7.91	0.000
Language performance	57.38	(7.92)	64.73	(8.23)	7.35	0.000
Math performance	56.24	(7.76)	64.09	(7.48)	7.85	0.000
Time spent on school-time PA	91.84	(23.65)	92.12	(25.02)	0.28	0.867
Time spent on leisure-time PA	202.65	(47.78)	182.24	(44.81)	-20.40	0.000
Wearing time of accelerometer during school time (min./day)	308.75	(31.47)	325.62	(32.40)	16.87	0.000
Wearing time of accelerometer during leisure time (min./day)	546.82	(70.43)	529.76	(69.61)	-17.06	0.001
Age when tests were taken (months)	124.89	(8.27)	136.74	(8.11)	11.85	0.000
Age during PA measurement (months)	128.36	(8.34)	140.31	(8.34)	11.95	0.000
Gender (1 = boy)	0.42	(0.49)	0.42	(0.49)	0.00	1.000
Cohort (1 = 5th graders, 0 = 4th graders)	0.52	(0.50)	0.52	(0.50)	0.00	1.000
Outside temperature during PA measurement (Celsius)	12.46	(5.72)	16.80	(3.36)	4.34	0.000
Amount of sunshine during PA measurement week (hours/day)	2.59	(1.62)	4.96	(2.39)	2.37	0.000
Rain during PA measurement week (1=yes)	0.44	(0.50)	0.51	(0.50)	0.07	0.039
Number of children with school performance and school-time PA data	422		422			
Number of children with school performance and leisure-time PA data	398		398			

Notes: Descriptive statistics for the school performance data are based on the estimation sample of the analyses presented in column 3 of Table 1.3 in the main text. Descriptive statistics for the PA data are based on the estimation sample of the analyses presented in columns 3 and 6 of Table 1.4 in the main text. Descriptive statistics for the merged data are based on the estimation sample of the analyses presented in Figure A1.7 in Appendix 1.4.

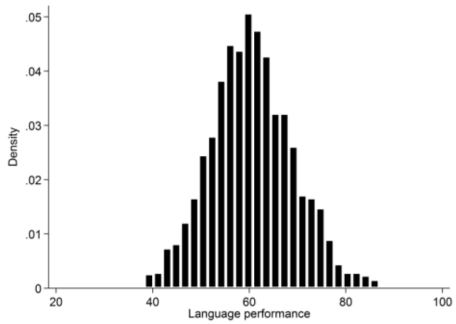
The weather conditions were slightly better during the PA measurements in the treatment year compared to the PA measurements in the pre-treatment year: outside temperatures were slightly higher and there were more hours of sunshine.

Figures A1.1 and A1.2 indicate that the pre-treatment measures of school performance and time spent on PA are approximately normally distributed.

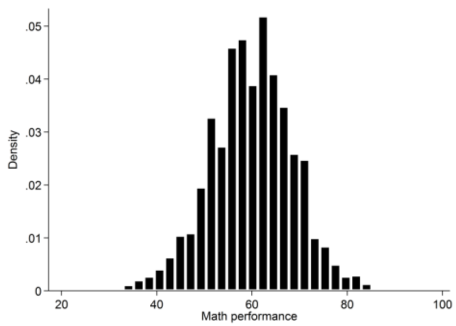
Figure A1.1. Distribution of School Performance in the Estimation Sample.
Panel A. Overall school performance



Panel B. Language

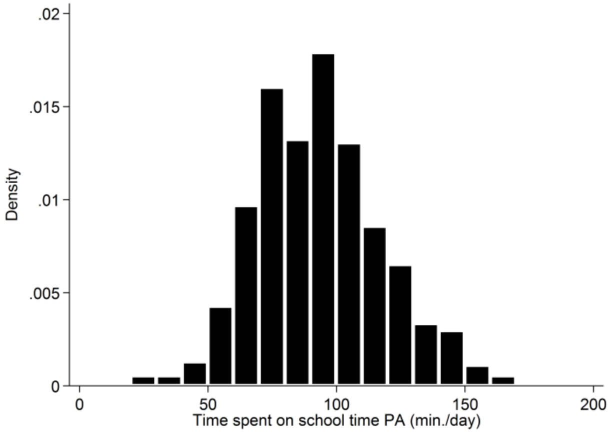


Panel C. Math

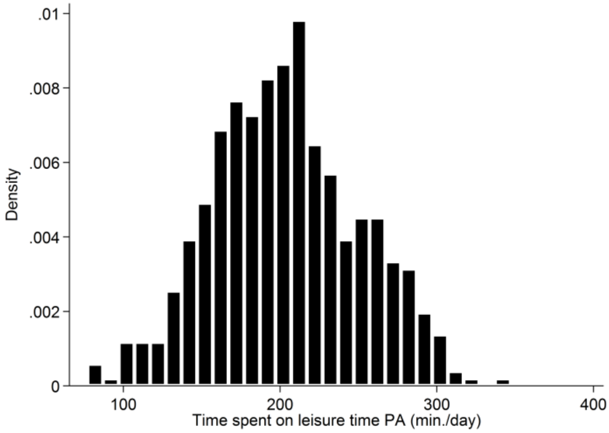


Notes: Calculations are based on the estimation sample of the analysis presented in column 3 of Table 1.3 in the main text.
Source: School performance data.

Figure A1.2. Distribution of Time Spent on Physical Activity in the Estimation Sample.
Panel A. School-time PA



Panel B. Leisure-time PA



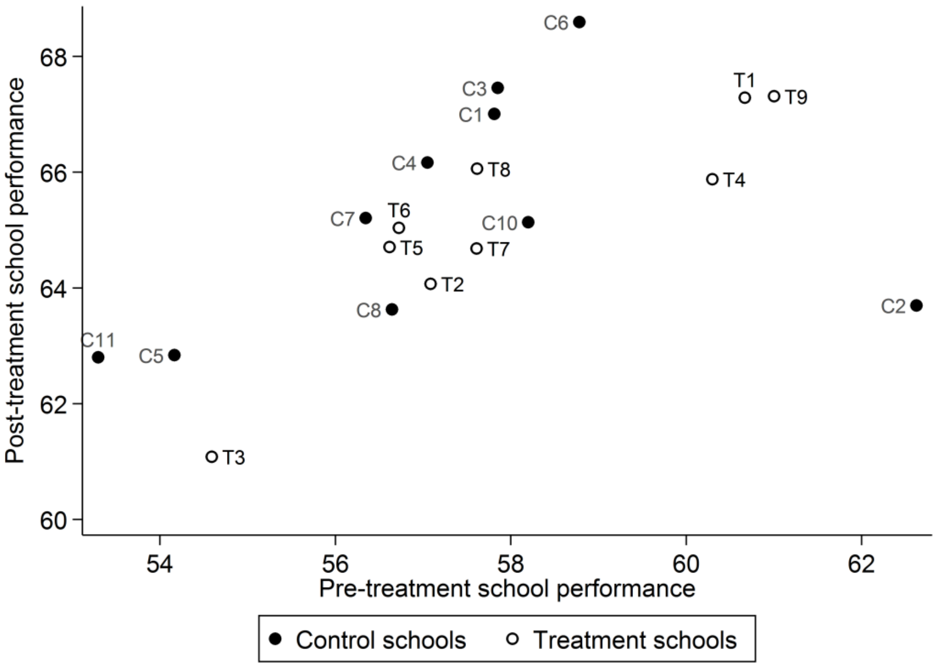
Notes: Calculations are based on the estimation sample of the analyses presented in columns 3 and 6 of Table 1.4 in the main text.
Source: PA data.

1.2. ROBUSTNESS CHECKS

1.2.1. Checking for Outliers

In our main text, we show that the pre-treatment trends in school performance appear to be parallel between the treatment and control groups. Here, we verify whether our results may be driven by an outlier school. This is of particular concern due to the relatively small number of schools included in this study and because the treatment is not uniform across the treatment schools. Figure A1.3 indicates that it is unlikely that the results are driven by any particular treatment school, i.e., any particular intervention package.

Figure A1.3. Average Pre- and Post-Treatment School Performance, by Active Living School.



Notes: The figure is based on the estimation sample of the analysis presented in Table 1.3, column 3. The numbers of the treatment schools correspond to those in Table 1.1 in the main text. The results in Tables A1.3 and A1.4 show that the difference-in-differences estimations of the effect of the Active Living Program on overall school performance remain robust when one of the Active Living schools is excluded from the analysis.
Source: School performance data.

We next run regressions excluding one different school each time (see Tables A1.3 and A1.4). Considering that the results remain robust across these regressions, we conclude that outlier treatment or control schools do not drive the estimates.

Table A1.3. Difference-in-differences estimations of the effect of the Active Living Program on school performance—one-by-one deletion of treatment schools.

Dependent variable: overall school performance	(1) Excl. T1	(2) Excl. T2	(3) Excl. T3	(4) Excl. T4	(5) Excl. T5	(6) Excl. T6	(7) Excl. T7	(8) Excl. T8	(9) Excl. T9
Treatment * Post	-1.052	-1.161	-1.075	-1.096	-1.325	-1.314	-1.111	-1.329	-1.083
	(0.543)	(0.556)	(0.562)	(0.547)	(0.561)	(0.555)	(0.553)	(0.550)	(0.550)
Adjusted p-value	[0.073]	[0.056]	[0.076]	[0.065]	[0.033]	[0.033]	[0.064]	[0.030]	[0.069]
Post (1 =2013/2014, 0 =2012/2013)	2.632	2.006	4.334	3.174	2.517	3.016	1.514	3.418	2.373
	(3.996)	(3.972)	(4.240)	(3.853)	(3.907)	(3.848)	(3.490)	(3.821)	(3.824)
Age when tests were taken (months)	1.045	1.085	0.881	0.938	0.949	0.954	0.980	0.993	0.966
	(0.416)	(0.411)	(0.452)	(0.407)	(0.448)	(0.408)	(0.404)	(0.413)	(0.401)
Age squared	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.001	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Child fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,904	1,958	1,896	1,978	1,826	1,874	1,918	1,898	1,942
R-squared	0.813	0.813	0.815	0.814	0.807	0.810	0.823	0.811	0.814
Number of children in the estimation sample	952	979	948	989	913	937	959	949	971
Mean school performance in estimation sample	61.042	61.245	61.457	61.174	61.267	61.249	61.226	61.179	61.092
Standard deviation	(8.911)	(8.921)	(8.814)	(8.898)	(8.847)	(8.888)	(8.947)	(8.863)	(8.980)
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: The table shows the results of nine ordinary least squares regressions. The dependent variable in each column is overall school performance. Tn = Treatment school (number), which is excluded from the analysis. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

Table A1.4. Difference-in-differences estimations of the effect of the Active Living Program on school performance—one-by-one deletion of control schools.

Dependent variable: overall school performance	(1) Excl. C1	(2) Excl. C2	(3) Excl. C3	(4) Excl. C4	(5) Excl. C5	(6) Excl. C6	(7) Excl. C7	(8) Excl. C8	(9) Excl. C9	(10) Excl. C10	(11) Excl. C11
Treatment * Post	-1.142	-1.294	-0.890	-1.145	-1.158	-1.044	-1.122	-1.373	-1.170	-1.515	-1.051
	(0.589)	(0.538)	(0.508)	(0.562)	(0.575)	(0.564)	(0.621)	(0.564)	(0.544)	(0.492)	(0.571)
Adjusted p-value	[0.073]	[0.030]	[0.102]	[0.061]	[0.064]	[0.085]	[0.092]	[0.029]	[0.048]	[0.008]	[0.087]
Post (1 = 2013/2014, 0 = 2012/2013)	3.024	2.298	2.197	3.324	2.859	4.497	2.571	6.034	2.950	1.782	2.987
	(3.919)	(3.806)	(4.471)	(4.077)	(3.814)	(3.773)	(3.938)	(3.234)	(3.798)	(4.168)	(3.741)
Age when tests were taken (months)	1.089	0.997	0.865	0.983	1.013	0.923	0.927	0.686	0.972	1.217	0.955
	(0.412)	(0.409)	(0.428)	(0.410)	(0.407)	(0.409)	(0.417)	(0.344)	(0.400)	(0.408)	(0.400)
Age squared	-0.002	-0.002	-0.001	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Child fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,944	2,012	1,892	1,988	1,958	1,950	1,874	1,892	2,028	1,868	1,932
R-squared	0.809	0.819	0.814	0.810	0.809	0.809	0.804	0.816	0.812	0.820	0.807
Number of children in the estimation sample	972	1,006	946	994	979	975	937	946	1,014	934	966
Mean school performance in estimation sample	61.170	61.206	61.133	61.214	61.319	61.123	61.258	61.300	61.222	61.183	61.379
Standard deviation	(8.883)	(8.914)	(8.712)	(8.901)	(8.855)	(8.895)	(8.952)	(8.918)	(8.890)	(8.982)	(8.805)
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: The table shows the results of 11 ordinary least squares regressions. The dependent variable in each column is overall school performance. Cn = Control school (number), which is excluded from the analysis. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

1.2.2. Configuration of Design Choices

As Table 1.1 in the main text indicates, there is considerable heterogeneity in the implemented interventions across treatment schools. Having such heterogeneity in implementation is, in some respects, a strength because it likely mimics a real-world implementation. However, this also makes it harder to interpret the treatment effects. Although Figure A1.3 and Table A1.3 both indicate that the treatment effects on school performance are relatively similar between treatment schools, the question remains whether the intervention packages that cause the strongest treatment effects have something in common. To answer this question, we run the difference-in-differences regressions while each time including only one treatment school (and all control schools), instead of excluding them one by one as we did in Table A1.3. This allows us to identify which configuration of design choices (i.e., intervention package) is most strongly associated with the observed negative effect on school performance. However, this strategy reduces the number of observations drastically: we go from having test scores of 1,014 children to having scores from only 500 children, which makes the interpretation of insignificant effects problematic due to a potential lack of power.

Based on the results shown in Table A1.5, we first conclude that none of the design choices lead to a positive effect on school performance. The strongest negative treatment effect is found in treatment schools 1, 4, and 9. The negative treatment effect on school performance is smallest in treatment schools 5, 6, and 8. However, the configuration of design choices is so heterogeneous across schools that it is not possible to identify a particular intervention (or combination of interventions) that may drive these results.

Table A1.5. Difference-in-differences estimations of the effect of Active Living on school performance—one-by-one inclusion of treatment schools.

Dependent variable: Overall school performance	(1) Incl. T1	(2) Incl. T2	(3) Incl. T3	(4) Incl. T4	(5) Incl. T5	(6) Incl. T6	(7) Incl. T7	(8) Incl. T8	(9) Incl. T9
Treatment * Post	-2.167	-1.347	-1.591	-2.612	-0.507	-0.321	-1.647	-0.085	-2.194
	(0.529)	(0.527)	(0.486)	(0.423)	(0.498)	(0.513)	(0.536)	(0.469)	(0.466)
Adjusted p-value	[0.005]	[0.038]	[0.014]	[0.000]	[0.343]	[0.551]	[0.018]	[0.862]	[0.002]
Post (1 = 2013/2014, 0 = 2012/2013)	-0.161	1.937	-0.107	0.267	1.402	0.622	2.264	-0.065	0.525
	(4.609)	(4.616)	(4.299)	(4.737)	(4.574)	(4.670)	(5.319)	(4.624)	(4.760)
Age when tests were taken (months)	1.028	0.872	1.069	1.038	1.049	1.049	1.094	1.031	1.058
	(0.603)	(0.627)	(0.575)	(0.649)	(0.548)	(0.630)	(0.650)	(0.614)	(0.649)
Age squared	-0.001	-0.001	-0.001	-0.001	-0.002	-0.001	-0.002	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Child fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,094	1,040	1,102	1,020	1,172	1,124	1,080	1,100	1,056
R-squared	0.829	0.827	0.825	0.827	0.836	0.832	0.811	0.831	0.826
Number of children in the estimation sample	547	520	551	510	586	562	540	550	528
Mean school performance in estimation sample	61.412	61.050	60.695	61.182	61.037	61.056	61.090	61.173	61.334
Standard deviation	8.978	8.967	9.126	9.014	9.075	9.017	8.919	9.064	8.857
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: The table shows the results of nine ordinary least squares regressions. The dependent variable in each column is overall school performance. Tn = Treatment school (number), which is included in the analysis. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

1.2.3. Comparisons to Non-Active Living Schools

We cannot completely rule out the possibility that control schools changed certain attitudes or policies toward PA. However, we can analyze whether there were changes in school performance in control schools relative to all other schools in our data that did not participate in the Active Living Program as a treatment or control group (i.e., the non-Active Living schools). By doing so, we can verify whether the negative treatment effects we find are indeed due to a decrease in school performance in the treatment group, and not due to an (unexpected) increase in school performance in the control group.

When we compare the school performance of children in the treatment group to those of children in non-Active Living schools, the difference-in-differences model we estimate is similar to equation (1.1) in the main text, with the exception that the *Treat* variable now has the value of zero for children in the non-Active Living schools and is set to "missing" for children in the control schools.

When we compare the children in the control schools to those in the non-Active Living schools, the difference-in-differences model we estimate is as follows:

$$Score_{ist} = \alpha_{11} + \alpha_{12}(Control_{is} * Post_t) + \alpha_{13}Post_t + \alpha_{14}age_{ist} + \alpha_{15}age^2{}_{ist} + \gamma_{4is} + \varepsilon_{4ist} \quad \text{(A1.1)}$$

Control is an indicator variable with a value of one for children in control schools and zero for children in non-Active Living schools. The other variables are identical to those in equation (1.1) in the main text. Although pre-treatment differences in school performance between control schools and non-Active Living schools exist because all Active Living schools are located in relatively deprived areas (as shown in Table A1.2 in Appendix 1.1), these differences should not change significantly over time. In other words, α_{12} should not be significantly different from zero, for otherwise it indicates that the control group may (unexpectedly) be affected by the Active Living Program, which would complicate the interpretation of any effects of the Active Living Program we find. However, if we indeed cannot reject the hypothesis that the difference-in-differences estimator α_{12} is equal to zero, we have strong reasons to believe that the effects of the Active Living Program are the result of the Active Living interventions in the treatment schools. This conclusion becomes even stronger if the point estimate of the effect of the Active Living Program is similar to our main result (i.e., negative) when we compare the treatment group to the non-Active Living schools.

The results in Table A1.6 show that the school performance of children in the control group slightly increases due to the Active Living Program (0.07 standard deviations), but as expected, this effect is not significantly different from zero. Therefore, even if control schools changed their attitudes or policies towards PA, this does not appear to have had a significant effect on school performance, either in statistical or economic terms. It is therefore unlikely that unintentional and unobserved changes within control schools bias our estimates of the treatment effect on school performance.

The effect of the Active Living Program on the treatment group's school performance remains negative and significantly different from zero at the 10 percent level in the main specification, although it is now smaller in size (-0.06 standard deviations). This smaller effect may be explained by the positive, though insignificant, "treatment" effect on the control group's school performance. Although caution is needed when interpreting results that are not statistically significant, this suggests that the estimated effects as presented in Table 1.3 (main text) might be slightly biased upward, and that the more conservative estimate of the average effect of the Active Living Program on school performance is 5.9 percent of a standard deviation.[51]

[51] We also ran a regression of school performance on participating in the Active Living Program as a treatment school in which we control for baseline school performance instead of including the *Treat*Post* interaction.

Table A1.6. Difference-in-differences estimations of the effect of the Active Living Program on school performance—Children in Active Living schools compared to children in non-Active Living schools.

Dependent variable: overall school performance	(1)	(2)	(3)	(4)	(5)	(6)
	Control group			Treatment group		
	OLS	FE	FE	OLS	FE	FE
Control * Post	0.342	0.765	0.664			
	(0.501)	(0.453)	(0.466)			
Adjusted p-value	[0.496]	[0.094]	[0.157]			
Treatment * Post				-0.090	-0.325	-0.546
				(0.880)	(0.334)	(0.301)
Adjusted p-value				[0.919]	[0.331]	[0.072]
Control (1 = Control group, 0 = non–Active Living)	-1.056					
	(0.606)					
Treatment (1 = Treatment group, 0 = non–Active Living)				-0.246		
				(0.715)		
Post (1 = 2013/2014, 0 = 2012/2013)	8.247	7.683	-1.351	8.247	7.683	-1.080
	(0.199)	(0.133)	(1.465)	(0.199)	(0.133)	(1.487)
Age when tests were taken (months)			1.566			1.571
			(0.190)			(0.185)
Age squared			-0.003			-0.003
			(0.001)			(0.001)
Child fixed effects	No	Yes	Yes	No	Yes	Yes
Constant	57.916			57.916		
	(0.313)			(0.313)		
Observations	11,361	10,350	10,348	11,480	10,438	10,436
R-squared	0.197	0.788	0.801	0.195	0.784	0.797
Number of children in the estimation sample	6,186	5,175	5,174	6,261	5,219	5,218
Mean school performance in estimation sample	62.311	61.665	61.667	62.367	61.687	61.689
Standard deviation	(9.313)	(9.173)	(9.173)	(9.309)	(9.143)	(9.143)
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: The table shows the results of six ordinary least squares regressions. The dependent variable in each column is overall school performance. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

The estimated treatment effect using this specification is -0.06 standard deviations, significantly different from zero at the 10 percent level.

1.3. EXTERNAL VALIDITY

Although the internal validity of our findings appears to be strong, it is uncertain to what extent our findings are externally valid (generalizable) beyond the setting of this study, i.e., 8- to 12-year-olds living in low-SES regions in the Netherlands. Does encouraging informal PA also have a negative effect on school performance (1) in high-SES regions in the Netherlands, (2) in low- or high-SES regions in other countries, and (3) among children in different age categories? Future research should address these questions. In Chapter 1, we can only provide suggestive evidence.

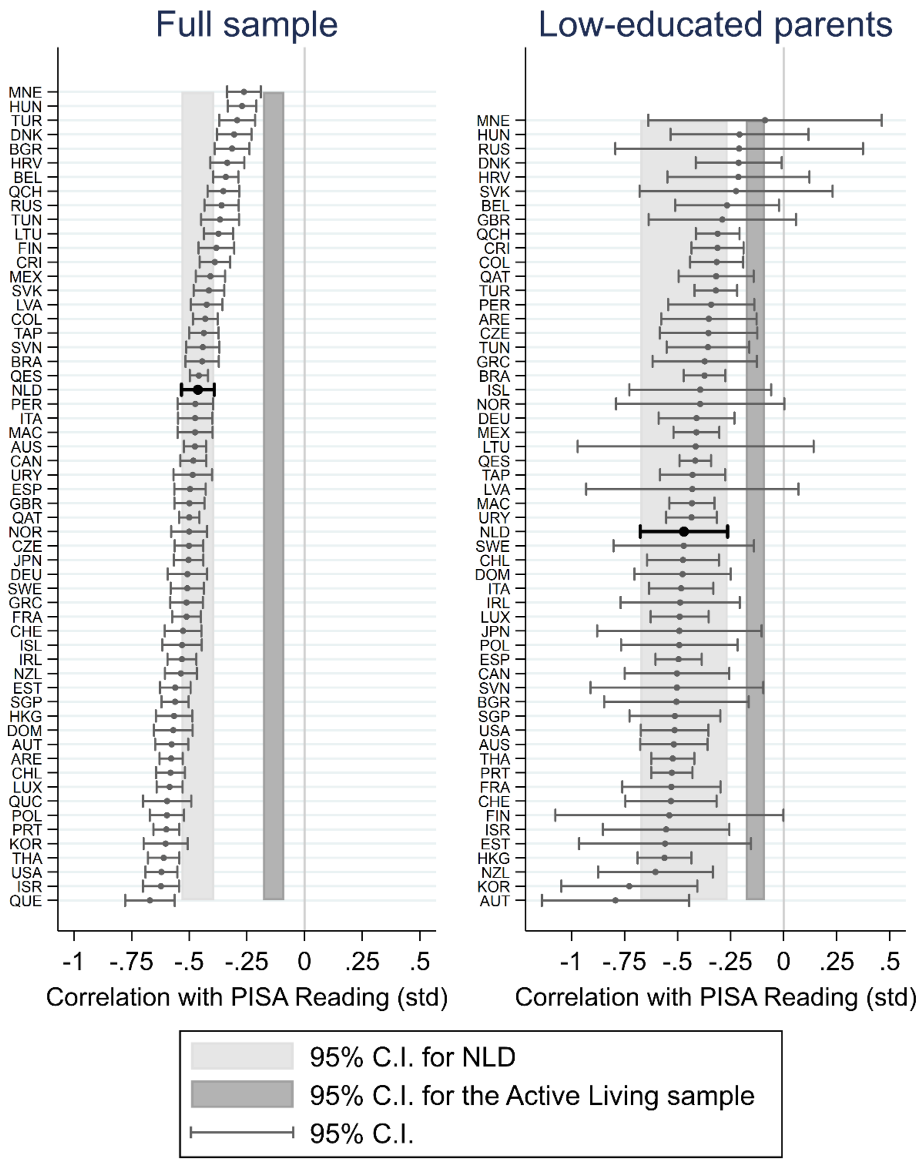
For instance, data from the 2015 OECD Programme for International Student Assessment (PISA) show that, within 60 countries and economies across the world, exercising or practicing a sport before going to school is negatively correlated with 15-year-olds' student performance (see Figure C1). Although the PA is not necessarily informal in nature, the time at which it takes place is comparable to the time of day at which the Active Living Program has the strongest positive effect on PA (i.e., in the mornings of schooldays). Moreover, the correlation is negative in all countries, and it is even more negative in the full sample than from within the sample of children from low-educated parents.[52]

Evidently, these analyses are not causal in nature. However, the similarity between our "naïve" pre-treatment OLS estimates (see Table 1.2 in the main text) and the Wald estimator (see the section on Instrumental Variables Estimation in the main text) indicate that the OLS might be biased toward zero. Figure A1.4 therefore suggests that we may expect that increasing PA in the mornings of school days can also have negative effects on school performance in other high-income countries such as the United States, Sweden, or Japan, as well as in middle-income countries such as Brazil, Colombia, or Tunisia. Additionally, it suggests that the results are generalizable outside of low-SES regions (within and outside of the Netherlands), and among children in different age categories (i.e., 8- to 12-year-olds as well as 15-year-olds).

However, it remains unlikely that we can generalize our findings to Physical Education classes, high school athletics, or PA after school hours including participation in sports clubs. It is possible that the difference in the effect that formal and informal PA have on school performance is related to the potential mechanism we find. Children may find it harder to remain calm and quiet shortly after their activities than they do shortly before potential activities. More research is needed to identify the mechanisms for the effects of different types of PA on school performance.

[52] The results for children of low-educated parents are similar to the results for children in the bottom quartile of the SES indicator as derived by the OECD. Moreover, the results are qualitatively robust for controlling for gender and age, and for different measures of student performance (i.e., PISA Mathematics and PISA Science).

Figure A1.4. Correlation Between Physical Activity Before Going to School and Reading Performance.



Notes: The figure shows the within-country correlation between students' answer to the question "On the most recent day you attended school, did you do any of the following before going to school? Exercise or practice a sport" (ST076Q11NA) and PISA Reading (standardized within the country-level estimation samples). "95 percent C.I. for the Active Living sample" refers to the pre-treatment correlation between spending 30 minutes on physical activity during school time and performance on language tests. Low-educated parents are those whose highest obtained degree (combined) is at most ISCED level 2 (lower secondary education/second stage of basic education).
Source: Authors' calculations based on PISA 2015.

1.4. ADDITIONAL TABLES AND FIGURES

Textbox A1.1. Additional information on the selection on treatment and control schools.

> First, municipal support for participation was essential because costs for environmental adaptations in public spaces (i.e., school surroundings) had to be covered by municipalities. In November 2011, municipal development plans of all 19 Southern-Limburg municipalities were checked to see if they contained formal references to the themes of "youth" and "overweight prevention." Municipalities were excluded from the list of potential treatment areas if they did not state these topics in their development plans because this implied that no budget was reserved for environmental adaptations for these purposes. In total, 12 out of 19 municipalities had formulated targets either for youth or overweight prevention or both. Municipal health officers in these 12 municipalities were visited and informed about the project and conditions for participation. This resulted in six municipalities wanting to become involved as an intervention area. In the six municipalities that agreed to participate, four school corporations were identified, three of which agreed to recruit schools falling under their responsibility. To participate in the Active Living project as a treatment school, schools were checked for eligibility according to predefined inclusion criteria: (1) located in a deprived area; (2) at least 25 students enrolled in grades 4 and 5; (3) no plans to merge with another school or plans to relocate in the upcoming 3 years; and (4) willing to actively participate and to form an "Active Living" working group at school. Within the municipalities and school corporations that consented, 37 primary schools were identified in deprived areas. Municipalities and corporations were asked to recommend schools that were most eligible to participate from their perspective. We visited 13 eligible schools and informed them about the project and conditions for participation; 10 schools (76.9 percent) agreed to participate as a treatment school (see Figure A2.6).
>
> After recruiting treatment schools, we defined the potential control areas as the municipalities that were excluded in the process of treatment school selection, i.e. the seven municipalities without development plans including references to the themes of youth and/or overweight prevention, and the six municipalities that were not interested in participating as a treatment area or not able to cover the expenditures of environmental adaptations. From these municipalities, control schools were chosen based on matching on the level of SES and degree of urbanization. Control schools were offered Public Health Services (PHS) support to implement effective elements of the project after the end of the effectiveness study.

Note: This textbox is based on Van Kann et al. (2015).

Picture A1.1. Pictures of Implemented Interventions.

Panel A. Traffic circle in schoolyard environment

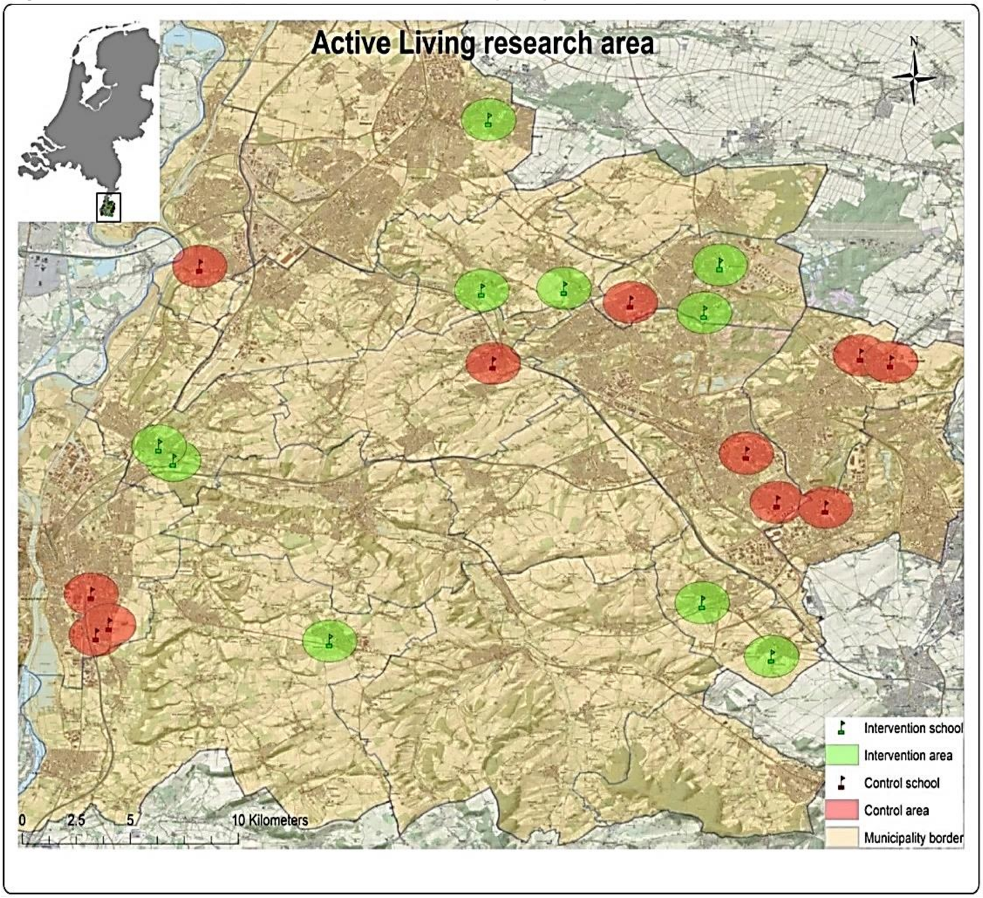
Panel B. Establish training circuit (before and after intervention)

Picture A1.1 Continued. Pictures of Implemented Interventions.

Panel C. New ball backstop besides railway and ball game area (before and after intervention)

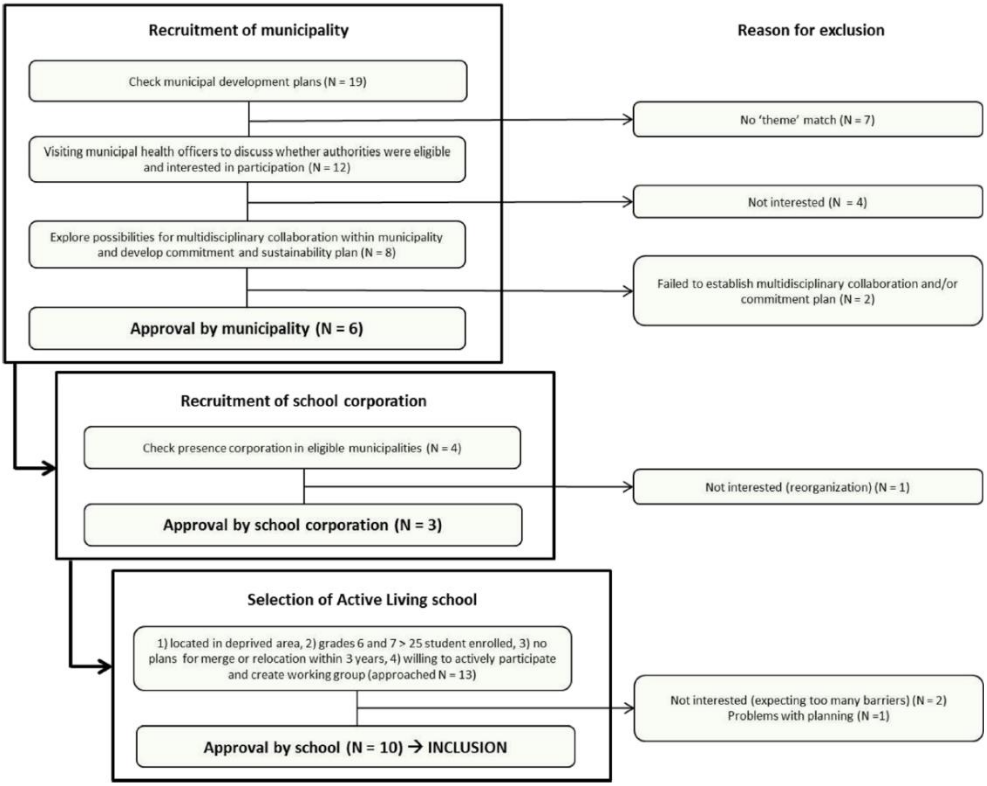
Source: Pictures taken by the research team.

Figure A1.5. Research Area—Southern-Limburg Region, The Netherlands.



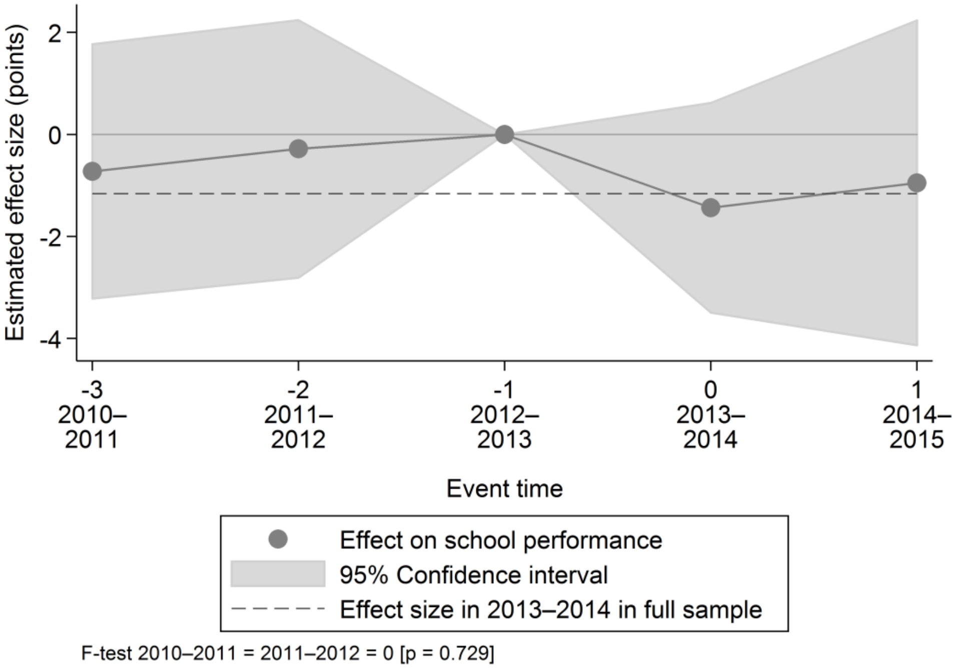
Source: Van Kann et al. (2015).

Figure A1.6. Recruitment of Treatment Schools.



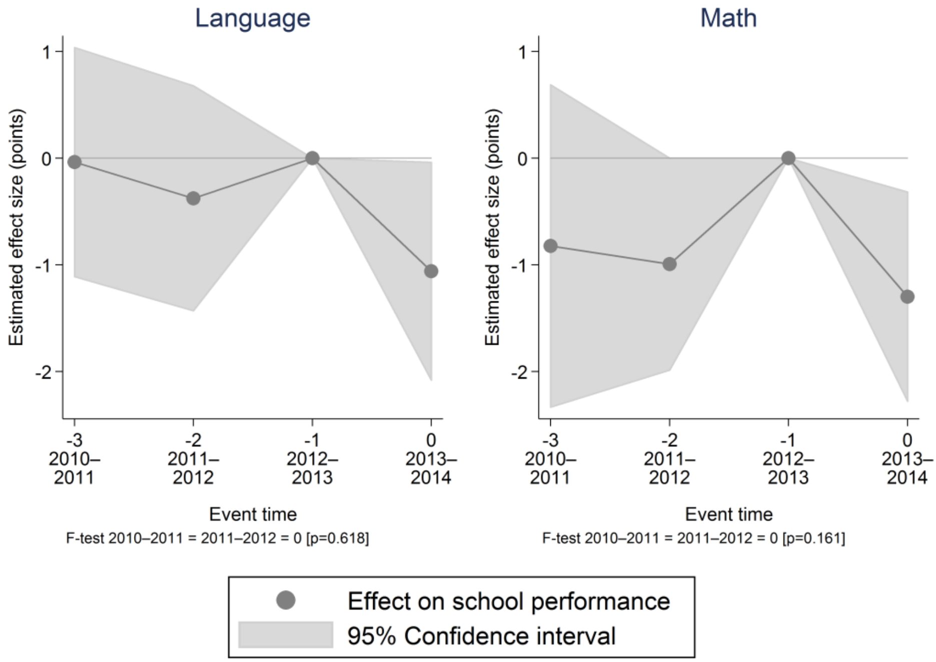
Source: Van Kann et al. (2015).

Figure A1.7. The Long-Run Effects of the Active Living Program on School Performance.



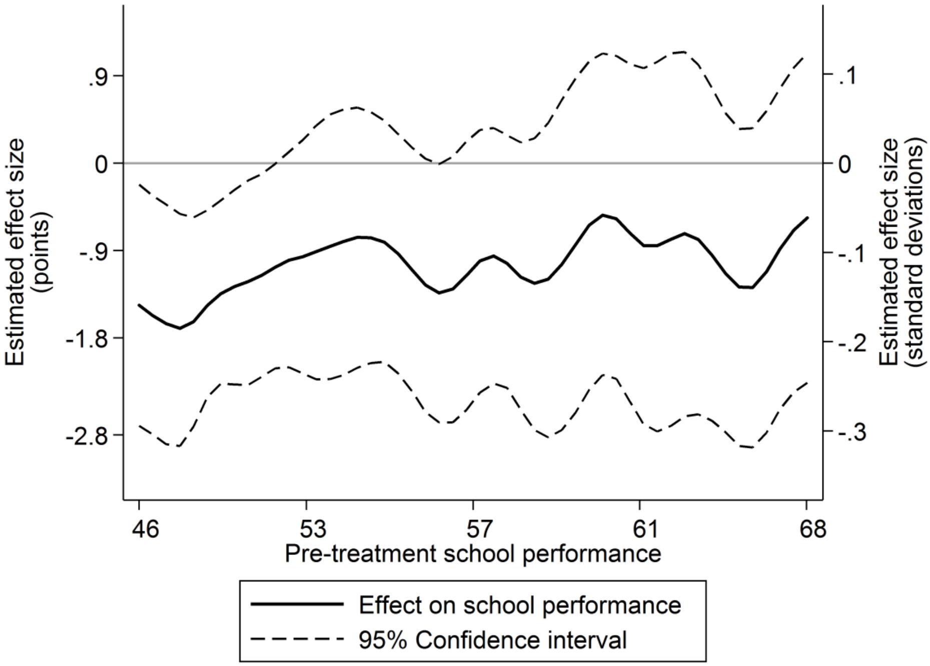
Notes: The figure shows the results of an ordinary least squares regression including only children who were enrolled in grade 4 in the pre-treatment year 2012/2013. The dependent variable is overall school performance. The estimated effect sizes based on a difference-in-differences estimation are plotted on the y-axis. Independent variables are year fixed effects, age, age squared, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: School performance data.

Figure A1.8. Event Study Analyses of the Effect of the Active Living Program on Language and Math Tests.



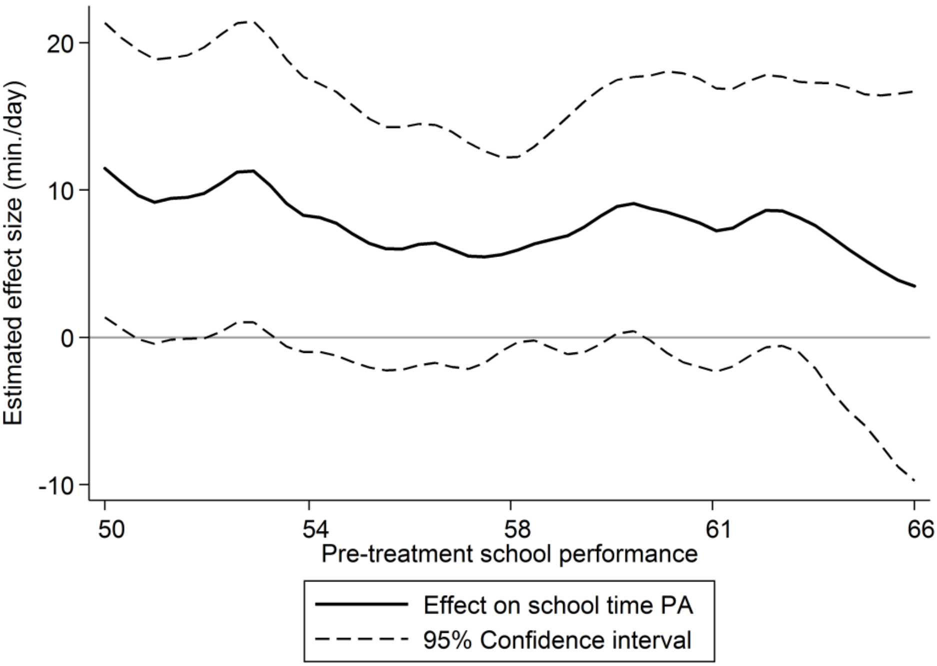
Notes: The figure shows the results of two ordinary least squares regressions. The dependent variable in the left panel is language performance, and the dependent variable in the right panel is math performance. The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axes. Independent variables are year fixed effects, age, age squared, and child fixed effects. Adjusted 90 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The figure depicts the 90 percent confidence interval, because the results in Table A1.9 show that the treatment effects in the treatment year are significant at the 90 percent level.
Source: School performance data.

Figure A1.9. The Effect of the Active Living Program on School Performance, by Pre-Treatment School Performance.



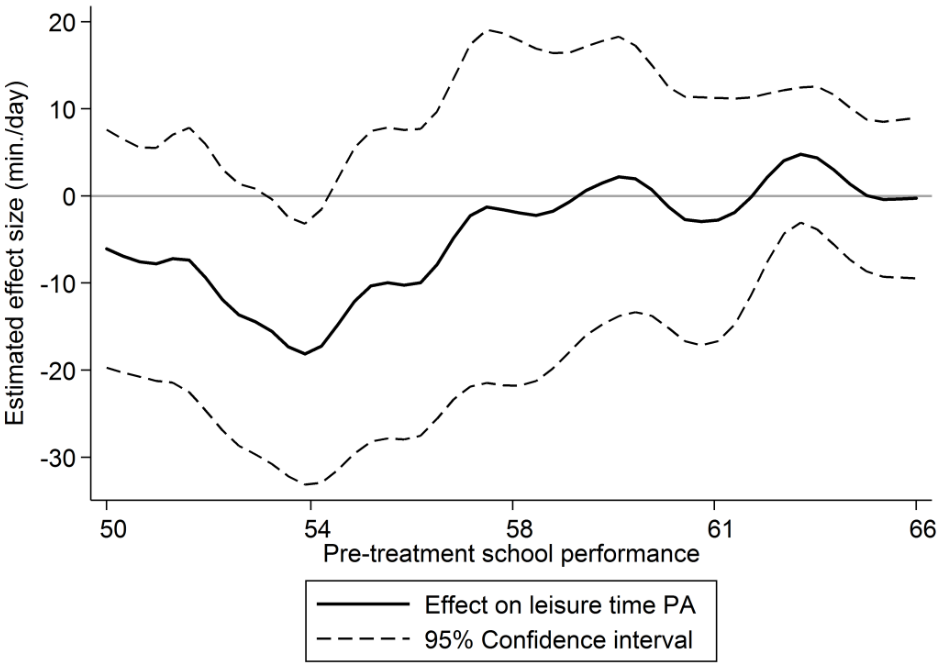
Notes: The figure shows the smoothed local polynomial graph of the results of ordinary least squares regressions across the pre-treatment school performance distribution in 2012/2013. Pre-treatment school performance is a moving window of 200 children. The dependent variable in each regression is overall school performance. The estimated effect size based on difference-in-differences estimations is plotted on the y-axis. Independent variables are a year dummy, age, age squared, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The results in the right tail of the distribution should be interpreted with caution because of the differences in pre-treatment trends in school performance in this part of the distribution (see Figure 1.4 in the main text).
Source: School performance data.

Figure A1.10. The Effect of the Active Living Program on Time Spent on Physical Activity During School Time, by Pre-Treatment School Performance.



Notes: The figure shows the smoothed local polynomial graph of the results of ordinary least squares regressions across the pre-treatment school performance distribution in 2012/2013. Pre-treatment school performance is a moving window of 200 children. The dependent variable in each regression is time spent on physical activity during school-time. The estimated effect size based on difference-in-differences estimations is plotted on the y-axis. Independent variables are a year dummy, wearing time of the accelerometer, wearing time squared, weather variables, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The results in the right tail of the distribution should be interpreted with caution because of the differences in pre-treatment trends in school performance in this part of the distribution (see Figure A1.8, panel B).
Source: Merged school performance and PA data.

Figure A1.11. The Effect of the Active Living Program on Time Spent on Physical Activity During Leisure Time, by Pre-Treatment School Performance.



Notes: The figure shows the smoothed local polynomial graph of the results of ordinary least squares regressions across the pre-treatment overall school performance distribution. Pre-treatment school performance is a moving window of 200 children. The dependent variable in each regression is time spent on physical activity during leisure time. The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axes. Independent variables are a year dummy, wearing time of the accelerometer, wearing time squared, weather variables, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The results in the right tail of the distribution should be interpreted with caution because of the differences in pre-treatment trends in school performance in this part of the distribution (see Figure A1.8, panel B).
Source: Merged school performance and PA data.

Table A1.7. Interventions in the treatment schools.

Treatment school 1		
Active transportation to school	• School pedestrian crossing indicators	• Put up signs at busy or unsafe roads in the school environments to indicate to children where they can cross the road safely on foot. These signs also signal to car drivers that they should pay extra attention because children may cross the road
	• Create traffic circle in schoolyard environment	• Area in the schoolyard where children can learn how to safely participate in different traffic environments, including one-way streets, roundabouts, zebra crossings, and traffic signs (see Picture D1, panel A)
	• Sticker competition for active transport	• Provide stickers to children who traveled to school on foot or by bike
PA in school	• Establish ball game area	• Markings as indicator of a ball games area and regulations where to play ball games in the schoolyard
	• Put a ball backstop beside railway	• Re-facilitate the use of balls at the schoolyard by placing tall fences (see Picture D1, panel C) that prevent balls from falling on the railroad tracks next to the schoolyard
	• Additional sports day in schoolyard	• Introduce one additional sports day per school year
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
	• Prize contest for best idea to encourage PA	• Children asked to submit ideas about what they think could encourage them and/or their classmates to become more physically active
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	• Establish out-of-school PA program	• Multiple sports activities provided directly after school, e.g., gymnastics, soccer

Treatment school 2		
Active transportation to school	• Availability of bicycle racks	• Provide bicycle racks.
	• School pedestrian crossing indicators	• Put up signs at busy or unsafe roads in the school environments to indicate to children where they can cross the road safely on foot. These signs also signal to car drivers that they should pay extra attention because children may cross the road

	• Create traffic circle in schoolyard environment	• Area in the schoolyard where children can learn how to safely navigate different traffic patterns, including one-way streets, roundabouts, zebra crossings, and traffic signs. (see Picture D1, panel A)
	• Sticker competition for active transport	• Provide stickers to children who traveled to school on foot or by bike
PA in school	• New fixed equipment in schoolyard	• Place new soccer goals
	• New loose equipment in schoolyard	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment.
	• Playground markings	• Paintings, hopscotch markings, twister field
	• Establish ball game area	• Markings as indicator of a ball games area on the schoolyard
	• Sound equipment in schoolyard environment	• Install music equipment in the schoolyard to facilitate dancing
	• Additional sports day in schoolyard	• Introduce one additional sports day per school year
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
	• Use of schoolyard games	• Media cards that provide information to children on how to play games
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	• Establish out-of-school PA program	• Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer

Treatment school 3		
Active transportation to school	• Development safe route to school	• A marked route between the school and the closest public playground that avoids dangerous crossings and provides a guided crossing where the crossing of a dangerous road could not be avoided
	• Mobilize crossing guards	• Approximately 15 minutes before and after school hours, crossing guards help children cross the road at the nearest (busy) road in the school environment
	• Adapt unsafe intersection in school environment	• Establish a new priority situation at an unsafe intersection close to the school
	• Create safer parking situation around school	• Redesign parking lots into parallel parking lots

	• Sticker competition for active transport • School stimulation documentation on safe active transport • Speed check action performed by children	• Provide stickers to children who traveled to school on foot or by bike • Develop an active school transportation policy • Children check the speed of car users in their school environment with the help of the local police. They also provide feedback to the drivers about their driving behavior
PA in school	• New loose equipment in schoolyard • Additional sports day in schoolyard • Sports clinics in recess • Use of schoolyard games	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment • Introduce two additional sports days per school year • Local sports clubs provide trial lessons and information during recess • Media cards that provide information to children on how to play games
PA in leisure-time	• Establish training circuit • Active Living Games • Establish out-of-school PA program • Establish school soccer team • Establish PA activities by children for local residents	• A "green" track next to the schoolyard (see Picture D1, panel B) • An additional sports day for all participating treatment schools (funded by the Province of Limburg) • Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer • Establishment of a school team that regularly plays soccer after school, available for all interested children, irrespective of their gender • Establishment of "Neighborhood in Action" group that encourages physical activity by performing family (physical) activities, such as walks through parks

Treatment school 4		
Active transportation to school	• Develop safe route to school • Mobilize crossing guards	• A marked route between the school and the closest public playground that avoids dangerous crossings and provides guided crossings where the crossing of a dangerous road could not be avoided • Approximately 15 minutes before and after school hours, crossing guards help children cross the road at the nearest (busy) road in the school environment

	• Sticker competition for active transport	• Provide stickers to children who traveled to school on foot or by bike
	• School stimulation documentation on safe active transport	• Develop an active school transportation policy
	• Speed check action performed by children	• Children check the speed of car users in their school environment with the help of the local police. They also provide feedback to the drivers about their driving behavior
PA in school	• New fixed equipment in schoolyard • Additional sports day in schoolyard • Sports clinics in recess	• Place a climbing structure in the schoolyard • Introduce two additional sports days per school year • Local sports clubs provide trial lessons and information during recess
PA in leisure-time	• Active Living Games • Establish out-of-school PA program	• An additional sports day for all participating treatment schools (funded by the Province of Limburg) • Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer

Treatment school 5		
Active transportation to school	• Develop safe route to school • Lessons to improve bicycle skills	• A marked route between the school and the closest public playground that avoids dangerous crossings and provides guided crossings where the crossing of a dangerous road could not be avoided • Several skill lessons on how to safely ride a bike
PA in school	• New fixed equipment in schoolyard • New loose equipment in schoolyard • Sports clinics in recess	• Place new soccer goals in the schoolyard • Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment • Local sports clubs provide trial lessons and information during recess
PA in leisure-time	• Active Living Games • Establish out-of-school PA program	• An additional sports day for all participating treatment schools (funded by the Province of Limburg) • Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer

Treatment school 6		
Active transportation	• Availability of bicycle racks	• Provide bicycle racks in front of the school

to school	• School pedestrian crossing indicators • Introduce "Walk/Bike to school day"	• Put up signs at busy or unsafe roads in the school environments to indicate to children where they can cross the road safely on foot. These signs also signal to car drivers that they should pay extra attention because children may cross the road • Mark one day in the week or month as active transportation day
PA in school	• New fixed equipment in schoolyard • New loose equipment in schoolyard • Playground markings • Establish ball game area • Sports clinics in recess	• Place a table tennis court in the schoolyard • Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment • Paint hopscotch markings on the schoolyard • Introduce regulations regarding where to play ball games in the schoolyard • Local sports clubs provide trial lessons and information during recess
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)

Treatment school 7		
Active transportation to school	• Sticker competition for active transport	• Provide stickers to children who traveled to school on foot or by bike
PA in school	• New loose equipment in schoolyard • Playground markings • Use of schoolyard games	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment • Paintings, hopscotch markings • Media cards that provide information to children on how to play games
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)

Treatment school 8		
Active transportation to school	-	-
PA in school	• New fixed equipment in schoolyard • Establish ball game area	• Place soccer goals, a climbing structure, and a table tennis court in the schoolyard • Introduce regulations regarding where to play ball games in the schoolyard

	• Additional sports day in schoolyard	• Introduce one additional sports day per school year
	• Use of schoolyard games	• Media cards that provide information to children on how to play games
	• Prize contest for best idea to encourage PA	• Children asked to submit ideas about what they think could stimulate them and/or their classmates to become more physically active
PA in leisure-time	• Establish out-of-school PA program	• Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer

Treatment school 9		
Active transportation to school	• Develop safe route to school	• A marked route between the school and the closest public playground that avoids dangerous crossings and provides guided crossings where the crossing of a dangerous road could not be avoided
	• Lessons to improve bicycle skills	• Several skill lessons on how to safely ride a bike
PA in school	• New fixed equipment in schoolyard	• Place a climbing structure in the schoolyard
	• New loose equipment in schoolyard	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	• Establish out-of-school PA program	• Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer

Table A1.8. Differences between children in treatment and control schools in the pre-treatment year (2012/2013).

	Treatment		Control		Diff.	P-value
	mean	sd	mean	sd		
Panel A. School performance data						
Number of tests taken	4.92	(0.69)	5.09	(1.12)	-0.17	0.003
Overall school performance	57.67	(7.73)	56.86	(8.06)	0.81	0.103
Language performance	57.31	(7.77)	56.63	(8.02)	0.68	0.169
Math performance	56.49	(7.72)	55.56	(8.27)	0.93	0.065
Age when tests were taken (months)	125.51	(8.05)	126.75	(8.63)	-1.24	0.018
Gender (1 = boy)	0.48	(0.50)	0.47	(0.50)	0.00	0.946
Cohort (1 = 5th graders, 0 = 4th graders)	0.55	(0.50)	0.60	(0.49)	-0.05	0.095
Panel B. PA data						
Number of days the accelerometer worn	3.78	(1.32)	4.07	(1.27)	-0.28	0.000
Time spent on PA during school time (min./day)	85.29	(25.14)	100.28	(25.44)	-14.99	0.000
Time spent on PA during leisure time (min./day)	201.92	(52.33)	200.51	(47.24)	1.41	0.712
Age during PA measurement (months)	128.14	(8.21)	128.47	(8.27)	-0.33	0.468
Gender (1 = boy)	0.49	(0.50)	0.41	(0.49)	0.08	0.002
Cohort (1 = 5th graders, 0 = 4th graders)	0.47	(0.50)	0.50	(0.50)	-0.03	0.266
Wearing time of accelerometer (min./day)	419.87	(137.63)	429.56	(128.12)	-9.69	0.176

Notes: The table shows results from *t*-tests on the equality of means.
Sources: School performance data and PA data.

Table A1.9. Difference-in-differences estimations of the effect of the Active Living Program on school performance on language and math tests.

Dependent variable: school performance on language or math tests	(1)	(2)	(3)	(4)	(5)	(6)
		Language			Math	
	OLS	FE	FE	OLS	FE	FE
Treatment * Post	-1.231	-1.004	-1.072	-1.191	-1.216	-1.309
	(0.595)	(0.575)	(0.586)	(0.568)	(0.549)	(0.511)
Adjusted p-value	[0.055]	[0.099]	[0.087]	[0.052]	[0.041]	[0.022]
Treatment (1 = Treatment group)	0.682			0.929		
	(0.821)			(0.859)		
Post (1 = 2013/2014, 0 = 2012/2013)	7.979	7.832	3.892	8.409	8.444	-0.448
	(0.479)	(0.459)	(3.844)	(0.480)	(0.450)	(3.442)
Age when tests were taken (months)			0.803			1.250
			(0.418)			(0.285)
Age squared			-0.002			-0.002
			(0.001)			(0.001)
Child fixed effects	No	Yes	Yes	No	Yes	Yes
Constant	56.630			55.561		
	(0.574)			(0.455)		
Observations	2,064	2,028	2,028	2,057	2,016	2,016
R-squared	0.175	0.751	0.754	0.200	0.783	0.789
Number of children in the estimation sample	1,050	1,014	1,014	1,049	1,008	1,008
Mean school performance in estimation sample	60.722	60.640	60.640	60.015	59.938	59.938
Standard deviation	(8.814)	(8.784)	(8.784)	(8.731)	(8.724)	(8.724)
Mean school performance in full sample	61.623	61.623	61.623	60.964	60.964	60.964
Standard deviation	(9.248)	(9.248)	(9.248)	(8.882)	(8.882)	(8.882)

Notes: The table shows the results of six ordinary least squares regressions. The dependent variable in columns (1) through (3) is school performance on language tests. The dependent variable in columns (4) through (6) is school performance on math tests. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

Table A1.10. Difference-in-differences estimations of the effect of the Active Living Program on time spent on PA by boys.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	School time			Leisure time			Total		
	OLS	FE	FE	OLS	FE	FE	OLS	FE	FE
Treatment * Post	29.616	26.113	6.714	6.007	0.893	-2.754	31.169	26.218	6.197
	(10.846)	(10.409)	(6.468)	(10.645)	(9.697)	(7.050)	(12.324)	(11.263)	(7.260)
Adjusted p-value	[0.013]	[0.021]	[0.312]	[0.579]	[0.928]	[0.700]		[0.031]	[0.404]
Post (1=2013/2014, 0=2012/2013)	-15.139	-12.627	-13.780	-18.183	-19.210	-28.711	-28.571	-27.703	-40.127
	(6.794)	(6.554)	(4.392)	(6.850)	(6.948)	(11.305)	(8.052)	(7.988)	(9.621)
Wearing time of accelerometer (min./day)			1.215			0.532			0.269
			(0.443)			(0.348)			(0.558)
Wearing time squared			-0.001			-0.000			0.000
			(0.001)			(0.000)			(0.000)
Outside temperature during PA measurement week (Celsius*10)			-0.308			0.242			0.033
			(0.161)			(0.314)			(0.242)
Temperature squared			0.001			-0.000			0.000
			(0.001)			(0.001)			(0.001)
Amount of sunshine during PA measurement week (hours/day)			11.269			11.835			16.474
			(3.310)			(9.429)			(8.883)
Sunshine squared			-1.001			-0.783			-1.081
			(0.440)			(1.059)			(0.974)
Rain during PA measurement week (1=yes)			1.919			-0.461			2.516
			(3.942)			(10.124)			(9.177)
Treatment (1=Treatment group)	-16.127			-8.902			-19.611		
	(5.135)			(8.187)			(8.059)		
Child-fixed-effects	No	Yes	Yes	No	Yes	Yes		Yes	Yes

Table A1.10 Continued. Difference-in-differences estimations of the effect of the Active Living Program on time spent on PA by boys.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	School time			Leisure time			Total		
	OLS	FE	FE	OLS	FE	FE	OLS	FE	FE
Constant	104.548			203.381			268.632		
	(3.766)			(5.330)			(4.064)		
Observations	612	538	466	580	494	434	612	538	464
R-squared	0.066	0.118	0.566	0.024	0.094	0.385	0.028	0.074	0.359
Number of children in the estimation sample		269	233		247	217		269	232
Mean time spent on PA in estimation sample	96.155	97.321	99.438	191.201	193.196	193.818		255.135	257.874
Standard deviation	28.977	28.419	26.210	52.440	51.932	50.740		56.043	54.039
Mean time spent on PA in full sample	96.155	96.155	96.155	191.201	191.201	191.201		252.192	252.192
Standard deviation in full sample	28.977	28.977	28.977	52.440	52.440	52.440		57.602	57.602

Notes: The table shows the results of nine ordinary least squares regressions. The dependent variable in columns (1) through (3) is time spent on PA during school time (min./day). The dependent variable in columns (4) through (6) is time spent on PA during leisure time (min./day). The dependent variable in columns (7) through (9) is total time spent on PA (min./day). Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: PA data.

Table A1.11. Difference-in-differences estimations of the effect of the Active Living Program on time spent on PA by girls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	School time			Leisure time			Total		
	OLS	FE	FE	OLS	FE	FE	OLS	FE	FE
Treatment * Post	21.781	19.283	9.974	-10.201	-13.159	-5.898	13.242	9.407	-2.466
	(6.899)	(7.560)	(2.139)	(8.193)	(7.705)	(4.601)	(7.978)	(7.527)	(4.606)
Adjusted p-value	[0.005]	[0.020]	[0.000]	[0.228]	[0.104]	[0.215]		[0.227]	[0.599]
Post (1=2013/2014, 0=2012/2013)	-15.808	-13.986	-12.467	-18.954	-18.779	-19.626	-34.553	-32.149	-27.146
	(5.430)	(6.151)	(3.111)	(5.530)	(4.391)	(3.317)	(6.623)	(5.978)	(3.766)
Wearing time of accelerometer (min./day)			0.438			0.338			0.960
			(0.305)			(0.209)			(0.415)
Wearing time squared			-0.000			0.000			-0.000
			(0.000)			(0.000)			(0.000)
Outside temperature during PA measurement week (Celsius*10)			-0.228			0.188			-0.006
			(0.063)			(0.190)			(0.247)
Temperature squared			0.001			-0.000			0.000
			(0.000)			(0.001)			(0.001)
Amount of sunshine during PA measurement week (hours/day)			7.185			9.011			14.042
			(2.545)			(4.830)			(5.995)
Sunshine squared			-0.761			-0.937			-1.419
			(0.281)			(0.542)			(0.678)
Rain during PA measurement week (1=yes)			1.745			-3.362			-1.598
			(1.962)			(5.319)			(5.039)
Treatment (1=Treatment group)	-15.292			9.396			-4.331		
	(5.878)			(6.399)			(8.445)		
Child-fixed-effects	No	Yes	Yes	No	Yes	Yes		Yes	Yes

Table A1.11 Continued. Difference-in-differences estimations of the effect of the Active Living Program on time spent on PA by girls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	School time			Leisure time			Total		
	OLS	FE	FE	OLS	FE	FE	OLS	FE	FE
Constant	97.339			199.014			263.292		
	(4.936)			(4.718)			(6.993)		
Observations	720	650	590	698	610	568	718	646	584
R-squared	0.070	0.126	0.481	0.071	0.183	0.492	0.080	0.199	0.480
Number of children in the estimation sample		325	295		305	284		323	292
Mean time spent on PA in estimation sample	87.409	87.323	88.606	192.434	191.763	192.036		246.778	248.913
Standard deviation	24.507	24.408	23.243	46.897	46.107	43.649		50.398	46.899
Mean time spent on PA in full sample	87.409	87.409	87.409	192.434	192.434	192.434		247.918	247.918
Standard deviation in full sample	24.507	24.507	24.507	46.897	46.897	46.897		50.825	50.825

Notes: The table shows the results of nine ordinary least squares regressions. The dependent variable in columns (1) through (3) is time spent on PA during school time (min./day). The dependent variable in columns (4) through (6) is time spent on PA during leisure time (min./day). The dependent variable in columns (7) through (9) is total time spent on PA (min./day). Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: PA data.

Table A1.12. Regressions of overall school performance on mechanism variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	Music		Chess		Arts and crafts	
	No controls	Incl. controls	No controls	Incl. controls	No controls	Incl. controls
Music	2.847***	1.900***				
	(0.404)	(0.386)				
Chess			0.779	1.102		
			(1.491)	(1.551)		
Arts and crafts					-1.537***	-1.473***
					(0.506)	(0.510)
Parental education						
Lower secondary or below (≤ ISCED 2)		Ref.		Ref.		Ref.
Upper secondary / lower vocational (ISCED 3/4)		1.974***		2.084***		2.068***
		(0.683)		(0.679)		(0.676)
Higher vocational / university (≥ ISCED 5)		4.957***		5.250***		5.221***
		(0.683)		(0.685)		(0.680)
Missing		1.607**		1.766***		1.764***
		(0.661)		(0.659)		(0.658)
Age when tests were taken (months)		1.910**		1.921**		1.925**
		(0.902)		(0.916)		(0.917)
Age squared		-0.008**		-0.008**		-0.008**
		(0.003)		(0.003)		(0.003)
Gender (1=boy)		0.198		0.147		0.070
		(0.311)		(0.311)		(0.313)
Constant	68.238***	-49.581	68.633***	-50.021	68.747***	-50.058
	(0.282)	(64.371)	(0.280)	(65.399)	(0.286)	(65.484)
Observations	3,116	3,115	3,116	3,115	3,116	3,115
R-squared	0.016	0.131	0.000	0.124	0.002	0.126

Notes: For this table, only non-Active Living schools are selected. Results are from six ordinary least squares regressions. The dependent variable in each column is overall school performance in sixth grade in 2013/2014. Robust standard errors, clustered at the school level, are reported in parentheses, , *** p<0.01, ** p<0.05, * p<0.1.
Sources: School performance data and Onderwijs Monitor Limburg questionnaire 2014.

Appendix 2

Appendix to Chapter 2

2.1. Diagnostics criteria for ADHD[53]

A. A persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development, as characterized by (1) and/or (2):

1. **Inattention**: Six or more of the following symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level and that negatively impacts directly on social and academic/occupational activities:

   Note: The symptoms are not solely a manifestation of oppositional behavior, defiance, hostility, or failure to understand tasks or instructions.

a. Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).

b. Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).

c. Often does not seem to listen when spoken to directly (e.g., mind seems elsewhere, even in the absence of any obvious distraction).

d. Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., starts tasks but quickly loses focus and is easily sidetracked).

e. Often has difficulty organizing tasks and activities (e.g., difficulty managing sequential tasks; difficulty keeping materials and belongings in order; messy, disorganized work; has poor time management; fails to meet deadlines).

f. Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework; for older adolescents and adults, preparing reports, completing forms, reviewing lengthy papers).

g. Often loses things necessary for tasks or activities (e.g., school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).

h. Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).

i. Is often forgetful in daily activities (e.g., doing chores, running errands; for older adolescents and adults, returning calls, paying bills, keeping appointments).

2. **Hyperactivity and impulsivity**: Six or more of the following symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level and that negatively impacts directly on social and academic/occupational activities:

[53] Source: American Psychiatric Association (2013)

Note: The symptoms are not solely a manifestation of oppositional behavior, defiance, hostility, or a failure to understand tasks or instructions.

a. Often fidgets with or taps hands or feet or squirms in seat.
b. Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place).
c. Often runs about or climbs in situations where it is inappropriate. (Note: In adolescents or adults, may be limited to feeling restless.)
d. Often unable to play or engage in leisure activities quietly.
e. Is often “on the go,” acting as if “driven by a motor” (e.g., is unable to be or uncomfortable being still for extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).
f. Often talks excessively.
g. Often blurts out an answer before a question has been completed (e.g., completes people’s sentences; cannot wait for turn in conversation).
h. Often has difficulty waiting his or her turn (e.g., while waiting in line).
i. Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people’s things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).

B. Several inattentive or hyperactive-impulsive symptoms were present prior to age 12 years.

C. Several inattentive or hyperactive-impulsive symptoms are present in two or more settings (e.g., at home, school, or work; with friends or relatives; in other activities).

D. There is clear evidence that the symptoms interfere with, or reduce the quality of, social, academic, or occupational functioning.

E. The symptoms do not occur exclusively during the course of schizophrenia or another psychotic disorder and are not better explained by another mental disorder (e.g., mood disorder, anxiety disorder, dissociative disorder, personality disorder, substance intoxication or withdrawal).

Table A2.1. DSM-5diagnostic criteria for ADHD and the related OML variables

DSM-5 criteria for ADHD	Related OML variable	Pre-treatment cohorts of sixth graders					Treated cohorts of sixth graders	
		08/09	09/10	10/11	11/12	12/13	13/14	14/15
1. Inattentiveness								
a. Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).	How good are you at: Finishing and handing in my work neatly and tidily				x	x	x	x
	I am conscientious	x	x	x				
b. Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading). d. Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., starts tasks but quickly loses focus and is easily sidetracked).	How good are you at: Focusing on something				x	x	x	x
d. Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., starts tasks but quickly loses focus and is easily sidetracked).	How good are you at: Doing exactly what the teacher asks me to do				x	x	x	x
	When I start something I finish it	x	x	x				

Table A2.1 Continued. DSM-5diagnostic criteria for ADHD and the related OML variables

DSM-5 criteria for ADHD	Related OML variable	Pre-treatment cohorts of sixth graders					Treated cohorts of sixth graders	
		08/09	09/10	10/11	11/12	12/13	13/14	14/15
2. Hyperactivity and impulsivity								
b. Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place).	How good are you at: Remaining seated when the teacher wants me to				x	x	x	
a. Often fidgets with or taps hands or feet or squirms in seat. c. Often runs about or climbs in situations where it is inappropriate. d. Often unable to play or take part in leisure activities quietly. f. Often talks excessively.	How good are you at: Being calm and quiet when the teacher wants me to				x	x	x	x
d. Often unable to play or engage in leisure activities quietly. i. Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).	How good are you at: Being studious and not being a pest				x	x	x	

Sources: American Psychiatric association (2013) and Onderwijs Monitor Limburg questionnaires 2009-2015.

2.2. SUMMARY STATISTICS

Table A2.2. Time spent on PA among boys and girls in the pre-treatment year 2012/2013

	Boys	Girls	Diff.	*p*-value
Time spent on PA (min./day)				
Total	260.92	261.77	-0.84	0.831
During school-time	97.25	90.95	6.30	0.001
During leisure-time	199.59	203.42	-3.83	0.312

Source: PA data

2.3. DIFFERENT CLUSTER CORRECTIONS FOR TABLE 2.2

Table A2.3. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit) – clustered standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.013	Ref.	0.002	0.001
	(0.016)		(0.012)	(0.011)
6 (sufficient)	0.030		0.004	0.001
	(0.036)		(0.027)	(0.025)
7 (more than sufficient)	-0.004		-0.001	-0.000
	(0.006)		(0.004)	(0.003)
8 (good)	-0.039		-0.005	-0.002
	(0.047)		(0.036)	(0.033)
Observations	2,251		2,251	2,251
Focusing on something				
5 (insufficient)	-0.002	Ref.	0.010	0.008
	(0.008)		(0.008)	(0.011)
6 (sufficient)	-0.005		0.025	0.021
	(0.022)		(0.022)	(0.029)
7 (more than sufficient)	-0.001		0.006	0.005
	(0.005)		(0.005)	(0.007)
8 (good)	0.009		-0.040	-0.034
	(0.036)		(0.035)	(0.046)
Observations	3,740		3,740	3,740
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.008	Ref.	0.004	0.008
	(0.006)		(0.005)	(0.005)
6 (sufficient)	0.059		0.030	0.061
	(0.045)		(0.042)	(0.038)
7 (more than sufficient)	-0.007		-0.003	-0.007**
	(0.007)		(0.005)	(0.004)
8 (good)	-0.060		-0.030	-0.062
	(0.045)		(0.043)	(0.041)
Observations	2,253		2,253	2,253

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.4. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit) – bootstrapped standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.013	Ref.	0.002	0.001
	(0.017)		(0.012)	(0.011)
6 (sufficient)	0.030		0.004	0.001
	(0.037)		(0.028)	(0.026)
7 (more than sufficient)	-0.004		-0.001	-0.000
	(0.006)		(0.004)	(0.003)
8 (good)	-0.039		-0.005	-0.002
	(0.048)		(0.036)	(0.034)
Observations	2,251		2,251	2,251
Focusing on something				
5 (insufficient)	-0.002	Ref.	0.010	0.008
	(0.008)		(0.008)	(0.011)
6 (sufficient)	-0.005		0.025	0.021
	(0.022)		(0.022)	(0.029)
7 (more than sufficient)	-0.001		0.006	0.005
	(0.005)		(0.005)	(0.007)
8 (good)	0.009		-0.040	-0.034
	(0.035)		(0.036)	(0.046)
Observations	3,740		3,740	3,740
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.008	Ref.	0.004	0.008
	(0.006)		(0.006)	(0.005)
6 (sufficient)	0.059		0.030	0.061
	(0.044)		(0.047)	(0.040)
7 (more than sufficient)	-0.007		-0.003	-0.007*
	(0.007)		(0.005)	(0.004)
8 (good)	-0.060		-0.030	-0.062
	(0.044)		(0.048)	(0.043)
Observations	2,253		2,253	2,253

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level and block-bootstrapped with 5,000 replications, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.5. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit) – clustered standard errors with the Donald&Lang (2007) correction

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.013	Ref.	0.002	0.001
	(0.016)		(0.012)	(0.011)
6 (sufficient)	0.030		0.004	0.001
	(0.036)		(0.027)	(0.025)
7 (more than sufficient)	-0.004		-0.001	-0.000
	(0.006)		(0.004)	(0.003)
8 (good)	-0.039		-0.005	-0.002
	(0.047)		(0.036)	(0.033)
Observations	2,251		2,251	2,251
Focusing on something				
5 (insufficient)	-0.002	Ref.	0.010	0.008
	(0.008)		(0.008)	(0.011)
6 (sufficient)	-0.005		0.025	0.021
	(0.022)		(0.022)	(0.029)
7 (more than sufficient)	-0.001		0.006	0.005
	(0.005)		(0.005)	(0.007)
8 (good)	0.009		-0.040	-0.034
	(0.036)		(0.035)	(0.046)
Observations	3,740		3,740	3,740
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.008	Ref.	0.004	0.008
	(0.006)		(0.005)	(0.005)
6 (sufficient)	0.059		0.030	0.061
	(0.045)		(0.042)	(0.038)
7 (more than sufficient)	-0.007		-0.003	-0.007*
	(0.007)		(0.005)	(0.004)
8 (good)	-0.060		-0.030	-0.062
	(0.045)		(0.043)	(0.041)
Observations	2,253		2,253	2,253

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level and corrected with the Donald & Lang correction, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

2.4. DIFFERENT CLUSTER CORRECTIONS FOR TABLE 2.3

Table A2.6. Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity-impulsivity (marginal effects after ordered probit) – clustered standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	0.000	Ref.	-0.001	-
	(0.008)		(0.008)	
6 (sufficient)	0.002		-0.006	-
	(0.033)		(0.032)	
7 (more than sufficient)	0.001		-0.003	-
	(0.015)		(0.014)	
8 (good)	-0.003		0.009	-
	(0.055)		(0.054)	
Observations	1,670		1,670	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	0.004	Ref.	0.017**	0.017
	(0.014)		(0.008)	(0.012)
6 (sufficient)	0.010		0.048**	0.047
	(0.037)		(0.021)	(0.033)
7 (more than sufficient)	-0.000		-0.001	-0.001
	(0.001)		(0.003)	(0.003)
8 (good)	-0.014		-0.064**	-0.063
	(0.050)		(0.029)	(0.045)
Observations	2,255		2,255	2,255
Being studious and not being a pest				
5 (insufficient)	0.014	Ref.	-0.001	-
	(0.019)		(0.015)	
6 (sufficient)	0.033		-0.003	-
	(0.043)		(0.035)	
7 (more than sufficient)	-0.008		0.001	-
	(0.011)		(0.008)	
8 (good)	-0.039		0.003	-
	(0.051)		(0.043)	
Observations	1,671		1,671	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.7. Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity-impulsivity (marginal effects after ordered probit) – bootstrapped standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	0.000	Ref.	-0.001	-
	(0.008)		(0.008)	
6 (sufficient)	0.002		-0.006	-
	(0.034)		(0.034)	
7 (more than sufficient)	0.001		-0.003	-
	(0.015)		(0.015)	
8 (good)	-0.003		0.009	-
	(0.057)		(0.057)	
Observations	1,670		1,670	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	0.004	Ref.	0.017**	0.017
	(0.015)		(0.008)	(0.013)
6 (sufficient)	0.010		0.048**	0.047
	(0.040)		(0.022)	(0.035)
7 (more than sufficient)	-0.000		-0.001	-0.001
	(0.001)		(0.003)	(0.003)
8 (good)	-0.014		-0.064**	-0.063
	(0.054)		(0.029)	(0.047)
Observations	2,255		2,255	2,255
Being studious and not being a pest				
5 (insufficient)	0.014	Ref.	-0.001	-
	(0.019)		(0.015)	
6 (sufficient)	0.033		-0.003	-
	(0.044)		(0.036)	
7 (more than sufficient)	-0.008		0.001	-
	(0.011)		(0.008)	
8 (good)	-0.039		0.003	-
	(0.052)		(0.043)	
Observations	1,671		1,671	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level and block-bootstrapped with 5,000 replications, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.8. Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity-impulsivity (marginal effects after ordered probit) – clustered standard errors with the Donald&Lang (2007) correction

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	0.000	Ref.	-0.001	-
	(0.008)		(0.008)	
6 (sufficient)	0.002		-0.006	-
	(0.033)		(0.032)	
7 (more than sufficient)	0.001		-0.003	-
	(0.015)		(0.014)	
8 (good)	-0.003		0.009	-
	(0.055)		(0.054)	
Observations	1,670		1,670	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	0.004	Ref.	0.017**	0.017
	(0.014)		(0.008)	(0.012)
6 (sufficient)	0.010		0.048**	0.047
	(0.037)		(0.021)	(0.033)
7 (more than sufficient)	-0.000		-0.001	-0.001
	(0.001)		(0.003)	(0.003)
8 (good)	-0.014		-0.064**	-0.063
	(0.050)		(0.029)	(0.045)
Observations	2,255		2,255	2,255
Being studious and not being a pest				
5 (insufficient)	0.014	Ref.	-0.001	-
	(0.019)		(0.015)	
6 (sufficient)	0.033		-0.003	-
	(0.043)		(0.035)	
7 (more than sufficient)	-0.008		0.001	-
	(0.011)		(0.008)	
8 (good)	-0.039		0.003	-
	(0.051)		(0.043)	
Observations	1,671		1,671	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level and corrected with the Donald & Lang correction, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

2.5. DIFFERENT CLUSTER CORRECTIONS FOR TABLE 2.4

Table A2.9. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit), boys only – no clustering

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.016	Ref.	0.044	0.017
	(0.035)		(0.036)	(0.035)
6 (sufficient)	0.020		0.053	0.020
	(0.041)		(0.042)	(0.042)
7 (more than sufficient)	-0.014		-0.039	-0.015
	(0.030)		(0.031)	(0.031)
8 (good)	-0.022		-0.058	-0.022
	(0.045)		(0.047)	(0.046)
Observations	1,061		1,061	1,061
Focusing on something				
5 (insufficient)	-0.003	Ref.	0.010	0.015
	(0.012)		(0.013)	(0.013)
6 (sufficient)	-0.008		0.028	0.041
	(0.034)		(0.036)	(0.036)
7 (more than sufficient)	-0.001		0.004	0.006
	(0.005)		(0.005)	(0.005)
8 (good)	0.012		-0.042	-0.061
	(0.051)		(0.055)	(0.054)
Observations	1,797		1,797	1,797
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.016*	Ref.	0.015*	0.022**
	(0.009)		(0.009)	(0.010)
6 (sufficient)	0.118*		0.112*	0.160**
	(0.063)		(0.065)	(0.065)
7 (more than sufficient)	-0.045*		-0.043*	-0.061**
	(0.025)		(0.025)	(0.026)
8 (good)	-0.089*		-0.085*	-0.121**
	(0.048)		(0.049)	(0.049)
Observations	1,063		1,063	1,063

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Standard errors are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.10. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit), boys only – clustered standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.016	Ref.	0.044	0.017
	(0.047)		(0.031)	(0.036)
6 (sufficient)	0.020		0.053	0.020
	(0.056)		(0.036)	(0.042)
7 (more than sufficient)	-0.014		-0.039	-0.015
	(0.041)		(0.028)	(0.031)
8 (good)	-0.022		-0.058	-0.022
	(0.061)		(0.039)	(0.047)
Observations	1,061		1,061	1,061
Focusing on something				
5 (insufficient)	-0.003	Ref.	0.010	0.015
	(0.014)		(0.013)	(0.011)
6 (sufficient)	-0.008		0.028	0.041
	(0.039)		(0.037)	(0.031)
7 (more than sufficient)	-0.001		0.004	0.006
	(0.006)		(0.005)	(0.004)
8 (good)	0.012		-0.042	-0.061
	(0.059)		(0.056)	(0.045)
Observations	1,797		1,797	1,797
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.016	Ref.	0.015*	0.022***
	(0.013)		(0.009)	(0.008)
6 (sufficient)	0.118		0.112*	0.160***
	(0.093)		(0.065)	(0.060)
7 (more than sufficient)	-0.045		-0.043*	-0.061***
	(0.037)		(0.024)	(0.022)
8 (good)	-0.089		-0.085*	-0.121**
	(0.070)		(0.050)	(0.048)
Observations	1,063		1,063	1,063

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.11. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit), boys only – bootstrapped standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.016	Ref.	0.044	0.017
	(0.049)		(0.033)	(0.038)
6 (sufficient)	0.020		0.053	0.020
	(0.057)		(0.037)	(0.044)
7 (more than sufficient)	-0.014		-0.039	-0.015
	(0.043)		(0.029)	(0.032)
8 (good)	-0.022		-0.058	-0.022
	(0.063)		(0.041)	(0.049)
Observations	1,061		1,061	1,061
Focusing on something				
5 (insufficient)	-0.003	Ref.	0.010	0.015
	(0.015)		(0.022)	(0.011)
6 (sufficient)	-0.008		0.028	0.041
	(0.042)		(0.062)	(0.031)
7 (more than sufficient)	-0.001		0.004	0.006
	(0.006)		(0.009)	(0.004)
8 (good)	0.012		-0.042	-0.061
	(0.063)		(0.093)	(0.046)
Observations	1,797		1,797	1,797
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.016	Ref.	0.015*	0.022**
	(0.014)		(0.009)	(0.009)
6 (sufficient)	0.118		0.112*	0.160***
	(0.097)		(0.067)	(0.061)
7 (more than sufficient)	-0.045		-0.043*	-0.061***
	(0.039)		(0.025)	(0.022)
8 (good)	-0.089		-0.085	-0.121**
	(0.073)		(0.052)	(0.049)
Observations	1,063		1,063	1,063

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level and block-bootstrapped with 5,000 replications, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

2.6. DIFFERENT CLUSTER CORRECTIONS FOR TABLE 2.5

Table A2.12 Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity and impulsivity (marginal effects after ordered probit), boys only – no clustering

How good are you at…	(1) ‘11/’12	(.) ‘12/’13	(2) ‘13/’14	(3) ‘14/’15
Remaining seated when the teacher wants me to				
5 (insufficient)	-0.004	Ref.	0.007	-
	(0.013)		(0.014)	
6 (sufficient)	-0.016		0.024	-
	(0.047)		(0.048)	
7 (more than sufficient)	-0.003		0.005	-
	(0.009)		(0.010)	
8 (good)	0.023		-0.036	-
	(0.069)		(0.071)	
Observations	785		785	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	0.027	Ref.	0.043*	0.047**
	(0.022)		(0.023)	(0.023)
6 (sufficient)	0.061		0.096*	0.105**
	(0.049)		(0.051)	(0.050)
7 (more than sufficient)	-0.014		-0.022*	-0.024*
	(0.012)		(0.012)	(0.013)
8 (good)	-0.075		-0.117*	-0.128**
	(0.060)		(0.062)	(0.061)
Observations	1,066		1,066	1,066
Being studious and not being a pest				
5 (insufficient)	0.016	Ref.	0.016	-
	(0.022)		(0.022)	
6 (sufficient)	0.037		0.036	-
	(0.051)		(0.052)	
7 (more than sufficient)	-0.017		-0.017	-
	(0.024)		(0.024)	
8 (good)	-0.035		-0.035	-
	(0.049)		(0.050)	
Observations	785		785	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Standard errors are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author’s calculations based on OML questionnaires 2012-2015.

Table A2.13 Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity and impulsivity (marginal effects after ordered probit), boys only – clustered standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	-0.004	Ref.	0.007	-
	(0.018)		(0.014)	
6 (sufficient)	-0.016		0.024	-
	(0.064)		(0.050)	
7 (more than sufficient)	-0.003		0.005	-
	(0.013)		(0.010)	
8 (good)	0.023		-0.036	-
	(0.096)		(0.074)	
Observations	785		785	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	0.027	Ref.	0.043**	0.047*
	(0.028)		(0.019)	(0.025)
6 (sufficient)	0.061		0.096**	0.105*
	(0.065)		(0.046)	(0.055)
7 (more than sufficient)	-0.014		-0.022*	-0.024*
	(0.016)		(0.012)	(0.013)
8 (good)	-0.075		-0.117**	-0.128*
	(0.077)		(0.054)	(0.067)
Observations	1,066		1,066	1,066
Being studious and not being a pest				
5 (insufficient)	0.016	Ref.	0.016	-
	(0.029)		(0.022)	
6 (sufficient)	0.037		0.036	-
	(0.066)		(0.052)	
7 (more than sufficient)	-0.017		-0.017	-
	(0.032)		(0.025)	
8 (good)	-0.035		-0.035	-
	(0.063)		(0.049)	
Observations	785		785	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.14 Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity and impulsivity (marginal effects after ordered probit), boys only – bootstrapped standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	-0.004	Ref.	0.007	-
	(0.019)		(0.014)	
6 (sufficient)	-0.016		0.024	-
	(0.066)		(0.051)	
7 (more than sufficient)	-0.003		0.005	-
	(0.014)		(0.010)	
8 (good)	0.023		-0.036	-
	(0.099)		(0.075)	
Observations	785		785	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	0.027	Ref.	0.043**	0.047*
	(0.030)		(0.022)	(0.027)
6 (sufficient)	0.061		0.096*	0.105*
	(0.069)		(0.051)	(0.061)
7 (more than sufficient)	-0.014		-0.022*	-0.024*
	(0.016)		(0.013)	(0.014)
8 (good)	-0.075		-0.117*	-0.128*
	(0.082)		(0.061)	(0.074)
Observations	1,066		1,066	1,066
Being studious and not being a pest				
5 (insufficient)	0.016	Ref.	0.016	-
	(0.031)		(0.024)	
6 (sufficient)	0.037		0.036	-
	(0.070)		(0.056)	
7 (more than sufficient)	-0.017		-0.017	-
	(0.033)		(0.027)	
8 (good)	-0.035		-0.035	-
	(0.067)		(0.054)	
Observations	785		785	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level and block-bootstrapped with 5,000 replications, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

2.7. THE EFFECT OF THE ACTIVE LIVING PROGRAM ON GIRLS' ADHD-LIKE SYMPTOMS

2.7.1. Inattentiveness

Table A2.15 Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit), girls only – no clustering

	(1)	(.)	(2)	(3)
How good are you at…	'11/'12	'12/'13	'13/'14	'14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.011	Ref.	-0.013	-0.004
	(0.011)		(0.012)	(0.011)
6 (sufficient)	0.039		-0.047	-0.013
	(0.039)		(0.042)	(0.040)
7 (more than sufficient)	0.016		-0.020	-0.005
	(0.016)		(0.018)	(0.017)
8 (good)	-0.067		0.080	0.022
	(0.067)		(0.071)	(0.068)
Observations	1,190		1,190	1,190
Focusing on something				
5 (insufficient)	-0.001	Ref.	0.010	0.004
	(0.012)		(0.013)	(0.012)
6 (sufficient)	-0.002		0.025	0.010
	(0.030)		(0.032)	(0.030)
7 (more than sufficient)	-0.001		0.008	0.003
	(0.010)		(0.011)	(0.010)
8 (good)	0.004		-0.044	-0.017
	(0.051)		(0.056)	(0.053)
Observations	1,943		1,943	1,943
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.003	Ref.	-0.004	-0.000
	(0.007)		(0.007)	(0.007)
6 (sufficient)	0.019		-0.026	-0.001
	(0.043)		(0.045)	(0.044)
7 (more than sufficient)	0.004		-0.006	-0.000
	(0.009)		(0.010)	(0.009)
8 (good)	-0.026		0.035	0.002
	(0.059)		(0.062)	(0.060)
Observations	1,190		1,190	1,190

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Standard errors are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.16 Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit), girls only – clustered standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.011	Ref.	-0.013	-0.004
	(0.009)		(0.012)	(0.014)
6 (sufficient)	0.039		-0.047	-0.013
	(0.035)		(0.043)	(0.048)
7 (more than sufficient)	0.016		-0.020	-0.005
	(0.015)		(0.019)	(0.020)
8 (good)	-0.067		0.080	0.022
	(0.059)		(0.074)	(0.081)
Observations	1,190		1,190	1,190
Focusing on something				
5 (insufficient)	-0.001	Ref.	0.010	0.004
	(0.011)		(0.010)	(0.016)
6 (sufficient)	-0.002		0.025	0.010
	(0.026)		(0.025)	(0.040)
7 (more than sufficient)	-0.001		0.008	0.003
	(0.009)		(0.009)	(0.014)
8 (good)	0.004		-0.044	-0.017
	(0.046)		(0.044)	(0.069)
Observations	1,943		1,943	1,943
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.003	Ref.	-0.004	-0.000
	(0.006)		(0.009)	(0.006)
6 (sufficient)	0.019		-0.026	-0.001
	(0.040)		(0.054)	(0.040)
7 (more than sufficient)	0.004		-0.006	-0.000
	(0.009)		(0.011)	(0.008)
8 (good)	-0.026		0.035	0.002
	(0.055)		(0.073)	(0.054)
Observations	1,190		1,190	1,190

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.17 Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit), girls only – bootstrapped standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.011	Ref.	-0.013	-0.004
	(0.010)		(0.013)	(0.014)
6 (sufficient)	0.039		-0.047	-0.013
	(0.035)		(0.046)	(0.048)
7 (more than sufficient)	0.016		-0.020	-0.005
	(0.015)		(0.020)	(0.020)
8 (good)	-0.067		0.080	0.022
	(0.058)		(0.078)	(0.081)
Observations	1,190		1,190	1,190
Focusing on something				
5 (insufficient)	-0.001	Ref.	0.010	0.004
	(0.011)		(0.010)	(0.016)
6 (sufficient)	-0.002		0.025	0.010
	(0.027)		(0.026)	(0.041)
7 (more than sufficient)	-0.001		0.008	0.003
	(0.009)		(0.009)	(0.014)
8 (good)	0.004		-0.044	-0.017
	(0.048)		(0.044)	(0.071)
Observations	1,943		1,943	1,943
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.003	Ref.	-0.004	-0.000
	(0.006)		(0.010)	(0.007)
6 (sufficient)	0.019		-0.026	-0.001
	(0.043)		(0.060)	(0.042)
7 (more than sufficient)	0.004		-0.006	-0.000
	(0.010)		(0.012)	(0.009)
8 (good)	-0.026		0.035	0.002
	(0.059)		(0.081)	(0.058)
Observations	1,190		1,190	1,190

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level and block-bootstrapped with 5,000 replications, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.18 Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness (marginal effects after ordered probit), girls only – clustered standard errors with the Donald&Lang (2007) correction

	(1)	(.)	(2)	(3)
How good are you at…	'11/'12	'12/'13	'13/'14	'14/'15
Finishing and handing in my work neatly and tidily				
5 (insufficient)	0.011	Ref.	-0.013	-0.004
	(0.009)		(0.012)	(0.014)
6 (sufficient)	0.039		-0.047	-0.013
	(0.035)		(0.043)	(0.048)
7 (more than sufficient)	0.016		-0.020	-0.005
	(0.015)		(0.019)	(0.020)
8 (good)	-0.067		0.080	0.022
	(0.059)		(0.074)	(0.081)
Observations	1,190		1,190	1,190
Focusing on something				
5 (insufficient)	-0.001	Ref.	0.010	0.004
	(0.011)		(0.010)	(0.016)
6 (sufficient)	-0.002		0.025	0.010
	(0.026)		(0.025)	(0.040)
7 (more than sufficient)	-0.001		0.008	0.003
	(0.009)		(0.009)	(0.014)
8 (good)	0.004		-0.044	-0.017
	(0.046)		(0.044)	(0.069)
Observations	1,943		1,943	1,943
Doing exactly what the teacher asks me to do				
5 (insufficient)	0.003	Ref.	-0.004	-0.000
	(0.006)		(0.009)	(0.006)
6 (sufficient)	0.019		-0.026	-0.001
	(0.040)		(0.054)	(0.040)
7 (more than sufficient)	0.004		-0.006	-0.000
	(0.009)		(0.011)	(0.008)
8 (good)	-0.026		0.035	0.002
	(0.055)		(0.073)	(0.054)
Observations	1,190		1,190	1,190

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level and corrected with the Donald & Lang correction, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

2.7.2. Hyperactivity-impulsivity

Table A2.19 Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity and impulsivity (marginal effects after ordered probit), girls only – no clustering

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	0.004	Ref.	-0.008	-
	(0.008)		(0.008)	
6 (sufficient)	0.018		-0.038	-
	(0.039)		(0.042)	
7 (more than sufficient)	0.012		-0.025	-
	(0.025)		(0.027)	
8 (good)	-0.034		0.071	-
	(0.072)		(0.077)	
Observations	885		885	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	-0.008	Ref.	0.004	0.001
	(0.013)		(0.014)	(0.013)
6 (sufficient)	-0.026		0.011	0.003
	(0.042)		(0.044)	(0.043)
7 (more than sufficient)	-0.005		0.002	0.001
	(0.009)		(0.009)	(0.009)
8 (good)	0.039		-0.017	-0.004
	(0.063)		(0.067)	(0.065)
Observations	1,189		1,189	1,189
Being studious and not being a pest				
5 (insufficient)	0.016	Ref.	-0.014	-
	(0.018)		(0.019)	
6 (sufficient)	0.033		-0.030	-
	(0.038)		(0.040)	
7 (more than sufficient)	0.001		-0.001	-
	(0.003)		(0.003)	
8 (good)	-0.050		0.045	-
	(0.057)		(0.059)	
	0.016		-0.014	
Observations	(0.018)		(0.019)	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Standard errors are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.20 Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity and impulsivity (marginal effects after ordered probit), girls only – clustered standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	0.004	Ref.	-0.008	-
	(0.008)		(0.010)	
6 (sufficient)	0.018		-0.038	-
	(0.041)		(0.053)	
7 (more than sufficient)	0.012		-0.025	-
	(0.027)		(0.034)	
8 (good)	-0.034		0.071	-
	(0.076)		(0.097)	
Observations	885		885	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	-0.008	Ref.	0.004	0.001
	(0.013)		(0.012)	(0.012)
6 (sufficient)	-0.026		0.011	0.003
	(0.041)		(0.039)	(0.040)
7 (more than sufficient)	-0.005		0.002	0.001
	(0.009)		(0.008)	(0.008)
8 (good)	0.039		-0.017	-0.004
	(0.062)		(0.059)	(0.060)
Observations	1,189		1,189	1,189
Being studious and not being a pest				
5 (insufficient)	0.016	Ref.	-0.014	-
	(0.020)		(0.020)	
6 (sufficient)	0.033		-0.030	-
	(0.040)		(0.045)	
7 (more than sufficient)	0.001		-0.001	-
	(0.003)		(0.003)	
8 (good)	-0.050		0.045	-
	(0.061)		(0.067)	
Observations	886		886	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.21 Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity and impulsivity (marginal effects after ordered probit), girls only – bootstrapped standard errors

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	0.004	Ref.	-0.008	-
	(0.009)		(0.011)	
6 (sufficient)	0.018		-0.038	-
	(0.042)		(0.057)	
7 (more than sufficient)	0.012		-0.025	-
	(0.028)		(0.037)	
8 (good)	-0.034		0.071	-
	(0.079)		(0.105)	
Observations	885		885	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	-0.008	Ref.	0.004	0.001
	(0.014)		(0.013)	(0.013)
6 (sufficient)	-0.026		0.011	0.003
	(0.044)		(0.041)	(0.043)
7 (more than sufficient)	-0.005		0.002	0.001
	(0.009)		(0.008)	(0.009)
8 (good)	0.039		-0.017	-0.004
	(0.067)		(0.061)	(0.064)
Observations	1,189		1,189	1,189
Being studious and not being a pest				
5 (insufficient)	0.016	Ref.	-0.014	-
	(0.019)		(0.021)	
6 (sufficient)	0.033		-0.030	-
	(0.039)		(0.047)	
7 (more than sufficient)	0.001		-0.001	-
	(0.003)		(0.004)	
8 (good)	-0.050		0.045	-
	(0.059)		(0.071)	
Observations	886		886	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level and block-bootstrapped with 5,000 replications, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.22 Difference-in-differences estimations of the effect of the Active Living Program on indicators of hyperactivity and impulsivity (marginal effects after ordered probit), girls only – clustered standard errors with the Donald&Lang (2007) correction

How good are you at…	(1) '11/'12	(.) '12/'13	(2) '13/'14	(3) '14/'15
Remaining seated when the teacher wants me to				
5 (insufficient)	0.004	Ref.	-0.008	-
	(0.008)		(0.010)	
6 (sufficient)	0.018		-0.038	-
	(0.041)		(0.053)	
7 (more than sufficient)	0.012		-0.025	-
	(0.027)		(0.034)	
8 (good)	-0.034		0.071	-
	(0.076)		(0.097)	
Observations	885		885	-
Being calm and quiet when the teacher wants me to				
5 (insufficient)	-0.008	Ref.	0.004	0.001
	(0.013)		(0.012)	(0.012)
6 (sufficient)	-0.026		0.011	0.003
	(0.041)		(0.039)	(0.040)
7 (more than sufficient)	-0.005		0.002	0.001
	(0.009)		(0.008)	(0.008)
8 (good)	0.039		-0.017	-0.004
	(0.062)		(0.059)	(0.060)
Observations	1,189		1,189	1,189
Being studious and not being a pest				
5 (insufficient)	0.016	Ref.	-0.014	-
	(0.020)		(0.020)	
6 (sufficient)	0.033		-0.030	-
	(0.040)		(0.045)	
7 (more than sufficient)	0.001		-0.001	-
	(0.003)		(0.003)	
8 (good)	-0.050		0.045	-
	(0.061)		(0.067)	
Observations	886		886	-

Notes: The table shows the marginal effects after three ordered probit regressions. All regressions include a treatment dummy, year fixed effects, and controls for parental education, age, and gender. Robust standard errors, clustered at the school-level, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

2.8. ROBUSTNESS CHECKS

Table A2.23. Difference-in-differences estimations of the effect of the Active Living Program on 'doing exactly what the teacher asks me to do' (marginal effects after ordered probit), one-by-one exclusion of treatment schools, boys only

How good are you at: Doing exactly what the teacher asks me to do (boys)	(1) Excl. T1	(2) Excl. T2	(3) Excl. T3	(4) Excl. T4	(5) Excl. T5	(6) Excl. T6	(7) Excl. T7	(8) Excl. T8	(9) Excl. T9
2011/2012									
5 (insufficient)	0.017	0.017	0.012	0.014	0.018	0.013	0.021	0.008	0.023*
	(0.013)	(0.014)	(0.014)	(0.014)	(0.015)	(0.014)	(0.013)	(0.012)	(0.013)
6 (sufficient)	0.121	0.122	0.090	0.107	0.124	0.104	0.153	0.059	0.169*
	(0.093)	(0.095)	(0.098)	(0.103)	(0.104)	(0.106)	(0.092)	(0.088)	(0.088)
7 (more than sufficient)	-0.046	-0.046	-0.035	-0.039	-0.044	-0.039	-0.058	-0.022	-0.067*
	(0.037)	(0.037)	(0.040)	(0.039)	(0.039)	(0.041)	(0.037)	(0.033)	(0.036)
8 (good)	-0.092	-0.093	-0.067	-0.082	-0.098	-0.079	-0.115	-0.046	-0.126*
	(0.070)	(0.071)	(0.073)	(0.078)	(0.081)	(0.080)	(0.068)	(0.067)	(0.066)
2012/2013	Ref.								
2013/2014									
5 (insufficient)	0.016*	0.016*	0.012	0.014	0.015	0.012	0.020***	0.013	0.018*
	(0.009)	(0.009)	(0.009)	(0.008)	(0.010)	(0.008)	(0.007)	(0.009)	(0.009)
6 (sufficient)	0.112	0.111	0.090	0.107	0.106	0.096	0.150**	0.101	0.127*
	(0.065)	(0.065)	(0.069)	(0.067)	(0.071)	(0.069)	(0.054)	(0.070)	(0.070)
7 (more than sufficient)	-0.042*	-0.042*	-0.035	-0.039	-0.037	-0.036	-0.057***	-0.037	-0.050*
	(0.024)	(0.024)	(0.027)	(0.024)	(0.024)	(0.025)	(0.020)	(0.025)	(0.026)
8 (good)	-0.085	-0.085	-0.067	-0.082	-0.084	-0.073	-0.114**	-0.078	-0.095*
	(0.050)	(0.050)	(0.052)	(0.052)	(0.057)	(0.053)	(0.043)	(0.055)	(0.053)

Table A2.23 Continued. Difference-in-differences estimations of the effect of the Active Living Program on ‘doing exactly what the teacher asks me to do’ (marginal effects after ordered probit), one-by-one exclusion of treatment schools, boys only

How good are you at: Doing exactly what the teacher asks me to do (boys)	(1) Excl. T1	(2) Excl. T2	(3) Excl. T3	(4) Excl. T4	(5) Excl. T5	(6) Excl. T6	(7) Excl. T7	(8) Excl. T8	(9) Excl. T9
2014/2015									
5 (insufficient)	0.021**	0.023**	0.019**	0.022**	0.026***	0.018**	0.021**	0.020**	0.025**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
6 (sufficient)	0.152**	0.163**	0.136**	0.172**	0.181***	0.141**	0.157**	0.148**	0.180***
	(0.060)	(0.061)	(0.062)	(0.061)	(0.059)	(0.063)	(0.063)	(0.064)	(0.061)
7 (more than sufficient)	-0.058**	-0.061**	-0.053**	-0.063***	-0.064***	-0.052**	-0.060**	-0.054**	-0.071***
	(0.022)	(0.022)	(0.023)	(0.022)	(0.021)	(0.022)	(0.023)	(0.022)	(0.021)
8 (good)	-0.116**	-0.124**	-0.102**	-0.132**	-0.144***	-0.107**	-0.119**	-0.114**	-0.134**
	(0.048)	(0.049)	(0.048)	(0.050)	(0.048)	(0.050)	(0.050)	(0.053)	(0.050)
Observations	1,049	1,040	1,009	1,014	968	988	1,014	990	988

Notes: The table shows the marginal effects after nine ordered probit regressions when selecting only boys, and when excluding treatment schools one by one. All regressions include a treatment dummy, year fixed effects, and controls for parental education, and age. Robust standard errors, clustered at the school-level and corrected with the Donald&Lang-correction, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author’s calculations based on OML questionnaires 2012-2015.

Table A2.24. Difference-in-differences estimations of the effect of the Active Living Program on 'being calm and quiet when the teacher wants me to' (marginal effects after ordered probit), one-by-one exclusion of treatment schools, boys only

How good are you at: Being calm and quiet when the teacher wants me to (boys)	(1) Excl. T1	(2) Excl. T2	(3) Excl. T3	(4) Excl. T4	(5) Excl. T5	(6) Excl. T6	(7) Excl. T7	(8) Excl. T8	(9) Excl. T9
2011/2012									
5 (insufficient)	0.028	0.027	0.022	0.028	0.022	0.014	0.041	0.022	0.042
	(0.028)	(0.028)	(0.031)	(0.030)	(0.032)	(0.027)	(0.026)	(0.033)	(0.025)
6 (sufficient)	0.062	0.061	0.047	0.064	0.047	0.033	0.087	0.047	0.101
	(0.065)	(0.066)	(0.067)	(0.070)	(0.071)	(0.065)	(0.061)	(0.071)	(0.060)
7 (more than sufficient)	-0.014	-0.013	-0.010	-0.013	-0.010	-0.007	-0.019	-0.011	-0.025
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.014)	(0.015)	(0.018)	(0.015)
8 (good)	-0.076	-0.075	-0.058	-0.079	-0.059	-0.040	-0.109	-0.058	-0.118
	(0.077)	(0.079)	(0.082)	(0.085)	(0.089)	(0.079)	(0.072)	(0.086)	(0.070)
2012/2013	Ref.								
2013/2014									
5 (insufficient)	0.043**	0.043**	0.039	0.038*	0.031	0.039*	0.055***	0.049**	0.049**
	(0.019)	(0.019)	(0.023)	(0.020)	(0.020)	(0.021)	(0.017)	(0.022)	(0.020)
6 (sufficient)	0.096*	0.095*	0.084	0.087*	0.067	0.091*	0.117**	0.103*	0.118**
	(0.046)	(0.046)	(0.051)	(0.048)	(0.043)	(0.052)	(0.043)	(0.051)	(0.049)
7 (more than sufficient)	-0.021*	-0.021*	-0.019	-0.018	-0.014	-0.018	-0.026*	-0.025*	-0.029**
	(0.012)	(0.012)	(0.013)	(0.011)	(0.010)	(0.012)	(0.012)	(0.014)	(0.013)
8 (good)	-0.118**	-0.117**	-0.105	-0.108*	-0.084	-0.112*	-0.147***	-0.128**	-0.138**
	(0.055)	(0.054)	(0.061)	(0.058)	(0.053)	(0.062)	(0.049)	(0.060)	(0.057)

Table A2.24 Continued. Difference-in-differences estimations of the effect of the Active Living Program on 'being calm and quiet when the teacher wants me to' (marginal effects after ordered probit), one-by-one exclusion of treatment schools, boys only

How good are you at: Being calm and quiet when the teacher wants me to (boys)	(1) Excl. T1	(2) Excl. T2	(3) Excl. T3	(4) Excl. T4	(5) Excl. T5	(6) Excl. T6	(7) Excl. T7	(8) Excl. T8	(9) Excl. T9
2014/2015									
5 (insufficient)	0.045*	0.049*	0.046	0.050*	0.036	0.041	0.061***	0.046	0.047*
	(0.025)	(0.025)	(0.027)	(0.025)	(0.026)	(0.025)	(0.021)	(0.028)	(0.025)
6 (sufficient)	0.101*	0.108*	0.098	0.112*	0.078	0.096	0.131**	0.097	0.114*
	(0.056)	(0.056)	(0.059)	(0.056)	(0.054)	(0.060)	(0.048)	(0.059)	(0.061)
7 (more than sufficient)	-0.023	-0.024*	-0.022	-0.023*	-0.016	-0.020	-0.028**	-0.023	-0.028*
	(0.013)	(0.013)	(0.014)	(0.013)	(0.011)	(0.013)	(0.013)	(0.015)	(0.015)
8 (good)	-0.124*	-0.133*	-0.122	-0.139*	-0.098	-0.118	-0.163**	-0.120	-0.133*
	(0.068)	(0.068)	(0.073)	(0.069)	(0.069)	(0.073)	(0.058)	(0.072)	(0.072)
Observations	1,052	1,043	1,011	1,017	971	989	1,016	993	991

Notes: The table shows the marginal effects after nine ordered probit regressions when selecting only boys, and when excluding treatment schools one by one. All regressions include a treatment dummy, year fixed effects, and controls for parental education, and age. Robust standard errors, clustered at the school-level and corrected with the Donald&Lang-correction, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Table A2.25. Implemented interventions in Treatment school 7

Category	Intervention	Description	Also implemented by:
Active transportation to school	Sticker competition for active transport	Provide stickers to children who traveled to school on foot or by bike	T1, T2, T3, T4
PA in school	New loose equipment in schoolyard	Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment	T2, T3, T5, T8, T9
	Playground markings	Paintings, hopscotch markings	T2, T6
	Use of schoolyard games	Media cards that provide information to children on how to play games	T2, T3, T8
PA in leisure-time	Active Living Games	An additional sports day for all participating treatment schools (funded by the Province of Limburg)	T1, T2, T3, T4, T5, T6, T9

Table A2.26. Difference-in-differences estimations of the effect of the Active Living Program on indicators of inattentiveness and hyperactivity/impulsivity when comparing control schools and treatment schools to Non-Active Living schools (marginal effects after ordered probit), boys only

	(1)		(2)	(3)	(4)		(5)	(6)
	Control vs. Non-Active Living				Treatment vs. Non-Active Living			
	'11/'12	'12/'13	'13/'14	'14/'15	'11/'12	'12/'13	'13/'14	'14/'15
Doing exactly what the teacher asks me to do								
5 (insufficient)	-0.017*	Ref.	-0.011	-0.012	0.005	Ref.	0.011	0.019**
	(0.010)		(0.009)	(0.009)	(0.016)		(0.009)	(0.009)
6 (sufficient)	-0.078*		-0.049	-0.055	0.024		0.049	0.082**
	(0.044)		(0.039)	(0.039)	(0.068)		(0.038)	(0.037)
7 (more than sufficient)	0.024*		0.015	0.017	-0.008		-0.016	-0.027**
	(0.014)		(0.012)	(0.012)	(0.022)		(0.013)	(0.012)
8 (good)	0.071*		0.045	0.051	-0.021		-0.044	-0.074**
	(0.041)		(0.036)	(0.036)	(0.062)		(0.035)	(0.033)
Observations	8,113		8,113	8,113	8,064		8,064	8,064
Being calm and quiet when the teacher wants me to								
5 (insufficient)	-0.026		-0.017*	-0.012	0.001		0.024	0.033**
	(0.016)		(0.010)	(0.017)	(0.022)		(0.016)	(0.016)
6 (sufficient)	-0.059		-0.038*	-0.027	0.001		0.056	0.076**
	(0.037)		(0.023)	(0.039)	(0.051)		(0.038)	(0.037)
7 (more than sufficient)	0.010		0.006*	0.004	-0.000		-0.010	-0.014**
	(0.006)		(0.004)	(0.006)	(0.009)		(0.007)	(0.007)
8 (good)	0.075		0.049*	0.035	-0.002		-0.070	-0.095**
	(0.048)		(0.029)	(0.050)	(0.064)		(0.047)	(0.046)
Observations	8,108		8,108	8,108	8,064		8,064	8,064

Notes: The table shows the marginal effects after six ordered probit regressions when selecting only boys. All regressions include a control dummy (columns 1-3) or a treatment dummy (columns 4-6), year fixed effects, and controls for parental education, and age. Robust standard errors, clustered at the school-level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations based on OML questionnaires 2012-2015.

Appendix 3
Appendix to Chapter 3

3.1. Data sources

3.1.1. Data sources for the main results

Organ-patient mortality rates and deceased-donor transplantation rates

In every country, donated organs are scarce, and being on a waiting list for organ transplantation is usually the only way to legally receive an organ from a deceased donor. Therefore, being registered on the waiting list for organ transplantation is a good indicator of being an organ patient.

We use data on the number of patients who died while on the waiting list for a kidney, liver, heart, or lung transplantation from "Newsletter Transplant." These newsletters also serve as the data source for the number of deceased-donor kidney, liver, heart, and lung transplantations[54] per country. The newsletters are published jointly by the Organización Nacional de Trasplantes (ONT) and the European Directorate for the Quality of Medicines & HealthCare of the Council of Europe (EDQM). National data in these documents are provided by representatives from national and international transplant networks.[55] The resulting data set contains information from 79 countries across a period of 15 years (2001–2015, unbalanced panel). We calculate organ-patient mortality and deceased-donor transplantation rates by dividing the absolute numbers by the country's population size (in millions). Tables A3.1 and A3.2 list the description of the key variables and the countries included in the analyses. We define "organ patient mortality" as death while on the waiting list for a particular organ.

Table A3.1. Description of key variables included in the analysis

Variable	Description
Number of waiting list deaths	Number of patients who died while on the waiting list during the corresponding year.
Transplantations from deceased donors / deceased-donor transplantations	Transplantations from donors after brain death (i.e., death has been determined by neurological criteria) and donors after circulatory death (i.e., death has been determined by circulatory and respiratory criteria). Includes the transplantations of the corresponding organ with or without the simultaneous transplant of a different type of organ(s). Note that every donor can provide multiple organs, and that organs will not be removed from a potential donor for the purpose of transplantation unless there is a recipient.

Source: Newsletter Transplant 2016.

[54] To avoid confusion, please note that there is a difference between transplantations and donors; a donor is the person from whom at least one organ is obtained in order to transplant that organ into an organ patient. There is no use in becoming an organ donor (living or deceased) if there is no potential recipient (who is a good match and has a high probability of survival). This means that by construction, each donor equals at least one, but potentially multiple, transplantation(s).

[55] The data are available in pdf documents (http://www.ont.es/publicaciones/Paginas/Publicaciones.aspx).

Table A3.2. List of countries included in the Newsletter Transplant for the first time

Year	**Countries**
2001	Australia, Austria, Belgium, Canada, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Ireland, Israel, Italy, Kuwait, Latvia, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States of America
2002	Bulgaria, Lithuania
2003	Argentina, Brazil, Bolivia, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, Honduras, Iceland, Mexico, New Zealand, Panama, Paraguay, Peru, Puerto Rico, Dominican Republic, Uruguay, Venezuela
2008	Macedonia, Moldova, Nicaragua
2010	Algeria, Lebanon, Libya, Morocco, Palestine, Syria, Tunisia
2011	Belarus, Egypt, Russia
2012	Azerbaijan, Armenia, Ukraine
2013	Bosnia Herzegovina, Montenegro
2014	Iran, Malaysia, Sudan, Vietnam

Notes: Puerto Rico is unincorporated territory of the United States and part of the United Network for Organ Sharing (UNOS).
Source: Newsletter Transplant 2002–2016.

Consent systems

We obtain data on the countries' consent systems for organ donation from previous research articles, government websites, professional organizations, and legal documents (see Table A3.3). For ten of the 79 countries, no (reliable) information on the consent system could be found. These countries are therefore excluded from the analysis.

We verify whether the countries' consent system ever changed, and if so in which year, and adjust the opt-out variable accordingly. When no information is available for the year of introduction of the current consent system or whether the consent system changed between 2001 and 2015, we assume in our main analyses that the country has had the same consent system in all years. In our main analyses, we exclude countries that changed their consent system after 1999.

Table A3.3. Consent systems per country

Country	Current system	Since	Sources			
Algeria	opt-in		*5*			
Argentina	opt-out		*4*	*6*	*7*	*8*
Armenia	opt-in	2002	*3*			
Australia	opt-in	1980s	*3*	*4*	*9*	
Austria	opt-out	1982	*2*	*3*	*4*	
Azerbaijan	opt-out		*3*	*10*		
Belarus	opt-out		*3*			
Belgium	opt-out	1986	*2*	*3*	*4*	*9*
Bolivia	opt-in		*8*	*11*	*12*	
Bosnia and Herzegovina	opt-out		*13*			
Brazil	opt-in	1997	*3*	*4*	*7*	*8*
Bulgaria	opt-out	2007	*2*	*4*	*9*	*14*
Canada	opt-in	1980s	*3*	*4*	*9*	
Chile	opt-out		*3*	*4*		
Colombia	opt-out		*3*	*4*	*8*	
Costa Rica	opt-out	1994	*3*	*4*	*8*	*15*
Croatia	opt-out		*3*	*4*	*9*	
Cuba	opt-in	1988	*3*	*4*	*8*	
Cyprus	opt-out		*9*	*16*		

Table A3.3 Continued. Consent systems per country

Country	Current system	Since	Sources			
Czech Republic	opt-out	2002	*3*	*4*	*9*	
Denmark	opt-in	1990	*3*	*4*	*9*	*16*
Dominican	opt-out		*8*	*16*		
Ecuador	opt-out	1994	*3*	*8*		
Estonia	opt-out	2002	*2*	*4*	*9*	*16*
Finland	opt-out	2001	*3*	*9*		
France	opt-out	1976	*3*	*4*	*9*	
Georgia	opt-in		*4*	*14*		
Germany	opt-in	1997	*3*	*9*		
Greece	opt-out	1999	*4*	*9*		
Guatemala	opt-in		*4*	*8*	*15*	
Hungary	opt-out	1997	*4*	*9*		
Iceland	opt-in	1991	*3*	*4*		
Iran	opt-in		*18*			
Ireland	opt-out	2012	*3*	*9*		
Israel	opt-in		*3*	*4*	*9*	
Italy	opt-out	1999	*3*	*9*		
Latvia	opt-out	1995	*4*	*9*		
Lithuania	opt-in	1999	*3*	*4*	*9*	
Luxembourg	opt-out	1982	*3*	*4*	*9*	
Macedonia	opt-out		*13*			
Malaysia	opt-in	1974	*3*	*19*		
Malta	opt-in		*3*	*4*	*20*	
Mexico	opt-in		*3*	*8*		
Moldova	opt-out		*4*	*21*		
Montenegro	opt-in		*2*			
Netherlands	opt-in	1996	*3*	*4*	*9*	
New Zealand	opt-in		*3*	*4*	*9*	
Norway	opt-out	1973	*3*	*9*	*16*	
Panama	opt-in		*15*			
Paraguay	opt-out	1998	*3*	*8*		
Peru	opt-out		*8*			
Poland	opt-out	1995	*3*	*4*	*9*	
Portugal	opt-out	1994	*4*	*9*		
Romania	opt-in		*3*	*9*	*16*	
Russian F.	opt-out	1992	*3*	*4*		
Slovakia	opt-out	2004	*3*	*4*	*9*	
Slovenia	opt-out	2000	*3*	*9*		
Spain	opt-out	1979	*3*	*4*	*9*	
Sudan	opt-in		*22*	*23*		
Sweden	opt-out	1995	*3*	*4*	*9*	
Switzerland	opt-in	2007	*3*	*9*		
Syria	opt-in		*23*			
Turkey	opt-out	1979	*3*	*9*		
United Kingdom	opt-in	2006	*3*	*4*	*9*	
Ukraine	opt-in		*24*			
Uruguay	opt-out		*25*			
USA	opt-in		*3*	*4*	*9*	
Venezuela	opt-in	1992	*3*	*4*	*8*	
Vietnam	opt-in		*26*			

Notes: Numbers in the table refer to the literature in the reference list.

Confounders

In our main specifications, we control for countries' gross domestic product (GDP) per capita, health expenditures per capita, general mortality rates (all causes; age standardized), and religion. Table A3.4 provides descriptive statistics for these variables by consent system for kidneys. Tables A3.5-A3.7 show the descriptive statistics for the samples used to estimate the opt-out effects for livers, hearts, and lungs. These tables show that there are no statistically significant differences in the characteristics between opt-in and opt-out countries.

Table A3.4. Means and standard deviations of key variables by consent system in the estimation sample for kidneys

	(1) Total	(2) Opt-in	(3) Opt-out	(4) Diff.	(5) [*p*-value]
Population size (millions)	35.12	50.80	23.67	-27.13	[0.119]
	(57.46)	(78.68)	(32.10)		
Western countries (%)	53.33	47.37	57.69	10.32	[0.504]
	(50.45)	(51.30)	(50.38)		
Total deaths all causes (pmp, age-standardized)	54.45	51.48	56.12	4.64	[0.373]
	(14.78)	(10.99)	(16.54)		
GDP per capita (current US$/10,000)	21,434.12	20,983.99	21,763.06	779.07	[0.902]
	(20,552.31)	(17,540.11)	(22,839.46)		
Health expenditures per capita (current US$/1,000)	1,940.21	2,177.59	1,785.00	-392.59	[0.541]
	(2,024.45)	(2,189.79)	(1,937.27)		
Religious denomination (%)					
None	19.93	19.29	20.45	1.16	[0.857]
	(16.78)	(17.54)	(16.69)		
Muslim	9.74	11.29	8.47	-2.82	[0.743]
	(22.45)	(20.14)	(24.74)		
Roman Catholic	32.06	25.78	37.16	11.38	[0.338]
	(31.24)	(24.96)	(35.51)		
Orthodox	15.68	12.49	18.27	5.79	[0.624]
	(30.82)	(29.07)	(32.88)		
Protestant	6.43	6.80	6.13	-0.67	[0.892]
	(12.94)	(9.30)	(15.60)		
Missing values (%)					
Total deaths all causes	31.88	37.46	27.80	-9.66	[0.402]
	(37.69)	(43.90)	(32.73)		
GDP per capita	2.22	5.26	0.00	-5.26	[0.037]
	(8.41)	(12.49)	(0.00)		
Health expenditures per capita	4.44	10.53	0.00	-10.53	[0.095]
	(20.84)	(31.53)	(0.00)		
Religious denomination	35.56	31.58	38.46	6.88	[0.643]
	(48.41)	(47.76)	(49.61)		
Number of countries	45	19	26		

Notes: Based on the estimation sample for the main results for kidneys (Table 3.2, column 1).

Table A3.5. Means and standard deviations of key variables by consent system in the estimation sample for livers

	(1) Total	(2) Opt-in	(3) Opt-out	(4) Diff.	(5) [*p*-value]
Population size (millions)	36.94	51.78	26.85	-24.93	[0.177]
	(58.32)	(82.38)	(32.01)		
Western countries (%)	52.38	52.94	52.00	-0.94	[0.954]
	(50.55)	(51.45)	(50.99)		
Total deaths all causes (pmp, age-standardized)	53.23	50.83	54.70	3.87	[0.424]
	(14.06)	(10.84)	(15.76)		
GDP per capita (current US$/10,000)	21,061.55	23,952.02	19,096.03	-4,856.00	[0.393]
	(17,817.41)	(17,525.14)	(18,100.60)		
Health expenditures per capita (current US$/1,000)	1,974.20	2,526.18	1,643.02	-883.16	[0.168]
	(1,947.62)	(2,198.52)	(1,743.49)		
Religious denomination (%)					
None	21.66	21.54	21.74	0.19	[0.974]
	(15.66)	(16.84)	(15.26)		
Muslim	8.08	7.60	8.43	0.83	[0.916]
	(21.23)	(16.89)	(24.36)		
Roman Catholic	35.24	26.79	41.33	14.54	[0.200]
	(30.85)	(24.08)	(34.30)		
Orthodox	13.12	12.50	13.57	1.07	[0.919]
	(28.32)	(29.06)	(28.61)		
Protestant	6.36	6.80	6.05	-0.75	[0.873]
	(12.57)	(9.30)	(14.74)		
Missing values (%)					
Total deaths all causes	25.92	25.57	26.15	0.58	[0.956]
	(32.65)	(36.02)	(30.92)		
GDP per capita	0.79	1.96	0.00	-1.96	[0.082]
	(3.59)	(5.54)	(0.00)		
Health expenditures per capita	4.76	11.76	0.00	-11.76	[0.082]
	(21.55)	(33.21)	(0.00)		
Religious denomination	26.19	23.53	28.00	4.47	[0.754]
	(44.50)	(43.72)	(45.83)		
Number of countries	42	17	25		

Notes: Based on the estimation sample for the main results for livers (Table 3.2, column 2).

Table A3.6. Means and standard deviations of key variables by consent system in the estimation sample for hearts

	(1) Total	(2) Opt-in	(3) Opt-out	(4) Diff.	(5) [*p*-value]
Population size (millions)	40.73	63.73	27.31	-36.42	[0.076]
	(61.05)	(88.47)	(32.60)		
Western countries (%)	60.53	64.29	58.33	-5.95	[0.726]
	(49.54)	(49.72)	(50.36)		
Total deaths all causes (pmp, age-standardized)	53.34	50.98	54.51	3.53	[0.500]
	(13.90)	(11.61)	(15.03)		
GDP per capita (current US$/10,000)	21,936.88	25,755.68	19,709.25	-6,046.44	[0.329]
	(18,173.39)	(18,138.93)	(18,201.22)		
Health expenditures per capita (current US$/1,000)	2,101.86	2,920.15	1,692.72	-1,227.43	[0.080]
	(1,983.02)	(2,217.43)	(1,763.07)		
Religious denomination (%)					
None	22.38	21.94	22.67	0.73	[0.916]
	(16.42)	(18.18)	(15.80)		
Muslim	7.08	6.57	7.42	0.85	[0.924]
	(21.36)	(15.01)	(25.23)		
Roman Catholic	37.20	26.03	44.65	18.62	[0.155]
	(31.75)	(23.55)	(34.97)		
Orthodox	9.08	9.32	8.92	-0.40	[0.968]
	(24.23)	(26.92)	(23.24)		
Protestant	7.02	7.92	6.42	-1.50	[0.797]
	(13.85)	(10.34)	(16.09)		
Missing values (%)					
Total deaths all causes	25.76	27.50	24.74	-2.76	[0.810]
	(33.45)	(39.59)	(30.18)		
GDP per capita	2.19	5.95	0.00	-5.95	[0.043]
	(8.81)	(14.03)	(0.00)		
Health expenditures per capita	5.26	14.29	0.00	-14.29	[0.059]
	(22.63)	(36.31)	(0.00)		
Religious denomination	34.21	28.57	37.50	8.93	[0.588]
	(48.08)	(46.88)	(49.45)		
Number of countries	38	14	24		

Notes: Based on the estimation sample for the main results for hearts (Table 3.2, column 3).

Table A3.7. Means and standard deviations of key variables by consent system in the estimation sample for lungs

	(1) Total	(2) Opt-in	(3) Opt-out	(4) Diff.	(5) [*p*-value]
Population size (millions)	46.81	67.93	32.74	-35.19	[0.159]
	(66.47)	(94.09)	(35.83)		
Western countries (%)	66.67	66.67	66.67	0.00	[1.000]
	(47.95)	(49.24)	(48.51)		
Total deaths all causes (pmp, age-standardized)	51.87	50.14	52.92	2.79	[0.608]
	(13.84)	(11.90)	(15.13)		
GDP per capita (current US$/10,000)	26,740.60	30,775.91	24,050.40	-6,725.51	[0.328]
	(18,124.07)	(16,518.04)	(19,095.73)		
Health expenditures per capita (current US$/1,000)	2,474.96	3,033.90	2,102.33	-931.56	[0.220]
	(2,011.66)	(2,178.10)	(1,861.74)		
None	24.22	24.94	23.67	-1.26	[0.858]
	(16.21)	(16.61)	(16.56)		
Muslim	8.21	7.77	8.56	0.79	[0.938]
	(23.28)	(18.66)	(27.06)		
Roman Catholic	34.93	27.16	40.91	13.74	[0.304]
	(31.11)	(22.48)	(36.14)		
Orthodox	6.14	0.75	10.29	9.54	[0.240]
	(18.95)	(0.92)	(24.80)		
Protestant	7.13	7.75	6.65	-1.10	[0.860]
	(14.39)	(10.42)	(17.25)		
Missing (total deaths all causes)	19.00	16.65	20.58	3.93	[0.662]
	(23.53)	(27.14)	(21.47)		
Missing (GDP)	0.00	0.00	0.00		
	(0.00)	(0.00)	(0.00)		
Missing (health expenditures)	0.00	0.00	0.00		
	(0.00)	(0.00)	(0.00)		
Missing (religious denomination)	23.33	16.67	27.78	11.11	[0.498]
	(43.02)	(38.92)	(46.09)		
Number of countries	30	12	18		

Notes: Based on the estimation sample for the main results for hearts (Table 3.2, column 4).

The GDP of a country is known to be a potential confounder in the relationship between consent systems and deceased-donor transplantation rates (e.g., Rithalia et al. 2009). It is likely to be related to organ-patient mortality rates because economically developed countries may, for example, be more able to invest in life-extending medical equipment. We obtain data on countries' annual GDP per capita (in current US$) from the World Bank.[56]

Besides GDP, health spending is also an indicator of a country's wealth, life expectancy, and access to advanced medical equipment. Countries' health expenditures may therefore be related to organ patient mortality as well as their transplantation capacity. Moreover, previous research has found some differences between consent systems with respect to health expenditures per capita (e.g., Abadie & Gay 2006). We obtain data on countries' annual health expenditures per capita (in current US$) from the global health expenditure database from the World Health Organization (WHO).[57]

[56] https://data.worldbank.org/indicator/NY.GDP.PCAP.CD (retrieved on June 1, 2018).
[57] http://apps.who.int/nha/database/Select/Indicators/en (retrieved on June 1, 2018). For all countries, this measure of health expenditures includes all health care goods and services used or consumed during a year,

It is important to control for general mortality rates, because, similar to the rest of the population, organ patients face the risk of dying from another cause besides organ failure. Moreover, previous research has shown that mortality rates affect deceased-donor transplantation rates (see, e.g., Coppen et al. 2005). Annual data on countries' age-standardized death rates per 100,000 world standard population (total deaths, all causes) are obtained from the WHO mortality database.[58]

Previous research has shown that deceased-donor transplantation rates and people's willingness to donate are related to religion (e.g., Gimbel et al. 2003). More specifically, it has been suggested that Catholicism may be associated with favorable attitudes towards organ donation, because the religion officially recognizes organ transplantation as a "service of life" (Rithalia et al. 2009). We obtain country-level data on people's religious denomination from the World Values Survey. This survey is conducted every five years, using a common questionnaire. It is nationally representative of all people aged 18 and older (regardless of their nationality, citizenship, or language) residing within private households. Unfortunately, not every country is included in each wave. To minimize the amount of missing religiosity data, we calculate each country's average share of each religious denomination across waves 4–6 (i.e., 2000–2014). We can do this because religious denominations are relatively stable over time at the country level. In our analyses, we control for the percentage of people in a country who consider themselves to be part of the Roman Catholic denomination.

3.1.2. Data sources for the mechanism section

Transplantations from living donors

We obtain data on the number of kidney and liver transplantations from living donors from the Newsletter Transplant (2001–2015, unbalanced panel). Living heart transplantations are medically not possible and living lung transplantations (i.e., two living donors each provide a lobe to one recipient) are extremely rare.

Removals from the waiting list

The number of patients who died while on a waiting list does not necessarily reveal the total number of patients who die due to organ failure because patients who become too sick to be transplanted are often taken off the waiting list (see, e.g., Charpentier and Mavanur 2008). The probability that these patients die due to organ failure after being taken off the waiting list is high. Therefore, in Section 3.2 on Mechanisms, we use additional data on waiting list removals, which we obtain from various transplantation organizations, for a selection of 16 countries (see Table A3.8). Although the main reason for waiting list removal is that patients became unfit for

i.e., the sum of the domestic general government health expenditures, the domestic private health expenditures, and the health expenditures from external sources. External sources comprise direct foreign transfers and foreign transfers distributed by governments encompassing all financial inflows into the health system from outside the country.

[58] http://apps.who.int/healthinfo/statistics/mortality/whodpms/ (retrieved on June 1, 2018).

transplantation due to health deterioration, patients may also be taken off the list for other reasons. The data on waiting list removals therefore provide an upper bound of the number of deaths after being taken off the waiting list.

Table A3.8. Countries for which information on removals from the waiting list is available

Australia (2010–2016), Austria (2007–2016), Belgium (2007–2016), Croatia (2007–2016), Denmark (2011–2016), France (2010–2015), Finland (2011–2016), Germany (2007–2016), Hungary (2007–2016), Netherlands (2007–2016), New Zealand (2010–2016), Norway (2011–2016), Slovenia (2007–2016), Sweden (2011–2016), United Kingdom (2015), USA (2007–2016).

Sources: (*27, 28, 29, 30, 31, 32*).

Length of the waiting list

We calculate the relative organ-patient mortality rates and deceased-donor transplantation rates by dividing these variables by the number of patients on the waiting list on December 31 in the previous year (per million population) plus the number of patients who entered the waiting list throughout the year. These data are from the Newsletter Transplant (2001–2015, unbalanced panel).

3.2. DATA SELECTION

Due to lack of (reliable) information on the national consent system for organ donation, we had to exclude Egypt, El Salvador, Honduras, Lebanon, Libya, Nicaragua, Palestine, Serbia, and Tunisia from all analyses. Puerto Rico is part of UNOS and deleted to avoid double counting.

In our main specifications, we select countries that did not change their consent system after 1999, because including them likely leads to biased estimates of the opt-out effects on transplantation rates and organ-patient mortality rates. Additionally, we only include country-year observations for which both the number of deceased-donor transplantations and the number of waiting list deaths are non-missing. By doing so, we ensure the same estimation sample for both outcomes, which improves the comparability of the results.

The data on transplantations and waiting list deaths contain many zeros and missing values. Although for some countries it may be true that in some years nobody died while on a waiting list or there were no transplantations, it is unclear whether a reported zero is a "true zero" when data for that country are missing in all other years. We therefore set all values of variable y for organ i in country c to missing when only zeros and missing values are reported.

We also replaced extreme outliers by the average value of variable y for organ i in country c. As a robustness check in the analyses, we truncate variables at the 99th percentile for each organ separately.

The data on our control variables do not always cover all countries and years in our sample. We therefore create a dummy variable for each control variable, which has a value of 1 if country c has a missing value on control variable x in year t, and 0 otherwise, and we replace the missing value in the control variables with the value 0. In all regressions in which we include control variable $x_¸$ we also include the corresponding missing dummy.

In one of the robustness checks, we exclude countries with reports of high levels of organ trafficking. Among the countries included in our data, there have been reports of high incidences of organ trafficking, either as an organ-importing or organ-exporting country, in Australia, Bolivia, Brazil, Canada, Colombia, Egypt, Iran, Israel, Moldova, Peru, Turkey, Ukraine, and the United States (Shimazono, 2007; Shepherd, O'Carroll & Ferguson, 2014).

Table A3.9. Additional information for countries excluded from the sample

Countries	**Additional information**
Egypt, El Salvador, Honduras, Lebanon, Libya, Nicaragua, Palestine, Serbia, Tunisia	Excluded from all analyses due to lack of (reliable) information on the national consent system for organ donation (*4*).
Bulgaria, Czech Republic, Estonia, Finland, Ireland, Slovakia, Slovenia, Switzerland, United Kingdom[59]	Changed their consent system after 1999 (*2-3*). Excluded from the sample in the main analyses, but included in the sample as a robustness check
Australia, Bolivia, Brazil, Canada, Colombia, Egypt[60], Iran, Israel, Moldova, Peru, Turkey, Ukraine, USA	Reports of high levels of organ trafficking as either an organ-importing or organ-exporting country (*4, 38*). Included in the sample in the main analyses, but excluded from the sample as a robustness check.
Algeria, Argentina, Armenia, Azerbaijan, Belarus, Bolivia, Bosnia Herzegovina, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, Egypt,[7] El Salvador,[7] Georgia, Guatemala, Honduras,[7] Iran, Israel, Kuwait, Lebanon, Libya, Macedonia, Malaysia, Mexico, Moldova, Montenegro, Morocco, Nicaragua,[7] Palestine,[7] Panama, Paraguay, Peru, Russia, Serbia,[7] Sudan, Syria, Tunisia,[7] Ukraine, Uruguay, Venezuela, Vietnam	Non-western countries (Latin American, Asian and African countries). Excluded from the sample as a robustness check.

Notes: Puerto Rico is part of UNOS and only listed as a separate entity in Newsletter Transplant 2003. In order to avoid double counting, it is excluded from all analyses.

[59] In 2006, the Human Tissue Act was introduced in Scotland, which made it illegal to remove or store human tissue without consent.

[60] Already excluded from all analyses due to a lack of (reliable) information on the national consent system for organ donation.

3.3. ADDITIONAL TABLES

Table A3.10. The relationship between consent systems and the number of transplantations from deceased and living donors (pmp), and the number of patients who died while on the waiting list plus the number of removals from the waiting list (pmp), by organ

		(1) Kidney	(2) Liver
Total TX	α_{34}: Opt-out (1 = yes)	5.148	7.032***
		(4.381)	(2.572)
	Total deaths all causes (pmp, age-standardized)	-0.148	-0.243*
		(0.162)	(0.141)
	GDP pc (current US$/10,000)	-2.518	-4.141***
		(1.532)	(1.386)
	Health expenditures pc (current US$/1,000)	4.484***	3.390***
		(1.139)	(0.721)
	Religious denomination (% Roman Catholic)	-0.040	-0.200
		(0.164)	(0.131)
	Constant	38.442**	28.220**
		(17.823)	(11.742)
Mortality incl. removals	β_{34}: Opt-out (1 = yes)	-3.550	1.038
		(2.352)	(1.660)
	Total deaths all causes (pmp, age-standardized)	0.138	-0.149**
		(0.117)	(0.072)
	GDP pc (current US$/10,000)	-6.253***	-2.961***
		(1.100)	(0.692)
	Health expenditures pc (current US$/1,000)	6.867***	2.590***
		(0.759)	(0.390)
	Religious denomination (% Roman Catholic)	-0.016	0.043
		(0.102)	(0.059)
	Constant	3.030	12.601*
		(10.787)	(6.461)
Chi^2 test: $\alpha_{34}+\beta_{34}=0$ [*p*-value]		[0.807]	[0.044]
Observations		101	100
Number of countries in the estimation sample		12	13
Mean total TX rate in estimation sample		44.088	15.241
	Standard deviation	*9.675*	*6.895*
Mean organ-patient mortality rate incl. waiting list removals in estimation sample		10.617	5.441
	Standard deviation	*8.712*	*3.428*

Notes: Each column presents the results of a seemingly unrelated OLS regression. The dependent variables in each column are the number of transplantations from deceased and living donors per million population (Total TX) and the number of organ patients who died while on the waiting list per million population plus the number of patients who were removed from the waiting list per million population (Mortality includes removals). All regressions include year fixed effects and dummy variables for missing values on the independent variables. We select countries that did not change their consent system after 1999, and we only include country-year observations for which both the number of transplantations from deceased donors, the number of transplantations from living donors, the number of waiting list deaths, and the number of removals from the waiting lists are non-missing. Robust standard errors clustered at the country level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

Table A3.11. The relationship between consent systems and the number of transplantations from deceased and living donors (pmp) and the number of patients who died while on the waiting list (pmp), relative to the length of the waiting lists, by organ

		(1) Kidney	(2) Liver
Relative total TX	α_{35}: Opt-out (1= yes)	0.013	0.122**
		(0.044)	(0.058)
	Total deaths all causes (pmp, age-standardized)	0.001	-0.001
		(0.001)	(0.003)
	GDP pc (current US$/10,000)	0.092***	0.128***
		(0.019)	(0.026)
	Health expenditures pc (current US$/1,000)	-0.060***	-0.077***
		(0.012)	(0.022)
	Religious denomination (% Roman Catholic)	-0.000	0.001
		(0.001)	(0.001)
	Constant	0.101	0.269
		(0.109)	(0.172)
Relative mortality	β_{35}: Opt-out (1 = yes)	-0.006	-0.017
		(0.006)	(0.013)
	Total deaths all causes (pmp, age-standardized)	-0.000	-0.000
		(0.000)	(0.001)
	GDP pc (current US$/10,000)	-0.002	-0.012**
		(0.004)	(0.005)
	Health expenditures pc (current US$/1,000)	0.001	-0.002
		(0.004)	(0.004)
	Religious denomination (% Roman Catholic)	-0.000	-0.000*
		(0.000)	(0.000)
	Constant	0.037**	0.179***
		(0.016)	(0.033)
Chi^2 test: $\alpha_{35}+\beta_{35}=0$ [*p*-value]		[0.863]	[0.078]
Observations		360	312
Number of countries in the estimation sample		42	36
Mean relative total TX rate in estimation sample		0.292	0.456
	Standard deviation	*0.254*	*0.237*
Mean relative organ-patient mortality rate in estimation sample		0.036	0.090
	Standard deviation	*0.044*	*0.061*

Notes: Each column presents the results of a seemingly unrelated OLS regression. The dependent variables in each column are the number of transplantations from deceased and living donors per million population divided by the number of patients on the waiting list on December 31 of the previous year plus the number of new waiting list entrants throughout the current year per million population (Relative total TX), and the number of organ patients who died while on the waiting list per million population divided by the number of patients on the waiting list on December 31 of the previous year plus the number of new waiting list entrants throughout the current year per million population (Relative mortality). All regressions include year fixed effects and dummy variables for missing values on the independent variables. We select countries that did not change their consent system after 1999, and we only include country-year observations for which both the relative total transplantation rates and the relative mortality rates are non-missing. Robust standard errors clustered at the country level are reported in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

Table A3.12. The relationship between consent systems and the number of transplantations from deceased and living donors (pmp) and the number of patients who died while on the waiting list (pmp), excluding zero values on the dependent variables, by organ

		(1) Kidney	(2) Liver
Total TX	α_{36} : Opt-out (1 = yes)	5.844**	5.108***
		(2.289)	(1.336)
	Total deaths all causes (pmp, age-standardized)	0.098	-0.111
		(0.094)	(0.075)
	GDP pc (current US$/10,000)	1.186	-0.637
		(1.729)	(1.029)
	Health expenditures pc (current US$/1,000)	4.102***	2.661***
		(1.304)	(0.679)
	Religious denomination (% Roman Catholic)	-0.069	0.025
		(0.047)	(0.031)
	Constant	16.796***	8.533*
		(6.371)	(4.525)
Mortality	β_{36}: Opt-out (1 = yes)	-0.738	0.437
		(1.457)	(0.493)
	Total deaths all causes (pmp, age-standardized)	0.008	-0.020
		(0.031)	(0.017)
	GDP pc (current US$/10,000)	-2.356***	-1.056***
		(0.881)	(0.294)
	Health expenditures pc (current US$/1,000)	2.342***	0.991***
		(0.821)	(0.238)
	Religious denomination (% Roman Catholic)	-0.035	-0.012
		(0.031)	(0.013)
	Constant	4.987*	3.668**
		(2.733)	(1.436)
Chi^2 test: $\alpha_{36}+\beta_{36}=0$ [*p*-value]		[0.124]	[0.001]
Observations		518	469
Number of countries in the estimation sample		44	39
Mean total TX rate in estimation sample		28.915	9.852
	Standard deviation	*15.621*	*7.452*
Mean organ-patient mortality rate in estimation sample		10.617	5.441
	Standard deviation	*8.712*	*3.428*

Notes: Each column presents the results of a seemingly unrelated OLS regression. The dependent variables in each column are the number of transplantations from deceased and living donors per million population (Total TX), and the number of organ patients who died while on the waiting list per million population (Mortality). All regressions include year fixed effects and dummy variables for missing values on the independent variables. We select countries that did not change their consent system after 1999, and we only include country-year observations for which both the total transplantation rates and the organ-patient mortality rates are non-missing. Additionally, we exclude zero values on the dependent variables. Robust standard errors clustered at the country level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

Table A3.13. The relationship between consent systems and the number of transplantations from deceased and living donors (pmp) and the number of patients who died while on the waiting list (pmp), by organ. Marginal effects after Tobit regressions.

		(1) Kidney	(2) Liver
Total TX	α_{37} : Opt-out (1 = yes)	4.921**	4.706***
		(2.315)	(1.225)
	Total deaths all causes (pmp, age-standardized)	0.005	-0.100*
		(0.132)	(0.058)
	GDP pc (current US$/10,000)	1.556	-0.668
		(1.798)	(0.866)
	Health expenditures pc (current US$/1,000)	3.691***	2.494***
		(1.411)	(0.575)
	Religious denomination (% Roman Catholic)	-0.043	0.024
		(0.052)	(0.026)
Mortality	β_{37}: Opt-out (1 = yes)	-0.614	0.429
		(1.102)	(0.437)
	Total deaths all causes (pmp, age-standardized)	-0.001	-0.023
		(0.026)	(0.016)
	GDP pc (current US$/10,000)	-1.783**	-0.934***
		(0.709)	(0.253)
	Health expenditures pc (current US$/1,000)	1.809***	0.894***
		(0.653)	(0.207)
	Religious denomination (% Roman Catholic)	-0.024	-0.008
Chi^2 test: $\alpha_{37}+\beta_{37}=0$ [*p*-value]		[0.200]	[0.001]
Observations		534	485
Number of countries in the estimation sample		44	39
Mean total TX rate in estimation sample		28.270	9.632
	Standard deviation	*15.879*	*7.482*
Mean organ-patient mortality rate in estimation sample		4.618	2.026
	Standard deviation	*5.906*	*1.711*

Notes: Each column presents the marginal effects obtained after a seemingly unrelated left-censored Tobit regression. The dependent variables in each column are the number of transplantations from deceased donors per million population plus the number of transplantations from living donors per million population (Total TX), and the number of organ patients who died while on the waiting list per million population (Mortality). All regressions include year fixed effects and dummy variables for missing values on the independent variables. We select countries that did not change their consent system after 1999, and we only include country-year observations for which both the number of transplantations from deceased donors, the number transplantations from living donors, and the number of waiting list deaths are non-missing. Robust standard errors clustered at the country level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

Table A3.14. The relationship between consent systems and the number of transplantations from deceased and living donors (pmp) and the number of patients who died while on the waiting list (pmp), excluding non-Western countries, by organ

		(1) Kidney	(2) Liver
Total TX	α_{38}: Opt-out (1 = yes)	5.553**	6.433***
		(2.510)	(1.357)
	Total deaths all causes (pmp, age-standardized)	-0.032	-0.331***
		(0.137)	(0.071)
	GDP pc (current US$/10,000)	-1.083	-2.992***
		(1.381)	(0.920)
	Health expenditures pc (current US$/1,000)	4.154***	3.312***
		(1.062)	(0.784)
	Religious denomination (% Roman Catholic)	-0.106**	0.033
		(0.041)	(0.032)
	Constant	28.739***	22.291***
		(8.706)	(4.358)
Mortality	β_{38}: Opt-out (1 = yes)	-1.485	0.758
		(1.516)	(0.589)
	Total deaths all causes (pmp, age-standardized)	-0.036	-0.039*
		(0.041)	(0.023)
	GDP pc (current US$/10,000)	-3.375***	-1.593***
		(0.889)	(0.252)
	Health expenditures pc (current US$/1,000)	2.494***	1.208***
		(0.791)	(0.199)
	Religious denomination (% Roman Catholic)	-0.070***	-0.011
		(0.017)	(0.014)
	Constant	10.581***	5.128***
		(3.092)	(1.864)
Chi^2 test: $\alpha_{38}+\beta_{38}=0$ [*p*-value]		[0.245]	[0.000]
Observations		337	309
Number of countries in the estimation sample		23	21
Mean total TX rate in estimation sample		35.803	13.154
	Standard deviation	*13.463*	*3.428*
Mean organ-patient mortality rate in estimation sample		10.617	5.441
	Standard deviation	*8.712*	*7.021*

Notes: Each column presents the results of an OLS regression. The dependent variables in each column are the number of transplantations from deceased donors per million population plus the number of transplantations from living donors per million population (Total TX), and the number of organ patients who died while on the waiting list per million population (Mortality). All regressions include year fixed effects and dummy variables for missing values on the independent variables. We select countries that did not change their consent system after 1999, and we only include country-year observations for which both the number of transplantations from deceased donors, the number transplantations from living donors, and the number of waiting list deaths are non-missing. Additionally, we only include European countries, Australia, Canada, New Zealand, and the United States (i.e., excluding Latin American, Asian, and African countries). Robust standard errors clustered at the country level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

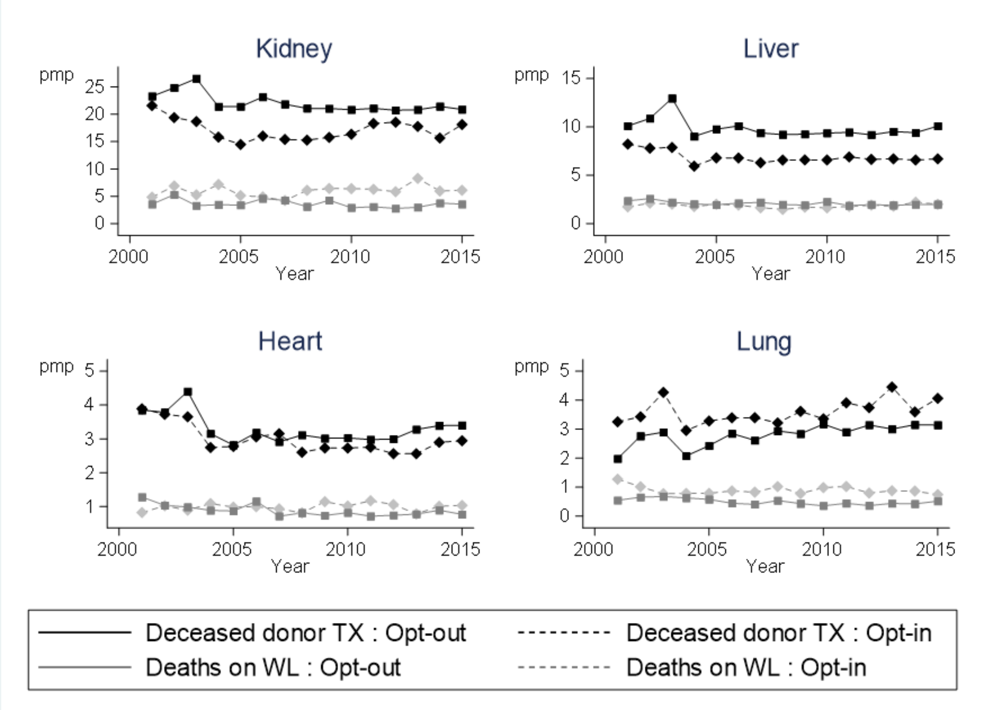
Table A3.15. The relationship between consent systems and the number of transplantations from deceased and living donors (pmp) and the number of patients who died while on the waiting list (pmp), by organ, truncated at the 99th percentile

		(1) Kidney	(2) Liver
Total TX	α_{39}: Opt-out (1 = yes)	4.815**	5.060***
		(2.309)	(1.347)
	Total deaths all causes (pmp, age-standardized)	0.011	-0.106*
		(0.133)	(0.063)
	GDP pc (current US$/10,000)	1.400	-0.701
		(1.693)	(0.935)
	Health expenditures pc (current US$/1,000)	4.016***	2.794***
		(1.252)	(0.613)
	Religious denomination (% Roman Catholic)	-0.037	0.031
		(0.050)	(0.028)
	Constant	20.496***	8.088**
		(7.753)	(3.954)
Mortality	β_{39}: Opt-out (1 = yes)	-0.558	0.483
		(1.247)	(0.494)
	Total deaths all causes (pmp, age-standardized)	-0.003	-0.026
		(0.032)	(0.018)
	GDP pc (current US$/10,000)	-2.301***	-1.020***
		(0.889)	(0.283)
	Health expenditures pc (current US$/1,000)	2.314***	0.974***
		(0.843)	(0.232)
	Religious denomination (% Roman Catholic)	-0.033	-0.009
		(0.029)	(0.012)
	Constant	5.448*	3.792**
		(2.908)	(1.526)
Chi2 test: $\alpha_{39}+\beta_{39}=0$ [*p*-value]		[0.178]	[0.001]
Observations		534	485
Number of countries in the estimation sample		44	39
Mean total TX rate in estimation sample		28.270	9.632
	Standard deviation	*15.879*	*3.428*
Mean organ-patient mortality rate in estimation sample		10.617	5.441
	Standard deviation	*8.712*	*7.482*

Notes: Each column presents the results of an OLS regression. The dependent variables in each column are the number of transplantations from deceased donors per million population plus the number of transplantations from living donors per million population, truncated at the 99th percentile (Total TX), and the number of organ patients who died while on the waiting list per million population, truncated at the 99th percentile (Mortality). All regressions include year fixed effects and dummy variables for missing values on the independent variables. We select countries that did not change their consent system after 1999, and we only include country-year observations for which both the number of transplantations from deceased donors, the number transplantations from living donors, and the number of waiting list deaths are non-missing. Robust standard errors clustered at the country level are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

3.4. ADDITIONAL FIGURE

Figure A3.1. The number of deceased-donor transplantations and waiting list deaths per million population, by organ, by year, and by consent system



Notes: The graphs are based on the estimation samples for the results shown in Table 3.2.
Source: Authors' calculations based on Newsletter Transplant 2002-2016.

REFERENCES

1. Organización Nacional de Trasplantes, *Publicaciones.* Retrieved from http://www.ont.es/publicaciones/Paginas/Publicaciones.aspx on Dec 9, 2016.
2. R. Bouwman, J. Lie, M. Bomhoff, R.D. Friele. *The ACTOR-study: Study on the set-up of organ donation and transplantation in the EU Member States, uptake and impact of the EU Action Plan on Organ Donation and Transplantation (2009-2015)*, Utrecht: NIVEL (2013).
3. A.M. Rosenblum, L.D. Horvat, L.A. Siminoff, V. Prakash, J. Beitel, A.M. Garg. The authority of next-of-kin in explicit and presumed consent systems for deceased organ donation: an analysis of 54 nations. *Nephrology Dialysis Transplantation* **0**, 1-14 (2011).
4. L. Shepherd, R.E. O'Carroll, E. Ferguson. An international comparison of deceased and living organ donation/transplant rates in opt-in and opt-out systems: a panel study. *BMC Medicine* **12** (131) (2014).
5. K. Habchi, A. Benziane, K. Toumi, B. Mohamed. Panorama of Transplantation in Algeria, In: ELPAT (ed), *4th ELPAT Congress: Ethical, Legal and Psychosocial Aspects of Transplantation, Global Challenges, Programme & Abstract Book*, 96 (2016).
6. G. Balbinotto Neto, E. Nunes da Silva, A.K. Campelo. The impact of presumed consent law on organ donation: an empirical analysis from quantile regression for longitudinal data. *Working Paper of Public Health* **3**, 1-22 (2012).
7. G. Balbinotto Neto, E. Nunes da Silva, A.K. Campelo. The impact of presumed consent law on organ donation: an empirical analysis from quantile regression for longitudinal data. *Berkely Program in Law & Economics* (2007).
8. R. Mizraji, I. Alvarez, R.I. Palacios, C. Fajardo, C. Berrios, F. Morales, E. Luna, C. Milanés, M. Andrade, E. Duque, F. Giron, J. Alfonso, S. Herra, C. Soratti, R. Ibar, V.D. Garcia. Organ donation in Latin America. *Transplantation Proceedings* **39** (2), 333–335 (2007).
9. D. Austen, E. Skewes. Organ Donation and Transplantation, In: D. Elliott, L. Aitken, W. Chaboyer (eds.) *ACCCN's Critical Care Nursing*, 746-762 (2012).
10. C. Bigg. 'Gutted like a pig': grieving mother takes on Russia's organ donation system, Eurasianet.org. Retrieved from http://www.eurasianet.org/node/77806 on Dec 19, 2016.
11. P. Flores. Culture hinders organ donations, transplants in Bolivia, Honolulu Star Advertiser (2015). Retrieved from http://www.staradvertiser.com/2015/02/21/breaking-news/culture-hinders-organ-donations-transplants-in-bolivia/ on 19 Dec 2016.
12. G.O. Phillips. *Legal Basis of Global Tissue Banking: A Proactive Clinical Perspective,* World Scientific (2015).
13. N. Hicko. Croation model the possible answer to lack of donors in Balkan countries? Balkan Science Report (2013). Retrieved from http://www.balkansciencereport.com/2013/09/croatian-model-possible-answer-to-lack.html. on Dec 19, 2016.
14. G. Kekenadze, I. Javakhishvili, D. Kekenadze, S. Rustaveli. Legal and Ethical Aspects of Organ Transplantology in Georgia. Global international scientific analytical project (2011). Retrieved from http://gisap.eu/node/1028#. on Dec 19, 2016.

15. R. Garcia-Gallont, R. Matesanz, F. L. Delmonico. Organ donation and transplantation in Central America. *Transplantation* **99** (3), 459-460 (2015).
16. Bundeszentrale für gesundheitliche Aufklärung. *The Law on Organ Donation in German: A Declaration-based Solution.* Retrieved from https://www.organspende-info.de/sites/all/files/files/BZGA-15-02525_Flyer_Entscheidungslosung_EN.pdf on Dec 19, 2016.
17. Global Observatory on Donation and Transplantation. Region: The Americas, Country: Dominican Republic. Retrieved from http://ont.es/americas/Organizational%20and%20legal%20aspects/Dominican%20Republic%20profile.pdf. on Dec 19, 2016.
18. S.M. Kazemeyni, A.R. Bagheri, A.R. Heidary. Worldwide cadaveric organ donation systems (transplant organ procurement). *Urology Journal* **1** (3), 157-164 (2004).
19. M. Tumin, K. Tafran, R. Hazli Zakaria, N.H. Mohd Satar, K. Peng Ng, S. Kun Lim. Do family members of dialysis patients have a positive attitude toward organ donation? *Annals of Transplantation* **20**, 752-756 (2015).
20. Parliamentary Secretariat for Health. *Proposal for Legislation on Organ and Tissue Donation: A Maltese Government White Paper* (2015) Retrieved from https://socialdialogue.gov.mt/en/Public_Consultations/MEH-HEALTH/Documents/L-05-2015%20-%20White%20Paper%20on%20Organ%20and%20Tissue%20donation/Organ%20Donation-White%20paper%20for%20consultation%20(2).pdf on Dec 19, 2016.
21. European Court of Human Rights. Case of Petrova v. Latvia. Application no. 4605/05 (2014).
22. E. Kelly. International organ trafficking crisis: solutions addressing the heart of the matter. *Boston College Law Review* **54** (3), 1317-1349 (2013).
23. UNESCO Cairo Office. Ethics and Law in Biomedicine and Genetics: An Overview of National Regulations in the Arab States. Cairo, Egypt: United Nations Education, Scientific and Cultural Organization (2011).
24. S.G. Stetsenko, V.Y. Stetsenko, I. C. Senyuta. *Medical law Ukraine* (2008)
25. Republica Oriental del Uruguay, *Donación y Transplante de Células, Órganos y Tejidos* (Ley Nº 18.968, 2012) Retrieved from http://200.40.229.134/Leyes/AccesoTextoLey.asp?Ley=18968&Anchor= on Dec 19, 2016.
26. Socialist Republic of Vietnam. Law on donation, removal and transplantation of human tissues and organs and donation and recovery of cadavers (2006).
27. Statistics Eurotransplant. Retrieved from http://statistics.eurotransplant.org/index.php?search_type=WL+removals&search_organ=by+organ&search_region=by+country&search_period=2007&search_characteristic=removal+reason&search_text=9023 on Oct 23, 2017.
28. Statistics Scandiatransplant. Retrieved from http://www.scandiatransplant.org/data. on Oct 23, 2017.

29. Statistics from the Agence Biomedicine. Retrieved from: https://www.agence-biomedecine.fr/Procurement-and-transplantation-of,530?lang=fr on Oct 23, 2017.
30. Statistics on UK organ donation. Retrieved from: https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/4496/section_5_kidney_activity.pdf on Oct 23, 2017.
31. Statistics from the OPTN. Retrieved from: https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/# on Oct 23, 2017.
32. Statistics on organ donation in New Zealand and Australia. Retrieved from: http://www.anzdata.org.au/anzod/ANZODReport/2016/2016-anzod-12-waitinglist_v2.0_20170114.pdf on Oct 23, 2017.
33. A. Rithalia, C. McDaid, S. Suekarran, L. Myers, A. Sowden. Impact of presumed consent for organ donation on donation rates: a systematic review. *BMJ* **338**, 1-8 (2009).
34. A. Abadie, S. Gay. The impact of presumed consent legislation on cadaveric organ donation: a cross-country study. *Journal of Health Economics* **25** (4), 599-620 (2006).
35. R. Coppen, R. Friele, R. Marquet, S. Gevers, S. Opting-out systems: no guarantee for higher donation rates. *Transplant International* **18**, 1275-1279 (2005).
36. R. Gimbel, M. Strosberg, S. Lehrman, E. Gefenas, F. Taft. Presumed consent and other predictors of cadaveric organ donation in Europe. *Progress in Transplantation* **13** (1), 17-23 (2003).
37. K. Charpentier, A. Mavanur. Removing patients from the liver transplant waiting list: a survey of US liver transplant programs. *Liver Transplantation* **14**, 303-307 (2008).
38. Shimazono, Y. (2007). The state of the international organ trade: a provisional picture based on integration of available information. Bulletin of the World Health Organization 85 (12), 901-980.

Samenvatting

Alle daden hebben gevolgen. Sommige gevolgen worden opzettelijk bereikt, terwijl andere onopzettelijk zijn, en sommige zijn positief voor individuen of de samenleving, terwijl andere nadelig zijn. Hetzelfde geldt voor overheidsbeleid en voor sociale interventies: hoewel het de bedoeling is om daarmee de samenleving te verbeteren, hebben deze daden soms ook onbedoelde gevolgen. De beoogde effecten worden de directe effecten genoemd, terwijl andere uitkomsten indirecte effecten worden genoemd. Niet alleen directe effecten, maar ook indirecte effecten van overheidsbeleid en interventies dienen zorgvuldig te worden gemonitord en geëvalueerd, zelfs als eerder onderzoek of gezond verstand suggereert dat de indirecte effecten wenselijk zijn. Dat is de belangrijkste conclusie van dit proefschrift, dat de directe en bepaalde indirecte effecten evalueert van een quasi-veldexperiment dat kinderen stimuleert om meer te bewegen in hun dagelijks leven (het Active Living Programma), en overheidsbeleid dat mensen stimuleert om na hun overlijden hun organen te doneren (opt-out toestemming voor orgaandonatie).

Het is belangrijk om te benadrukken dat beide beleidsmaatregelen de meeste van hun directe effecten succesvol bereiken: het Active Living Programma verhoogt met succes de tijd die kinderen tijdens schooltijd aan bewegen besteden, en opt-out toestemming voor orgaandonatie is gerelateerd aan significant meer nier-, lever-, en harttransplantaties van overleden donoren. Beide beleidsmaatregelen blijken echter ook ongewenste indirecte effecten te hebben.

Het Active Living Programma verslechtert onbedoeld schoolprestaties, vooral onder de slechtst presterende leerlingen en bij jongens (Hoofdstuk 1). Bovendien leidt het Programma tot een significante verhoging van de prevalentie van Attention-Deficit/Hyperactivity Disorder (ADHD)-achtige symptomen bij jongens (Hoofdstuk 2). De resultaten van deze twee hoofdstukken benadrukken dat beleidsmakers en onderzoekers zeer voorzichtig moeten zijn met het generaliseren van eerder gevonden gewenste (neven) effecten naar verschillende contexten en beleidsontwerpen: de studies in Hoofdstukken 1 en 2 zijn een van de eersten die de causale indirecte effecten identificeren van het stimuleren van lichamelijke activiteit in het dagelijks leven bij kinderen.

Hoofdstuk 3 analyseert of een beoogd indirect beleidseffect wordt bereikt door te onderzoeken of de sterftecijfers onder orgaanpatiënten inderdaad lager zijn in landen met een opt-out toestemmingssysteem voor orgaandonatie. Dit is een indirect effect, omdat opt-out systemen niet letterlijk levens kunnen redden: sterftecijfers kunnen alleen worden beïnvloed via andere mechanismes, zoals transplantaties van overleden donoren. De sterftecijfers van orgaanpatiënten blijken echter nauwelijks te verschillen tussen toestemmingssystemen. Bovendien betekent voor nieren en levers een extra transplantatie van een overleden donor niet per sé dat een leven van een orgaanpatiënt wordt gered. Dit duidt op het bestaan van factoren gerelateerd aan opt-out toestemmingssystemen die positief gerelateerd zijn aan nier- en leverpatiëntsterfte. Het hoofdstuk laat zien dat zelfs de schijnbaar voor de hand liggende relaties een expliciete evaluatie verdienen.

Résumé

Toutes les actions ont des conséquences. Certaines conséquences sont intentionnelles tandis que d'autres ne le sont pas, et certaines ont des conséquences bénéfiques pour les individus ou la société alors que d'autres non. Il en va de même pour les politiques publiques et pour les interventions sociales : bien que l'intention soit d'améliorer la société, parfois ces actions ont aussi des conséquences non-intentionnelles. Les effets escomptés sont appelés effets directs, tandis que les autres résultats sont appelés effets indirects. Non seulement les effets directs, mais aussi les effets indirects des politiques et interventions publiques doivent être soigneusement suivis et évalués, même lorsque des recherches antérieures ou le bon sens suggèrent que les effets indirects sont souhaitables. C'est le principal point à retenir de cette thèse de doctorat, qui évalue les effets directs et des effets indirects d'une quasi-expérience de terrain qui encourage les enfants à devenir plus actifs dans leur vie quotidienne (le *Programme Vie Active*), et une politique publique qui incite les gens à donner leurs organes à leur décès (*consentement présumé pour le don d'organes*).

Il est important de souligner que les deux politiques réussissent à atteindre la plupart de leurs effets directs : le Programme Vie Active augmente avec succès le temps que les enfants consacrent à l'activité physique pendant le temps scolaire, et le consentement présumé pour le don d'organes est lié à un nombre de transplantations des reins, foies et cœurs significativement plus élevé provenant de donneurs décédés. Cependant, ces deux programmes semblent également avoir des effets indirects indésirables.

Le Programme Vie Active diminue involontairement le rendement scolaire, en particulier chez les élèves les moins performants et chez les garçons (Chapitre 1). De plus, le Programme augmente significativement la prévalence des symptômes de type Trouble Déficitaire de l'Attention/Hyperactivité (TDAH) chez les garçons (Chapitre 2). Les résultats de ces deux chapitres soulignent que les responsables des politiques sociales et les chercheurs doivent être très prudents en généralisant les effets (secondaires) souhaités trouvés dans des recherches ultérieures à des contextes différents : les études des Chapitres 1 et 2 sont parmi les premières à identifier les effets causals indirects de l'encouragement de l'activité physique au quotidien chez les enfants.

Le Chapitre 3 analyse si un effet indirect escompté d'une politique publique est atteint, en étudiant si les taux de mortalité des patients d'organes sont effectivement inférieurs dans les pays qui utilisent un système de consentement présumé pour le don d'organes. Il s'agit d'un effet indirect, car les systèmes de consentement présumé ne peuvent pas littéralement sauver des vies : les taux de mortalité ne peuvent être affectés que par d'autres mécanismes, tels que les taux de transplantation de donneurs décédés. Cependant, les taux de mortalité des patients d'organes semblent à peine différer entre les systèmes de consentement. De plus, pour les reins et les foies, une transplantation supplémentaire de donneur décédé n'implique pas qu'une vie de patient d'organe soit sauvée. Cela indique l'existence de facteurs liés aux systèmes de consentement présumé qui sont positivement liés à la mortalité des patients du rein et du foie. Le chapitre démontre que même les relations apparemment évidentes méritent une évaluation explicite.

Curriculum Vitae

Annelore Verhagen was born on 17 March 1987 in Emmen (the Netherlands). She obtained her Bachelor of Science in Public Administration at Radboud University Nijmegen in 2011. She then pursued a Master of Science in Public Administration with a specialization in Policy Analysis and Advice at the same university. She graduated with honours in 2012, with a Master thesis on hospital staff's communication strategies to ask for consent for organ donation to next-of-kin. She wrote this thesis during a research internship at the Netherlands Institute for Health Services Research (NIVEL) in Utrecht.

Between April 2012 and June 2015, Annelore was employed as a researcher at the Research Centre for Education and the Labour Market (ROA) at Maastricht University. Her work focused on early school-leavers, youngsters not in employment education or training (NEETs) and school-to-work transitions of graduates from vocational education and training (VET).

In June 2015, Annelore became a PhD candidate at the Department of Economics at Maastricht University. Since September 2018, she is also employed as an economist in the Skills Team in the Directorate for Employment, Labour and Social Affairs (ELS) at the Organisation for Economic Co-operation and Development (OECD) in Paris. Her work in the team focuses on understanding which skills are needed in the workplace, how skill needs change due to trends such as increasing digitalization, globalization and population ageing, and how adults' regular participation in learning activities can prepare them for this changing world of work.