HEALTH, ENVIRONMENT, AND BEHAVIOR

Four Empirical Essays

DISSERTATION

zur Erlangung des Doktorgrades

"Dr. rer. pol."

der Fakultät für Wirtschaftswissenschaften

der Universität Duisburg-Essen

vorgelegt

von

Maryna Ivets

aus

Grodno, Belarus

Betreuerin:

Prof. Dr. Katharina Blankart

Lehrstuhl für VWL, inbs. empirische Gesundheitökonomik

Essen, Dezember 2022

DOI: 10.17185/duepublico/81759
URN: urn:nbn:de:hbz:465-20240320-064914-6

For my family.

Erstgutachter	Prof. Dr. Katharina Blankart
Zweitgutachter	Prof. Dr. Ansgar Wübker
Tag der mündlichen Prüfung:	16. März 2023

Acknowledgements

I would like to thank my supervisor and mentor, Katharina Blankart, who supported me, provided great encouragement, guidance, advice, and instilled confidence in moving this thesis forward. I am very thankful to her for this. I would also like to thank Ansgar Wübker for readily agreeing to be my second supervisor and for his invaluable feedback and discussion of my work.

I also would like to thank my coauthors Daniel Avdic, Martin Karlsson, Therese Nilsson, and Ieva Sriubaite, with special mention going to Nina Boberg-Fazlić and Linda Hirt-Schierbaum. I feel very fortunate that we had an opportunity to work together and I learned a great deal from each of you. Our collaboration was instrumental in building this thesis chapter by chapter. Additionally, I would like to thank Martin Karlsson for the opportunity to learn and grow as a researcher at the health economics research center CINCH.

Special mention and thanks also go to Stephan Sommer, Franziska Then, and Christiane Wuckel. Your feedback, advice, and friendship were very important to this project. I also would like to thank my colleagues at CINCH, RWI Leibniz Institute for Economics Research, and Ruhr Graduate School of Economics (RGS Econ) who made this time stimulating and fun. Special thanks go to Tim Brand, Natalia Bula, Simon Decker, Fang Duan, Eva Goetjes, Johann Han, Christoph Kronenberg, Sarah Okoampah, Francesca Paolini, and Nina Schwarz. The list can continue. Some of you became personal friends and I feel very fortunate about that. I truly thank you all.

I also would like to thank Eva Zielke for her always professional and friendly administrative support.

Last, but not least, I would like to thank my family and friends for their unconditional support, endless patience, and unwavering love. Thank you for always being there for me and for giving me space and time to pursue my goals. Heartfelt thanks go to my mom Zoya, my brother Mikhail and my grandma Lena. Though physically far away I always felt your presence, support and love by my side. I am especially thankful to my husband – my biggest advocate, supporter, and co-pilot – Bob, who has always been by my side no matter how out-of-the box my ideas have been that landed us in Germany; our two little passengers, Boris and Lucy, who joined us along the way and made this endeavor more exciting; my immediate and extended family in Belarus and in the US; and all those dear who saw the beginning of this work, but not the end. I love you and I am extremely grateful to you all.

From those I inadvertently forgot to mention I ask for forgiveness and understanding.

Contents

List of Tables

List of Figures

1 Introduction

In 2015 the United Nations (UN) member states agreed on 17 different Sustainable Development Goals (SDGs) to be reached by 2030 (United Nations, 2015). Each of these goals outlines an area where global improvements were deemed necessary for sustainable social, economic, and environmental development. Health and well-being was recognized as one of the critically important areas for humanity and the planet. A specific SDG was dedicated to this area.

The proposed SDGs highlight the interconnectedness of environmental, social, and economic aspects of sustainable development. The goal of good health and well-being requires addressing poverty, improving nutrition, ensuring availability of clean water and sanitation, investing in quality education, reducing inequities, and taking actions to combat climate change and to protect the natural environment. Simultaneously, attaining the goals related to education and economic growth relies on having healthy populations to carry them out. Thus, the SDGs underline the connection between our actions, our environment, and our health.

This connection implies the need to take an ecological perspective to health promotion and disease prevention. Specifically, we need to look at health determinants on both an individual and population level. Our health is determined by a range of individual, social, economic, and environmental factors, which include social and physical determinants of health – the conditions in the environments where we are born, grow, live, learn, work, and age; access to and quality of health services; individual health behavior; biological and genetic factors; and policymaking. The latter – if implemented correctly using a solid evidence base – can improve our health on the population level and our individual health behavior, allowing us to reach our full potential.

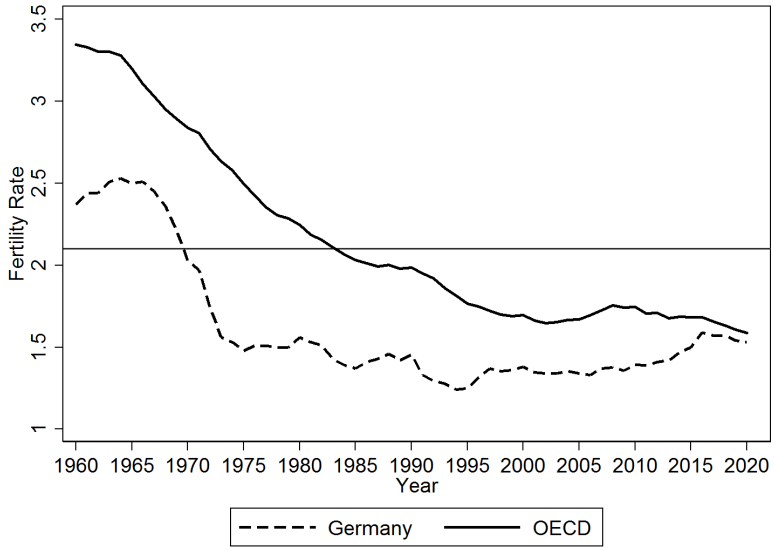
This thesis deals with the topic of health, environment, and behavior. The chapters are united by the practical value of the presented evidence. It could be used for better political and individual decision-making in real-world settings by those interested in ensuring economic growth and sustainability of social security systems, reducing inequalities, and providing equitable and efficient healthcare and early childcare.

To assist with these goals Chapter 2 looks at how the 1918-19 influenza pandemic influenced fertility behavior; Chapter 3 studies individual weight loss behavior in today's obesogenic environment; Chapter 4 explores how physicians' environment influences their treatment behavior; and Chapter 5 investigates how a change in social environment for young children due to early childcare policy affects behavior. The following provides details on the studies, their contributions and relevance, and contextualizes them in relation to current global challenges.

The first study (Chapter 2) of this thesis explores the consequences of a severe health shock on fertility using the 1918-19 influenza pandemic in Sweden. The study gives valuable insights that can be applied to the COVID-19 pandemic.

It is well established that the COVID-19 pandemic affected diverse areas from health to economy. It is also expected to affect demography beyond the immediate loss of life. Specifically, early assessments indicate that COVID-19 is expected to negatively influence fertility in high-income countries (see, e.g., Aassve *et al.*, 2020, 2021; Cohen, 2021; Sobotka *et al.*, 2021), but the lack of long-term data on this outcome presents challenges to decision-makers wanting to understand its future development. This is concerning, because pandemic-induced fertility change could influence future rates of population aging and economic growth. This makes a well-informed policy response particularly important, because it would have implications for the sustainability of pay-as-you-go social security systems in European countries, which already experience fertility rates below the replacement rate of 2.1 births per woman (OECD, 2022d) (see Figure 1.1).

Figure 1.1: Fertility rates (*OECD Statistics*).



Policymakers could look for insights in similar historical events to prepare an adequate policy response to the crisis. The 1918-19 influenza pandemic provides a unique opportunity to study fertility dynamics following a severe, unanticipated morbidity and mortality shock. We use rich administrative historical data from Sweden – a country during demographic transition – to analyze both the short- and long-term effects. We find evidence of a small baby boom following the pandemic's conclusion, but this effect was second-order compared to the negative long-term fertility effect. In this chapter we also explore compositional changes within the net fertility effect and look at the mechanisms responsible for the fertility change after the pandemic. We find evidence of a negative income effect and

increased economic uncertainty underlying the importance of economic conditions for the long-term fertility decline after the pandemic.

It became obvious that COVID-19 crisis threatened not only the SDG of health and well-being, but also other SDGs. The findings from Chapter 2 have implications for reaching the SDGs regarding future prospects of economic growth. Though there are no straightforward conclusions from this historical evidence due to a number of differences between the current and historical contexts,[1] however, it is nevertheless heuristically useful and provides insights into potential mechanisms and long-term effects. It implies that pandemic-induced economic uncertainty contributed to a change in family formation behavior. This can have far-reaching implications for future economic growth as pandemic could lead to fertility decline, which contributes to greater population aging and slower labor force growth and, as the result, to slower productivity and economic growth. Thus, the study provides important historical evidence that could be used by policymakers in concert with the emerging evidence on COVID-19 to predict potential long-term fertility effects from the recent pandemic, and to decide on the best policies to mitigate potentially negative demographic consequences.

Potential policy actions to address the negative fertility problem driven by the pandemic-induced economic uncertainty could be to reconsider existing policies aiming to support childbearing and childrearing by reducing its costs and by supporting families, i.e., extending public funding of family-related programs (child allowance, child benefit and parental allowance), strengthening early childcare services, and ensuring compatibility of family and work life.

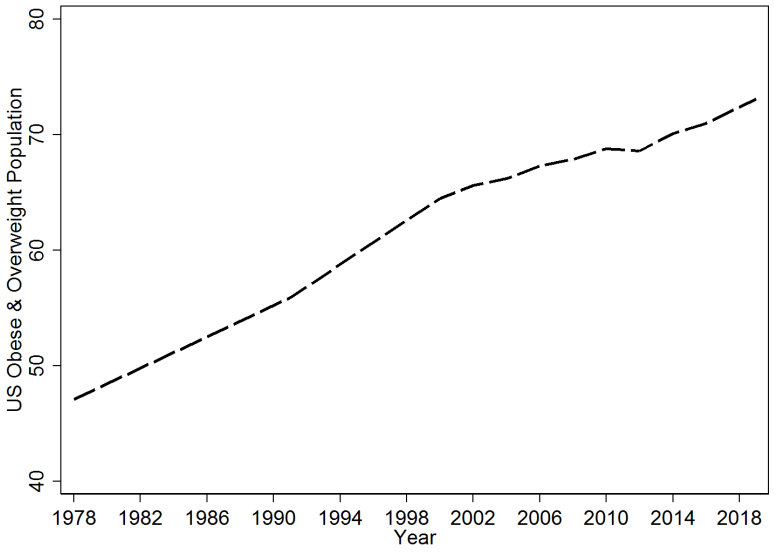
The second study (Chapter 3) deals with individual weight loss behavior in the context of an obesogenic environment. This term refers to how our surroundings, opportunities, and other life conditions influence obesity in individuals and populations (Swinburn *et al.*, 1999). The study provides evidence that could be used to address the obesity and overweight problem.

The steady rise in obesity and overweight rates at a population level (see, e.g., Figure 1.2 for the US) points toward the broad change in the environment we live in, i.e., the increased availability, affordability and intense marketing of energy-dense foods (fast foods), sedentary lifestyles, mechanized commuting, and increased urbanization (Allen, 2017; Swinburn *et al.*, 2011; WHO, 2021).

Obesity contributes heavily to a growing global non-communicable disease (NCD) burden. It is a major cause of cardiovascular diseases, cancer, diabetes, and related deaths.

[1] There are certain differences in contexts between COVID-19 and the 1918-19 influenza pandemic that should be considered. Specifically, socioeconomic conditions of early 20th century Sweden are different from high-income countries today, but could be similar to some low- or middle-income countries. Also, while the global economic prospects were already uncertain due to the COVID-19 pandemic, the war in Ukraine is likely to make the matters worse (OECD, 2022b). Additionally there are differences in health systems, access to contraceptives, family, and social security policies compared to more than 100 years ago. Additionally, the progress in developing effective vaccines and a limited mortality toll in working-age adults may imply a different scenario for COVID-19 regarding the persistence of the economic decline as compared to the 1918-19 influenza (see Beach *et al.*, 2022). This suggests that COVID-19's short- and long-term fertility response in various countries could be driven by different mechanisms (see Aassve *et al.*, 2020; Luppi *et al.*, 2020, for a discussion).

Figure 1.2: Obese & overweight population (USA) (*OECD Statistics*).



Treating diseases related to overweight and obesity is expected to cost the Organization for Economic Cooperation and Development (OECD) countries 201$ per person per year on average, which equates to 8.3% of total health spending (OECD, 2019). Thus, obesity-related heath conditions place a sizable financial burden on healthcare systems throughout the world, contributing to the increase in healthcare expenditures (HCEs). As such, obesity and overweight present a major global threat not only to health, but also to social and economic development by increasing inequalities – since they disproportionately affect vulnerable and socially disadvantaged populations (Robertson, 2014), threaten sustainability of social security and healthcare systems, inhibit economic growth and, thus, contribute to holding people and society in poverty. This makes actions that fight obesity and overweight pertinent to achieving SDGs.

While various environmental factors contribute to the obesity and overweight epidemic – potentially compromising individuals' ability to make better decision regarding diet and exercise – people are not powerless regarding their lifestyle choices. Individual behavior is an important factor for understanding why certain people with a preference for commitment to a healthier lifestyle might fail to commit efficiently. Chapter 3 investigates why some people fail to commit efficiently when attempting to lose weight. It adapts the results from a theoretical model that introduces heuristic bias in individuals' decision-making when using a self-commitment mechanism – an investment-payoff combination – in the setting of a real-world weight loss program DietBet, and tests the conclusions with its data. It suggests that commitment success depends on how accurately agents predict their future self-control costs and payoffs. Specifically, naifs and naive optimists (overconfident agents) are more likely to overestimate their self-control costs and payoffs – making them more likely to fail to commit efficiently; while sophisticates and naive pessimists (under-

confident agents) are more likely to estimate accurately or underestimate them – making them more likely to commit efficiently or overcommit. To distinguish agent types we apply "false hope syndrome" and "fresh start effect" to popular but infamously unsuccessful New Year's resolutions; the observation of people seeking commitment before Christmas to avoid holiday weight gain; and the observations of gender differences in overconfidence. We show that overconfident individuals can increase their chances of successful weight loss by making greater investments.

The study could inform actions to prevent obesity and overweight by targeting overconfident agents, i.e., men and people engaging in self-improvement activities in January, to prevent their potential exploitation by the market and help them to commit successfully to a positive lifestyle change. Also, encouraging commitment ahead of certain holidays associated with high risk of weight gain, e.g., before Christmas, could prevent people from gaining excessive weight. Furthermore, the study's insights could be used for assessing digital health technologies aiming to encourage positive behavioral change.

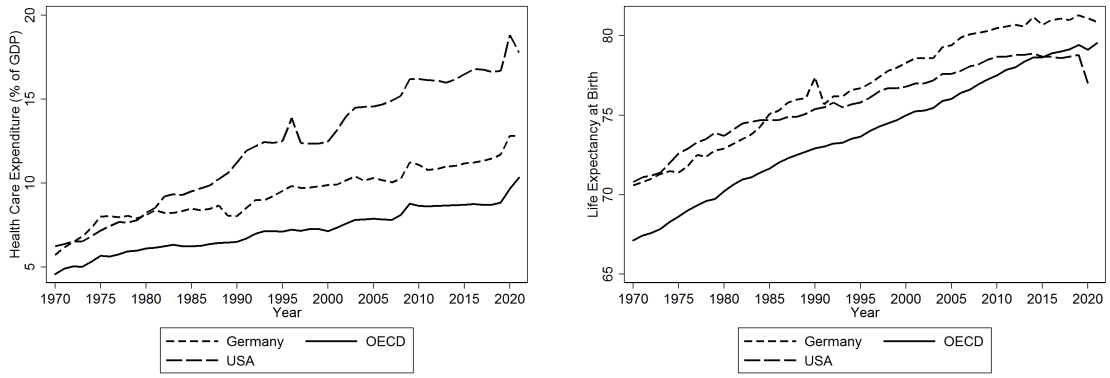
Although the results are important for designing policies and interventions to fight obesity and overweight, given the complexity of the problem, there is no single solution. While Chapter 3 provides insights into individual weight loss behavior in the context of the current obesogenic environment and shows that individuals are not powerless to make healthier lifestyle choices, policymakers should address the upstream environmental drivers of obesity and overweight. They should do this by creating environments conducive to making healthier lifestyle choices that are affordable and accessible to everyone by implementing evidence-based policies (WHO, 2021). In the meantime, until changes that reverse the obesity and overweight epidemic are implemented, the growing NCD burden makes it even more important to optimize healthcare services.

The third study (Chapter 4) looks at how physicians' practice environment affects their treatment behavior. This study adds to the evidence base about determinants of variation in healthcare use.

Healthcare expenditures have been steadily increasing over the last 50 years (see Figure 1.3a). For example, from 1970 to 2020 the percentage of GDP attributed to HCEs in the OECD countries more than doubled, constituting 9.7% or 4,272$ per capita. However, high HCEs are not always connected to better health outcomes. For example, in 1970 the US and Germany had similar HCEs of 6.2% (327$ per capita) and 5.7% of their GDP (252$ per capita) (OECD, 2022a), respectively; and similar life expectancy at 70.8 and 70.6 years (OECD, 2022c), respectively (see Figures 1.3a and 1.3b). By 2019 the US was spending 16.8% of their GDP (10,856$ per capita), while Germany – 11.7% (6,408$ per capita), but had a lower life expectancy than Germany at 78.8 vs 81.3 years, indicating that higher HCEs do not guarantee better results.

Great geographic variations in HCEs exist not only between countries, but also within them. For example, in Germany rates of cardiac procedures in high-intensity areas are twice those in low-intensity areas (OECD, 2014a). While some of these differences could be attributed to demand-side factors, they are usually unable to explain the substantial geographic variation in healthcare use (see, e.g., Chandra *et al.*, 2011; Cutler *et al.*, 2019; Finkelstein *et al.*, 2016; Skinner, 2011; Skinner *et al.*, 2011; Wennberg and Gittelsohn, 1973).

Figure 1.3: Healthcare expenditure (as % of GDP) and life expectancy at birth (*OECD Statistics*).



(a) Healthcare expenditures

(b) Life expectancy

On one hand, increasing HCEs could indicate greater investment in health and promote economic growth. However, the observed weak correlation between HCEs and health outcomes indicates certain inefficiency in the healthcare sector. Therefore, the growing HCEs represent a threat to social and economic development by compromising the sustainability of healthcare systems, potentially increasing inequalities, and inhibiting economic growth. This makes the actions addressing unwarranted HCEs pertinent to achieving SDGs.

Improving efficiency of healthcare delivery necessitates looking at the determinants of this unwarranted variation on the supply side. As a result, policymakers and researchers have increasingly shifted their focus to physicians' behavior and their role in healthcare delivery, since physicians are at the front line of medical decision-making.

In this study we look at physicians' behavior and how it is influenced by their environment. We separate physicians' environment into a physical (hospital where they work) and a social environment – influenced by the treatment styles of their peers. To disentangle how physicians' practice styles are influenced by hospital and peer group factors, we compare the stent choices of cardiologists moving between hospitals over time. We find that cardiologists rapidly adapt their stent choices to the new practice environment after relocation, and that they are equally driven by the hospital and peer environments. In contrast, treatment costs and quality of care are largely unchanged despite the altered treatment styles.

The finding that physicians equally adapt their treatment styles to their hospital and peer environments indicates that physicians' behavior is an important determinant of the geographical variation in treatment choices. The absence of an impact on patient outcomes is also in line with the puzzling observation of a weak correlation between the regional variation in HCEs and quality of care. Potential policy advice could relate to harmonizing information about treatment choices across healthcare professionals, i.e., ensuring greater adherence to clinical guidelines, and investing in additional physician training.

The results are relevant for policymakers concerned about increasing HCEs, and those interested in equitable and efficient provision of healthcare services, especially in the land-

scape of growing financial pressures on social security systems. Reducing the unwarranted variation in HCEs that are not always explainable by demand-side factors like patient preferences and needs, might be possible by better understanding physicians' behavior and how it is affected by their environment.

The fourth and final study (Chapter 5) looks at how an early childcare (ECC) policy shift in Germany changed children's early social environment, and how these changes subsequently influenced their behavior. The study provides evidence that could be used to improve ECC policies and parental decision-making regarding ECC use.

ECC policies are becoming increasingly popular, with many countries expanding access to ECC for children under age three. This development dramatically changes the early social environment for children by having them spend extensive amounts of time in out-of-family childcare from an early age. Public policies, such as the German policy studied, induce these changes, which can influence children's development and well-being.

Early childhood and its associated conditions have a notable effect on child development. It is recognized that the first postnatal years are very important for children's development due to rapid brain growth and its sensitivity to external influences (CDC, 2022), with lifelong effects. As such, ECC policy represents an environmental change during these early years that can play a vital role in children's development, highlighting the importance of studying the effects of ECC policy.

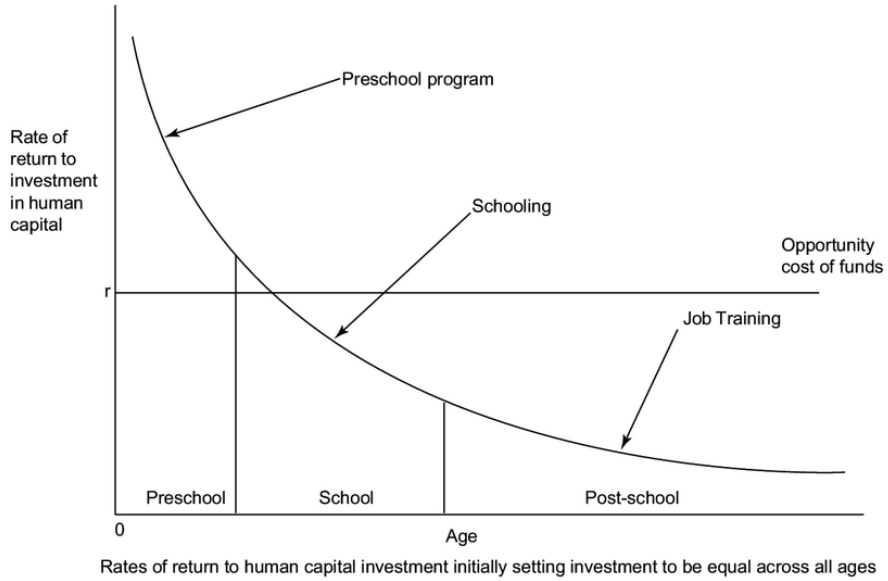
Chapter 5 looks specifically at how ECC expansion in Germany affected children's behavior. While there is extensive research in non-economic and economic fields on ECC, it provides puzzling mixed results. There is little connection between the research in these fields, giving limited guidance to parents and policymakers. Therefore, the fourth study addresses this limitation and incorporates several research areas by synthesizing findings from non-economic literature on developmental neuroscience, child psychology, and child development with economic literature that evaluates ECC programs and policies. This allows me to create a bigger picture of the potential ECC effects on child development. The study highlights the importance of considering the cumulative lifetime quantity of childcare – the combination of the ECC starting age and mean weekly hours of care. By focusing on ECC quantity, a consistent theory and evidence-supported hypothesis emerges: early and extensive childcare before the age of one and a half for girls and two for boys is harmful for children's socioemotional and behavioral development. The adverse effects of ECC before age two should be especially pronounced among boys and cohorts with more extensive ECC access at earlier ages. Chapter 5 tests this hypothesis by exploring the 2005 ECC reform in Germany that substantially expanded supply of childcare places to children under age three, and finds supporting evidence.

The findings indicate that early, extensive and continuous childcare could harm children's non-cognitive skills and undermine their human capital base since non-cognitive skills promote the formation of cognitive skills (Cunha and Heckman, 2008; Heckman *et al.*, 2006). Non-cognitive skills are also predictive of labor market success (see, e.g., Barón and Cobb-Clark, 2010; Brunello and Schlotter, 2011; Caliendo *et al.*, 2015; Heckman *et al.*, 2013; Prevoo and ter Weel, 2015); and emotional health and social behavior are the most important predictors of adult life satisfaction (Layard *et al.*, 2014). This suggests that early and extensive ECC, instead of increasing the future workforce's productivity and reducing in-

equities, could actually do the opposite. This implies weakening future economic growth and an increase in inequalities in health and human capital, inhibiting the achievement of the SDGs. Taking this together with increasing budgetary constraints and interest in the efficiency of the ECC sector, it is important to develop policies that use public funds efficiently to improve children's skills, and avoid side effects that can adversely impact child development.

Additionally, declining fertility and aging population underline the importance of making greater investment in human capital development through the life cycle – especially during the earliest ages due to the greater rate of return on such investments (Figure 1.4) – to increase future workers' productivity and to ensure sustainable economic growth. Thus, when introducing public policies aiming to boost human capital development, it is crucial that policymakers understand what kind of investment (parental vs out-of-family ECC), at what age and for whom will be the most beneficial to avoid unintended dis-investments in human capital formation during the earliest ages (implying a negative mirror effect in skills' formation).

Figure 1.4: Rates of Return to Human Capital Investment Initially Setting Investment to be Equal Across All Ages.



Notes: Source: Based on Carneiro and Heckman (2003), Figure 2.6.

To prevent negative effects on children's non-cognitive skills and well-being, policymakers could implement a policy similar to academic red-shirting based on a child's individual disposition, where age-eligible children postpone the start of ECC to allow for socioemotional maturation. For this, policymakers should provide greater support for families with children under age two, i.e., building awareness and providing parents with skills to support positive parenting. Thus, parents should be informed about potential negative effects

of early and extensive ECC to allow them to make well-informed decisions on whether, when and for how long to send their children to ECC. Additionally, providing parents with information on good parenting practices and reducing parental stress could ensure that parenting is not negatively affected.

To conclude, while ECC policies are increasingly popular, they substantially change the early social environment for young children, and not always to beneficial effect. Policy-makers and parents should be aware of potential negative effects of extensive ECC to implement better policies and to make better decisions that do not compromise children's well-being.

Summing up, the studies in this thesis focus on the topic of health, environment, and behavior. Each of the four studies has practical implications for health and well-being, directly or indirectly. The chapters aim to support well-informed decision-making by taking an evidence-based approach to real-world problems.

2 Disease and Fertility: Evidence from the 1918–19 Influenza Pandemic in Sweden*

Abstract. What are the consequences of a severe health shock like an influenza pandemic on fertility? Using rich administrative data and a difference-in-differences approach, we evaluate fertility responses to the 1918-19 influenza pandemic in Sweden. We find evidence of a small baby boom following the end of the pandemic, but we show that this effect is second-order compared to a strong long-term negative fertility effect. Within this net fertility decline there are compositional effects: we observe a relative increase in births to married women and to better-off families. Several factors – including disruptions to the marriage market and income effects – contribute to the long-term fertility reduction. The results are consistent with studies that find a positive fertility response following natural disasters, but we show that this effect is short-lived.

2.1 Introduction

A central line of inquiry in economic and demographic research concerns how fertility responds to changes in mortality. Yet, we have limited knowledge on the causal relationship between pandemics and fertility, and particularly few insights about the time dynamics of fertility responses to major health shocks. The event of a pandemic can cause major losses and in a globalized world where viruses can spread quickly, insights on whether, when, and why fertility changes together with mortality seem highly relevant.

A handful of studies examine fertility responses to pandemics. The results suggest there are immediate negative effects with fewer births six to nine months after the mortality peak, pointing to increased miscarriages, stillbirths and preterm deliveries (Bloom-Feshbach *et al.*, 2012; Chandra and Yu, 2015b,a; Chandra *et al.*, 2018; Guimbeau *et al.*, 2020), followed by increased fertility in the short run (Donaldson and Keniston, 2015; Mamelund, 2004).[1] This short-run positive response aligns with findings in the literature on the fertility effects of mortality following wars and natural disasters (see, e.g., Nobles *et al.* (2015) on the tsunami in South-East Asia; Pörtner (2008) on hurricanes in Guatemala; Finlay (2009)

*This chapter is co-authored with Nina Boberg-Fazlić, Martin Karlsson, and Therese Nilsson and published as *Boberg-Fazlic, N., Ivets, M., Karlsson, M., & Nilsson, T. (2021). Disease and fertility: Evidence from the 1918-19 influenza pandemic in Sweden. Economics & Human Biology, 43, 101020.* https://doi.org/10.1016/j.ehb.2021.101020

[1] Fletcher (2018) examines a different question focusing on family formation effects to in utero exposure to the influenza pandemic. Results suggest a small effect on the number of children that exposed women have in relation to unexposed cohorts.

on severe earthquakes; Lindstrom and Berhanu (1999) and Agadjanian and Prata (2002) on war) that shows that birth rates tend to increase in the short term. Short-run positive fertility effects can be explained by postponement or replacement fertility, but in theory such effects should no longer be present in the longer run.

This paper uses detailed information on the 1918-19 influenza pandemic in Sweden to study its effects on subsequent fertility rates using a difference-in-differences approach. The influenza pandemic was unforeseen and provides a unique opportunity to study fertility dynamics following a severe morbidity and mortality shock. Assembling administrative information from parish records, censuses, chief medical officer reports and midwife journals, we create a purpose-built historical database of high-quality data for a country that was neutral during World War I (WWI). As discussed by Beach *et al.* (2022) WWI is a potential confounder when studying fertility effects as the war likely affected the marriage market, but also incomes and women's participation in the labor market in countries taking part in the war. Our data and design allow us to study immediate, short-term and long-term fertility responses to the pandemic, and to asses the plausibility of various mediating factors.

A major contribution of the paper is that we contrast fertility effects observed over different time horizons, and we are able to show that different observation windows may lead to very different conclusions about the impact of the pandemic. We also seek to analyze possible mechanisms behind the observed fertility changes beyond biological effects. Despite its relevance for improving our understanding of the relationship between mortality and fertility, the role of mechanisms has often been overlooked in the empirical literature. A pandemic may have psychological effects but also alter economic conditions, introduce uncertainty and influence fertility decisions by disrupting family structures and marriage markets.[2] We examine mechanisms of economic character and marriage market effects, but also provide insights to whether certain groups changed their fertility behavior more than others.

The paper is the result of a vast data collection effort, combining various individual-level and aggregate data covering the entire Swedish population over a 13-year period. Our analytical sample includes the number of deaths from all causes, births, stillbirths, influenza and pneumonia cases and various mother and birth characteristics for about 400 urban and rural health districts located within 25 counties. The comprehensiveness of the data allows us to make a number of additional contributions to the literature. First, we can carefully assess the plausibility of the identifying assumptions, which leads to less concern about confounding factors biasing estimates. Second, covering the entire population implies high external validity compared to studies providing specific sub-population effects. This also implies that we have data from both rural and urban areas and can explicitly investigate different dynamics in different types of districts.[3] Third, while previous empirical studies generally focus on overall mortality, we consider both adult and child mortality as well as morbidity which allow for different mechanisms operating in the immediate, short

[2]Since marriage is traditionally seen as a proxy of fertility (Bongaarts, 1978) and was the main setting for childbearing in the early 20th century this may be an important pathway in the context of the study.

[3]A rural-urban divide seems highly relevant in a historical context, but seems to matter also in contemporary settings: Aassve *et al.* (2020) note possibly different post-pandemic fertility trajectories by income level and rural or urban area for the COVID-19 pandemic.

and long run. Finally, we can explicitly deal with internal migration, which otherwise confounds any analysis on the effects of a mortality shock.

After a short dip in conception rates during the pandemic, we find evidence of a small baby boom in rural areas *after* the peak of the pandemic. These results corroborate the fertility response noted after natural disasters. We further show that the positive short-term effect is driven by high social status parents: married couples, higher socioeconomic groups and mothers who already have at least one child contribute more than proportionally to the short-term increase in conceptions. This finding is interesting *per se*, but also of relevance for the large and widely cited literature on the fetal origins hypothesis following Barker (1990). Numerous studies show that in utero exposure to a health shock has consequences for health and socioeconomic status later in life (see, e.g., Almond *et al.*, 2018; Helgertz and Bengtsson, 2019). These results rest on the assumption that people conceived during a health shock do not differ from those conceived shortly after, other than through exposure. Some recent research revisits the literature which evaluates in utero exposure to the 1918-19 influenza and assesses the role of parental selection for the exposed cohorts (Beach *et al.*, 2018; Brown and Thomas, 2018).[4]

We find a shift towards higher social status parents after the pandemic. If children conceived shortly after the shock have better predisposition than those conceived during the pandemic, adverse health and income effects of an in utero shock will be overestimated. The same caveat applies to results on intergenerational effects of (c.f. Veenendaal *et al.*, 2013) and parental responses (Almond and Mazumder, 2013) to prenatal exposure. Parman (2015) demonstrates by example of the 1918-19 influenza, that the negative effects of in utero exposure can be further reinforced by parents reallocating resources towards older siblings, not affected in utero by the shock. This emphasizes how a large mortality shock can disrupt family structures and the allocation of resources among children. We demonstrate, that a large mortality shock will also affect decisions regarding family size.

Notably, the short-term fertility increase is swamped by a strong negative effect in the longer term. Areas greatly affected by the pandemic experience decreased fertility rates for years after the pandemic. Moving from the quartile of districts least affected by the flu in terms of adult mortality to the quartile of districts most affected associates with a decline in the monthly conception rate by about 10.5 percent in the long run. We show that this negative fertility effect goes beyond the 'mechanical' effect of those adults lost to the flu not having children, and rather represents behavioral and economic effects, including disruptions in the marriage market (a persistent reduction in the proportion of married individuals in the population). The noted composition effect is exacerbated by a disproportionate reduction in fertility among unmarried people: with the number of children to married couples decreasing less than general fertility in the disease aftermath, whereby the noted marriage market disruptions explain a substantial part of the fertility drop.

All in all, the results suggest that a deadly pandemic will be felt decades later and that the long-run effects may be very different from the short-term effects. The historical context

[4]Both studies focus on the US, which participated in WWI. While Brown and Thomas (2018) find no significant flu effects on later life outcomes after including proxies for parental characteristics of the 1919 cohort, the results of Beach *et al.* (2018) are more in line with Almond (2006) and are largely unaffected by controlling for parental socioeconomic status.

corresponds to a country during the fertility transition which makes our findings pertinent to many contemporary epidemic settings.[5] Our findings contribute to the understanding of the mortality-fertility link and show that the effects go well beyond those of direct exposure.

2.2 The 1918–19 Influenza Pandemic

The first recorded case of the influenza in Sweden was in June 1918. Initially, the seemingly mild flu caused little concern, but this situation soon changed. Influenza-related mortality and morbidity rates were particularly high from August 1918 to February 1919, peaking in October and November. A milder wave appeared in March 1919 and a final wave in early 1920.[6] Knowledge about the virus was limited. Flu vaccines were yet to be invented and the only effective measures were rest, hot blankets, cold compresses for headaches and drinking plenty of water (Mamelund, 2011).

According to official sources around 10 percent of the Swedish population was infected (Richter and Robling, 2013) and nearly one percent died (Karlsson *et al.*, 2014), but death rates varied considerably across age groups and across the country.[7] The most heavily affected counties experienced death rates almost three times higher than the least affected counties (Åman, 1990). Despite a clear north/south county gradient, with higher mortality in the north, there was considerable heterogeneity across districts within each county: Figure 4.A1 shows district-level influenza and pneumonia morbidity, all-cause adult and child mortality rates for the period August 1918 to March 1919 (per 1,000 inhabitants).[8]

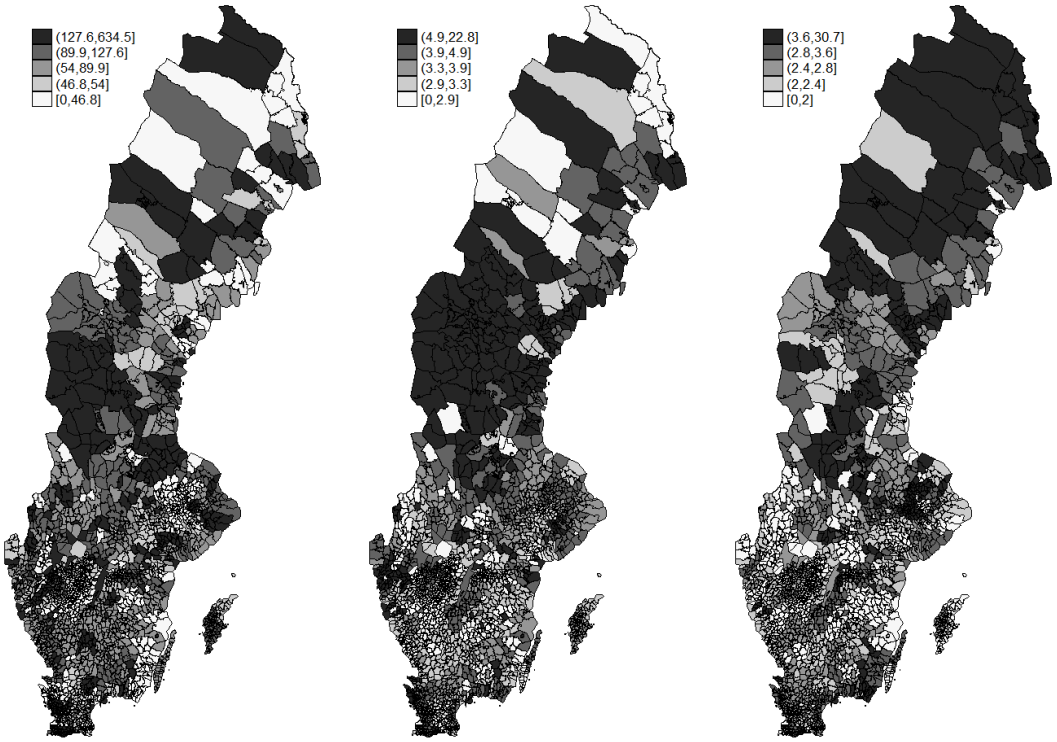
[5] In Sweden, fertility began to decline around 1880 when the number of children to married women was above four. The fertility transition to below two children per woman was completed by the mid-1930s (c.f. Strulik and Vollmer, 2015). According to Bengtsson and Dribe (2014) fertility started to decline at about the same time for older and younger women.

[6] Mamelund *et al.* (2016) find evidence for an early spring wave with high morbidity rates in the US and in Norway. We do not find evidence for this in the Swedish context, see Figures 2.B3, 2.B4 and 2.C6 in the Appendix.

[7] During the pandemic period, August 1918 to March 1919, 1.45 percent of the population in the age-group 20-40 died in Sweden: 1.62 percent of males and 1.28 percent of females.

[8] Geographic heterogeneity in 1918-19 influenza-related mortality rates is also noted in other countries, and some studies try to identify possible determinants. For example, Clay *et al.* (2019) show significant cross-city variation in excess mortality in the US and find high poverty and poor health levels contributed to pandemic severity. Similarly, the ongoing COVID-19 pandemic shows a very unequal impact of the virus across countries and regions (see, e.g., Fenoll and Grossbard (2020) on variation in deaths across EU countries and US states and the role of intergenerational coresidence, and Fielding-Miller *et al.* (2020) on variation in deaths across US states and the role of social determinants).

Figure 2.1: Influenza and pneumonia morbidity and overall mortality rates in Sweden during August 1918 to March 1919 (per 1,000 inhabitants)



(a) Morbidity **(b)** Adult mortality **(c)** Child mortality

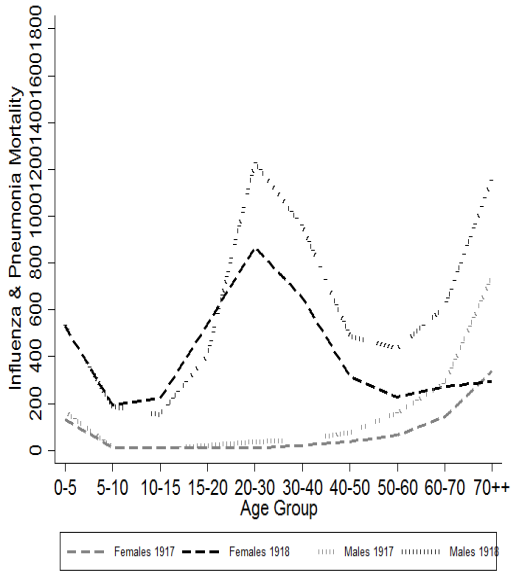
Note: Data correspond to the health district level. Legend categories represent quintiles.

The pandemic had several characteristics. First, in its most virulent form the influenza struck swiftly and unexpectedly. Most people died within 6 to 11 days after contracting the illness (Taubenberger and Morens, 2006). Second, the influenza affected the bronchus and the lungs which induced more pneumonia deaths (Morens and Fauci, 2007). Third, the pandemic was unique in that it primarily killed adults aged 20 to 40. Figure 2.2 illustrates the age distribution of mortality during the pandemic in different ways. Figure 2.2a compares influenza and pneumonia mortality rates by age and separated by gender in 1918 and 1917. Figure 2.2b shows the elevation of overall mortality during the influenza period at the health district level. Figures 2.2c and 2.2d show that the share of adults aged 20 to 40 in total mortality in relation to child mortality was many times higher during the pandemic than before and after. Research suggests that the reason for this mortality pattern was cytokine shock, an overreaction of the immune system (Kobasa *et al.*, 2007) such that a strong immune system was a liability rather than an asset, and possibly a lack of prior exposure to similar viruses (Mamelund, 2011).

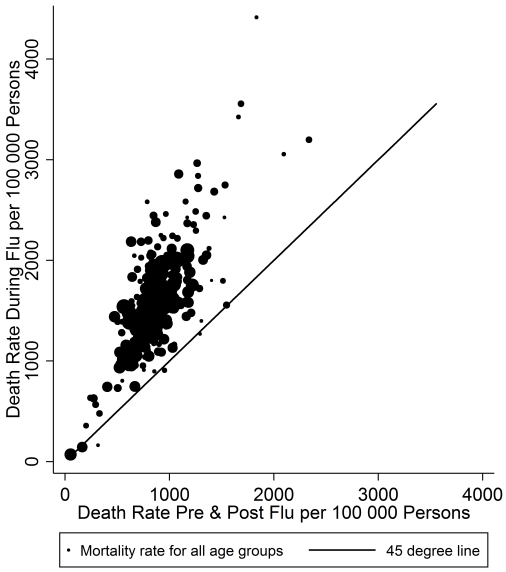
Given that the most deadly wave of the pandemic was unanticipated and short, it is unlikely that people adjusted fertility behavior in anticipation.[9] It was also impossible

[9] Appendix Table 2.C1 illustrates the time trend for conception rates in heavily and less affected districts.

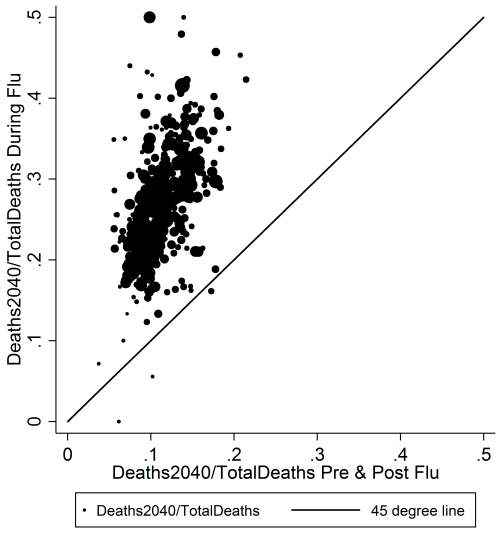
Figure 2.2: Mortality rates in Sweden during the pandemic and in adjacent periods



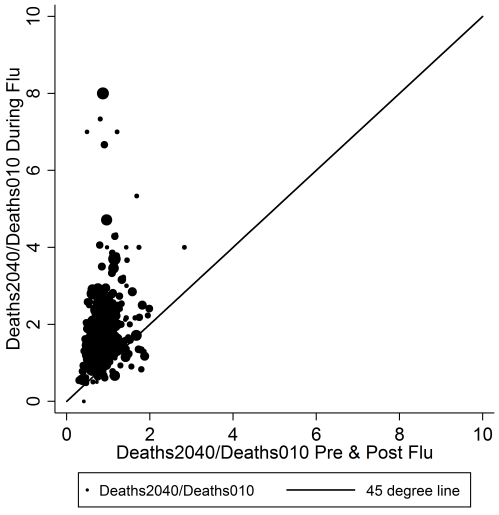
(a) Age profile, influenza and pneumonia mortality (100,000 people)



(b) All-cause mortality rate, all ages



(c) All-cause deaths age 20-40 to all deaths



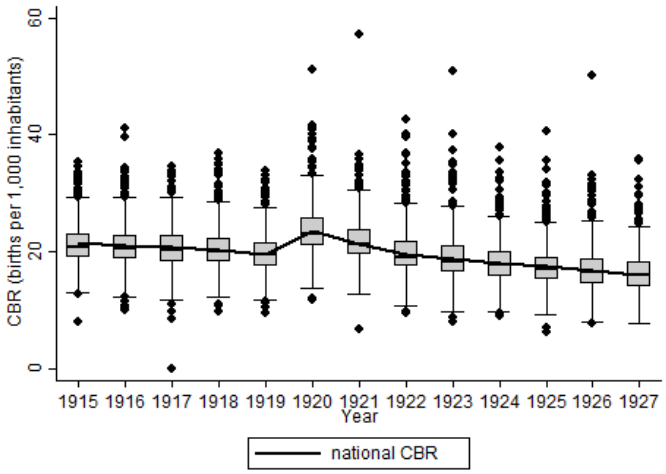
(d) All-cause deaths age 20-40 to deaths age 0-10

Note: Observations are weighted by the district population. Each dot in Figure b–d corresponds to a health district.

to foresee who had a higher infection risk. Men exhibited slightly higher mortality rates than women (see Table 2.B1 in the Appendix), but some evidence suggests that pregnant women in the last trimester were especially susceptible to the influenza, leading to early termination of pregnancy (Barry, 2004; Bland, 1919; Bloom-Feshbach *et al.*, 2012).[10]

Several European countries experienced a baby boom in the 1920s, commonly ascribed to the ending of WWI. The U.K. birth rate jumped from 18.3 births per 1,000 population in 1918 to more than 23 in 1919, but neutral countries like Sweden and Norway also exhibited elevated birth rates in the 1920s. Despite being neutral, these countries may of course still have been affected by the war ending, but it is notable that they did not experience any wartime fertility dip (Chesnais, 1992). Swedish fertility rates declined linearly from 1911 to 1919, and WWI neither accelerated nor decelerated this decrease (Statistics Sweden, 1999). The 1920 baby boom has, therefore, also been linked to the influenza pandemic (Mamelund, 2004). Figure 2.3 shows the crude birth rate (CBR) for Sweden from 1915 to 1927, along with the CBR distribution across all health districts in each year. Fertility rates generally declined throughout the period, but a clear deviation from the trend appears in 1920-21.

Figure 2.3: Crude birth rate, Sweden 1915-1927



Note: The solid line refers to yearly crude birth rate (CBR).The box plots refers to the CBR distribution across health districts in each specific year.

[10]This should be kept in mind when interpreting the empirical results, especially for the period August 1918 to March 1919, as live birth numbers may be depressed due to an increased number of stillbirths and miscarriages.

World War I and Economic Conditions

Sweden was neutral during WWI. Mortality rates were normal in the years prior to the pandemic, and morbidity and mortality record keeping was uninterrupted. Still, the war affected the Swedish economy. The U.K. naval blockade and German naval belligerence hurt the country's import trade (Jörberg and Krantz, 1978) and price controls and food rationing were introduced. A poor harvest in the fall of 1916 led to food shortages and social unrest for a short period, but in general, the wartime period was characterized by adequate food supply (Nyström, 1994), and historical sources suggest that the economic impact of WWI was generally even across the country (Östlind, 1945).

Some sectors of the economy benefited from the war. Raw material exports increased and Swedish agriculture did well because of the lack of competitive imports (Schön, 2010), leading to a large trade surplus (Magnusson, 1996). Conversely, these sectors experienced a downturn after the war. After a period of growth, the economy experienced a brief decline in 1920-21, where GDP dropped by 5 percent in one year and unemployment increased, but the country recovered quickly. Real wages were also positively affected, due in part to the introduction of the eight-hour working day (Jörberg and Krantz, 1978).

As pointed out by Beach *et al.* (2022), studying the fertility effect of the 1918-19 influenza in countries participating in WWI is difficult, as the war may have affected fertility through, e.g., changing marriage markets since more men than women died in the war. Another issue related to WWI, pointed out by Brown and Thomas (2018) and Beach *et al.* (2018), is that there seems to have been a shift in fertility in the U.S. during the war, whereby cohorts exposed to the pandemic in utero were born to fathers of lower socioeconomic status compared to earlier cohorts. The shift followed from a social gradient in WWI mobilization and enlistment and from men being stationed outside the U.S. and likely also because of active family planning. With Sweden being neutral in WWI, there was no change in the male-to-female sex ratio following war-related deaths and war mobilization is less of a concern. Swedish men were not stationed outside the country, but families may have delayed births following the uncertainty surrounding the war.[11] Such behavior would depress birth rates before the influenza pandemic outbreak, which implies that estimated fertility effects will present a lower bound. Since we use the variation in birth rates over time and across districts, such behavior is only a problem if districts reacting to the war were also those being more affected by the pandemic.

2.3 Conceptual Framework

This section provides a short outline of the theory guiding our analysis on fertility effects with a special focus on mechanisms and time dynamics. An important starting point is that a pandemic is only a temporary shock and should therefore not have an impact on *desired* fertility in the long run, unless it also affects determinants of desired fertility.

In terms of pathways, we may first consider *biological* effects following a pandemic. Fertility may change if a pandemic reduces sexual activity or the ability to conceive. Infections

[11] See Richter and Robling (2013) for a discussion.

may lead to pregnancy termination, and a spouse's death clearly reduces fertility prospects for the surviving spouse. Regarding the dynamics of this relationship, we expect an immediate negative fertility effect. While the effect stemming from infections is expected to fade out over time, a negative effect may also remain in the long run following spousal deaths, and with partner matching and remarriage taking time.[12] Such long-run negative fertility effect may also follow as individuals not directly affected through the death of a spouse may face consequences of a mortality shock on the marriage market as the sex ratio determines chances of finding a spouse Becker (1960, 1973, 1974) and as the population composition in terms of other traits, important for marriage market outcomes through assortative mating, may be affected.

Second, there may be important *behavioral effects* following a pandemic affecting fertility. These effects can be classified as *psychological* or *economic*. Psychological effects may be distinguished as either postponement or replacement fertility. Postponement fertility refers to delaying fertility decisions due to uncertainty about survival prospects or fear of contagion (Castro *et al.*, 2015; Lee, 1981; Menken *et al.*, 1981). With such postponement, fertility will decrease in the immediate and then increase right after the peak of a pandemic, as couples who would have conceived anyway and couples who postponed fertility will conceive in the short run. Replacement fertility stems from couples losing children who then conceive again to replace a lost offspring (see Preston, 1978). It has also been shown that high mortality events may even trigger a society-wide action of population rebuilding, leading to new conceptions at the intensive and extensive margins (see Conning and Udry, 2007; Geertz, 1963; Grimard, 1993; Townsend, 1994). This kind of replacement effect is possibly stronger in more rural settings where communities are closer. In terms of time dynamics, replacement fertility increases conceptions in the short term,[13] but the short-run positive fertility effect following replacement and postponement should not be present in the long run.

A major epidemic may also have *economic effects* impacting the fertility decision, triggered by changes in relative prices and opportunity costs, but also by introducing uncertainty. Mortality within a family likely reduces incomes which may delay fertility as children are costly (Alam and Pörtner, 2018). Economic theory also suggests that the death of young adults will increase wages and wealth in the economy as labor supply sharply decreases, and fixed factors such as land and capital are shared by fewer people (see Boucekkine *et al.*, 2009; Herlihy, 1997; Young, 2005). The substitution effect associated with such wage increases will reduce fertility as female labor supply likely increases and having children becomes relatively more costly. At the same time, the income effect may increase fertility, as agents can afford to have more children (c.f. Del Bono *et al.*, 2015), although Galor and Weil (1996) show that the substitution effect may dominate if women's

[12] A fading out of an immediate negative effect also aligns with a story where women experiencing a termination of their pregnancy soon after can get pregnant again.

[13] An alternative view attributes increased fertility after a mortality shock to a *hoarding* effect: parents have more children than initially intended because a recent mortality shock instills doubt about their children's survival prospects (LeGrand *et al.*, 2003; National Research Council, 1998; Palloni and Rafalimanana, 1999; Preston, 1978; Rosenzweig and Schultz, 1983). This is more pertinent for long-duration events and less pertinent for a short-term mortality shock following a pandemic. This mechanism would only be relevant if the 1918-19 influenza pandemic shifted the expectations of children's future survival over the longer term.

relative wages increase. A major pandemic may also fuel general perceptions of uncertainty about future economic conditions. If individuals avoid making long-term commitments following such uncertainty family formation may be negatively affected. In terms of dynamics, the economic effects are expected to have an impact on fertility in the short and the long run.

2.4 Data and Empirical Strategy

We build a unique dataset combining data from several official administrative sources collected from archives and public libraries.[14] To create the dataset, we combine individual-level data with aggregate information corresponding to three administrative partitions. The smallest geographical unit is a *parish* (around 2,500 at the time). The next unit is a *health district*, grouping together several parishes served by the same medical personnel. There were about 400 health districts at the time of varying sizes and populations. The largest administrative unit is the *county*, of which there were 25 at the time.

The main unit of our analysis is the health district level, but some of the data refers to the parish level. We, therefore, map parishes to health districts and track changes in the allocation and borders of health districts.[15] We aggregate health districts to obtain units with stable borders over the entire study period. This leaves us with a total of 396 districts, including 65 aggregated districts. The empirical analysis examines rural and urban health districts separately as fertility dynamics are likely to be different in these contexts. The division used is the contemporary classification of districts into *extra provincial*, *provincial*, *municipal district*, and *city* in the source material. We group the first two categories into *rural* and the latter two into *urban* health districts.[16]

Sources

A central source is the parish church books recording all deaths in Sweden. Already in 1686 local priests were obliged to record all births, deaths and marriages in a parish into church books that today are publicly available in local archives (Wicksell, 1922). The Federation of Swedish Genealogical Societies has digitized church records as the *Swedish Death Index*, which includes parish location, birth and death dates for all individuals who died in Sweden between 1901 and 2013. For a majority of individuals the civil status at the time of death is also recorded. We use this source to calculate the monthly death numbers across age groups. We also use the *Swedish Death Index* to derive the monthly birth numbers for each health district. Some people in the cohorts of interest were still alive in 2013 and we

[14] Most of the information was scanned from hard copies and digitized by the authors and research assistants.

[15] The initial allocation is based on an official list of health districts and which parishes they include from 1930. Changes are identified using information from royal decrees, http://sara.moricz.se/Kommungränskonverterare/ (*kommun*) and individual web searches.

[16] The distinction between extra provincial and provincial was usually one of timing, where a newly formed district would start as an extra provincial district which was later turned into a provincial district if the separation proved viable. The urban category mainly corresponds to smaller towns. Our results are robust to defining only *city* districts as urban.

thus do not observe them in the *Swedish Death Index*. We identify those individuals by using the 1950 Census, which includes people born between 1915 and 1927 who were still alive in 2013 (and, therefore, also alive in 1950). The census reports their data and parish of birth, which we use to supplement the birth numbers from the *Swedish Death Index*.

A second source is historical records from the National Medical Board, who collected monthly data from physicians on district morbidity, which we digitize. This variable correlates strongly with influenza mortality at the local level in Sweden (Karlsson *et al.*, 2014), but there is an ongoing debate regarding its accuracy, especially in periods of high influenza mortality (Bloom-Feshbach *et al.*, 2012; Mamelund *et al.*, 2016). Doctors were obliged to report verified cases of the flu (Influensakommittén, 1924), and historical records suggest that people did visit healthcare centers when they had the flu and that the pandemic clearly increased the demand for GPs (see, e.g., Influensabyrån, 1919). But morbidity is likely under-reported, and more so in rural compared to urban areas, as a sick patient had to visit a physician to get recorded and the distance to healthcare was longer.[17] Reporting consistency across districts may also be a potential issue when it comes to morbidity data: despite that the symptoms of the influenza were well known, there were no microbiological testing.[18] As influenza was often complicated by pneumonia, we combine information on influenza and pneumonia incidents in our morbidity measure.[19] The historical records from the National Medical Board also include demographic information and the number of inhabitants at the beginning of each year in each health district. We digitize this information and combine it with the monthly birth and death numbers to calculate the monthly district population.

Appendix Figure 2.B3 shows time trends in influenza and pneumonia morbidity and all-cause mortality for different age groups from 1915 to 1927. All series exhibit very pronounced spikes in the autumn of 1918. The timing and severity of the increase in deaths in late 1918 suggest that it is reasonable to assume that the majority of the excess deaths in this period were caused by the pandemic.

Our third main source is midwife journals. Swedish midwifery was professionalized early on. Trained midwives attended around 80 percent of births by the turn of the twentieth century, while less than 10 percent of women gave birth in hospitals (Högberg *et al.*, 1986).[20] Midwives had to keep diaries on all attended births and reported them annually to the main district physician (Bhalotra *et al.*, 2017). We digitize the information from the

[17] As discussed by Mamelund *et al.* (2016) under-reporting could also follow from a shortage of doctors. A general under-reporting of morbidity is corroborated by sickness reports for workplaces across Sweden suggesting higher morbidity rates, see, e.g., Helgertz and Bengtsson (2019).

[18] It is worth noting the long tradition and the well-defined responsibilities of the main district physician likely improved consistency in reporting. Disease control was one of the main responsibilities of the main district physicians already in the 19th century (Edvinsson, 2011) and the district physician had an obligation to make reports regarding monthly cases of epidemic disease to the National Medical Board using standardized forms separating between disease types (Jonsson, 2009), which is one of the main reasons why Sweden is one of the few countries that have historical monthly morbidity data by type.

[19] Using measures combining influenza and pneumonia incidents should also better capture any eventual early spring or summer wave of the influenza (Andreasen *et al.*, 2008). Morbidity data correspond to all influenza and pneumonia incidents in a district and are not available for separate age groups.

[20] By 1819, every parish was required to employ a licensed and trained midwife. In 1870 the ratio of midwives to doctors was 3.1 in Sweden, compared to 1.4 in the rest of Scandinavia (Romlid, 1997) and 1.2 in France (Thomson, 1997).

midwife journals from 1915 to 1927, including data on the number of midwives in each district, birth type (live births, stillbirths, and miscarriages, and the number of preterm and full-term births), and mother characteristics (the number of births to married, unmarried or widowed mothers and whether the woman was a first-time mother).[21]

Finally, we use annual information on local poverty rates, income and capital income. Poverty rates, taken from the annual publication on poor relief Statistics Sweden, are defined as the proportion of the population living in public poorhouses. Income includes all taxable earnings reported to the tax authorities, and capital income includes asset yields, rents and dividends, and comes from the yearbook of municipalities (Statistics Sweden, 1920). In heterogeneity analyses and in balancing tests we also use information from municipality yearbooks on private property assessed value, public revenue, public assets, public debts and population density. We also use data on the number of railway stations in a district in 1918 from Olofsson (1921). Appendix A provides definitions of all variables used in the following.

Appendix Table 2.B1 provides summary statistics on all variables for the periods before, during and after the pandemic. Notably there is considerable variation in the pandemic across districts, with an overall mortality rate ranging between 3.85 and 46 deaths per 1,000 inhabitants during the influenza period, and a corresponding morbidity rate ranging between 0 to 635 infections per 1,000 inhabitants.

Main Variable Definitions

Since it is the conditions at the time of conception that matter for the fertility decision, we specify the model in terms of conceptions rates. Conception rates are estimated based on the universe of live births, which are observed at the individual level and aggregated up to the health district-month level. With the exact number of conceptions unobserved, due to stillbirths and miscarriages that are not observed with the same frequency, we impute the following measure:

$$\begin{aligned} Conceptions_{im} = \\ DeathIndexBirths_{im+9} \quad &+\frac{1}{3}\sum_{\phi=7}^{9}\left[\frac{Stillbirths_{it}}{MidwiveBirths_{it}} \times \frac{DeathIndexBirths_{it}}{12}\right]_{im+\phi} \\ &+\frac{1}{3}\sum_{\theta=4}^{6}\left[\frac{Miscarriages_{it}}{MidwiveBirths_{it}} \times \frac{DeathIndexBirths_{it}}{12}\right]_{im+\theta} \end{aligned} \tag{2.1}$$

where i represents a health district, m a month, and t the corresponding year. We, thus, lag the number of live births by nine months and adjust this number for stillbirths and miscarriages.

We only have information on stillbirths and miscarriages as reported by midwives on an annual level, and therefore assume an equal distribution of stillbirths and miscarriages

[21] The midwife journals also provide a possibility to compare birth numbers from a different source. Appendix Figure 2.B1 compares derivations of births based on the *Swedish Death Index* data with birth numbers from the midwife journals.

throughout the year.[22] We also assume that we observe the correct *share* of stillbirths and miscarriages as a share of total births in the data reported by midwifes and then calculate the 'true' number of stillbirths and miscarriages by assuming the same share of stillbirths and miscarriages for the births observed in our main *Swedish Death Index* source and the 1950 Census. Stillbirths include pregnancy losses in months seven to nine. Hence, we lag one third of the calculated number of stillbirths occurring in month m by seven months, one third by eight months and one third by nine months. A miscarriage is a pregnancy loss occurring less than seven months into the pregnancy, but likely only miscarriages after three months of pregnancy are noted in the data. We, thus, lag one third of the calculated number of miscarriages in month m by four, five and six months, respectively. As early miscarriages are likely to have increased during and shortly after the flu (see Bloom-Feshbach *et al.*, 2012; Chandra and Yu, 2015a,b), our results will represent a lower bound, especially in the short run.[23, 24]

The conception rate is calculated by dividing the number of conceptions in district i in month m by the corresponding monthly population.[25]

$$ConceptionRate_{im} = \frac{Conceptions_{im}}{Population_{im}} \tag{2.2}$$

Ideally, we would define the conception rate with respect to the population at risk (women in ages 15 to 49), but this information is not available on the district level. On the other hand, we measure influenza exposure with reference to the same population number, which means that our estimates may be interpreted as *elasticities*. This will prove useful when we consider the cumulative net impact of the pandemic. Nevertheless, we carefully assess the extent to which compositional changes induced by the pandemic might be driving some of the results.

We apply an extended difference-in-differences framework and use variation in pandemic severity across districts and variation in conception rates over time within districts. For flu intensity, the influenza period is defined as August 1918 to March 1919.[26] We allow for persistent effects of the pandemic but rule out anticipation effects. Therefore, our treat-

[22] Some studies suggest seasonality in stillbirth rates (see, e.g., Auger *et al.*, 2017; Barnett and Dobson, 2010; Bruckner *et al.*, 2014; Eriksson and Fellman, 2000; Strand *et al.*, 2012) with higher stillbirth rates during summer and/or winter when temperatures are at extremes. We find no evidence of seasonality in stillbirths in our data, see Appendix Figure 2.B2.

[23] Around one in four pregnant women experience a miscarriage, with the vast majority occurring well before week 12 of gestation. One could argue that miscarriages are part of the natural process of pregnancy and should not be included in the conception numbers. In our data, miscarriages constitute on average around 4.1% of all annual conceptions. Our results do not change when we exclude miscarriages.

[24] Current research on COVID-19 also links the ongoing pandemic to increased risk of preterm births (Delahoy, 2020). Also, a study by Khalil *et al.* (2020) documents that the overall stillbirth rate has increased during the COVID-19 pandemic.

[25] Monthly population is calculated by using the population numbers as of January 1st for each year from the demographic data – provided in the health district yearbooks – and adding/subtracting the monthly number of births/deaths. Migration is thus attributed to the last month of the year.

[26] Appendix Figure 2.B4 shows the distribution of the peak month of morbidity and mortality across districts, defined as the month with the highest increase in incidents/deaths compared to the average morbidity/mortality between January 1916 and December 1917 in a district. The vast majority of districts have their peak within our defined peak flu period.

ment variable *FluIntensity* is a *cumulative* influenza intensity measure capturing all-cause deaths or influenza and pneumonia morbidity up to conception month m in district i. This implies that only mortality/morbidity incidents in August 1918 are assumed to matter for conceptions in this month, whereas the sum of incidents in August and September 1918 matters for conceptions in September 1918, etc.:

$$FluIntensity_{im} = \begin{cases} 0, & \text{if } m \in [Jan1915, Aug1918). \\ \frac{\sum_{j=Aug1918}^{m} Incidents_{ij}}{Population_{i1917}}, & \text{if } m \in [Aug1918, Mar1919]. \\ \frac{\sum_{j=Aug1918}^{Mar1919} Incidents_{ij}}{Population_{i1917}}, & \text{if } m \in (Mar1919, Dec1927]. \end{cases} \tag{2.3}$$

As adult and child mortality may affect the fertility decision differently, we calculate age-specific mortality rates. Adult mortality is the sum of deaths in the 20–40 age group representing potential parents, and child mortality is the sum of deaths in the 0–10 age group.

Figure 2.4 shows how conception rates and the three influenza variables evolve over time. Conception rates were at their lowest in September to November 1918 and drop with the increase in mortality and morbidity. Figure 2.4b further shows that conceptions gradually increased after the influenza peak. As outlined above, we expect fertility effects to differ during different periods. Our analysis focuses on three time periods: *Peak* (August to November 1918), where we expect a negative effect on conceptions from the beginning of the pandemic up to its peak due to biological effects and/or postponement fertility; *After* (December 1918 to December 1920), where we expect an increase in conception rates due to postponement and/or replacement fertility leading to a baby boom in 1920-21;[27] and *Later* (1921-27), where we expect a negative effect mainly stemming from long-term economic effects.[28]

Econometric Approach

For our main analysis, we specify the following model:

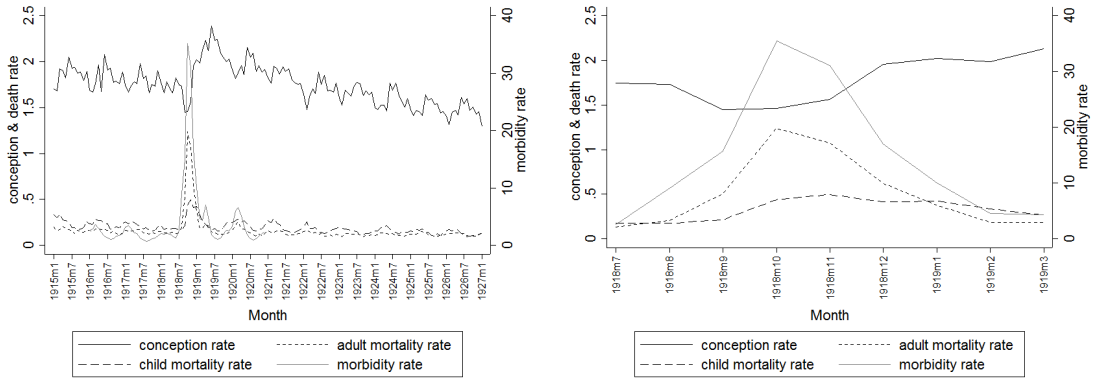
$$\begin{aligned} ConceptionRate_{im} = {} & \alpha_i + \beta_1 FluIntensity_{im} \times D_{Peak} + \beta_2 FluIntensity_{im} \times D_{After} \\ & + \beta_3 FluIntensity_{im} \times D_{Later} + \delta_1 D_{Peak} + \delta_2 D_{After} + \delta_3 D_{Later} + X_{im}\Gamma + \lambda_m + \epsilon_{im} \end{aligned} \tag{2.4}$$

for district i in period $m \in [1915m1, 1927m12]$. Our main specification includes district fixed effects (α_i) and month-year fixed effects (λ_m). The dummy variables D_{Peak}, D_{After}

[27] During this time period a few districts have experienced second or third flu waves and we might expect them to depress their fertility. Excluding these districts from the analysis leaves results qualitatively similar.

[28] We do not include the fourth wave in 1920 as it was mild and concentrated in the north. Our results do not change when we exclude northern districts from the analysis. Similarly, our results do not change when excluding districts where influenza morbidity and mortality increased already in July 1918, or when including July in the *Peak* period and in the treatment variable calculation in equation (2.3). Results also remain qualitatively the same when redefining the pre-flu reference period and let it last until May 1918 – the month when the first media reports on the influenza came from Spain.

Figure 2.4: Conception, morbidity and overall mortality rates (per 1,000 people)



(a) Entire period, 1915–1927 **(b)** During the pandemic

and D_{Later} indicate whether period m falls within the influenza peak period (immediate effects), in the one to two years following the pandemic (short-term effects), or later years (long-term effects). The reference period is the pre-influenza period ranging from January 1915 to July 1918.

The coefficients of interest are β_1, β_2 and β_3. With treatment defined as the degree of influenza exposure, β_1 corresponds to the differential effect of greater influenza intensity at the district level on conception rates during the peak period, β_2 captures the short-term effect after the peak, while β_3 corresponds to the long-term effect. We consider the overall effect, but also split the analysis by rural and urban districts.

Specification (2.4) represents a difference-in-differences model with variable treatment intensity. The crucial identifying assumption is that in the absence of the pandemic, conception rates in differently affected districts would have followed a common time trend. Appendix B provides evidence supporting this assumption: graphical evidence suggests parallel trends in the years preceding the pandemic regarding conceptions; balancing tests show that local observables were unrelated to excess mortality during the pandemic. Appendix B also presents event study graphs showing the coefficients of a flexible difference-in-differences model interacting the treatment variable with quarterly dummies. The flexible estimation allows for a placebo test, assuring that the flu had no effect on conceptions before it happened. All estimates before August 1918 are insignificant.[29]

As a robustness check, we include a set of control variables X_{it} and county-specific linear trends. The control variables include per capita earnings and capital income (both normalized by 1917 prices, in logs), the poverty rate and the log of the number of midwives proxying the local medical infrastructure.[30] Notably, some of the control variables can be seen as bad controls due to endogeneity (Angrist and Pischke, 2008) and some caution is

[29] As mentioned above, we do not find evidence for a pronounced early summer wave of the pandemic in Sweden. This finding is further supported by the insignificance of the coefficients before August 1918.

[30] The pandemic strained the healthcare system and financial means to cope with the flu fell short in some districts (Holtenius and Gillman, 2014). This may also have had an impact on the medical care in those districts afterwards, which in turn may influence the decision to have children or not.

required when interpreting estimates in specifications with controls. We therefore show results with and without controls.

Clearly, our outcome variable will react to changes in the composition of the population. The pandemic represents a shock to the population and may thus cause a mechanical change in the conception rates. We return to this issue in the next section, both by keeping population constant at 1917 levels and by quantifying the estimated effects relative to mechanical effects.

2.5 Analysis

Fertility Effects

Table 2.1 presents estimates of the pandemic's impact on fertility for reported influenza and pneumonia incidents (Panel A) and adult and child mortality (Panels B and C). For morbidity, we note a small negative effect on conceptions during the peak period, completely driven by rural areas. This immediate response is in line with biological effects where women have difficulties conceiving if they or their husbands are ill or psychological effects of not wanting to conceive in uncertain times. There are no significant short- and long-term effects for either rural or urban areas and thus no indication of postponement fertility due to high morbidity. The lack of effects in the *Later* period corroborates the idea that morbidity primarily measures biological effects, mainly expected to be present during the influenza peak and some time after.

Turning to adult mortality (Panel B), there is again an immediate negative fertility effect, evident in both urban and rural areas. After the peak period, fertility bounces back in rural areas, but fertility is then depressed in the long run. This negative long-term pattern is also very pronounced in urban areas, where no bounce-back is observed right after the pandemic peak.[31]

We, thus, find evidence of a small baby boom in the *After* period (December 1918 to December 1920), which can be explained by postponement fertility in the *Peak* period and later catch-up in the *After* period, but this should not be unique to rural areas. Yet, there could be differences across rural and urban contexts regarding replacement fertility.[32] In many respects rural societies were culturally and socially more close-knit than urban Sweden in the early 20th century. For example, households were interdependent during sowing and harvesting periods, tightening social ties. Also, divorces were predominantly an urban phenomenon (Sandström, 2011). Such knit may have initiated community rebuilding in rural districts that lost many adults, increasing collective fertility. An alternative expla-

[31] A potential concern for the observed difference between rural and urban areas is differences in measurement errors. Regressing the mortality rate on the morbidity rate and including an interaction term with an urban dummy, the interaction term is, however, not significant, suggesting that there were no significant differences in reporting influenza and pneumonia cases between rural and urban areas.

[32] Table 2.B8 shows results for including both adult and child mortality in the regression. The results indicate that the noted positive short-run effects stem from adult mortality, indicating general replacement rather than child replacement. It should, however, also be noted that the correlation between child and adult mortality is very high – at 0.78. It is, thus, difficult to gauge the true effect of one over the other.

Table 2.1: Fertility effects of the influenza pandemic.

	All				Rural		Urban	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. INFLUENZA MORBIDITY								
Peak × morbidity	-0.0009**	-0.0009**	-0.0013***	-0.0012***	-0.0013***	-0.0012***	0.0002	0.0001
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0010)	(0.0010)
After × morbidity	0.0002	0.0003	-0.0001	-0.0000	-0.0001	-0.0000	0.0008	0.0008
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0006)	(0.0006)
Later × morbidity	-0.0004	-0.0002	-0.0003	-0.0001	-0.0003	-0.0001	-0.0001	-0.0003
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0008)	(0.0007)
Peak × morbidity × *urban*			0.0015	0.0013				
			(0.0011)	(0.0010)				
After × morbidity × *urban*			0.0009	0.0009				
			(0.0006)	(0.0007)				
Later × morbidity × *urban*			0.0002	-0.0002				
			(0.0009)	(0.0008)				
B. ADULT MORTALITY								
Peak × adult mortality	-0.076***	-0.109***	-0.057**	-0.084***	-0.057**	-0.064***	-0.093***	-0.148***
	(0.019)	(0.033)	(0.024)	(0.025)	(0.024)	(0.017)	(0.026)	(0.048)
After × adult mortality	0.022	0.019	0.058***	0.052***	0.058***	0.053***	-0.020	-0.024
	(0.023)	(0.023)	(0.014)	(0.015)	(0.014)	(0.015)	(0.035)	(0.030)
Later × adult mortality	-0.088***	-0.093***	-0.079***	-0.087***	-0.079***	-0.087***	-0.095**	-0.104***
	(0.025)	(0.024)	(0.030)	(0.027)	(0.030)	(0.027)	(0.041)	(0.030)
Peak × adult mortality × *urban*			-0.036	-0.036				
			(0.035)	(0.035)				
After × adult mortality × *urban*			-0.079**	-0.076**				
			(0.038)	(0.034)				
Later × adult mortality × *urban*			-0.017	-0.019				
			(0.051)	(0.041)				
C. CHILD MORTALITY								
Peak × child mortality	-0.151***	-0.164***	-0.085**	-0.101**	-0.085**	-0.113***	-0.180***	-0.160***
	(0.026)	(0.024)	(0.042)	(0.043)	(0.042)	(0.042)	(0.022)	(0.040)
After × child mortality	-0.023	-0.027	0.053***	0.047***	0.053***	0.047***	-0.052**	-0.046*
	(0.031)	(0.031)	(0.015)	(0.013)	(0.015)	(0.013)	(0.026)	(0.023)
Later × child mortality	-0.099***	-0.110***	-0.040*	-0.062*	-0.040*	-0.062**	-0.117***	-0.107***
	(0.019)	(0.018)	(0.024)	(0.032)	(0.024)	(0.032)	(0.016)	(0.017)
Peak × child mortality × *urban*			-0.095**	-0.074				
			(0.047)	(0.047)				
After × child mortality × *urban*			-0.105***	-0.092***				
			(0.030)	(0.026)				
Later × child mortality × *urban*			-0.078***	-0.043				
			(0.029)	(0.037)				
Add. controls	No	Yes	No	Yes	No	Yes	No	Yes
County trend	No	Yes	No	Yes	No	Yes	No	Yes
N	46,861	46,861	46,861	46,861	35,200	35,200	11,661	11,661
N (cluster)	367	367	367	367	270	270	97	97
Baseline	1.81	1.81	1.81	1.81	1.68	1.68	2.14	2.14

Monthly data on health district level. N refers to the number of health districts x the number of time periods. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. The dependent variable is conception rate. All regressions include district and month-year fixed effects. Standard errors in parentheses, clustered at the district level. Additional control variables include the log of the number of midwives, the log of earnings and the log of capital income in 1917 prices, and the poverty share. The *Peak* period includes August 1918 to November 1918; *After* includes December 1918 to December 1920; *Later* includes January 1921 to December 1927. Morbidity and mortality rates are calculated as the cumulative sum of influenza cases/all-cause deaths occurring during the flu period, normalized by the district population in 1917.

nation to why conception rates in urban districts did not rebound in the same manner as in rural areas is that the incentives to have children differed across these settings. In rural areas children represented an investment good, as they provided labor on the farm and care for the parents during old age, while children were more of a consumption good in urban areas. With costs and potential pay-offs of having children being different in the two settings, and if the influenza increased uncertainty, the decision to have a child or not could go in opposite directions.

In the long term (*Later* period), both rural and urban districts that exhibited high adult mortality decreased their fertility compared to less affected districts. In the full sample without additional controls, each additional adult death per 1,000 people reduced the monthly conception rate in the long-term period by 0.09. With a baseline monthly conception rate of 1.81, this translates to about 5 percent fewer conceptions. Comparing the quartile of districts least affected (the 25th percentile) in terms of adult mortality with the districts most affected (the 75th percentile) with an adult mortality rate of 3.13 and 5.24 deaths per 1,000 people respectively, the difference between these two districts correspond to a 10.5 percent reduction in the monthly conception rate. This pattern is in line with economic effects including negative income effects and changes on the marriage market, as shown in greater detail in the next section.[33]

Clearly, population size depends on mortality. Therefore, especially the short-run positive effect on conceptions may stem from a mechanical effect of reduced population. Appendix Table 2.B2 provides results when keeping population constant at 1917 levels in the calculation of conceptions rates. The previously noted short-run positive effects also appear with this specification.

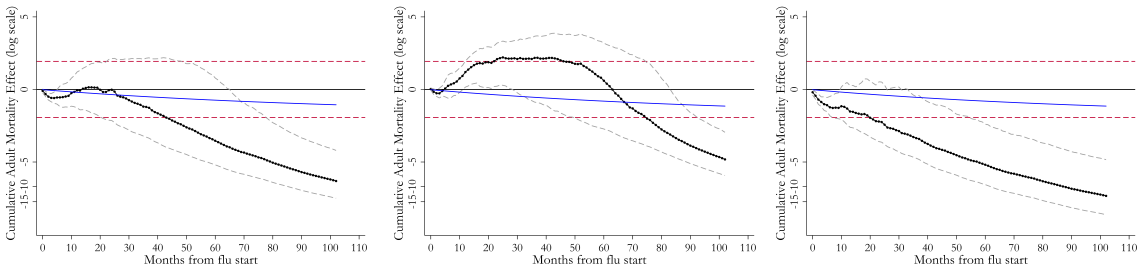
We may also be concerned about a mechanical fertility effect following the death of potential parents. To assess whether our estimates go beyond mechanical fertility effects in the long run, we estimate the number of conceptions that would have happened if adults killed by the pandemic had remained alive and reproduced at pre-pandemic (1911-17) rates. This estimate is given by

$$MissedConceptions_{im} = \phi FluIntensity_{im} \times (240 - m) \,/240 \tag{2.5}$$

where ϕ is the monthly fertility rate in the population of reproductive age in the 1911–17 period (derived from and calculated based on Statistics Sweden, 1929), $FluIntensity_{im}$ is adult influenza mortality measured according to equation (2.3), and the last term adjusts for the fact that the 1918 population of reproductive age gradually moved out of that age bracket (we normalize $m = 0$ at the outbreak of the pandemic so that $m = 240$ after 20 years have passed). Figure 2.5 (a)-(c) graphs the resulting cumulative fertility effect using our point estimates (illustrated by the solid black line) from columns (1), (5), and (7) of panel B (adult mortality) of Table 2.1, respectively. Confidence intervals are estimated analogously based on the estimated covariance matrix of coefficients. The dashed horizontal lines in each of the figures correspond to 1 and -1, which are useful benchmarks as 1 represents a situation where the pandemic is completely undone in the sense that there is an additional conception for each individual dying. The figures thus demonstrate the net cumulative effect of the sometimes conflicting short- and long-term responses. Figure 2.5b

[33]Results using child mortality (Panel C) are qualitatively similar to the results from adult mortality.

Figure 2.5: Cumulative fertility effects, adult mortality



(a) All districts **(b)** Rural districts **(c)** Urban districts

Note: The figure shows cumulative fertility effects (the solid black line) from regressions using adult mortality in all, rural and urban districts. The horizontal dashed lines corresponds to 1 and -1, respectively, where the upper dashed line represents complete replacement. The blue solid curve represents "mechanical effects" arising from potential parents being killed in the pandemic. A base-10 log scale is used for the Y axis.

shows that the short initial decline in rural areas is offset by the rebound in the medium term: 16 months after the onset of the pandemic (in December 1919), the cumulative effect is one new conception for each adult killed in the pandemic. This replacement is, however, completely undone 63 months after the beginning of the pandemic (in November 1923) after which the cumulative effect turns negative. In the urban areas the cumulative effect is always negative (Figure 2.5c). In the pooled sample (Figure 2.5a), the cumulative effect becomes significantly negative after approximately 83 months (by the fall of 1925).

The blue solid lines in Figures 2.5 (a)-(c) represent the cumulative "mechanical" effect of missing conceptions, calculated according to equation (2.5). In the pooled sample, the initial dip and the long-term decline are both significantly larger than the predicted mechanical effect. In rural areas, also the intermediate increase in fertility occurring in the aftermath of the pandemic is significantly different from the mechanical effect. Thus, we conclude that our analysis demonstrates behavioral responses that go well beyond mechanical effects driven by deceased potential parents.

As the cumulative effect is strongly negative in both rural and urban areas, we aggregate the periods *Peak*, *After*, and *Later* into one *Post* period and focus on the total effect of cumulative mortality during the flu months in the following analyses.[34] We will focus on adult mortality as the effects do not vary across mortality measures and morbidity only exhibits temporary effects.[35] As noted in Table 2.1 results are insensitive to the inclusion of covariates, therefore we exclude additional controls in the following.

[34] For consistency we also provide tables with results for the three different periods in Appendix B.

[35] Some districts may have experienced high morbidity but low mortality, or high adult mortality and low child mortality, or vice versa. To gauge the relative importance of the three influenza variables, we run regressions jointly including morbidity and mortality measures. Results do not change when including morbidity and mortality in the same specification.

Treatment Effect Heterogeneity

Next, we conduct a heterogeneity analysis and explore whether the impact of adult mortality on conception rates differs across district types. For this exercise we use baseline district characteristics in 1917 collected from official yearbooks. To classify districts, we generate dummy variables indicating whether the district was *above* the median for a specific characteristic in 1917 and interact this with our treatment variable. We also check whether rural districts with different shares of Sami people (measured in 1910) responded differently.[36] We further include a measure for how well connected a district was. Here, we use information on the number of railway stations in a district and generate a dummy variable taking the value one if a district had at least one railway station in 1918.[37]

Table 2.2 presents results for the whole country and for rural and urban areas separately. Mainly three characteristics associate with the fertility effect: high poverty rates, low population density and worse railway connection. The interaction term attains statistical significance for poverty in the urban sample and population density, which can also be interpreted as a measure of poverty in rural settings, in the rural sample. Urban areas with above-median poverty rates experienced disproportionate declines in fertility rates and more densely populated rural areas experienced smaller declines. We also note that fertility declines induced by the pandemic were less pronounced in rural districts connected to the railway network.

Appendix Table 2.B3 provides the results from a complementary heterogeneity analysis using continuous variables instead of a median cut-off. A possible concern with such a specification is that it gives disproportionate weight to districts that are at the extremes of the distribution of the interaction variables. However, the results in Table 2.2 are generally insensitive to this change in specification.

Taken together the heterogeneity analysis suggests that the fertility declines induced by the pandemic were particularly pronounced in relatively poor areas. This finding underlines the importance of adverse economic conditions in fertility declines, and is interesting in the light of previous research that suggests that the 1918-19 pandemic had negative economic impacts. For example, Karlsson *et al.* (2014) find that the pandemic led to a significant increase in poorhouse rates in Sweden. Also Barro *et al.* (2020), Correia *et al.* (2022) and Guimbeau *et al.* (2020) find higher influenza-related mortality associated with persistent economic decline.

[36] Being an indigenous population, the Sami people could exhibit divergent fertility behavior and previous work has shown that the local Sami population was an important predictor of influenza mortality in Norway (Mamelund, 2003).

[37] The Swedish state started to build a national railway network around 1850, and the country soon had an extensive network of overland transport routes (Hedin, 1967). Around 60 percent of districts had a station in 1918.

Table 2.2: Heterogeneity analysis

	(1) Midw.	(2) Poverty	(3) Priv Prop.	(4) Popdens	(5) Taxinc	(6) Conceptions	(7) Sami	(8) Railway
ALL DISTRICTS								
Exposure	-0.061**	-0.031	-0.045	-0.055**	-0.055*	-0.012	-0.062***	-0.057**
	(0.029)	(0.027)	(0.032)	(0.027)	(0.029)	(0.015)	(0.022)	(0.028)
Exposure $\times$ *Variable*	-0.056	-0.029	-0.003	-0.006	-0.053*	-0.053*	0.239	0.001
	(0.044)	(0.041)	(0.039)	(0.042)	(0.030)	(0.031)	(0.275)	(0.043)
N (clust)	367	367	367	367	367	367	367	367
N	46,861	46,861	46,861	46,861	46,861	46,861	46,861	46,861
RURAL DISTRICTS								
Exposure	-0.054**	-0.012	-0.021	-0.055***	-0.054**	-0.005	-0.044**	0.063***
	(0.023)	(0.016)	(0.013)	(0.019)	(0.024)	(0.012)	(0.020)	(0.019)
Exposure $\times$ *Variable*	0.041	-0.042	-0.033	0.070**	0.044	-0.043	0.222	0.058**
	(0.028)	(0.027)	(0.028)	(0.027)	(0.028)	(0.028)	(0.287)	(0.024)
N (clust)	270	270	270	270	270	270	270	270
N	35,200	35,200	35,200	35,200	35,200	35,200	35,200	35,200
URBAN DISTRICTS								
Exposure	-0.062	-0.031	-0.055	-0.048	-0.051	-0.025		-0.050
	(0.058)	(0.039)	(0.051)	(0.056)	(0.057)	(0.041)		(0.050)
Exposure $\times$ *Variable*	-0.020	-0.154**	-0.057	-0.057	-0.055	-0.052		-0.082
	(0.074)	(0.076)	(0.059)	(0.065)	(0.065)	(0.061)		(0.057)
N (clust)	97	97	97	97	97	97		97
N	11,661	11,661	11,661	11,661	11,661	11,661		11,661

Note: Monthly data on health district level. N refers to the number of health districts x the number of time periods. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. The dependent variable is conception rate. All regressions include district and month-year fixed effects. Standard errors in parentheses, clustered at the district level. *Exposure* is used for readability and is defined as $Exposure_{it} := FluIntensity_{it} \times D_{Post}$, where $D_{Post} = 1$ if $t \in [Aug1918, Dec1927]$, otherwise 0. *Exposure* $\times X$ denotes the interaction of *Exposure* with the variable in the column heading. All interaction variables in specifications (1)–(7) are dummy variables taking on the value one for districts being above the median for the specific variable in 1917. The Sami share is taken from the 1910 census and since its median equals zero, the actual share is used in the interaction. Railway is a dummy variable taking the value one if the district had at least one railway station in 1918.

Mechanisms

Marriage Market

Given the observed long-term negative fertility response it is natural to look at changes in nuptiality as a potential pathway. We start by discussing the implications for the marriage market stemming from a mortality shock. Although a loss of 1 percent of the population may seem irrelevant, there is large variation in mortality across districts and the historical literature provides plenty of stories about families falling apart.[38] The flu was especially hard on individuals between 20 and 40 years of age and the pandemic likely broke up existing marriages by the death of a spouse. Remarriage was common after widowhood in early 20th century Sweden, but this process could take time (Lundh, 2007).[39] In fact, widowers were not unattractive on the marriage market as they generally could offer an established household. For women it was often a necessity to remarry to support themselves and their children. Young widows generally had better prospects of remarriage, but also stronger incentives to remarry as older widows could expect support from their adult children (Dribe *et al.*, 2007; Lundh, 2007).

Also individuals not directly affected through the death of a spouse may face the consequences of a large mortality shock on the marriage market. Following Becker (1960, 1973, 1974), the sex ratio determines the chances of finding a spouse in a monogamous society for obvious reasons, but also population composition in terms of other traits plays an important role for marriage market outcomes through assortative mating (see, for example, Abramitzky *et al.* (2011); Angrist (2002), and Dribe and Lundh (2005) for an account on assortative mating in 19th-century Sweden). There are also reasons to expect that the marriage market may be affected by the economic uncertainty and psychological effects that followed the pandemic. Research on family formation during economic downturns has found adverse economic shocks to have negative effects on nuptiality and consequent childbearing (Neels, 2010),[40] and research in psychology provides theoretical grounds for a large mortality shock affecting the marriage market.[41]

[38] See, for example, Lundgren (1989, 1991) on the story of a family and its survivors in Arjeplog, a parish in northern Sweden most severely hit by the flu. The local newspaper even had a special category on tragic family stories during the pandemic (*Norrbottens-Kuriren*, 1918-1920. Available from the archive of Norrbottens museum: https://norrbottensmuseum.se/arkivcentrum/arkiv-bibliotek/tidningsarkiv.aspx)

[39] The Protestant Church accepted remarriage but imposed a mourning period of six months on men and one year on women.

[40] See also Comolli (2017); Matysiak *et al.* (2021); Percheski and Kimbro (2014) and Sobotka *et al.* (2011) on a link between economic recessions and fertility. A negative relationship between both births and marriage rates and economic crises has also been observed in historical studies (Bavel, 2001; Bengtsson *et al.*, 2004; Lee, 1990; Teitelbaum, 2014; Tzannatos and Symons, 1989).

[41] On the one hand, *stress theory* suggests that community-wide exposure to mortality brought by disasters and pandemics has a negative psychological impact, in turn reducing marriage rates (Cohan and Cole, 2002; Goldmann and Galea, 2014). Research shows that adverse psychological effects are present years after a community shock (Bland *et al.*, 1996; Bolton *et al.*, 2000; Bonanno *et al.*, 2008; DiGrande *et al.*, 2010; Jalloh *et al.*, 2018) and that enduring psychological damage is typically observed in up to 30% of exposed individuals (see Goldmann and Galea (2014) for a review). Depression and anxiety may also increase following stigmatization and discrimination of epidemic survivors (see, e.g., Karafillakis *et al.*, 2016; O'Leary *et al.*, 2018; Rabelo *et al.*, 2016). On the other hand, *attachment theory* suggests that marriage rates will increase af-

With information on the last civil status of a deceased person and the date of the last change in civil status, we estimate the number of people getting married or becoming widowed (those changing to the status 'married' or 'widow' for the last time before death) in each district following the pandemic. Appendix Figure 2.B5 shows the evolution of these series over time, and Appendix Figure 2.C2 shows trends for the highest and lowest district quartiles in terms of influenza exposure, suggesting that there were no significant differences in the trends before the flu. Some caveats should be kept in mind. First, our data come from the *Swedish Death Index* and we do not observe the civil status of individuals that are still alive.[42] Second, we do not know the order of a marriage, i.e., whether it is a first or second marriage. Also, we do not know in which parish the marriage/widowhood took place or with whom. We use the birth parish whenever the birth date is closer to the marriage/widowhood date and the death parish whenever the death date is closer to this date. This leads to an assignment of the birth parish in around 70 percent of the cases. Comfortingly, incorrect assignments will largely be reduced by the aggregation of parishes to health districts, as birth and death parish lie within the same health district in almost half of all cases.

In a first step, we estimate whether districts, which were particularly hard hit by the pandemic (in terms of adult mortality) experienced a change in marriage rates and/or widow rates afterwards. We, therefore, estimate the following model:

$$CivilStatusRate_{it} = \alpha_i + \beta_1 FluIntensity_{it} \times D_{Post} + D_{Post} + \lambda_t + \epsilon_{it} \tag{2.6}$$

where $CivilStatusRate_{it}$ is marriage or widowed rate relative to population numbers, in district i in period $t \in [1915, 1927]$. $D_{Post} = 1$ if $t \in [1918, 1927]$, otherwise 0 and $FluIntensity_{it}$ is cumulative adult mortality equal to mortality between August and December 1918 if t=1918 and equal to mortality between August 1918 and March 1919 if $t \in [1919, 1927]$.

In a second step, we examine if changes in widow and/or marriage rates stemmed from changes in the sex ratio induced by differential mortality rates among men and women. We calculate the absolute difference between adult male and female deaths normalized by the 1917 population and use this as the treatment variable in equation (2.6) instead of *FluIntensity*. This variable, *GenderDistortion*, measures whether more men than women (or vice versa) died in the district, possibly making it more difficult to find a (new) partner. Table 2.3 presents the results.

The upper panel of Table 2.3 shows the marriage market effects of adult mortality. Columns (1), (3), and (5) suggest that the widow(er) rate increased significantly following the pandemic. Notably, examining the dynamics of this effect, this increase only originate from the immediate and short term (the 1918-19 and 1920-21 time windows) and is not significant and negative in the 1922-27 (*Later*) time period.[43] Table 2.3 also shows that the

ter a large mortality shock, as a society-wide pandemic will bring survivors closer together (Bowlby, 1973, 1988; Hazan and Shaver, 1994; Hill and Hansen, 1962). Empirical evidence on the two hypotheses and their link to marriage rates is however scarce and mixed (see Nobles *et al.*, 2015).

[42] A clear majority of the cohort population 1915–1927 were dead by 2013. Unfortunately we do not have access to any administrative information on civil status for people still alive.

[43] For results separated by the three time periods *Peak*, *After* and *Later*, see Appendix Table 2.B4.

Table 2.3: Marriage market

	All		Rural		Urban	
	Widowed (1)	Married (2)	Widowed (3)	Married (4)	Widowed (5)	Married (6)
EFFECTS OF ADULT MORTALITY						
Exposure	0.029***	-0.006	0.015*	-0.017**	0.035**	0.004
	(0.009)	(0.006)	(0.007)	(0.007)	(0.014)	(0.008)
Baseline	0.403	0.313	0.381	0.301	0.468	0.343
N	3,700	3,700	2,767	2,767	933	933
EFFECTS OF GENDER DISTORTION						
Gender Distortion	0.037	-0.014*	0.030	-0.024***	0.024	-0.004
	(0.026)	(0.008)	(0.028)	(0.007)	(0.047)	(0.017)
Baseline	0.403	0.313	0.381	0.301	0.468	0.343
N	3,700	3,700	2,767	2,767	933	933

Note: Annual data on health district level. N refers to the number of health districts x the number of time periods. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. The dependent variables are marriage rate and widow rate. *Exposure* is used for readability and is defined as $Exposure_{it} := FluIntensity_{it} \times D_{Post}$, where $D_{Post} = 1$ if $t \in [1918, 1927]$, otherwise 0. All regressions include district and year fixed effects. Standard errors in parentheses, clustered at the district level.

pandemic lowered marriage rates in rural areas (column 4) that likely had a less dynamic marriage market than urban areas.[44] Examining dynamics here (see Table 2.B4 in the Appendix) there are negative effects on marriage rates during the pandemic, but the main effect stems from depressed marriage rates in the long run. The pandemic, hence, caused a one-off shock to the marriage market during its peak, which was not compensated in later time periods. Instead marriage rates declined further.

The noted widowhood effect of 0.03 translates to a 7.4 percent increase above the pre-flu mean in the 1918-21 period, and the reduction in share of marriages added to the disturbance in the marriage market. With a baseline annual marriage rate of 3.01 per 1,000 people in rural areas, the estimate of 0.017 implies around 5.6 percent fewer marriages. Accordingly the long-term decline in marriage rates, rather than pandemic-induced couple-disruptions, seems decisive to the overall marriage market effect.

The lower panel of Table 2.3 suggests that imbalances in the sex ratio played a role in the decreased marriage rates in rural areas. Including both treatment variables (*FluIntensity* and *GenderDistortion*) at the same time, however, only *FluIntensity* remains significant with an unchanged coefficient of 0.017, a result which provides support for the importance of economic conditions and behavioral changes driving the decline in nuptiality, rather than mechanical effects.

[44] Stockholm was exceptional in its acceptance of fertility and co-habitation without marriage (see Matovic, 1986). Our results on marriages will not capture this. For fertility effects, however, we find a reduction in both legitimate and illegitimate births in urban areas (see Table 2.4). It could thus well be that also this form of family formation was distorted by the pandemic.

In conclusion, it seems that a substantial part of the observed fertility effect stems from the marriage market, in particular long-term declining marriage rates.[45] This result is interesting from a contemporary perspective. For the COVID-19 pandemic mortality of potential parents is not a viable mechanism for fertility changes, but pandemic-induced economic and psychological uncertainty may well change family formation behavior.

Mother Characteristics

Given that we find the decline in birth rates to be substantially driven by a decrease in nuptiality, it is also relevant to investigate *who* changed their fertility behavior. In this and the following section, we are, thus, examining compositional changes *within* the reduced number of births we identified in Table 2.1. We, therefore, change the specification from rates to a logarithmic specification in this subsection to identify changes in birth characteristics – given the knowledge that highly affected districts experienced lower birth numbers after the pandemic. Information on mother characteristics from the midwife journals gives a unique opportunity to answer this question. This analysis is of interest in itself, but is also motivated by the fact that changes in birth characteristics due to the flu would have great consequences for the interpretability of results in studies examining later life effects of in utero influenza exposure. If children conceived shortly after the mortality shock have better predisposition than those conceived during the pandemic, adverse health and income effects of being in utero during a shock will be overestimated.

In examining compositional changes we look for differences as compared to the 'normal' years 1915 to 1917. With annual data we focus on the time of actual *birth* and specify the following model:

$$ln(MotherType_{it}) = \alpha_i + \beta_1 ln(FluIncidents_{it}) \times D_{Post} + D_{Post} + \gamma ln(births_{it}) + \lambda_t + \epsilon_{it} \tag{2.7}$$

where $ln(MotherType_{it})$ is the natural logarithm of the number of births in year t to married, single, first-time or not first-time mothers. $ln(FluIncidents_{it})$ is the logarithm of the cumulative number of deaths between August 1918 and March 1919. We also include the log of the total number of births in district i in year t, $ln(births_{it})$, to account for the fact that fertility was reduced in districts heavily affected by the pandemic.[46] We, thus, only examine which type of births was or was not reduced disproportionately to the general fertility decline caused by the pandemic. As the dependent variables represent interdependent states we allow the error terms to be correlated across regressions. Appendix Figures 2.C3 and 2.C4 – showing the time trends for the variables married, single, first-time and not first-time mothers in the highest and lowest district quartiles in terms of influenza exposure, respectively – indicate no significant difference in trends before the flu.

[45] The quartile of rural districts least and most affected by the pandemic exhibited adult mortality of 3.1 and 5.0 per 1,000 people, respectively. The difference between these two types of districts correspond to a 10.6 percent reduction in the annual marriage rate. With a long-term decrease in conception rates of 10.2 percent in rural areas it seems that a substantial part of the observed fertility decrease stems from reduced marriage rates.

[46] This is, of course, a bad control variable. However, results are unchanged when not including it.

Table 2.4: Mother type

	All		Rural		Urban	
	(1)	(2)	(3)	(4)	(5)	(6)
MARITAL STATUS						
	ln (*married*)	ln (*single*)	ln (*married*)	ln (*single*)	ln (*married*)	ln (*single*)
Exposure	0.008	-0.016	0.025***	-0.002	-0.044**	-0.086***
	(0.007)	(0.013)	(0.006)	(0.014)	(0.017)	(0.031)
Baseline	0.889		0.891		0.885	
N	4,033		2,996		1,037	
PARITY						
	ln (*firstbirth*)	ln (*notfirst*)	ln (*firstbirth*)	ln (*notfirst*)	ln (*firstbirth*)	ln (*notfirst*)
Exposure	-0.006	0.017	0.018	0.030	-0.074***	-0.043**
	(0.021)	(0.024)	(0.029)	(0.034)	(0.024)	(0.017)
Baseline	0.241		0.234		0.263	
N	4,214		3,177		1,037	

Note: Annual data on health district level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. Results from estimating SUR models for married/unmarried and first birth/not first birth separately, standard errors in parentheses. All regressions include district and year fixed effects and the log of the total number of births. *Exposure* is used for readability and is defined as $Exposure_{it} := ln(FluIncidents_{it}) \times D_{Post}$, where $D_{Post} = 1$ if $t \in [1918, 1927]$, otherwise 0.

The upper panel of Table 2.4 shows a fertility shift to married mothers in rural areas. This indicates that more stable families had children after the flu, in line with a shift into higher social status parents (Richter and Robling, 2013). In urban areas the reduction in fertility is more evenly distributed, with negative effects for both married and single mothers, with stronger effects for single mothers. We also note relatively fewer births to first-time mothers, which is consistent with the postponement of would-be-parents of their first births during economic downturns documented in the literature (Goldstein *et al.*, 2013; Lanzieri, 2014; Neels, 2010). This indicates that first-time mothers delayed births, again implying a shift of the remaining births into existing families. Overall, we find an indication that urban areas were more affected by economic uncertainty in their fertility decisions.[47]

Social Gradient

As a third way to examine potential mechanisms we investigate the social gradient in changed fertility behavior. This investigation is also motivated by the results in Brown and Thomas (2018) and Beach *et al.* (2018), who note a shift towards lower social status parents shortly after the pandemic in the US.

With the available data we cannot directly observe socioeconomic status (SES) as income or occupational data for those born during our period are not available on the individual or the district level. Instead, we take advantage of having information on individuals'

[47] For results separated by the three time periods *Peak*, *After* and *Later*, see Appendix Table 2.B5.

last names and follow Clark (2014), who shows in a detailed study on several countries, including Sweden, that last names provide a good measure of social position. We classify individuals into social groups according to their last name. Here, we define two social groups: (1) nobility/high social status (aristocratic and Latin names) and bourgeoisie (names including or ending on Lund/-lund, Berg/-berg, Gren/-gren, -quist, -ström) and (2) others (including names ending on -son or -dotter). The vast majority of our individuals (76.4 percent) falls into the second category. 22.2 percent are born into the category 'bourgeoisie' and 1.4 percent constitute children of nobility/high social status parents.[48] We create a dummy variable for being born with 'high social status' (*HighSES*) taking the value one when the last name is 'noble' or 'bourgeoisie', and zero otherwise. We lag the date of birth by nine months to approximate the date of conception and estimate a linear probability model (LPM) of the probability of being conceived in a family with high social status.[49] As the number of births to high-SES parents varies considerably between months and across districts and is often zero for a particular month-district combination, we use the individual-level data in this subsection. Thereby, we are able to estimate the likelihood of a person born in month m in year t to have high-SES parents, again given the lower number of births due to the pandemic.

$$HighSES_{yim} = \alpha_i + \beta_1 FluIntensity_{im} \times D_{Post} + D_{Post} + \lambda_m + \epsilon_{im} \tag{2.8}$$

where $HighSES_{yim}$ indicates whether an individual y is born with high social status in district i in period (month-year) m. $D_{Post} = 1$ if $m \in [Aug1918, Dec1927]$, otherwise 0. *FluIntensity* is defined as above. Appendix Figure 2.C5 shows that the time trend for births to high-SES parents in the highest and lowest district quartiles in terms of influenza exposure was not different before the flu.

Table 2.5 shows no differential effect in the overall and rural sample. In urban areas, however, we observe a clear shift towards parents of higher social status, with a higher proportion of individuals with high-status names being conceived, after the flu. In the previous section, we found a shift towards more stable families in rural areas with negative results across the board for urban areas. Social status seems to be a more relevant indicator for urban areas, where clearly high social status parents were less affected by economic conditions and uncertainty and therefore did not reduce fertility as a consequence of the flu.[50]

Robustness Checks

In this section, we present several robustness checks to address potential concerns with our analysis. The main issue of concern is that the observed negative fertility effect may follow from migration. If life became more difficult in severely affected districts, people might choose to move away. Also, one spouse may move temporarily to avoid the risk of

[48] These numbers mirror official statistics and census data on the share of high-SES individuals in fertile age quite well. The annual publication *Befolkningsrrelsen* provides statistics on the occupation of fathers to newborns, and suggests that about 30 percent of fathers in the period 1911-1919 were classified as high-SES individuals.

[49] A logit model produces similar results.

[50] For results separated by the three time periods *Peak*, *After* and *Later*, see Appendix Table 2.B6

Table 2.5: Social gradient in conceptions

	All (1)	**Rural** (2)	**Urban** (3)
Exposure	0.0286	-0.0007	0.0982***
	(0.026)	(0.026)	(0.029)
Baseline	0.240	0.229	0.261
N	1,209,203	771,663	437,540

Note: Monthly data on the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. Dependent variable: dummy variable taking on the value one if born with a surname representing high social status, and 0 otherwise. All regressions include district and month-year fixed effects. Standard errors in parentheses, clustered at the district level. *Exposure* is used for readability and is defined as $Exposure_{it} := FluIntensity_{it} \times D_{Post}$, where $D_{Post} = 1$ if $t \in [Aug1918, Dec1927]$, otherwise 0.

infection if the other was ill, restricting the possibilities of conception. Although we are looking at conception *rates* it could be the case that young people in fertile age migrated more, which would bias our measure of conception rates downwards.

The left part of Table 2.6 presents estimates for the impact of the pandemic on annual migration rates.[51] All estimates in columns 1-3 are insignificant, but suggest that, if anything, there was an *inflow* into heavily affected areas. The right part of Table 2.6 assesses the importance of this inflow. In column 4 we use larger geographical units and repeat the analysis of equation (2.4) on the county level, reducing the number of geographical units from 367 to 25. Results are very similar to the main analysis.[52] Columns 5-7 drop counties that were characterized by particularly high outmigration (*Blekinge*, *Västmanland*, and *Kronoberg*). Again, the results are similar to Table 2.1. All in all, we conclude that selective migration does not represent a major threat to identification.

Furthermore, biological effects may be present for longer than we assume, i.e., beyond the *Peak* and *After* periods. The literature does not have a clear answer to how long women, and possibly also men, are negatively affected in their ability to reproduce following an influenza infection (Wiwanitkit, 2010). The positive effect on marital fertility in 1920–21 in rural areas contradicts this notion. Also, possible negative health effects would affect women who were infected but survived the infection. We would therefore expect such fertility effects to stem from morbidity, not mortality. Table 2.1 illustrates that this is not the case.

[51] We calculate the net migration rate for every district using population numbers, number of deaths and births. For every year the residual provides a measure of how many people moved in or out of the district, subject to random measurement error.

[52] The exercises where we aggregate and run the analysis on the county level is also informative regarding the potential challenge that pandemic intensity in one district influences pandemic intensity in a district close by, and should also handle potential outlier districts.

Table 2.6: Robustness checks

	Effect on Migration			**Adjusting for Migration**			
				County-level	*Dropping High-Migration Counties*		
	All	Rural	Urban	All	All	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
morbidity	-0.0101	-0.0164	0.0265	-0.0033*	-0.0001	-0.0002	0.0004
	(0.024)	(0.029)	(0.040)	(0.002)	(0.000)	(0.000)	(0.001)
adult mortality	1.2600	3.2865	-0.2909	-0.0857*	-0.0675***	-0.0399**	-0.0995**
	(1.327)	(2.030)	(0.893)	(0.045)	(0.022)	(0.020)	(0.040)
child mortality	0.3575	2.2688	-0.2512	-0.1425*	-0.0889***	-0.0134	-0.1138***
	(0.783)	(2.615)	(0.484)	(0.079)	(0.018)	(0.017)	(0.013)
N	3,788	2,842	946	3,732	42,765	32,388	10,377
N (cluster)	366	270	96	25	337	249	88
Baseline	1.51	-2.45	11.84	1.62	1.81	1.69	2.11

Note: Columns (1)-(3) use annual data on health district level, column (4) uses monthly data on county level, columns (5)-(7) use monthly data on health district level. N refers to the number of health districts/counties x the number of time periods. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. In columns (1)-(3) the dependent variable is net migration rates (positive numbers representing inmigration) defined as migrants per 1,000 population. In columns (4)-(7) the dependent variable is the conception rate. All regressions include district and month-year fixed effects. Standard errors in parentheses, clustered at the district level.

To take account of possible spatial correlation across districts, we also estimate our main results using Conley standard errors.[53] Estimates using cut-off levels of 50 km and 100 km are presented in Appendix Table 2.B7 and show that results are unaffected.

Finally, we check the sensitivity of results to changes in district borders over time, by including dummy variables which take the value one for the year of a border change and thereafter, and to changes in the urban-rural classification. These results can be found in Table 2.B9 in the Appendix. The results are unchanged. In summary, the robustness checks support the research design and the validity of the findings.

2.6 Conclusion

In this paper we have examined fertility response to the 1918-19 influenza pandemic in Sweden, which implied a great mortality and morbidity shock in a country that was neutral in WWI. We show that the pandemic affected fertility rates not only in the short term, but even a decade later. Specifically, we find some evidence of a positive fertility response in rural areas following the pandemic. However, this short-term effect is of second-order importance and is overshadowed by a large fertility reduction in the long run. Furthermore, in urban areas the effect of the pandemic on fertility is negative throughout the whole

[53] We use the procedure written by Hsiang (2010) and the extension written by Thiemo Fetzer.

study period. Our results, thus, suggest that the often noted positive fertility response to mortality shocks and pandemics is short-lived.

Examining heterogeneity effects we find that poor underdeveloped districts largely drive the negative long-term effect, suggesting a negative income effect on fertility. We further identify changes on the marriage market as an important mechanism. These changes on the marriage market represent mechanical effects due to the need to find a new partner, but more importantly behavioral and economic effects following increased uncertainty and reduced incomes. Overall, the mortality shock increased the cost of having children and, thus, reduced fertility in the long run.

We also find compositional effects: within the net fertility decline we observe a relative increase in births to married women and parents of higher social status. This result on parental composition is interesting in itself, but may also have implications for how we interpret the often noted later life effects of in utero exposure to health shocks. A recent literature assesses the implications of the observation that cohorts with in utero exposure to the 1918-19 influenza pandemic were born to lower socioeconomic households in the US (Beach *et al.*, 2018; Brown and Thomas, 2018). Our results suggest that the composition of cohorts born after the pandemic may also be important to consider.

It is all together evident that a deadly pandemic can have fertility effects that go far beyond the infection period itself. Putting the noted negative effect on family size in perspective of a quantity-quality trade-off, we may expect parents to invest more into the education of those (fewer) children born after the pandemic in highly affected districts. The fact that we find compositional effects in favor of parents with high socioeconomic status may further reinforce this effect. According to the results presented by Parman (2015), these effects may even hold true for older siblings, if resource allocation within the family changes due to the pandemic. For future research, it would, therefore, be interesting to examine educational outcomes of the children of those families who altered their fertility behavior due to the flu.

As stated in the beginning of this paper the event of a pandemic can cause major losses, and the past year has reminded us that deadly viruses may spread very quickly across countries. The are many similarities between the 1918-19 influenza and the COVID-19 pandemic, not least that both pandemics were caused by new and very contagious viruses and that transmission was similar. At the same time there are significant differences regarding who the two pandemics affected the most and society is very different compared to a century ago, which makes it difficult to draw straightforward conclusions from effects following the 1918-19 influenza for developed countries today. Nevertheless we believe that our results may still be informative for the sizeable population that lives under similar conditions to early 20^{th} century Sweden.

2.7 Appendix A: Variable Definitions

Information comes from church records digitized by the Federation of Swedish Genealogical Societies in the *Swedish Death Index*, the 1950 Census, and purposely digitized historical records from the National Medical Board, the historical midwife journals, the Swedish yearbook of municipalities and the annual publication on poor relief. The data on railway stations come from Olofsson (1921).

Adult mortality All cause deaths between August 1918 and March 1919 in the age group 20-40 up to conception month in district *i*.

Child mortality All cause deaths between August 1918 and March 1919 in the age group 0-10 up to conception month in district *i*.

Influenza morbidity All cases of influenza and pneumonia morbidity reported between August 1918 and March 1919 up to conception month in district *i*.

Conception rate Conceptions in district *i* in month *m* divided by population.

D_{Peak} Dummy variable taking on value one if the period falls within the influenza peak period, August 1918-November 1918, otherwise 0.

D_{After} Dummy variable taking on value one if the period falls within one to two years after the pandemic's peak, December 1918-December 1920, otherwise 0.

D_{Later} Dummy variable taking on value one if the period falls within 1921-1927, otherwise 0

D_{Post} Dummy variable taking on value one if the period falls within 1918-1927, otherwise 0

Rural Dummy variable taking on value one if a health district is classified as extra provincial or provincial, otherwise 0.

Urban Dummy variable taking on value one if a health district is classified as municipal or city, otherwise 0.

Midwives Numbers of midwives working in a health district (proxy of local medical infrastructure).

Married mothers Number of births in a year to married mothers in a district.

Single Number of births in a year to unmarried mothers in a district.

First birth Number of births in a year to first-time mothers in a district.

Not first Number of births in a year to not-first time mothers.

Poverty The share of the population living in public poorhouses in a district.

Taxable income Per capita taxable earnings as reported to tax authorities, normalized by 1917 prices.

Capital income Per capita asset yields, rents and dividends in a district as reported to tax authorities, normalized by 1917 prices.

Private property Per capita assessed value of private properties in a district, normalized by 1917 prices.

Local revenue Per capita public revenue in a district.

Local assets Per capita value of public assets by December 31 in a district.

Local debt Per capita public debt by December 31 in a district.

Population density Population per hectar of area in a district.

Sami Share of population belonging to the Sami people in 1910 in rural districts.

Widowed rate Widow rate (incidence of new widowhood) relative to population in a health district.

Marriage rate Marriage rate (incidence of new marriages) relative to population in a health district.

High SES Dummy variable taking on value one, if an individual is born with a last name defined as belonging to the nobility (aristocratic and Latin names) or bourgeoisie (last names including or ending with lund, berg, gren, quist or ström), otherwise 0.

Migration Net migration rates (positive numbers representing in-migration), migrants per 1,000 population.

Stillbirths Number of stillbirth per 1,000 births in a district.

Miscarriage Number of miscarriages per 1,000 births in a district .

Births Total number of births in a district.

Railway Dummy taking on value one if the district had a railway station in 1918, otherwise 0.

Nr of railway stations Number of railway stations in the district in 1918.

2.8 Appendix B: Tables and Figures

Table 2.B1: Descriptive statistics, all districts.

	All Periods					Specific Periods - mean			
	N	Min	Max	Mean	SD	*Before*	*Peak*	*After*	*Later*
MONTHLY DATA PER 1,000 PEOPLE, DISTRICT LEVEL									
Conception rate	46,861	0	20.48	1.78	1.15	1.81	1.82	2.07	1.64
Morbidity rate*	29,551	0	392.28	3.16	8.84	1.79	15.75	3.86	0.98
Death rate (20-40)	46,861	0	9.03	0.16	0.26	0.15	0.54	0.18	0.12
Children death rate (0-10)	46,861	0	5.52	0.20	0.26	0.22	0.35	0.24	0.16
Male death rate (20-40)	46,861	0	4.65	0.08	0.16	0.08	0.30	0.10	0.06
Female death rate (20-40)	46,861	0	5.80	0.08	0.15	0.07	0.24	0.08	0.06
Population	46,861	0.86	98.4	14.01	11.29	13.72	13.81	13.78	14.28
Net migration rate	42,747	-136.01	63.92	-0.24	0.76	0.19	-0.01	-0.43	-0.37
DATA FROM 1917-18, DISTRICT LEVEL									
Midwives p. birth	4,092	0	0.11	0.03	0.01				
Poverty rate	4,080	0.002	0.045	0.014	0.006				
Private property	4,080	93.61	5513.22	1241.40	544.89				
Population density	4,080	0.002	28.03	1.29	3.37				
Taxable income	4,080	20.42	1929.38	376.77	298.57				
Conceptions (rate 1917)	4,092	0	14.83	1.81	1.15				
Sami share (1910)	3,960	0	0.20	0.003	0.018				
Railway (dummy)	4,092	0	1	0.59	0.49				
Nr. of railway stations	4,092	0	39	2.55	4.15				
ANNUAL DATA, DISTRICT LEVEL									
Widowed rate**	3,700	0	6.11	0.66	0.51	0.40	0.74	0.64	0.76
Marriage rate**	3,700	0	2.35	0.31	0.18	0.31	0.32	0.34	0.29
Gender distortion	3,700	0	4.42	0.55	0.69	0	0.67	0.72	0.75
Married mothers	3,917	4	1,117	215.47	170.31	226.64	223.87	246.77	193.64
Single mothers	3,917	0	339	28.36	27.35	27.43	26.02	34.48	27.41
First time mothers	3,917	0	443	63.85	52.27	61.03	66.32	78.45	58.63
Not first time mothers	3,917	4	930	181.23	143.92	194.09	185.31	204.19	163.50
Miscarriage rate (per 1,000 births)	3,673	0	363.64	41.44	26.32	37.72	42.73	37.40	44.01
Stillbirths rate (per 1,000 births)	3,887	0	157.89	22.77	15.48	24.42	22.40	21.82	22.41
Number of midwives	3,917	1	43	7.42	6.38	7.35	7.50	7.49	7.40
Number of births	3,917	5	1,368	245.05	192.81	255.14	251.38	282.61	222.19
High SES***	1,296,073	0	1	0.24	0.42	0.24	0.24	0.24	0.23
COUNTY-LEVEL CONTROL VARIABLES									
Poverty share (%)	324	2.08	8.89	4.51	1.32	4.23	4.10	3.99	4.99
Earnings (SEK/capita)	324	138.46	1592.76	412.93	170.71	265.81	287.95	419.83	524.20
Capital income (SEK/capita)	324	4.17	274.43	38.85	27.29	18.31	24.70	49.85	49.94

The table shows the descriptive statistics for the variables used, and means for *Before* ($t \geq$January 1915 & $t <$August 1918), *During* ($t \geq$August 1918 & $t \leq$November 1918), *After* ($t \geq$December 1918 & $t \leq$December 1920) and *Later* ($t \geq$January 1921 & $t \leq$ December 1927) of the 1918–19 flu pandemic. *– morbidity data is only available from 1916 through 1921. **– rate per 1,000 people. The marriage and widow rates are incidence rates (for new marriages and widowhood). ***– individual-level data.

Figure 2.B1: Birth numbers from different sources compared, as percent of births recorded in official population statistics

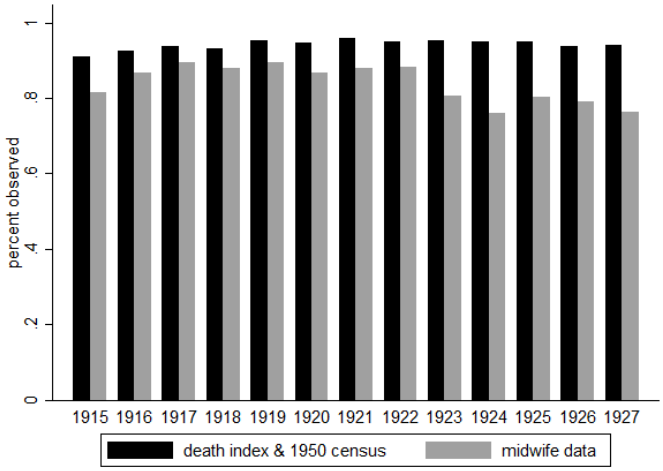
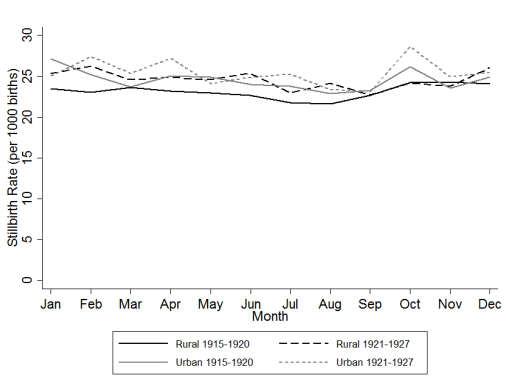
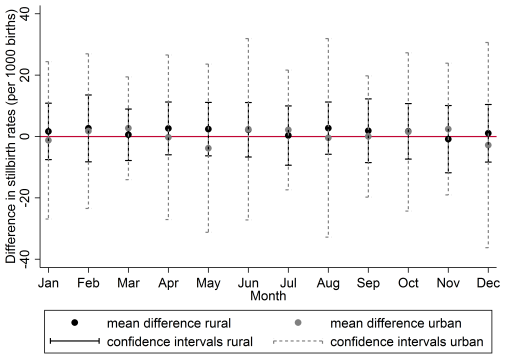


Figure 2.B2: Monthly stillbirth rates (per 1,000 births) at the county level

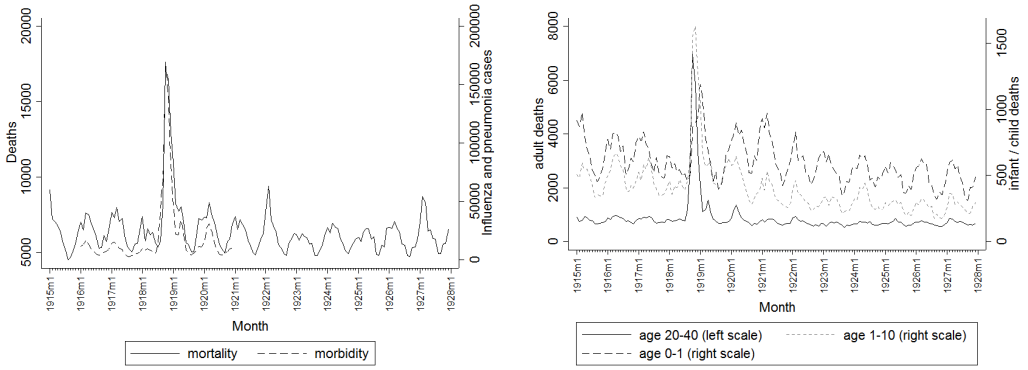


(a) Stillbirth rates 1915–20 and 1921–27.



(b) Changes in stillbirth rates from 1911–20 to 1921–27

Figure 2.B3: Monthly influenza and pneumonia morbidity (1916–21) and overall mortality (1915–27) in Sweden



(a) Overall deaths and flu incidents **(b)** Deaths by age group

Figure 2.B4: Density of peak month in morbidity, child and adult mortality

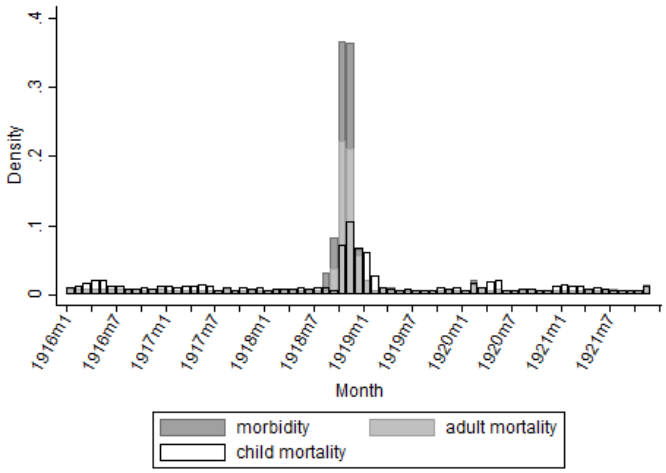
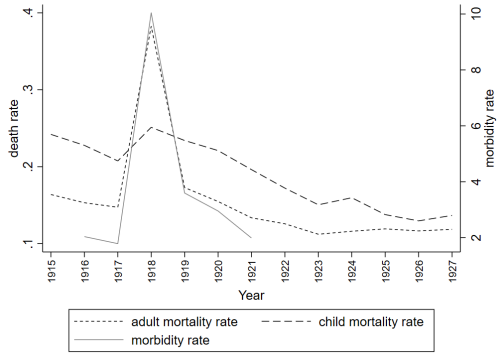
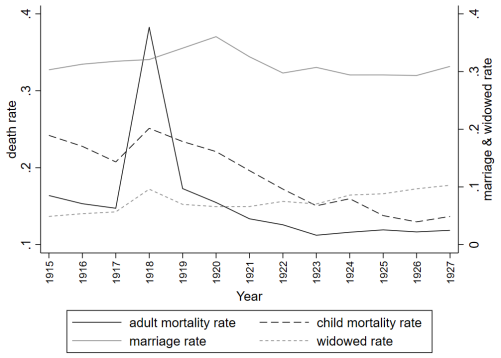


Figure 2.B5: Marriage, widow, morbidity and mortality rates



(a) Morbidity and overall mortality rates



(b) Marriage, widow, and overall mortality rates

Table 2.B2: Fertility effects of the influenza pandemic, using 1917 population for outcome variable.

	All				Rural		Urban	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. INFLUENZA MORBIDITY								
Peak × morbidity	-0.0008**	-0.0008*	-0.0012***	-0.0011***	-0.0012***	-0.0011***	0.0000	0.0000
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0010)	(0.0010)
After × morbidity	0.0005**	0.0005**	0.0001	0.0001	0.0001	0.0001	0.0007	0.0007
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0006)	(0.0005)
Later × morbidity	-0.0000	0.0001	-0.0002	-0.0000	-0.0002	-0.0000	0.0001	0.0001
	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0006)	(0.0006)
During × morbidity × *urban*			0.0012	0.0011				
			(0.0011)	(0.0011)				
After × morbidity × *urban*			0.0006	0.0006				
			(0.0006)	(0.0006)				
Later × morbidity × *urban*			0.0003	0.0001				
			(0.0007)	(0.0006)				
B. ADULT MORTALITY								
Peak × adult mortality	-0.0640***	-0.0716**	-0.0597**	-0.0604**	-0.0597**	-0.0373*	-0.0737**	-0.1109**
	(0.0191)	(0.0335)	(0.0238)	(0.0268)	(0.0238)	(0.0190)	(0.0305)	(0.0542)
After × adult mortality	0.0452***	0.0446***	0.0599***	0.0567***	0.0599***	0.0570***	0.0185	0.0126
	(0.0152)	(0.0159)	(0.0104)	(0.0118)	(0.0104)	(0.0119)	(0.0268)	(0.0239)
Later × adult mortality	-0.0287*	-0.0283	-0.0219*	-0.0252**	-0.0219*	-0.0257***	-0.0434	-0.0564**
	(0.0174)	(0.0177)	(0.0121)	(0.0097)	(0.0121)	(0.0096)	(0.0325)	(0.0266)
Peak × adult mortality × *urban*			-0.0140	-0.0185				
			(0.0385)	(0.0375)				
After × adult mortality × *urban*			-0.0414	-0.0447*				
			(0.0286)	(0.0267)				
Later × adult mortality × *urban*			-0.0215	-0.0328				
			(0.0345)	(0.0283)				
C. CHILD MORTALITY								
Peak × child mortality	-0.1405***	-0.1336***	-0.0891**	-0.0901**	-0.0891**	-0.0971**	-0.1626***	-0.1297***
	(0.0268)	(0.0247)	(0.0423)	(0.0437)	(0.0423)	(0.0423)	(0.0297)	(0.0391)
After × child mortality	0.0111	0.0048	0.0696***	0.0566***	0.0696***	0.0565***	-0.0125	-0.0114
	(0.0292)	(0.0293)	(0.0123)	(0.0125)	(0.0123)	(0.0125)	(0.0282)	(0.0252)
Later × child mortality	-0.0403**	-0.0562***	0.0231	-0.0144	0.0231	-0.0141	-0.0629***	-0.0600***
	(0.0176)	(0.0157)	(0.0159)	(0.0152)	(0.0159)	(0.0151)	(0.0137)	(0.0127)
Peak × child mortality × *urban*			-0.0736	-0.0485				
			(0.0516)	(0.0491)				
After × child mortality × *urban*			-0.0820***	-0.0675**				
			(0.0306)	(0.0274)				
Later × child mortality × *urban*			-0.0860***	-0.0448**				
			(0.0209)	(0.0203)				
Add. controls	No	Yes	No	Yes	No	Yes	No	Yes
County trend	No	Yes	No	Yes	No	Yes	No	Yes
N	46,861	46,861	46,861	46,861	35,200	35,200	11,661	11,661
N (cluster)	367	367	367	367	270	270	97	97
Baseline	1.81	1.81	1.81	1.81	1.68	1.68	2.14	2.14

Monthly data on health district level. N refers to the number of health districts x the number of time periods. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. The dependent variable is conception rate. All regressions include district and month-year fixed effects. Standard errors in parentheses, clustered at the district level. Additional control variables include the log of the number of midwives, the log of earnings and the log of capital income in 1917 prices and the poverty share. The *Peak* period includes August 1918 to November 1918; *After* includes December 1918 to December 1920; *Later* includes January 1921 to December 1927. Morbidity and mortality rates are calculated as the cumulative sum of influenza cases/all-cause deaths occurring during the flu period, normalized by the district population in 1917.

Table 2.B3: Heterogeneity analysis using continuous variables

	(1) Midw.	(2) Poverty	(3) Priv Prop.	(4) Popdens	(5) Taxinc	(6) Conceptions	(7) Sami	(8) Railway
ALL DISTRICTS								
Exposure	-0.095	-0.013	0.282	-0.044	-0.026	-0.080**	-0.062***	-0.050*
	(0.069)	(0.066)	(0.325)	(0.029)	(0.129)	(0.034)	(0.022)	(0.027)
Exposure $\times Variable$	1.530	-1.890	-0.048	-0.024	-0.014	-0.077***	0.242	-0.013
	(2.325)	(1.589)	(0.045)	(0.024)	(0.022)	(0.025)	(0.292)	(0.024)
N (clust)	367	367	367	367	367	367	367	367
N	46,861	46,861	46,861	46,861	46,861	46,861	46,861	46,861
RURAL DISTRICTS								
Exposure	-0.068	0.031	0.283	-0.062***	-0.165	0.041**	-0.044**	0.060***
	(0.054)	(0.032)	(0.250)	(0.023)	(0.116)	(0.018)	(0.020)	(0.019)
Exposure $\times Variable$	1.027	-1.747**	-0.046	0.166**	0.025	-0.053***	0.232	0.038**
	(1.540)	(0.880)	(0.037)	(0.069)	(0.020)	(0.009)	(0.303)	(0.016)
N (clust)	270	270	270	270	270	270	270	270
N	35,200	35,200	35,200	35,200	35,200	35,200	35,200	35,200
URBAN DISTRICTS								
Exposure	-0.095	0.009	0.270	-0.035	0.010	0.101		-0.046
	(0.111)	(0.078)	(1.162)	(0.075)	(0.478)	(0.065)		(0.049)
Exposure $\times Variable$	1.043	-2.393	-0.048	-0.029	-0.013	-0.082		-0.057**
	(3.770)	(1.863)	(0.158)	(0.043)	(0.071)	(0.042)		(0.024)
N (clust)	97	97	97	97	97	97		97
N	11,661	11,661	11,661	11,661	11,661	11,661		11,661

Note: Monthly data on health district level. N refers to the number of health districts x the number of time periods. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. The dependent variable is conception rate. All regressions include district and month-year fixed effects. Standard errors in parentheses, clustered at the district level. *Exposure* is used for readability and is defined as $Exposure_{it} := FluIntensity_{it} \times D_{Post}$, where $D_{Post} = 1$ if $t \in [Aug1918, Dec1927]$, otherwise 0. *Exposure* $\times X$ denotes the interaction of *Exposure* with the natural logarithm of the variable in the column heading.

Table 2.B4: Marriage market, divided into *Peak*, *After*, and *Later*

	All		Rural		Urban	
	Widowed (1)	Married (2)	Widowed (3)	Married (4)	Widowed (5)	Married (6)
EFFECTS OF ADULT MORTALITY						
Peak × adult mortality	0.038***	0.000	0.044***	-0.003	0.027**	0.002
	(0.010)	(0.003)	(0.014)	(0.003)	(0.013)	(0.006)
After × adult mortality	0.070***	-0.002	0.070***	-0.008***	0.065**	0.004
	(0.019)	(0.004)	(0.024)	(0.003)	(0.030)	(0.007)
Later × adult mortality	0.011	-0.009	-0.013	-0.023**	0.025	0.004
	(0.012)	(0.008)	(0.011)	(0.009)	(0.015)	(0.009)
Baseline	0.313	0.403	0.301	0.381	0.343	0.468
N	3,700	3,700	2,767	2,767	933	933
EFFECTS OF GENDER DISTORTION						
Peak × adult mortality	0.084**	-0.007	0.098***	-0.011	0.060	0.001
	(0.037)	(0.007)	(0.030)	(0.009)	(0.072)	(0.013)
After × adult mortality	-0.069*	-0.006	0.047	-0.018**	0.076	0.006
	(0.042)	(0.009)	(0.043)	(0.008)	(0.081)	(0.018)
Later × adult mortality	0.008	-0.020**	-0.000	-0.031***	-0.010	-0.010
	(0.027)	(0.010)	(0.034)	(0.009)	(0.045)	(0.021)
Baseline	0.313	0.403	0.301	0.381	0.343	0.468
N	3,700	3,700	2,767	2,767	933	933

Note: Annual data on health district level. N refers to the number of health districts x the number of time periods. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$. The dependent variables are marriage rate and widow rate. All regressions include district and year fixed effects. Standard errors in parentheses, clustered at the district level.

Table 2.B5: Mother type, divided into *Peak*, *After*, and *Later*

	All		**Rural**		**Urban**	
	(1)	(2)	(3)	(4)	(5)	(6)
MARITAL STATUS						
	ln (*married*)	ln (*single*)	ln (*married*)	ln (*single*)	ln (*married*)	ln (*single*)
Peak	0.008	0.003	0.014	-0.017	-0.011	-0.010
× adult mortality	(0.009)	(0.018)	(0.009)	(0.019)	(0.021)	(0.039)
After	0.028***	0.011	0.038***	0.010	-0.012	-0.039
× adult mortality	(0.009)	(0.018)	(0.009)	(0.019)	(0.022)	(0.039)
Later	-0.001	-0.038**	0.024***	-0.001	-0.081***	-0.154***
× adult mortality	(0.008)	(0.015)	(0.007)	(0.016)	(0.019)	(0.034)
Baseline	0.889		0.891		0.885	
N	4,033		2,996		1,037	
PARITY						
	ln (*firstbirth*)	ln (*notfirst*)	ln (*firstbirth*)	ln (*notfirst*)	ln (*firstbirth*)	ln (*notfirst*)
Peak	0.038	0.062*	0.055	0.078*	-0.024	0.002
× adult mortality	(0.028)	(0.032)	(0.039)	(0.046)	(0.031)	(0.021)
After	0.038	0.073**	0.059	0.089*	-0.044	-0.066
× adult mortality	(0.028)	(0.032)	(0.039)	(0.046)	(0.031)	(0.021)
Later	-0.046**	-0.029	-0.015	-0.013	-0.120***	-0.088***
× adult mortality	(0.023)	(0.027)	(0.031)	(0.037)	(0.027)	(0.018)
Baseline	0.889		0.891		0.885	
N	4,033		2,996		1,037	

Note: Annual data on health district level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. Results from estimating SUR models for married/unmarried and first birth/not first birth separately, standard errors in parentheses. All regressions include district and year fixed effects and the log of the total number of births.

Table 2.B6: Social gradient in conceptions, divided into *Peak*, *After*, and *Later*

	All	**Rural**	**Urban**
	(1)	(2)	(3)
Peak × adult mortality	-0.126	–0.072	-0.236***
	(0.158)	(0.187)	(0.179)
After × adult mortality	0.040	-0.009	0.159***
	(0.046)	(0.041)	(0.035)
Later × adult mortality	0.022	0.003	0.072**
	(0.021)	(0.024)	(0.034)
Baseline	0.240	0.229	0.261
N	1,209,203	771,663	437,540

Note: Monthly data on the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. Dependent variable: dummy variable taking on the value one if born with a surname representing high social status, and 0 otherwise. All regressions include district and month-year fixed effects. Standard errors in parentheses, clustered at the district level.

Table 2.B7: Fertility effects of the influenza pandemic, Conley standard errors for different cut-off levels.

	All		Rural		Urban	
	(1) 50km	(2) 100km	(3) 50km	(4) 100km	(5) 50km	(6) 100km
A. INFLUENZA MORBIDITY						
Peak × morbidity	-0.0009**	-0.0009**	-0.0013***	-0.0013***	-0.0002	-0.0002
	(0.0004)	(0.0003)	(0.0004)	(0.0004)	(0.0009)	(0.0009)
After × morbidity	0.0002	0.0002	-0.0001	-0.0001	-0.0008	-0.0008
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0006)	(0.0006)
Later × morbidity	-0.0004	-0.0004	-0.0003	-0.0003	-0.0001	-0.0001
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0007)	(0.0007)
B. ADULT MORTALITY						
Peak × adult mortality	-0.076***	-0.076***	-0.057**	-0.057**	-0.093***	-0.093***
	(0.0177)	(0.017)	(0.026)	(0.027)	(0.026)	(0.027)
After × adult mortality	0.022	0.022	0.058***	0.058***	-0.020	-0.020
	(0.022)	(0.022)	(0.016)	(0.016)	(0.033)	(0.033)
Later × adult mortality	-0.088***	-0.088***	-0.079***	-0.079***	-0.095**	-0.095***
	(0.021)	(0.021)	(0.025)	(0.025)	(0.034)	(0.034)
C. CHILD MORTALITY						
Peak × child mortality	-0.151***	-0.51***	-0.085**	-0.085**	-0.180***	-0.180***
	(0.025)	(0.025)	(0.034)	(0.036)	(0.036)	(0.036)
After × child mortality	-0.023	-0.023	0.053***	0.053***	-0.052**	-0.052**
	(0.028)	(0.028)	(0.016)	(0.016)	(0.024)	(0.024)
Later × child mortality	-0.099***	-0.099***	-0.040**	-0.040**	-0.117***	-0.117***
	(0.018)	(0.018)	(0.020)	(0.020)	(0.017)	(0.017)
N	46,861	46,861	35,200	35,200	11,661	11,661
N (cluster)	367	367	270	270	97	97
Baseline	1.81	1.81	1.68	1.68	2.14	2.14

Monthly data on health district level. N refers to the number of health districts x the number of time periods. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. The dependent variable is conception rate. All regressions include district and month-year fixed effects. Standard errors in parentheses, clustered at the district level. The *Peak* period includes August 1918 to November 1918; *After* includes December 1918 to December 1920; *Later* includes January 1921 to December 1927. Morbidity and mortality rates are calculated as the cumulative sum of influenza cases/all-cause deaths occurring during the flu period, normalized by the district population in 1917.

Table 2.B8: Fertility effects from adult and child mortality combined

	All (1)	Rural (2)	Urban (3)
Peak × adult mortality	-0.0358	-0.0431	-0.0307
	(0.0244)	(0.0338)	(0.0364)
After × adult mortality	0.0589***	0.0576***	0.0444
	(0.0188)	(0.0179)	(0.0330)
Later × adult mortality	-0.0543*	-0.0893***	0.0035
	(0.0302)	(0.0338)	(0.0355)
Peak × child mortality	-0.1135***	-0.0456	-0.1410***
	(0.0411)	(0.0621)	(0.0438)
After × child mortality	-0.0669**	0.0020	-0.0819***
	(0.0268)	(0.0176)	(0.0218)
Later × child mortality	-0.0610*	0.0330	-0.1196***
	(0.0315)	(0.0259)	(0.0092)
N	46,861	35,200	11,661
N (cluster)	367	270	97
Baseline	1.81	1.68	2.14

Dependent variable: conception rate. All regressions include district and month-year fixed effects. Standard errors are in parentheses, clustered at the district level. The *Peak* period includes August 1918 to November 1918; *After* includes December 1918 to December 1920; *Later* includes January 1921 to December 1927. Mortality rates are calculated as the cumulative sum of deaths occurring during the flu period, normalized by the district population in 1917. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$

Table 2.B9: Robustness check: Fertility effects controlling for district changes

	All (1)	Rural (2)	Urban (3)
Peak × adult mortality	-0.076***	-0.058**	-0.093***
	(0.019)	(0.024)	(0.026)
After × adult mortality	0.021	0.057***	-0.020
	(0.023)	(0.014)	(0.035)
Later × adult mortality	-0.090***	-0.083**	-0.095**
	(0.026)	(0.033)	(0.041)
N	46,756	35,109	11,647
N (cluster)	367	270	97
Baseline	1.81	1.68	2.14

Monthly data on health district level. Dependent variable: conception rate. All regressions include district and month-year fixed effects. Standard errors in parentheses, clustered at the district level. The *Peak* period includes August 1918 to November 1918; *After* includes December 1918 to December 1920; *Later* includes January 1921 to December 1927. Morbidity and mortality rates are calculated as the cumulative sum of influenza cases/deaths occurring during the flu period, normalized by the district population in 1917. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$.

2.9 Appendix C: Evidence Supporting Identification

Common time trend

In a difference-in-differences design the key identifying assumption is that fertility behavior in heavily and less affected areas would have followed a common time trend in the absence of the pandemic. This assumption is untestable, but having access to 43 months of pre-exposure data we assess its plausibility in different ways. Figure 2.C1 plots conception rates in the highest and lowest district quartiles in terms of influenza exposure. There is no significant difference in the trends before the flu (high mortality districts adjust the conception rates already in the spring of 1918, but the confidence intervals overlap) and a clearly diverging trend thereafter. The corresponding time trend graph for morbidity also shows no significant difference in trends before the influenza. We generate similar time trend graphs for morbidity and mortality rates for different age groups (available upon request). All provide very similar evidence to that of Figure 2.C1.

Figure 2.C1: Time trend for conception rate

Balancing tests

To further test the common time trend assumption we perform balancing tests and regress our influenza intensity measures on pre-flu values from the years 1916 and 1917. If the degree of influenza exposure is predicted by several baseline variables there is a concern that the intensity of the pandemic correlates with relevant unobservables. Guided by previous work on pandemics and seasonal influenza (see, e.g., Clay *et al.*, 2019; Markowitz *et al.*, 2019) we regress a number of different district pre-influenza characteristics on two different measures of the influenza intensity – defined as the absolute (left column) and relative (right column) increases in adult influenza mortality compared to the pre-influenza period – in Table 2.C1. For each observable characteristic, we also estimate differences in levels (the absence of which is not a requirement for identification) and in trends (which should

not be related to influenza exposure). The estimates show that heavily affected districts had slightly lower birth rates before the pandemic, but there are no systematic differences in trends with regard to birth rates, midwife density, infant mortality and overall mortality. The only exception in this regard is midwife density in urban areas, where the trend of heavily affected areas was more negative.

Table 2.C1: Balancing tests: Health and demography

	All Districts (N=369)				Rural Districts (N=272)				Urban Districts (N=97)			
	Levels 1917		Trends 1916–17		Levels 1917		Trends 1916–17		Levels 1917		Trends 1916–17	
	Abs	*Rel*	*Abs*	*Rel*	*Abs*	*Rel*	*Abs*	*Rel*	*Abs*	*Rel*	*Abs*	*Rel*
birthrate	-0.0549*	-0.0486*	0.0515	-0.0235	-0.0605*	-0.0623*	0.0561	-0.0221	-0.0004	0.0431	0.0020	0.0427
	(0.027)	(0.025)	(0.046)	(0.073)	(0.030)	(0.030)	(0.050)	(0.077)	(0.015)	(0.064)	(0.006)	(0.042)
infmort	-0.0242	-0.0294	0.0704	0.0565	-0.0266	-0.0389	0.0778	0.0756	0.0018	0.0106	-0.0087	0.0069
	(0.030)	(0.022)	(0.068)	(0.051)	(0.033)	(0.028)	(0.075)	(0.064)	(0.013)	(0.015)	(0.012)	(0.012)
midwifedens	-0.0176	-0.0102	-0.0168	-0.0077	-0.0236	-0.0270	-0.0091	-0.0003	0.0341	0.0788*	-0.0877**	-0.0913**
	(0.038)	(0.021)	(0.015)	(0.013)	(0.044)	(0.022)	(0.016)	(0.013)	(0.034)	(0.037)	(0.029)	(0.032)
deathrate	-0.0459	-0.0231	-0.0121	0.0114	-0.0503	-0.0276	-0.0134	0.0145	-0.0016	-0.0122	-0.0004	-0.0071
	(0.024)	(0.022)	(0.018)	(0.020)	(0.026)	(0.026)	(0.020)	(0.024)	(0.005)	(0.012)	(0.003)	(0.008)

Monthly data on health district level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. Labels *Abs* and *Rel* refer to whether excess mortality during the pandemic is described in absolute or relative terms. All dependent variables are based on health district data.

In Table 2.C2 we conduct the same balancing tests for some indicators of the local economy and public finances. Again, we find some evidence that heavily affected districts had different pre-influenza means of these variables, but the common time trend assumption cannot be rejected for average taxable earnings, local public revenue, local public assets, local public debt and local poverty rates. Only for property values there is some evidence of diverging trends in rural areas, where more affected districts had a more positive trend in the pre-influenza period. Taken as a whole, however, the evidence provided in Tables 2.C1 and 2.C2 supports the main identification strategy: out of 60 tests, only 2 are significant at the 5 percent level, and 5 are significant at the 10 percent level.

Table 2.C2: Balancing tests: Local finances

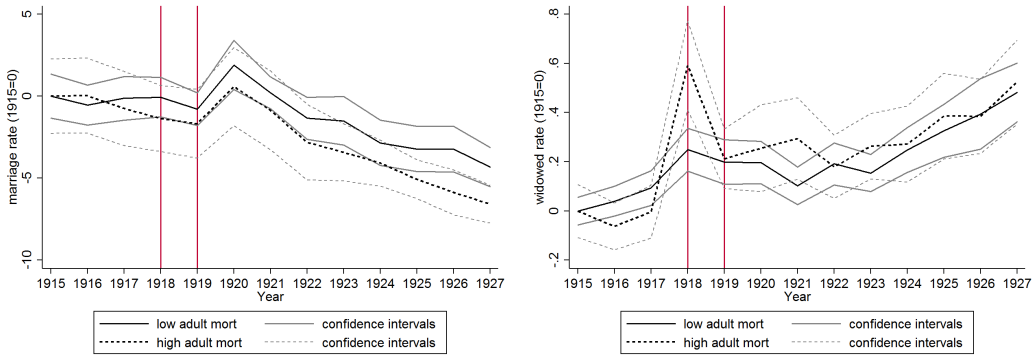
	All Districts (N=369)				Rural Districts (N=272)				Urban Districts (N=97)			
	Levels 1917		Trends 1916–17		Levels 1917		Trends 1916–17		Levels 1917		Trends 1916–17	
earnings	0.0236	-0.1291	-0.0304	0.0255	0.1805***	0.1204*	-0.0811	0.0108	-0.1267	-0.2565	0.0756	-0.2895
	(0.148)	(0.182)	(0.071)	(0.067)	(0.054)	(0.059)	(0.043)	(0.064)	(0.416)	(0.436)	(0.303)	(0.221)
propvalue	0.0334	0.0393	0.0860*	0.1602*	0.0561	0.1684**	0.0669	0.0992*	0.1886	-0.2811	-0.0278	0.1382
	(0.065)	(0.114)	(0.039)	(0.068)	(0.050)	(0.052)	(0.035)	(0.046)	(0.253)	(0.363)	(0.106)	(0.138)
locrevenue	-0.1034	-0.2206	0.0178	0.0425	0.0693**	0.0342	-0.0145	-0.0043	-0.3227	-0.3222	0.0967	0.1331
	(0.141)	(0.198)	(0.023)	(0.033)	(0.023)	(0.018)	(0.012)	(0.015)	(0.354)	(0.491)	(0.092)	(0.112)
locassets	-0.0756	-0.1314	0.0380	0.0608	0.0401**	0.0186*	-0.0125	0.0044	-0.2141	-0.0309	0.1603	0.0806
	(0.098)	(0.115)	(0.034)	(0.041)	(0.015)	(0.008)	(0.007)	(0.014)	(0.215)	(0.248)	(0.098)	(0.121)
locdebt	-0.0759	-0.1566	0.0541	0.1081	0.0285	-0.0031	-0.0152	-0.0035	-0.1296	-0.0563	0.2429	0.4184
	(0.100)	(0.132)	(0.041)	(0.064)	(0.019)	(0.009)	(0.010)	(0.014)	(0.263)	(0.332)	(0.186)	(0.242)
poverty	0.0379	0.0144	0.0215	0.2430	0.1156*	-0.0408	0.0051	-0.0150	-0.4860	0.3356	-0.4351	0.9867
	(0.104)	(0.098)	(0.102)	(0.179)	(0.046)	(0.043)	(0.024)	(0.041)	(0.601)	(0.636)	(0.671)	(0.718)

Monthly data on health district level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01. Labels *Abs* and *Rel* refer to whether excess mortality during the pandemic is described in absolute or relative terms. All dependent variables have been taken from municipality yearbooks and aggregated up to the health district level.

We also provide a set of figures illustrating the time trend in the highest and lowest district quartiles in terms of influenza exposure for the dependent variables used in the section

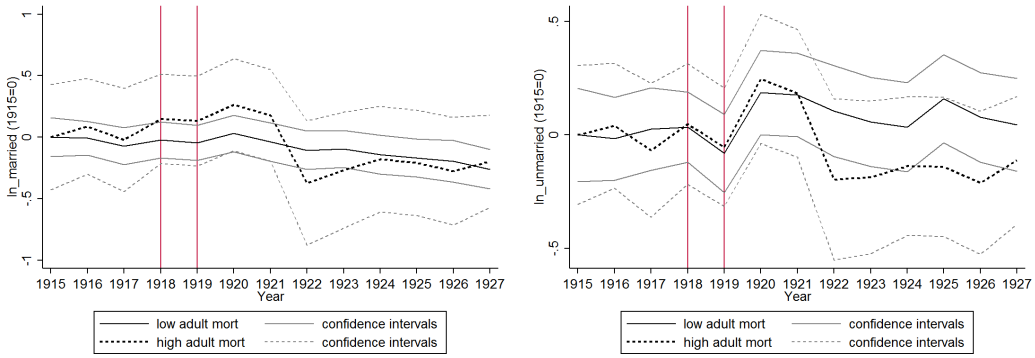
where we examine potential mechanisms, i.e., marriage rate, widowed rate, married mothers, single mothers, first births, higher order births, and births to high SES parents. In none of the cases we note any significant differences in the trends before the flu.

Figure 2.C2: Mortality rates in Sweden during the pandemic and in adjacent periods



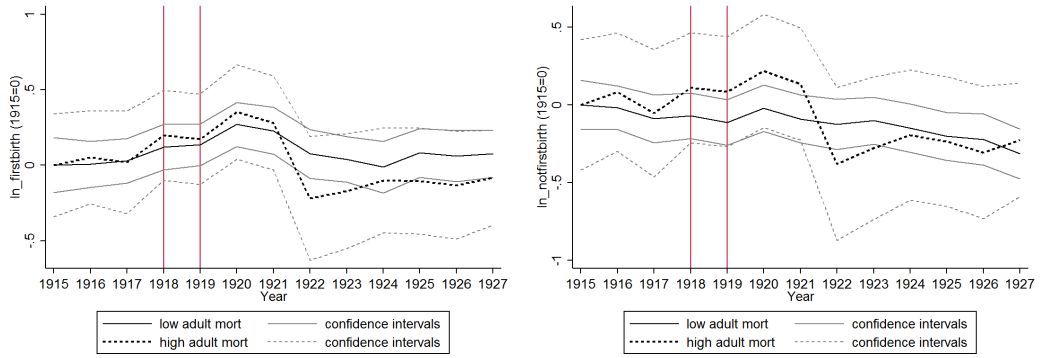
(a) Marriage rate **(b)** Widowed rate

Figure 2.C3: Time trend for married and single mothers



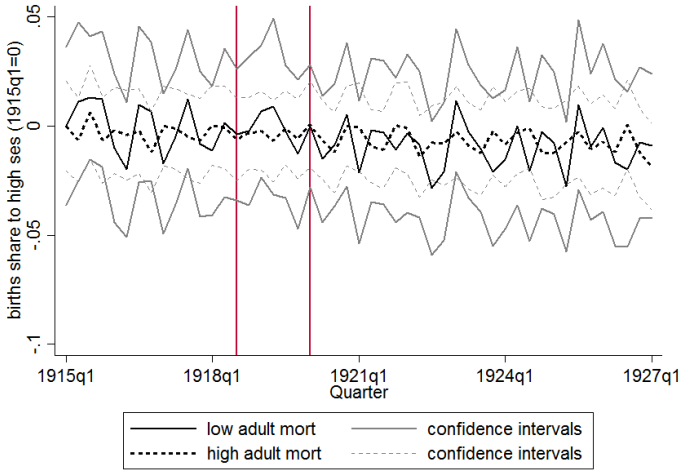
(a) ln(Married) **(b)** ln(Single)

Figure 2.C4: Time trend for first births and higher order births



(a) ln(First birth) **(b)** ln(Not first birth)

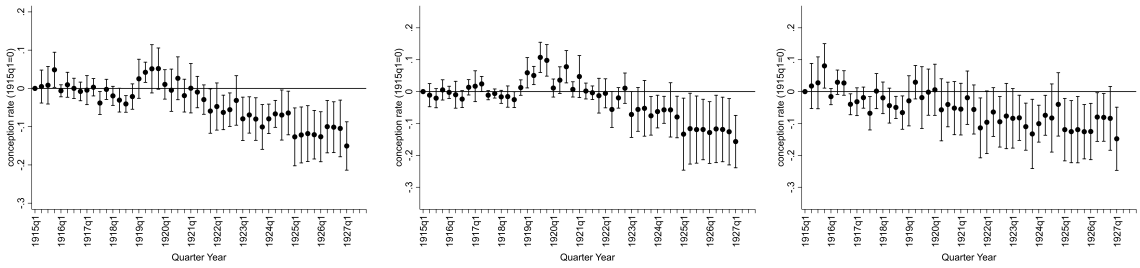
Figure 2.C5: Time trend for births to high SES parents



Event study graphs

Figure 2.C6 provides event study graphs where we estimate $\beta's$ in Eq. (2.4) for each quarter at a time. Clearly, there are no influenza effects before August 1918. The positive effect on conceptions after the influenza peak in rural areas is significant for a period of 19 months, but the estimate in fact stays positive for a total of 31 months. Around 1922 this trend is reversed and the districts most affected by the influenza pandemic exhibit lower conception rates than less affected districts. All together the event study graphs confirm and corroborate the findings of the regression analysis.

Figure 2.C6: Fertility effects, adult mortality



(a) All districts **(b)** Rural districts **(c)** Urban districts

3 You *Can* Win by Losing! Using Self-Betting as a Commitment Device: Evidence from a Weight Loss Program*

Abstract. Why do some people fail to commit efficiently when attempting to lose weight? This paper adapts the results from the theoretical model that introduces heuristic bias in agents' decision-making when using a self-commitment mechanism – an investment-payoff combination – to the setting of the real-world weight loss program DietBet, where players bet on the percentage of body weight they lose in a certain time, and tests the conclusions with its data. Our results suggest that the answer to the question lies in how accurately different agents predict their future self-control costs and payoffs. Specifically, naifs and naive optimists (overconfident agents) are more likely to overestimate their self-control costs and payoffs, while sophisticates and naive pessimists (underconfident agents) are more likely to estimate accurately or underestimate them. To distinguish agent types we apply "false hope syndrome" and "fresh start effect" to popular but infamously unsuccessful New Year's resolutions; the observation of people seeking commitment before Christmas to avoid holiday weight gain; and the observations of gender differences in overconfidence.

3.1 Introduction

Excessive weight is a global problem. Both high-income and low-income countries now have rising obesity rates. Overall, 13% of the adult population worldwide, including 11% of men and 15% of women, are obese (WHO, 2021).

Obesity and overweight negatively impact health (hypertension, diabetes, heart disease, several types of cancer, life expectancy) (Field *et al.*, 2001; Kopelman, 2007), and have direct economic costs: medical costs for the overweight increase by 20% and for the obese by 50% compared to non-obese peers (Thompson *et al.*, 1999). For the U.S., Cawley and Meyerhoefer (2012) estimate that obesity raises individual care costs by $2,741 (in 2005 USD)

*This chapter is co-authored with Linda Hirt-Schierbaum. An older version of this article is published as *Hirt-Schierbaum, L., Ivets, M. (2020). You can win by losing! Using self-betting as a commitment device: Evidence from a weight loss program (No. 881). Ruhr Economic Papers.* http://hdl.handle.net/10419/227105. We thank WayBetter Inc. for their cooperation; Katharina Blankart and Wolfgang Leininger for support, helpful comments and discussions. Special thanks to Stephan Sommer for his detailed comments and valuable remarks. We thank Daniel Avdic for feedback on an earlier draft. We also would like to thank the participants at various seminars and conferences for their constructive comments and suggestions. All remaining errors are ours.

annually. Considering *indirect* costs on productivity and human capital is likely to increase this estimate even further (see, e.g., Dee *et al.*, 2014; Trogdon *et al.*, 2008). Thus, direct and indirect costs indicate that individuals and policymakers have a significant economic rationale to reduce obesity-related externalities.

Simultaneously, 63% of US adults have tried to lose weight at least once, and 17% have never succeeded despite trying (Gallup, 2011). This demonstrates the consequences of a behavior that is said to result from self-control problems: people seek to reach a target, but often fail to follow through.

Emerging empirical evidence also shows that some people with a preference for commitment are failing to commit efficiently. This suggests that standard economic theory cannot explain this observed human behavior, underlining the importance of accounting for biases in individuals' subjective beliefs about their self-control abilities in their decision-making (Della Vigna and Malmendier, 2006).

Simultaneously, in the field of psychology there is ample literature on the psychological bias known as overconfidence (for an overview see, e.g., Glaser *et al.*, 2004). Overconfidence can be defined as an error in judgment or decision-making that leads to *overestimation* of one's abilities, performance or knowledge, and/or *underestimation* of skills, knowledge or abilities of one's opponents, difficulty of the task or possible risks.

The economic literature documenting heuristic biases in expectations and perceptions is still scarce. For example, Merkle and Weber (2011) show that the better-than-average beliefs that describe people's tendency to perceive their skills and virtues as above average are inconsistent with rational information processing, but are in accord with the psychological overconfidence bias. Others document overconfidence bias for business entry, longevity expectations, and stock index predictions (see, e.g., Camerer and Lovallo, 1999; d'Uva *et al.*, 2017; Kinari, 2016).

Small literature also links heuristic biases to health behaviors. For example, Arni *et al.* (2021) study the relationship between bias in health perception and risky health behaviors and find that people who overestimate their health are less likely to exercise and sleep enough, and are more likely to eat unhealthily and drink alcohol daily. Harris (2017) finds that people who overestimate their physical activity levels consume more calories.

To address the need to account for heuristic biases in individuals' decision-making and to understand what is driving such behaviors in agents with self-control problems and preferences for commitment Hirt-Schierbaum and Ivets (2021) - henceforth HSI - develop a theoretical model (called *Self-Commitment Decision Model* hereon). They introduce a heuristic bias into agent's decision-making when using a self-commitment mechanism – an investment-payoff combination – to help incentivize normatively-preferred behavior, i.e., behavior that would be observed in the absence of temptation.

In this paper we first adapt the *Self-Commitment Decision Model* from HSI to the setting of the online weight loss program *DietBet*, where players bet on the body weight percentage

they will lose in a certain time.[1] Next, we test the conclusions of the theoretical model with the *DietBet* data.

The model distinguishes different agent types based on *how accurately* the agents' predict their expected future self-control costs and payoffs: sophisticates (who are aware of their self-control issues and accurately predict their future self-control costs and payoffs), naifs (who are unaware of their self-control issues and of any changes in their future self-control costs, and overestimate future payoffs), and partially naive agents (who are aware of their self-control issues but under- or overestimate the severity of their problem and over- or underestimate their future payoffs) further separated into naive optimists and naive pessimists, respectively.

We hypothesize that naive and overconfident agents (naive optimists) are more likely to underinvest and default on their commitment, while sophisticated and underconfident agents (naive pessimists) are more likely to overinvest. Simultaneously, all agent types could benefit from placing higher wagers and participating in games with larger pots, especially naive and overconfident agents because they are more likely to underestimate their future self-control costs and overestimate their future payoffs.

To distinguish different agent types in our data we rely on the literature. Specifically, Dai *et al.* (2014) document the so-called "fresh start effect" in January that suggests that people use temporary landmarks, e.g., beginning of a new year, to define mental accounting periods, which delegate past imperfections and failures to the previous period and increase aspirational behaviors in the future period; while Ciccone (2011); Polivy (2001); Polivy and Herman (2000) also document "false hope syndrome", where people tend to have overly optimistic expectations in January (January Effect). Thus, it implies that January self-bettors – driven by their New Year's resolutions – could be more naive or overconfident (naive optimists).

Simultaneously, Christmas holidays represent a high-risk period for a weight gain, with the majority of annual weight being gained during this period (Helander *et al.*, 2016). Specifically, the weight gain pattern indicates that there is a smaller weight gain after Thanksgiving, peaking by the end of November. Afterwards, the weight slowly decreases until it plateaus ahead of Christmas holidays, while it steeply increases around Christmas time. Thus, this pattern indicates people seeking commitment in December before Christmas are loosing weight after Thanksgiving celebration and are trying to avoid the Christmas holiday weight gain. These self-bettors could be seen as being more sophisticated or of naive pessimist type.

Additionally, we rely on literature that documents gender differences in heuristic biases and related behaviors and finds that in uncertain situations, men tend to be more overconfident in self-assessment of their performance than women despite equivalent performance (Barber and Odean, 2001; Beyer, 1990; Beyer and Bowden, 1997; Deaux and Farris, 1977; Lichtenstein *et al.*, 1982; Lundeberg *et al.*, 1994; Niederle and Vesterlund, 2007). Simultaneously, the literature on behavioral response of weight loss to cash rewards indicates that females are more responsive to the prospect of larger rewards (Augurzky *et al.*, 2012), indi-

[1] *DietBet* has a parimutuel betting set up where players choose a monetary wager and join a game with other players. At the end of the game, the players who lost their targeted weight split the pot, while the players who did not lose their wagers.

cating that there is a potential asymmetry in naive optimistic expectations relating to future self-control costs and expected future payoffs.

In our empirical analysis we use people who participated in the weight loss program at least twice and rely on within-person variation. Individual fixed effects allow us to remove any game-invariant heterogeneity.

In line with our hypotheses we find that January and male bettors are more likely to underinvest and default on their commitments, while pre-Christmas and female bettors are more likely to overinvest. January and female bettors benefit more from games with larger pots due to overestimation of their future payoffs. Generally, agents who bet more and participate in games with larger pots are more successful in their commitments.

Thus, by combining new theory with the real-world data we provide evidence that it is important to account for heuristic biases in agents' decision-making regarding their self-control abilities and contribute an explanation to why certain agents with a preference for commitment might fail to commit efficiently.

Finally, since the theoretical model incorporates a self-bet as a commitment device and since we test it with the real-word data from a weight loss setting we also contribute to the literature on commitment devices and financial incentives in weight loss.

The commitment device literature (Ashraf *et al.*, 2006; Giné *et al.*, 2010; Thaler and Benartzi, 2004) and the commitment devices on the market[2] indicate public awareness of self-control problems and demand for external devices that aid commitment to normatively-preferred choices. Commitment contracts are one example.[3] However, they often have low take-up rates and weak effects (Giné *et al.*, 2010).

Recently, self-bet commitment mechanisms have been introduced and tested. Specifically, Lusher (2016) and Woerner (2021) apply a self-bet mechanism to education and exercise behavior, respectively. They find that the mechanism helps students achieve their educational goals, and has a significant positive effect on gym attendance.

While these studies document the effectiveness of self-bets with the help of lab or field randomized control trials (RCTs), the evidence on external validity of commitment mechanisms is still largely missing. We contribute by analyzing data from a real-world weight loss program *DietBet* that offers self-bets for commitment. Our study indicates that the self-bet mechanism can be used outside the RCT setting, supporting its external validity and showing that it could easily be adopted by policymakers to improve the effectiveness of policy interventions aimed at encouraging positive behavioral changes. Additionally, we investigate contributing factors of the self-bet mechanism – loss aversion and response to monetary rewards. While loss aversion is a main contributor of the mechanism's success (see, e.g., Lusher, 2016), the response to financial rewards in self-bets has been understudied. We contribute by showing that expected payoff could serve as an additional driver for successful commitment, especially for naive and overconfident agents (naive optimists).

[2]Some examples include `Beeminder.com`, `HealthWage.com`, `LazyJar.com`, `StickK.com` and `WayBetter.com`.

[3]A commitment contract is a binding agreement between an agent and a third party (a referee). To ensure the agent meets their pre-specified goals, they must invest money which will be returned in case of success, or forfeited in case of failure.

We also contribute to the literature on the use of financial incentives – deposit contracts and financial cash rewards – for weight loss (see, e.g., Augurzky *et al.*, 2012; Cawley and Price, 2011, 2013; Finkelstein *et al.*, 2007; John *et al.*, 2011; Relton *et al.*, 2011; Volpp *et al.*, 2008). With deposit contracts, people invest a certain amount at the program's start and are reimbursed if they reach their target weight. Financial rewards offer cash incentives to encourage weight loss. This literature indicates that financial incentives *can* encourage people to lose weight.

Policymakers, however, also want these interventions to be cost-effective. In a literature review on financial incentives for weight loss, Paul-Ebhohimhen and Avenell (2008) underline that there are no studies that justify the choice of the financial incentive amount. We contribute by providing evidence on an elasticity of weight loss to financial incentives (here, bet size) that can help design more efficient interventions.

The remainder of the paper is organized as follows. Section 3.2 adapts main findings from HSI and sets them in relation to the bet mechanism used in *DietBet*, and states the hypotheses we draw from the theory. Section 3.3 describes the data. Sections 3.4 and 3.5 present the empirical method and results. Section 3.6 provides discussion and Section 3.7 concludes.

3.2 Adaptation of Main Theoretical Results

This section first provides an overview of the main characteristics of the *Self-Commitment Decision Model* from HSI and and then adapts the main findings by setting them in relation to the *DietBet* weight loss program.

HSI develop a theoretical two-period decision model based on Gul and Pesendorfer (2001). An agent is facing a given menu over lotteries, A. In the first period the agent observes which menu he will be facing in period two, but does not face temptation yet. In period two the agent faces temptation and has to exercise self-control to make the ex-ante preferred choice, i.e., the choice he would have made in the absence of temptation, rather than succumb to temptation. His decision is highly dependent on his random time-variant degree of motivation (his temptation intensity), δ_i, $i = 1, 2$, which directly influences the cost of self-control he is facing.

HSI introduce a self-commitment mechanism based on an investment-payoff combination that can help agents commit successfully to their ex-ante preferred choices.

Definition 1 (Investment-Payoff Combination (HSI))**.** An *investment-payoff combination* is a self-commitment device where an investment is made before the action is taken. After the action is taken, a pre-defined payoff, at least the size of the investment, is rewarded if the pre-defined goal is reached. The investment is lost in the case of failure.

Definition 2 (Investment-Payoff Mechanism (HSI))**.** An *investment-payoff mechanism* is a self-commitment mechanism that utilizes an investment-payoff combination as a commitment device.

Agents have to choose a commitment in period one, *before* they face the actual temptation in period two. Thus, agents have to anticipate their future motivation and, therefore, their future costs of self-control. Based on *how accurately* the agents predict their future self-control costs, the *Self-Commitment Decision Model* distinguishes four types of agents: sophisticated, naive optimists, naive pessimists, and naive agents.

In the following analysis we present the theoretical results from HSI from the model with a random degree of motivation $\delta \in (0,1)$, where agents choose a commitment in period one. The motivation is revealed at the beginning of each period, so that agents know their period one motivation, but not their period two motivation. Here, δ_i, $i = 1,2$ is distributed on $(0,1)$. Suppose this distribution is well-behaved and denote its CDF $F(\cdot)$, with support $\text{supp}(F) = [0,1]$.

The period one utility function for $|A| \geq 2$ is given by

$$\mathbb{E}U_A(wp) := \mathbb{E}\left[\max_{x\in A}\left(u(x) - (\frac{1}{\delta_2} - 1)(v(y^M) - v(x)) + s(-w + \lambda p_2(x) - k)\right)\right], \quad (3.1)$$

where wp is the chosen investment-payoff combination (w, p_w).

The u and v are von Neumann-Morgenstern utilities, which describe the agent's normative preferences (*normative utility*) and how tempting an agent finds a lottery (*temptation utility*), respectively. The difference $\max_{y\in A} v(y) - v(x)$ describes self-control costs, where y is the most tempting item on the given menu A, and x is the chosen item from that menu. Following Gul and Pesendorfer (2001) we assume that players are only tempted by the most tempting item on the menu.

The *perceived (future) cost of self-control* $(\frac{1}{\mathbb{E}(\delta_2)} - 1)(\max_{y\in A} v(y) - v(x))$ is influenced by the random future (period two) degree of extrinsic motivation $\delta_2 \in (0,1)$. Based on *how accurately* the person perceives these future costs we can distinguish different types of agents.

Let $s : \mathbb{R} \to \mathbb{R}$ be *well-behaved*, i.e., it is defined, strictly monotonic, and twice continuously differentiable. With $s(-w) < 0$ for all $w > 0$ and $s(0) = 0$, defined over the investment w, the effort cost k and period two payoff p_2, which will be paid at the end of the period, and is discounted by $\lambda \in [0,1]$. s is upward-sloped; i.e., $s'(x) > 0$ for all $w \neq 0$ and $s'(0) = 0$. Furthermore, $s''(0) = 0$ and $s'''(0) \neq 0$; i.e., s is a (asymmetric) sigmoid function with reference point $s(0) = 0$. The payoff p_2 that is paid at the end of period two is defined as follows:

$$p_2(x) = \begin{cases} 0 & \text{if } x_2^* = y^M \neq x^M, \\ p_w & \text{if } x_2^* = x^M, \end{cases}$$

with $x^M := \arg\max_{x\in A}(u(x) + v(x))$ the normatively-preferred choice, $y^M := \arg\max_{y\in A} v(y)$ the most tempting item of the menu, and x_2^* the actual choice in period two. For example, if the agent succumbs to temptation his payoff is 0, and p_w otherwise.

To choose a welfare-enhancing and resistance-inducing commitment, period one agent has to solve the following equation:

$$u(y^M) - u(x^M) + (\frac{1}{\hat{\delta}_2} - 1)(v(y^M) - v(x^M)) < s(\lambda p_w - w - k). \tag{3.2}$$

Note that $\hat{\delta}_2$ – the biased expected future degree of motivation – differs for different types of agents, i.e., $\hat{\delta}_2$ is not necessarily equal to $\mathbb{E}(\delta_2)$. Based on *how accurately* these agents make these predictions, the *Self-Commitment Decision Model* introduces heuristic bias into the model and allows us to distinguish between four different types of agents.

Sophisticated agents *accurately* predict their future degree of motivation ($\hat{\delta}_2 = \mathbb{E}(\delta_2)$) and their future self-control costs. Naive agents, on the other hand, are completely *unaware* of their self-control problems ($\hat{\delta}_2 = 1$) and are unaware of any changes in their future degree of motivation and self-control costs. Partially naive agents of the optimistic type (naive optimists) neglect the possibility of a negative shock ($\hat{\delta}_2 > \mathbb{E}(\delta_2)$) and *underestimate* their future self-control costs. They are considered to be overconfident about their future self-control. Lastly, there are partially naive agents of the pessimistic type (naive pessimists) who underestimate the possibility of a positive shock ($\hat{\delta}_2 < \mathbb{E}(\delta_2)$) to their motivation and therefore *overestimate* their future self-control costs. They are considered to be underconfident about their future self-control.[4]

The main results from the model with random degree of motivation are summarized in Proposition 3 (HSI, p.26). It suggests the effect of the investment-payoff mechanism to vary between different types of agents. Overall it suggests the more sophisticated an agent the more successful the commitment.

In the following we adapt the model to bring the theory closer to the *DietBet* data. We then introduce a version of the proposition that summarizes the results we expect to see for different player types.

Given the game setting, *DietBet* players join a parimutuel bet to lose weight. This means the set over lotteries, A, consists, for example, of consumption choices or choices regarding exercise behavior (e.g., go to the gym; work out at home; watch TV) or a combination of both. All players set the same wager and the pot is split between winners at the end of a game. Contrary to what we assume in the theoretical model, players do not know their period two payoff in advance and, more importantly, cannot choose it. When entering the game, players can observe the current pot size which could still change with the entrance of additional players. At the game start players can observe the final pot size, but will only know which share the winners receive after the game ends.

Participants of open games do not know their fellow players, so there is no possible way to know or make an educated guess about other players' types (i.e., sophisticated, (partially) naive). On this basis, it is impossible to draw conclusions on their anticipated failure or success and therefore on the payoff an agent receives at the end of period two in case of success. Therefore, we analyze an agent's choice problem independently of what the other players do and incorporate only a dependence on the likelihood of winning that agents attribute to other players.

[4]For further details see HSI.

We consider a version of the model where players can only choose an investment (wager) w and build a belief about their future payoff, dependent on their own (biased) expected motivation.

Definition 3 (Degree of Naiveté)**.** Let ν be the *degree of naiveté* defined by

$$\nu := |\mathbb{E}(\delta_2) - \hat{\delta}_2|,$$

where $\mathbb{E}(\delta_2)$ is the (true) expectation of period two degree of motivation and $\hat{\delta}_2$ is the biased expected period two degree of motivation depending on an agent's type. We make use of the direction of bias: sophisticates do not have a bias ($\hat{\delta}_2 = \mathbb{E}(\delta_2)$); optimists overestimate their expected future motivation ($\hat{\delta}_2 > \mathbb{E}(\delta_2)$); pessimists underestimate their expected future motivation ($\hat{\delta}_2 < \mathbb{E}(\delta_2)$); naifs do not realize they are not fully motivated.

Definition 4 (Sign-dependent Degree of Naiveté)**.** Let $\tilde{\nu}$ be the *sign-dependent degree of naiveté* defined by

$$\tilde{\nu} := \hat{\delta}_2 - \mathbb{E}(\delta_2).$$

Following Gouveia and Clarke (2001), Kahnemann and Tversky (1979), Mansour *et al.* (2006), and Weinstein (1980), we assume that naive optimists tend to overestimate the expected value of their possible future payoff, while naive pessimists[5] underestimate it. Sophisticates correctly estimate that the chance of winning is 50% and that their expected period two payoff is given by $2 \cdot w$.

We use the sign-dependent degree of naiveté as a weight that influences the likelihood an agent attributes to himself and other players of winning a game. Equation (3.1) then changes to

$$\mathbb{E}U_A(w) := \mathbb{E}\left[\max_{x \in A}\left(u(x) - (\frac{1}{\delta_2} - 1)(v(y^M) - v(x)) + s(-w + \lambda \hat{p}_2(x) - k)\right)\right], \quad (3.3)$$

with

$$\hat{p}_2(x) = \frac{n}{(1+\tilde{\nu}) + (n-1)(1-\tilde{\nu})} \cdot \mathbb{E}(p_2(x))$$

$$= \begin{cases} 0 & \text{if } x_2^* = y^M \neq x^M, \\ \frac{2nw}{(1+\tilde{\nu})+(n-1)(1-\tilde{\nu})} & \text{if } x_2^* = x^M, \end{cases}$$

with n being the number of players in period one, when the agent joins the game.[6] For details on the definition of $\hat{p}_2(x)$ see Appendix A.

From the adapted model we can draw the following conclusions:

[5] We follow Abel (2002) in the notion that a pessimistic bias in individual beliefs is related to an underestimation of the probability of good outcomes and an overestimation of the probability of bad outcomes.

[6] This number might change after the player made his investment decision, which he can observe before he makes his choice in period two.

Proposition 1 (Bet Effect)**.**

i) An agent who chooses to join a bet with a wager $w > 0$ has a dominant investment strategy, given his beliefs.

ii) A sophisticated agent uses a bet successfully as a commitment device.

iii) If the pot size is perceived as exogenous constant, players are more likely to commit successfully when playing games with larger pot sizes.

iv) In a game with more than two players an optimistic agent

   a) is more likely to undercommit[7] when choosing an efficient wager given his belief. The higher the period one motivation, the more likely the undercommitment.

   b) profits from his biased expectation about the payoff, which decreases that effect.

v) In a game with more than two players a pessimistic agent

   a) is more likely to overcommit[8] when choosing an efficient wager, given his belief. The lower the period one motivation, the more likely the overcommitment.

   b) suffers from his biased expectation about the payoff, which increases that effect.

vi) Naive agents fail to use a bet as a commitment device, but might commit successfully by coincidence.

vii) Without a bet an agent with self-control problems is more likely to succumb to temptation.

Proof. See Appendix A. □

We use the results from the *Self-Commitment Decision Model* in HSI and the adapted version of the model as a theoretical basis for our empirical analysis. The main theoretical findings suggest that using a bet mechanism as a commitment device can increase the likelihood to resist temptation and can help agents to commit successfully to their predefined goals. Furthermore, it indicates that the likelihood of success depends heavily on an agent's sophistication/(partial) naiveté. We directly derive the following hypotheses:

Hypothesis 1. Agents that place higher wagers on themselves should be more successful in their commitments (win the games) and their weight loss.

Hypothesis 2. Subgroups with larger shares of naive and overconfident agents (naive optimists) should still have a positive effect from placing higher wagers,[9] but the effect size should be smaller because they are more likely to underinvest.

[7] An agent *undercommits* when his wager is not large enough to induce a binding commitment.
[8] An agent *overcommits* when his wager is larger than necessary to induce a binding commitment.
[9] By placing higher wagers optimistic (and naive) agents can "accidentally" commit themselves.

Hypothesis 3. Subgroups with larger shares of sophisticated and underconfident agents (naive pessimists) should have a positive effect from placing higher wagers,[10] but the effect size should be larger because they are more likely to overinvest.

Hypothesis 4. Agents that participate in games with larger pots should be more successful in their commitments (win the games) and their weight loss.

Hypothesis 5. Subgroups with larger shares of naive and overconfident agents (naive optimists) should have a positive effect from playing games with larger pot sizes, but the effect size should be larger because they are more likely to overestimate their future payoff.

Hypothesis 6. Subgroups with larger shares of sophisticated and underconfident agents (naive pessimists) should have no or a positive effect from playing games with larger pot sizes, but the effect size should be the same or smaller because they are more likely to estimate accurately or underestimate their future payoff.

3.3 DietBet Data

Waybetter Inc. and DietBet

This study utilizes data from WayBetter Inc. The company provides an online platform that offers people commitment opportunities in the form of self-bets to help them engage in healthier behaviors. We use data from DietBet program that offers self-bets as a commitment to promote weight loss. During enrollment, players bet money and join a game. The size of the game pot depends on the amount of the initial bet and the number of players. Players can join an existing game that has not yet started or can create their own game. They submit their initial weight within 48 hours of the game's start. After the game, players must submit their final weight within 48 hours.[11]

In this study, we focus on the *Kickstarter Bet*, where players bet to lose 4% of their initial body weight within 4 weeks (28 days). At the end of the 4-week period, weight loss is verified via official weigh-ins. Within each game, all winners (players who lose at least 4% of their initial weight) split the pool of money. Thus, there could be multiple winners. If no one lost 4% of their initial body weight, then a player who lost the most weight in percentage is rewarded the pool.[12]

DietBet adheres to the 'No Lose Guarantee' principle, which ensures that players who win DietBet will not lose money, i.e., the company will forfeit their cut to ensure that nobody loses money. Thus, in the worst case scenario, the players will lose weight for free. Figure 3.B1 in Appendix illustrates the screenshot of an example game a potential player could join.

[10] By placing higher wagers pessimistic and sophisticated agents will increase their chances of successful commitment.

[11] For more information on the weight verification process and referee review please refer to the DietBet website: https://www.dietbet.com/faq.

[12] Before paying the players, DietBet takes a portion of the initial gross pool to cover expenses. Thus, the players who do not win do not have to pay additional fees. The amount that is retained by DietBet depends on the amount of the initial bet. For more information please refer to the DietBet website: https://www.dietbet.com/faq.

Descriptive analysis

We use data from *Kickstarter Bet* from December 2011 until April 2017. Section 3.9 in Appendix presents summary statistics for the overall and panel DietBet samples (Tables 3.B1 and 3.B3, respectively). Figure 3.B7 shows the percent of bets placed by size. The most popular bets are \$30, \$35, and \$25. This is not surprising since DietBet offers many default bets for these amounts and they attract a lot of people. Moreover, people are less willing to participate in games where they have to put more money on the line as was also observed for deposit contracts (Jeffery, 2012).

People prefer to choose bets/prices that are round numbers that are multiples of five (Benartzi and Thaler, 2007; Lynn *et al.*, 2013), which corresponds with observed bet clustering at \$10, \$15, \$20, \$25, \$30, \$35, \$40, \$50, \$100, and \$150 values. There is a greater clustering of bets under \$50 with the next prominent cluster being at \$100. Figure 3.B7 shows that people make greater differentiation between smaller bets (e.g., bets under \$50), and that there is no such differentiation for bets between \$50 and \$100 or above \$100. This is consistent with the idea of cognitive biases and anchoring effect in pricing first identified by Tversky and Kahneman (1974). Here we can distinguish two anchors: \$50 and \$100 bets serving as natural reference points. For people who prefer lower bets, a \$50 bet serves as an anchor (upper bound of their willingness to pay) with bets under that amount seen as more attractive. Similarly, a \$100 bet is seen as an anchor by people who prefer higher bets. This anchor attracts bettors who are either willing to bet more than \$50 or those willing to bet more than \$100.

Figure 3.B5 presents scatter plots of shares of winners and average weight losses (in %) for each bet size (weighted by the number of observations). Here we see that this relationship is nonlinear and exhibits diminishing marginal returns. Specifically, there is a positive relationship between the bet size and the share of winners and weight loss. Simultaneously, this relationship exhibits diminishing marginal returns for higher bet stakes (around \$100). We plot the marginal effects of bets on our outcomes to examine this relationship in more detail (Figure 3.B6). The figure shows marginal effects for probability of winning and weight loss (%) from a quadratic fit. At first the marginal effects of higher bets are steep, but then exhibit diminishing marginal returns. The maximum is reached around \$225.

In our analysis we split bets into higher and lower stakes. We implement this split at the \$100 mark. The idea for this relates to the anchoring effect of the \$100 bet and prospect theory with loss aversion (Kahnemann and Tversky, 1979; Tversky and Kahneman, 1974). It suggests that people react more strongly to losses, and given that people might perceive bets of \$100 and over more costly than even the bets just under \$100, they can therefore exert extra effort to commit themselves.

The literature regarding the "fresh start effect" and "false hope syndrom" discussed in the section 3.1 suggests that different agent types participate in January and December games. Now we also look how they differ in our data.

Figure 3.B3 is based on DietBet data and shows that the most bets per month occur in January. This is in line with Google trends (Figure 3.B2) that show that there is a spike in

interest in dieting and weight loss right after the New Year.[13] Simultaneously, Figure 3.B4c shows that January bettors are, on average, heavier than bettors in other months, but their betting stakes (Figure 3.B4a) are similar to the average bets placed in other months (except December). These observations corroborate the literature that suggests that these games contain a larger share of naive and partially naive players of the optimistic type.

Figures 3.B2 and 3.B3 also show that December is the month with the least bets and the lowest interest in weight loss and dieting, with the lowest point in Google trend searches achieved just before Christmas (Figure 3.B2b). The literature suggests that people who participate in games before Christmas are those who would like to commit themselves ahead of the holidays and, thus, should contain a larger share of sophisticated and partially naive agents of pessimistic type. This is also supported by Figure 3.B4 where we see that December players place higher bets (Figure 3.B4a), while their initial weight, on average, is lower than that of January players (Figure 3.B4c).

3.4 Empirical Method

To estimate the relationship of interest between bet or pot size and successful commitment we could have relied on the OLS method in our empirical analysis. However, in this case there is an endogeneity problem since we could expect sophisticates or naive pessimists to systematically bet more and participate in games with larger pots to increase their commitment chances. Therefore, we would overestimate the effect of higher bets and larger pot sizes.

To deal with this potential endogeneity problem we use a multi-level data structure where individuals participate in multiple games[14] and explore within-person variation by using individual fixed effects. This allows us to remove any game-invariant heterogeneity between people. Table 3.B3 in Appendix B presents the sample's descriptive statistics.

To test hypothesis 1 we examine whether people who place high-stake bets ($100 or more) are more likely to win a game and lose weight. We specify the following model:

$$Y_{ig} = \alpha_i + \beta\, HighStakes_{ig} + X_{ig}\Gamma + Z_g\Theta + \epsilon_{ig} \tag{3.4}$$

where Y_{ig} is either probability of winning the game or weight loss in percent:

$$Y_{ig} := \begin{cases} Pr(Win) \\ \text{Weight Loss}(\%) \end{cases} \tag{3.5}$$

[13] Dieting/eating healthier, exercising more, and weight loss consistently appear as the top three New Year's resolutions (e.g., https://www.statista.com/chart/16500/top-us-new-years-resolutions/.

[14] One might argue that the selected sample of people participating more than once might differ from the overall population who try to lose weight. However, polls indicate that people usually attempt multiple times to lose weight (e.g., on average adults try to lose weight 5.3 times in their life (women vs men: 7 vs 3.6 times) (Gallup, 2011)). Moreover, we repeat the analysis on the full sample and find very similar results (Table 3.C9), indicating that the results are not sample-specific.

for player i in game g. $HighStakes_{ig}$ is equal to one if the wager of player i in game g is \$100 or more; and zero otherwise. α_i is player i's unobserved game-invariant characteristics that also include his initial degree of motivation (δ_{1i}). X_{ig} and Z_g represent other individual and game characteristics, respectively. Controls include game's pot size, player's starting weight, social engagement, number of weigh-ins, indicators whether the game is closed and a categorical variable for game order to control for participation experience. ϵ_{ig} is the error term that also contains player i's random motivation in period two, δ_{2ig}. We cluster standard errors at the individual level.

We expect that being a completer is associated with higher probability of winning the game and greater weight loss, since completing a final weigh-in is a precondition for being a winner, given our assumption that people who did not submit their final weigh-ins did not lose any weight. If we do not account for this in our analysis, we expect to identify an upper bound of the effect of bets on our outcomes, since completers bet more than noncompleters and noncompleters are automatically considered nonwinners by default. To account for this, we always include a dummy for being a completer in our main analysis and thus identify a lower effect bound. We repeat the analysis without controlling for completer status in Tables 3.C1 and 3.C2.

The coefficient of interest is β. It captures the difference in probability of winning and weight loss (%) for a person who places a high-stake wager vs low-stake wager.

We also look at marginal effects of bets and additionally specify the following model:

$$Y_{ig} = \alpha_i + \beta_1 \, Inter_{\geq \$100} + \beta_2 \, Bet_{<100\$,ig} + \beta_3 \, Bet_{\geq \$100,ig} + X_{ig}\Gamma + Z_g\Theta + \epsilon_{ig}. \tag{3.6}$$

Here, we utilize a linear spline regression to allow for different slopes for bets under and over \$100, since marginal effects of betting differ between high and low stakes (see Figures 3.B5 and 3.B6). $Inter_{\geq \$100}$ is the spline-specific intercept for bets of \$100 and more, which is equal to one if a bet is \$100 or more and zero otherwise; $Bet_{<\$100,ig}$ and $Bet_{\geq \$100,ig}$ are linear splines defined according to equations (3.7) and (3.8):

$$Bet_{<\$100,ig} = \begin{cases} Bet_{ig}, & \text{if } Bet_{ig} < \$100 \\ 99, & \text{otherwise} \end{cases} \tag{3.7}$$

$$Bet_{\geq \$100,ig} = \begin{cases} Bet_{ig} - 99, & \text{if } Bet_{ig} \geq \$100 \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

The coefficients of interest are β_2 and β_3 for the marginal effects of bets under and over \$100, respectively.

Hypotheses 2 and 3 state that higher bets should have different implications for different agent types. Since we expect that January games contain a larger share of naive and partially naive players of the optimistic type, January bettors should be less successful in their commitments compared to people who bet in other months, but we still expect high-stake January bettors to be more successful than low-stake January bettors. Simultaneously, we expect bets made in December before Christmas to contain larger shares of sophisticated and partially naive agents of the pessimistic type. Therefore, pre-Christmas December bet-

tors should be more successful in their commitments compared to people who bet in other months. Moreover, we expect high-stake pre-Christmas December bettors to be more successful than low-stake pre-Christmas December bettors. Finally, we expect larger shares of men to be of the naive optimist type about their future self-control, implying that male bettors should be less successful in their commitments compared to female bettors. However, we still expect high-stake male bettors to be more successful than low-stake male bettors.

To test these hypotheses, we fully interact models 3.4 and 3.6 separately with dummies for January, pre-Christmas December, and male bettors.

To test hypothesis 4 we look whether people who participate in games with pot sizes above the sample median are more successful in their commitment. We estimate the following model:

$$Y_{ig} = \alpha_i + \beta\, HighPots_{ig} + X_{ig}\Gamma + Z_g\Theta + \epsilon_{ig} \tag{3.9}$$

where Y_{ig} is either probability of winning the game or weight loss in percent for player i in game g. $HighPots_{ig}$ is equal to one if a game pot is above the sample median and zero otherwise.[15] X_{ig} and Z_g represent individual and game characteristics. Controls include bet size and bet size squared, player's starting weight, social engagement, number of weigh-ins, an indicator whether the game is closed and whether a person is a completer, and a categorical variable for game order to control for participation experience. ϵ_{ig} is an error term. We again cluster standard errors at the individual level.

The coefficient of interest is β. It captures the difference in probability of winning or weight loss (in %) for a person who participates in games with higher vs lower pot sizes.

Hypotheses 5 and 6 state that participation in games with larger pots should have different implications for different agent types. Here, again we look at January, pre-Christmas December, and male bettors (though now we expect that larger share of females are naive optimists about their future payoff). Here we expect that January players participating in games with larger pots should be more successful in their commitments compared to people who play in other months. Additionally, we expect high-pot January players to be more successful than low-pot January players. Simultaneously, we expect pre-Christmas December players to be as or more successful in their commitments compared to players in other months. Moreover, we expect high-pot pre-Christmas December players to be as or more successful than low-stake pre-Christmas December players. Finally, we expect that males should have a positive effect from participating in games with larger pots, but this effect should be smaller than for females. Additionally, we expect males participating in games with larger pots to be more successful than males participating in games with lower pots.

To test these hypotheses, we fully interact model 3.9 separately with dummies for January, pre-Christmas December, and male bettors.

[15] The median pot size in the panel sample is $24,200.

3.5 Results

Overall effect of high-stake bets and high pot sizes

We start by testing hypothesis 1 by estimating Eqs. (3.4) and (3.6). Table 3.1 presents the results.

Table 3.1: Effect of high-stake bets and marginal effects of bets on probability of winning and weight loss (%).

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High-Stake Bet	0.0504***	0.0530***	0.1302***	0.1389***
	(0.0016)	(0.0016)	(0.0062)	(0.0063)
	(1)	(2)	(3)	(4)
Bet Amount under $100	0.0020***	0.0017***	0.0061***	0.0050***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Bet Amount over $100	0.0001*	-0.0000	0.0004**	0.0001
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Add. controls	No	Yes	No	Yes
N	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

The results from columns (2) and (4) show that the relationship between high-stake bets ($100 or more) and probability of winning and weight loss is positive and statistically significant. High-stake bets are associated with 5.3 percentage points (pp) higher likelihood of winning and 0.14 pp higher weight loss. Regarding effect size, the increase in winning probability of 5.3 pp for high-stake bets corresponds to an increase of 10.6% with respect to winning probability for bets under $100 (5.3/49.8). Similarly, 0.14 pp higher weight loss for high-stake bets corresponds to 5.2% increase with respect to weight loss for bets under $100 (0.14/2.66).[16]

[16] These effect sizes represent a lower bound since we control for the completion status. If we do not control for it, the effect sizes for probability of winning and weight loss are 0.105 pp and 0.425 pp (corresponding to 21.1% and 16% increase relative to bets under $100), respectively (see Table 3.C1 in Appendix B).

Regarding marginal effects, we find that betting $10 more in bets under $100 is associated with 1.7 pp higher probability of winning and 0.05 pp more weight loss (as also evident from Figure 3.B6). However, the marginal effect of betting more money on probability of winning and weight loss is not significant once the person is a high-stake bettor. Thus, agents that place high-stake bets are more likely to win and lose more weight, but once they place such a bet, betting marginally more does not increase their likelihood of success.

Table 3.2 presents the results from participation in games with larger pots (tests hypothesis 4). On average, agents who participate in games with above median pot sizes are significantly more likely to win the game and lose more weight.[17] This supports our hypothesis 4 that agents who participate in games with larger pots should be more successful in their commitments (win the games) and their weight loss. The implication is that people expect larger payoffs when participating in games with larger pots.

Table 3.2: Effect of high game pots on probability of winning and weight loss (%).

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High Pot Size	0.0167***	0.0137***	0.0566***	0.0468***
	(0.0008)	(0.0008)	(0.0029)	(0.0029)
Add. controls	No	Yes	No	Yes
N	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). High Pot Size is a dummy variable equal to one if pot size is larger than the sample median ($24,200), and zero otherwise. All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, bet and bet squared, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Sophistication and naiveté

Here we test our second hypothesis. Table 3.3 presents the results from the fully interacted model with January games.

[17] Regarding the effect size, this corresponds to 2.9% and 2% (0.014/0.48 and 0.05/2.63) increase with respect to probability of winning and weight loss, respectively, for players in game with below median pot sizes. This, again, represents a lower bound. Table 3.C2 in Appendix B presents the results without controlling for completion status. The effect sizes for probability of winning and weight loss are 0.019 pp and 0.075 pp (corresponding to 4% and 2.9% increase relative to below median pots), respectively.

Table 3.3: Heterogeneity analysis of New Year's resolutions on probability of winning and weight loss (%).

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High-Stake Bet	0.0550***	0.0546***	0.1461***	0.1432***
	(0.0018)	(0.0018)	(0.0067)	(0.0068)
High-Stake Bet × NY Bet	-0.0222***	-0.0128***	-0.0730***	-0.0382***
	(0.0035)	(0.0036)	(0.0141)	(0.0142)
	(1)	(2)	(3)	(4)
Bet Amount under $100	0.0019***	0.0017***	0.0058***	0.0050***
	(0.0001)	(0.0001)	(0.0003)	(0.0003)
Bet Amount over $100	-0.0000	-0.0001*	0.0002	-0.0000
	(0.0001)	(0.0001)	(0.0003)	(0.0003)
Bet Amount under $100 × NY Bet	-0.0000	-0.0000	-0.0006	-0.0002
	(0.0002)	(0.0002)	(0.0006)	(0.0006)
Bet Amount over $100 × NY Bet	0.0002**	0.0002	0.0001	-0.0001
	(0.0001)	(0.0001)	(0.0004)	(0.0004)
Add. controls	No	Yes	No	Yes
N	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$

As anticipated, we see that people who participate in January games are on average less successful compared to players in other months. Although the high-stake January bettors are still more successful than low-stake January bettors,[18] the effect of high-stake January bets remains significantly smaller compared to other months. The marginal effects for January bets are not significantly different from marginal effects of bets placed in other months. Overall, the results indicate that even naive and partially naive agents of the optimistic type can increase their chances of successful commitment and achieve greater weight losses by placing high-stake bets.

To test hypothesis 3 we consider bets made in December before Christmas. Table 3.4 contains the results from the fully interacted model with pre-Christmas December games.

Table 3.4: Heterogeneity analysis of before Christmas bets on probability of winning and weight loss (%).

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High-Stake Bet	0.0501***	0.0527***	0.1297***	0.1383***
	(0.0016)	(0.0017)	(0.0063)	(0.0064)
High-Stake Bet× Before Xmas Bet	0.0272***	0.0153**	0.1023***	0.0577**
	(0.0074)	(0.0074)	(0.0283)	(0.0283)
	(1)	(2)	(3)	(4)
Bet Amount under $100	0.0020***	0.0017***	0.0060***	0.0050***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Bet Amount over $100	0.0001	-0.0000	0.0004*	0.0000
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Bet Amount under $100 × Before Xmas Bet	0.0009**	0.0007	0.0049***	0.0032**
	(0.0004)	(0.0004)	(0.0015)	(0.0016)
Bet Amount over $100 × Before Xmas Bet	-0.0001	-0.0000	0.0001	0.0007
	(0.0003)	(0.0004)	(0.0010)	(0.0012)
Add. controls	No	Yes	No	Yes
N	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size and a categorical variable for game order. 'Before Xmas Bet' is a dummy equal to 1 if a bet take place from December 1st until December 24th, and 0 otherwise. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

[18] For high-stake January bettors the effect is 0.042 for probability of winning and 0.105 for weight loss, significant at 1% significance level.

In Table 3.4 we see that the high-stake pre-Christmas bettors are more successful than low-stake pre-Christmas bettors.[19] We also see that pre-Christmas high-stake bettors are, on average, more successful compared to high-stake bettors in other months. For weight loss, the marginal effects for bets under $100 are also positive and significantly different from marginal effects of bets placed in other months. This indicates that even sophisticated and partially naive agents of the pessimistic type can increase their chances of successful commitment by placing higher bets.

Next, we investigate our hypotheses 5 and 6. The results from the fully interacted model are found in Table 3.5. We see that high-pot January and pre-Christmas players are more successful than low-pot January[20] and pre-Christmas[21] players. Also, players who participate in games with higher pots in January are more successful in their commitments, while players who participated in games with larger pots before Christmas are not significantly different in their commitment successes from high-pot players in other months. The results line up with our hypotheses, where we suggest that naifs and naive optimists would benefit more from participating in games with larger pots due to their overestimation of future payoffs, while sophisticates and naive pessimists would not benefit or would do so to a lesser extent.

Gender differences

We further consider gender differences in behavioral responses. Tables 3.6 and 3.7 (fully interacted model) preset the results.

We see that men participating in high-stake bets are more successful than men participating in low-stake bets (Table 3.6), but they are on average less successful in their commitments compared to high-stake female players (Table 3.7). This result is consistent with greater risk and loss aversion of women that has been documented in the literature[22] and that is sometimes attributed to gender differences in overconfidence.[23]

Next we look at behavioral gender differences in response to higher game pots.[24] Tables 3.8 and 3.9 present the results. Here we see that males have a positive effect from participation in games with larger pots (Panel A in Table 3.8), but this effect is smaller than for females (Panel B in Tables 3.8 and 3.9). This result can be attributed to greater responsiveness of women to financial incentives (i.e., by having a steeper s-shaped utility curve that is associated with higher utility from a given expected payoff) and indicates that there is a potential asymmetry in naive optimistic expectations relating to future self-control costs and expected future payoffs.

[19] For high-stake pre-Christmas bettors the effect is 0.068 for probability of winning and 0.196 for weight loss, significant at 1% significance level.

[20] 0.0181 for probability of winning and 0.0601 for weight loss, significant at 1% significance level

[21] 0.0201 for probability of winning and 0.0749 for weight loss, significant at 1% significance level

[22] Studies that examine gender differences in risk attitudes over monetary gambles find that women are either more risk averse than men (Gächter *et al.*, 2022; Rieger *et al.*, 2011; Schmidt and Traub, 2002) or that there are no gender differences (Byrnes *et al.*, 1999; Eckel and Grossman, 2008).

[23] Figures 3.C1b and 3.C1c in Appendix B show the percent of bets by size placed by males and females.

[24] Figures 3.C2b and 3.C2c in Appendix B show the distribution of pot sizes for male and female samples.

Table 3.5: Heterogeneity analysis of New Year's resolutions and participation in before Christmas games with high pot sizes on probability of winning and weight loss (%).

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High Pot Size	0.0121***	0.0101***	0.0479***	0.0339***
	(0.0009)	(0.0009)	(0.0031)	(0.0031)
High Pot Size	0.0074***	0.0079***	0.0167**	0.0262***
× NY Bet	(0.0020)	(0.0021)	(0.0070)	(0.0071)
	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High Pot Size	0.0171***	0.0128***	0.0518***	0.0424***
	(0.0008)	(0.0008)	(0.0029)	(0.0029)
High Pot Size	0.0068	0.0073	0.0306	0.0325
× Before Xmas Bet	(0.0069)	(0.0069)	(0.0247)	(0.0249)
Add. controls	No	Yes	No	Yes
N	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). High Pot Size is a dummy variable equal to one if pot size is larger than the sample median ($24,200), and zero otherwise. All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, bet and bet squared, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 3.6: Probability of winning and weight loss (%) with larger bets: Male sample.

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High-Stake Bet	0.0258***	0.0264***	0.0718***	0.0759***
	(0.0025)	(0.0025)	(0.0103)	(0.0104)
	(1)	(2)	(3)	(4)
Bet Amount under $100	0.0010***	0.0009***	0.0031***	0.0029***
	(0.0001)	(0.0001)	(0.0005)	(0.0005)
Bet Amount over $100	0.0000	-0.0000	0.0005*	0.0002
	(0.0001)	(0.0001)	(0.0003)	(0.0003)
Add. controls	No	Yes	No	Yes
N	87,992	87,992	87,992	87,992
N (clust)	19,591	19,591	19,591	19,591

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$

Table 3.7: Heterogeneity analysis by gender on probability of winning and weight loss (%) with higher bets.

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High Bet Stake	0.0580***	0.0612***	0.1497***	0.1603***
	(0.0020)	(0.0021)	(0.0077)	(0.0078)
High Bet Stake × Male	-0.0313***	-0.0339***	-0.0777***	-0.0840***
	(0.0031)	(0.0032)	(0.0126)	(0.0127)
	(1)	(2)	(3)	(4)
Bet Amount under 100$	0.0022***	0.0018***	0.0067***	0.0053***
	(0.0001)	(0.0001)	(0.0003)	(0.0003)
Bet Amount over 100$	0.0001	0.0000	0.0003	-0.0001
	(0.0001)	(0.0001)	(0.0003)	(0.0003)
Bet Amount under 100$ × Male	-0.0012***	-0.0009***	-0.0034***	-0.0023***
	(0.0002)	(0.0002)	(0.0006)	(0.0006)
Bet Amount over 100$ × Male	-0.0001	-0.0000	0.0003	0.0004
	(0.0001)	(0.0001)	(0.0004)	(0.0004)
Add. controls	No	Yes	No	Yes
N	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$

Table 3.8: Probability of winning and weight loss (%) with larger pots for males and females.

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
Panel A: Male Sample				
High Pot Size	0.0042***	0.0039**	0.0129**	0.0149**
	(0.0016)	(0.0016)	(0.0062)	(0.0062)
Add. controls	No	Yes	No	Yes
N	87,992	87,992	87,992	87,992
N (clust)	19,591	19,591	19,591	19,591
	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
Panel B: Female Sample				
High Pot Size	0.0188***	0.0151***	0.0638***	0.0511***
	(0.0010)	(0.00010)	(0.0033)	(0.0033)
Add. controls	No	Yes	No	Yes
N	543,352	543,352	543,352	543,352
N (clust)	134,564	134,564	134,564	134,564

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). High Pot Size is a dummy variable equal to one if pot size is larger than the sample median ($22,590 for Panel A and $24,675 for Panel B), and zero otherwise. All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, bet size, bet squared, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$

Table 3.9: Heterogeneity analysis by gender on probability of winning and weight loss (%) with larger pots.

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High Pot Size	0.0192***	0.0152***	0.0653***	0.0516***
	(0.0010)	(0.0010)	(0.0033)	(0.0033)
High Pot Size × Male	-0.0154***	-0.0113***	-0.0537***	-0.0376***
	(0.0018)	(0.0018)	(0.0067)	(0.0067)
Add. controls	No	Yes	No	Yes
N	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). High Pot Size is a dummy variable equal to one if pot size is larger than the sample median ($24,200), and zero otherwise. All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, bet size, bet squared, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$

Robustness checks

We start with a Placebo test, where we run the same regression for all other months besides January and December. There are other holidays that can influence people's preferences and incentives to lose weight. For example, April and November bets could be influenced by Easter and Thanksgiving.[25] During this time, some people might seek commitment before the holiday to avoid overconsumption, while others might want to lose extra holiday weight. Moreover, Madden (2017) documents a seasonality in weight loss contemplations with peaks in winter and summer. Indeed, Figures 3.B2 and 3.B3 show an increase in interest in weight loss during summer, peaking in July. The summer interest could be driven by differences between winter and summer clothing, since some surveys have indicated that appearance is among second or third most popular motives to lose weight (Gallup, 2014; O'Brien *et al.*, 2007).[26] Given this, people who participate in May and June bets could be seen as forward-looking because they start losing weight earlier and ahead of summer holidays.

People who participate in September and October games tend on average to weigh less (see Figure 3.B4c). After October, body weight begins to increase into early January due to

[25] Helander *et al.* (2016) and Yanovski *et al.* (2000) document that people tend to gain weight after certain holidays – the most prominent weight gains happen around Christmas, Thanksgiving and Easter.

[26] In Figure 3.B4c we see that the average initial weight of individuals participating in July and August games increases, indirectly supporting this proposition.

holiday festivities. Simultaneously, some agents participating in February games could be driven by their failed initial attempts to lose weight in January as part of their New Year's resolution.[27] However, it is difficult to conclude which types of agents dominate in any of these months in particular. The results are found in Tables 3.C3 and 3.C4 in Appendix C. We see that in most of these months the players are not different from players in other months of the year with respect to their commitment successes.[28] Table 3.C5 shows the results of larger pot sizes from the Placebo test by bet months.[29]

One potential concern regarding our model is that players' socioeconomic characteristics could be correlated with the betting stakes and their weight loss success. For example, agents with higher income can afford to place higher wagers and to buy better quality food, consult a nutritionist, and attend gyms with personal trainers. If these socioeconomic status characteristics stay constant between games, then the fixed effects regression should eliminate them. However, if there is a change in these unobserved variables, they will contribute to the omitted variable bias. To address this concern we focus only on bets placed by the agents within one calendar year. The underlying assumption is that education and income are unlikely to change within one year. The results are presented in Tables 3.C6 and 3.C7 in Appendix C. Results stay qualitatively the same; therefore, we conclude that this is not a concern.

Another concern could be that agents participate in more than one game at once.[30] In this case, our identifying assumption that the initial motivation (δ_{1i}) is game-invariant might not hold since the wager in the first game can influence motivation in the parallel game. To address this we exclude parallel games and concentrate only on games that were played sequentially without overlap. The results stay qualitatively the same (see Table 3.C8 in Appendix C).

We check the sensitivity of our results to using logit model instead of LPM for the probability of winning as an outcome. Logit model produces qualitatively similar results (available upon request). Finally, we repeat the main estimation of the bet and pot size effect on the full sample (i.e., also including people who only played once) to check that our results are not specific to our panel sample. The results are found in Tables 3.C9 and 3.C10. Here we see that the results are very similar to our panel sample, indicating that they are not sample-specific.

3.6 Discussion and Policy Relevance

Our empirical analysis concurs with the theoretical predictions from section 3.2 and shows that using self-bet can help different agents overcome self-control problems. When agents

[27] Figure 3.B2b shows that there is an increase in interest in weight loss in the second half of February which coincides with the time when most people fail their first initial attempts in New Year's resolutions.

[28] The coefficient on high-stake bets is positive and significant for April and July bets. This could be driven by Easter holidays and people trying to lose weight during the summer.

[29] Figure 3.C4 in Appendix C shows the distribution of pot sizes in each months.

[30] DietBet allows participation in up to three Kickstarter games at once: https://waybetter.com/dietbet/faq.

bet more, they are more likely to win the game and lose more weight. Also, agents benefit from games with larger pots.

The theoretical model uses the observation from behavioral economics that loss aversion is a significant motivator of human behavior (Kahneman and Tversky, 1979). The mechanism also utilizes agents' taste for rewards through payoffs for successful commitment and the observation that people often overestimate the likelihood of unlikely events (the size of future payoff) (Kahneman and Tversky, 1984).[31]

The empirical results add insights about the relative importance of loss aversion and response to monetary rewards as contributing factors of the bet mechanism's success. The effect size from high-stake bets was much larger than from higher pot sizes. However, the latter was still statistically significant. This indicates that while loss aversion does seem to be the main contributor to the mechanism's success, agent's response to monetary rewards can still provide an additional incentive and amplify the bet's effect.

The findings are also consistent with evidence from several behavioral weight loss programs involving deposit contracts and financial incentives.[32]

Policy relevance

While January represents a very popular time to engage in self-improvement, we show that some people might overestimate their abilities and fail. Simultaneously, marketeers of products intended for self-improvement, e.g., gym memberships, online commitment devices and weight loss programs, might appeal to potential customers at this time. This raises a concern of potential exploitation of (partially) naive agents.

This might call for public regulation. Policymakers can provide information and guidance to encourage successful commitment. It is also important to target men, as they participate less and have higher obesity rates (Flegal *et al.*, 2010; WHO, 2016). Men are also frequently overconfident about their future self-control and less responsive to monetary incentives, and therefore are at a higher risk of commitment default.

At the same time, policymakers, organizations, and companies could encourage commitment ahead of certain holidays associated with high risk of weight gain, e.g., before Christmas. Preventing weight gain during the holiday season is challenging even for those who consistently self-monitor (Phelan *et al.*, 2008), with obese and overweight individuals being more at risk of the greatest increase in weight and body fat (Díaz-Zavala *et al.*, 2017; Schoeller, 2014; Stevenson *et al.*, 2013). Literature also finds that the gained holiday weight

[31] Studies that compare the effectiveness of financial incentives structured as rewards vs deposit contracts find there is a higher take-up for rewards, but the deposit-based programs lead to greater behavioral change (Halpern *et al.*, 2015). The self-bet mechanism includes both: while participants put up their own money and therefore evoke loss aversion, there is also a chance for reward in case of successful commitment which could increase interest in the program compared to simple commitment or deposit contracts.

[32] Augurzky *et al.* (2012); Finkelstein *et al.* (2007); Jeffery *et al.* (1983, 1984); John *et al.* (2011), and Volpp *et al.* (2008) find that deposit contracts and financial incentives yield significant weight losses. Moreover, greater baseline deposits and cash rewards are associated with higher likelihood to reach weight loss goals (Augurzky *et al.*, 2012; Finkelstein *et al.*, 2007; Jeffery *et al.*, 1984). However, both types of financial incentives are found to exhibit diminishing marginal returns (Augurzky *et al.*, 2012; Finkelstein *et al.*, 2007; Jeffery *et al.*, 1983).

is not lost three months or more after the holiday season (Yanovski *et al.*, 2000). Therefore, it is important to promote commitment ahead of such holidays.

Finally, our study can be used when assessing new digital health technologies. For example, some countries have already introduced digital health applications to improve health behaviors.[33] Our results indicate that utilizing loss aversion can help a device encourage positive behavioral changes. For example, having a co-pay instead of free digital health apps could improve their behavioral health effects. This co-pay can be reimbursed in case of successful commitment. Moreover, since the digital health app directory provides comprehensive device information, this could include default probability and the importance of co-pay to increase likelihood of commitment. Simultaneously, it is also important to consider introducing a reward since it will amplify the positive behavioral effect. These considerations can help governmental agencies with device assessment and price negotiation.

3.7 Conclusion

In 2016 the global prevalence of obesity reached 650 million – nearly tripling from 1975 – while more than 1.9 billion adults were overweight (WHO, 2021). These numbers continue to grow with most countries experiencing substantial rises. At the same time, people often aim to lose weight, but many fail to achieve it. This behavior is sometimes attributed to preference reversal, but recent empirical evidence highlights the role of overconfidence bias in individual decision-making that could help explain the observed results. Combining new theory and empirical evidence that accommodates heuristic biases in individual's decision-making can improve our understanding of what drives such behaviors. This can help policymakers and individuals to create better policies and make better decisions.

In this paper we adapt the theory from *Self-Commitment Decision Model* from HSI to the real-world setting of DietBet. We then derive hypotheses about the real-world data from the theoretical results. The model incorporates heuristic bias in agents' decision-making when using self-commitment mechanism – a investment-payoff combination – and explains *why* certain agents with a preference for commitment sometimes fail to commit efficiently. The model distinguished different agent types based on *how accurately* they predict their future self-control costs and payoffs.

Our empirical results corroborate theoretical predictions from the *Self-Commitment Decision Model* and suggest that the observed weight loss commitment failure could be explained by naive and partially naive agents of the optimistic type – exemplified in our data by January bettors – overestimating their future self-control costs and underinvesting. They also show that some people are more likely to overinvest or invest efficiently, ensuring successful commitment (sophisticated and partially naive agents of the pessimistic

[33] One example is a Digital Healthcare Act that came into effect on December 19th 2019 in Germany and introduced "apps on prescription" as part of the care provided to statutory health insurance (SHI) patients. This program allows patients to receive healthcare through digital health apps that can be prescribed by physicians or psychotherapists and are reimbursed by health insurers. To be eligible for reimbursement, digital health apps should be approved by the Federal Institute for Drugs and Medical Devices (BfArM). Part of this assessment is examining the evidence of their positive healthcare effect.

type – as exemplified by pre-Christmas December bettors in our data. Additionally, we also document behavioral differences in gender responses, where males are more likely to be overconfident about their future self-control costs and underinvest, while females are more likely to overestimate their future payoffs facilitating successful commitment.

Theory and evidence suggest that, in general, greater investments (placing higher bets) could help all agent types to follow through with their normative intentions. Specifically, participants in high-stake bets ($100 or more) were more successful than participants in lower stake bets (under $100). While betting more money increased the chances of winning for bettors under $100, marginal effects for high-stake bets ($100 and over) were not statistically significant – implying diminishing marginal returns.

Finally, we also show that agents benefit from games with larger pots due to overestimation of future payoffs. This effect is much smaller than that of the high-stake bet, implying that loss aversion is the main contributor to the mechanism's success. However, agents' response to the prospective monetary reward is still significant and amplifies the bet effect.

3.8 Appendix A: Theoretical Extension

In our main model we consider an investment-payoff combination that is generally-defined and is not context-specific. Here, we specify the case of a monetary unmatched bet as is observed in the context of the online dieting program.

In this case an agent can observe how many players joined the bet before him. Let n be the number of players after the agent joined the game. $g \leq n$ of these players will win the bet. The possible payoff from the bet is then defined by $p_w = \frac{n*w}{g}$, i.e., the pot is split equally amongst winners.

This implies that the monetary payoff that winners receive is financed by all players and the bet is thus budget-balanced.

Proposition 2 (Budget Balancedness)**.** If the pot of an unmatched bet with n players, $g \leq n$ winners and bet amount $w > 0$ is split equally among winners, the bet is budget-balanced.

Proof. The value of the pot is $n * w$, each winner receives the payoff $p_w = \frac{n*w}{g}$. Thus, the net value of the pot is $n * w - g * p_w = 0$. □

Generally, players cannot foresee how many players will win the bet and g is unknown in advance. Therefore, at the beginning of periods one and two players have to maximize their utility given their believes about other players' motivation and the overall number of players, $\hat{g}$. If we were analyzing a matched bet, players with equal motivation would have been matched in one bet. In our case, however, players are not matched and therefore face a heterogeneous group of people. And since players do not have any information about the other participants of the bet, it is valid to assume that all players are equally likely to win or lose (i.e., $g = \frac{1}{2} \cdot n$).[34]

The true expected payoff at time t=1 is given by

$$\mathbb{E}(p_2(x)) = \frac{n \cdot w}{\sum_{i=1}^{n} \mathbb{E}(1_i)}, \qquad 1_i = \begin{cases} 1 & \text{if } x_2^* = x^M, \\ 0 & \text{if } x_2^* = y^M \neq x^M. \end{cases}$$

Given that the probability of winning is $\frac{1}{2}$, $\mathbb{E}(1_i) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0$. This yields

$$\mathbb{E}(p_2(x)) = \frac{n \cdot w}{\frac{1}{2} \cdot n} = 2 \cdot w.$$

We follow Abel (2002), Mansour *et al.* (2006), Shepperd *et al.* (2002), and Weinstein (1980) in the assumption that an agent attributes a certain likelihood to himself and others winning the bet. Sophisticates correctly anticipate that their chances are 50:50. Naive optimists (pessimists) overestimate (underestimate) their own motivation and their own likelihood

[34] Also, our data shows that on average players win 50% of their bets (see Table 3.B1), which consolidates this assumption.

of winning the bet ($\frac{1}{2} \cdot (1 + \tilde{v})$). At the same time, they underestimate (overestimate) the motivation of other players and their respective likelihood of winning ($\frac{1}{2} \cdot (1 - \tilde{v})$).[35]

We use the sign-dependent degree of naiveté as a distortion (weight) of the true winning probability and receive the following result for the biased expected payoff $\hat{p}_2(x)$ in case $x_2^* = x^M$:

$$\begin{aligned}\hat{p}_2(x) &= \frac{n \cdot w}{\underbrace{\frac{1}{2}(1+\tilde{v}) \cdot 1 + \frac{1}{2}(1-\tilde{v}) \cdot 0}_{\text{distorted probability that agent wins}} + \sum_{i=1}^{n-1} [\underbrace{\frac{1}{2}(1-\tilde{v}) \cdot 1 + \frac{1}{2}(1+\tilde{v}) \cdot 0}_{\text{distorted probability that other players win}}]} \\ &= \frac{n \cdot w}{\frac{1}{2}(1+\tilde{v}) + \frac{1}{2}(n-1)(1-\tilde{v})} = \frac{2 \cdot n \cdot w}{(1+\tilde{v}) + (n-1)(1-\tilde{v})}.\end{aligned}$$

Note that $\hat{p}_2(x) \in (w, nw)$ is bound by $\tilde{v} \in (-1, 1)$. We assume that the agent assigns reversed probability distributions to himself vs other players, so if he is an optimist he assumes other players to perform worse than himself and if he is a pessimist vice versa.

Sophisticates are assumed to have a degree of naiveté of $\tilde{v} = 0$, which yields $\hat{p}_2(x) = \mathbb{E}(p_2(x))$. This then results in naive optimists to overestimate their possible future payoff and naive pessimists to underestimate it.

Given their beliefs and following HSI agents only bet if ($\hat{8}$) from HSI is satisfied for $\hat{p}_2(x)$, i.e., if the expected payoff from investing is welfare-enhancing compared to not investing and it induces resistance in expectation:

$$\begin{aligned}&\underbrace{u(y^M) - u(x^M) + (\frac{1}{\hat{\delta}} - 1)(v(y^M) - v(x^M))}_{\widetilde{LHS}} \\ &\leq \underbrace{s(\lambda \frac{n \cdot w}{\frac{1}{2}(1+\tilde{v}) + \frac{1}{2}(n-1)(1-\tilde{v})} - w - k)}_{\widetilde{RHS}}. \end{aligned} \tag{$\tilde{8}$}$$

Definition 5 (Efficient Bet)**.** A wager is *efficient*, given a player's belief, whenever it induces equality in ($\tilde{8}$).

Before we start with the proof of Proposition 1 we need to consider an important factor. In case of parimutuel betting in general, players often receive information about the bet amount and current pot size. It is up to them to draw conclusions on how many players are involved in the game (pot size/wager). Although this seems like an easy mathematical task, the authors assume that there might be a range of players that will not do calculations like these and just ignore the fact that a larger pot size comes with a larger amount of play-

[35] See, for example, Shepperd *et al.* (2002) for optimism in comparative risk judgments.

ers. These type of players might simply associate a larger pot size with a higher expected payoff – in the (wrong) assumption the pot size was an exogenously given constant.[36]

In the special case of DietBet players are actively informed about how many players are in the game (see, e.g., Figure 3.B1), which should make this kind of ignorance a lot harder. But, given that we assume partially naive players to have a biased expectation about their own and other players' ability to win the bet we can still see a positive effect from a larger pot size (larger number of players in a game) on the biased expected payoff – see calculations below.

We leave the conclusion whether players realize the connection between pot size and number of players up to psychologists and simply observe that in both cases the expected payoff depends positively on the pot size/number of players:

1. the agent considers the pot size to be an exogenous constant, ρ. Then, $\frac{\partial \hat{p}_2(x)}{\partial \rho} = \frac{2}{(1+\tilde{v})+(n-1)(1-\tilde{v})} > 0$.
2. the agent considers the pot size to be dependent on the number of players. Then, $\frac{\partial \hat{p}_2(x)}{\partial n} = \frac{\partial \hat{p}_2(x)}{\partial \rho(n)} \cdot \frac{\partial \rho(n)}{\partial n}$

Just the size of this effect varies, but it is of no special interest in this theoretical approach.

Proof. **Proposition 1.**

i) This follows from the fact that agents only bet if Equation $\tilde{8}$ is satisfied.

ii) Sophisticates will solve their true problem and will only bet if resistance is induced and welfare is enhanced.

iii) This follows directly from the calculation above. $\frac{\partial \hat{p}_2(x)}{\partial \rho} = \frac{2}{(1+\tilde{v})+(n-1)(1-\tilde{v})} > 0$. So, the larger the pot size, the larger the expected payoff. This holds true independently of naiveté.

iv) a) Recall that naive optimists face $\tilde{v} > 0$. Now compare Equation $\tilde{8}$ with (8) from HSI (called (8_{HSI}) here, to avoid confusion):

$$\underbrace{u(y^M) - u(x^M) + (\frac{1}{\mathbb{E}\delta_2} - 1)(v(y^M) - v(x^M))}_{LHS} \leq \underbrace{s(\lambda 2w - w - k)}_{RHS} \qquad (8_{HSI})$$

For convenience we label the left-hand sides (right-hand sides) of inequalities $(\tilde{8})$ and (8_{HSI}) with $\widetilde{LHS}$ and LHS ($\widetilde{RHS}$ and RHS), respectively.

Given his belief, the monotonicity of s and for $n > 2$ we receive $\widetilde{LHS} < LHS$ and $RHS < \widetilde{RHS}$. Both of these relations follow directly from $\tilde{v} > 0$. If we assumed there was no naiveté about the expected payoff, the agent solved $\widetilde{LHS} = RHS$,

[36] Psychologists find that concerning lotteries people rather concentrate on the pot size than on the probability of winning and generally more people play when the pot size is larger (Griffiths and Wood, 2001).

i.e., he bets efficiently given his belief about his future degree of motivation while he estimates the expected payoff correctly (has no bias about his and his fellow players' likelihood to win), then $\widetilde{LHS} = RHS < LHS$. This implies he is more likely to undercommit. Due to the fact that he can only choose the efficient wager with respect to his expected motivation there is always the possibility that the random shock is positive and of such force that his actual motivation is even larger than his biased expected motivation ($\delta_2 > \hat{\delta}_2 > \mathbb{E}(\delta_2)$) which would imply that the commitment was binding.

b) If we include naiveté about the expected payoff it is possible that he ends up with $\widetilde{LHS} = RHS < LHS < \widetilde{RHS}$. This means, given the same wager his biased expectation would cancel out his undercommitment and he would win the bet nevertheless.

v) a) Analogously to iii) compare ($\tilde{8}$) and (8_{HSI}). Keep in mind that naive pessimists face $\tilde{v} < 0$ which reverses the relations stated above (for $n > 2$): $\widetilde{LHS} > LHS$ and $RHS > \widetilde{RHS}$.

Again, if we assume that there was no naiveté about the expected payoff, an agent would falsely choose the wager that solves $\widetilde{LHS} = RHS$. With $LHS < \widetilde{LHS} = RHS$ this implies he is more likely to overcommit in the sense that his bet is not efficient wrt. his true expected motivation. Due to the fact that he can only choose the efficient wager wrt. his expected motivation there is still the possibility that the random shock is negative and of such force that his actual motivation is smaller than his biased expected motivation ($\delta_2 < \hat{\delta}_2 < \mathbb{E}(\delta_2)$), which would imply that the commitment was not binding.

b) As above, include naiveté about the expected payoff. Generally, since $\widetilde{RHS} < RHS$ if he chose $\widetilde{LHS} = RHS$ he might end up with $LHS < \widetilde{RHS} < \widetilde{LHS} < RHS$ or $\widetilde{RHS} < LHS < \widetilde{LHS} < RHS$. This implies that, given the same bet amount, the distorted expectation about the pot size lets the chosen wager seem too small. To cancel this effect out (solving $\widetilde{LHS} = \widetilde{RHS}$) the agent would have to bet even more, which would make his overcommitment more severe.

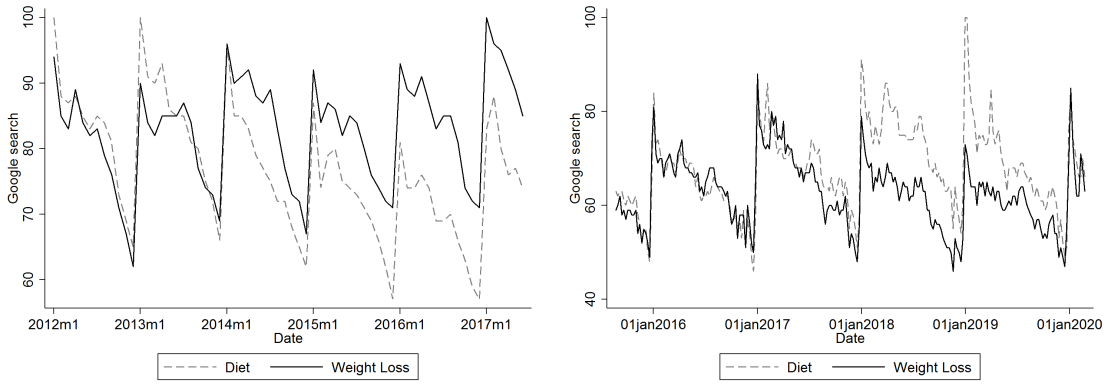
vi) This follows directly from ii)-v).

□

3.9 Appendix B: Descriptive Appendix

Figure 3.B1: Screenshot of a game example (taken on December 12th 2019)

Figure 3.B2: Data from worldwide Google trends for terms "diet" and "weight loss" (collected on August 8th 2020).



(a) Monthly searches (2012-2017)

(b) Weekly searches (2015-2020)

Figure 3.B3: Bets by month.

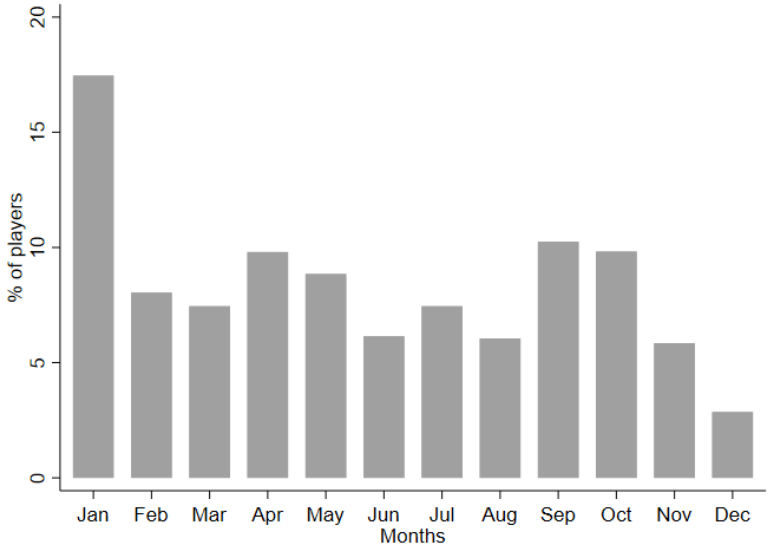
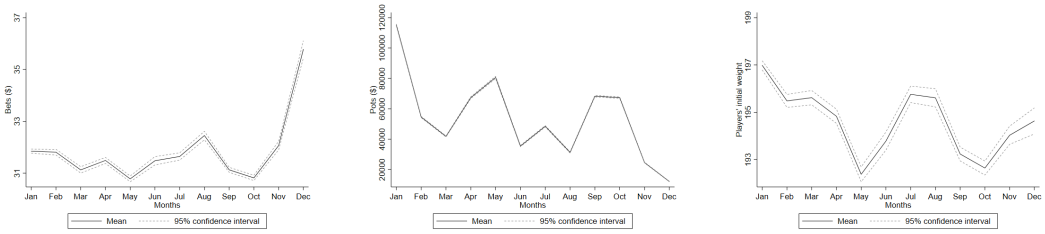
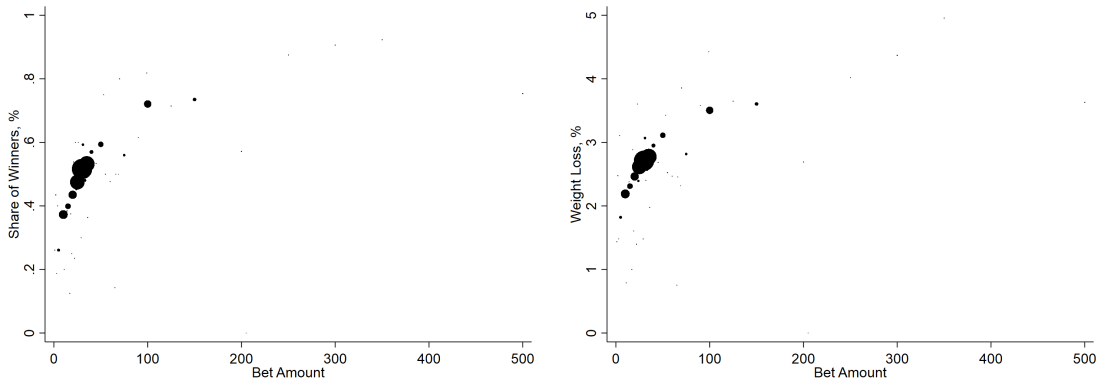


Figure 3.B4: Average bets, pots and players' starting weight by month.



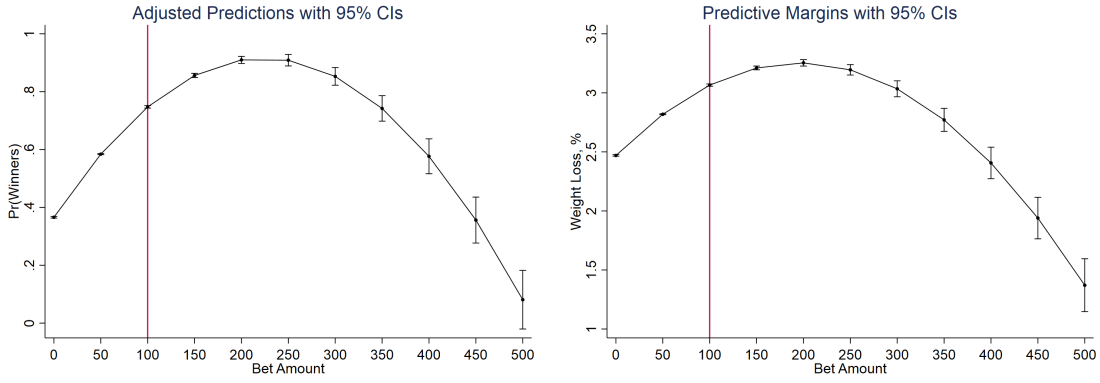
(a) Bets **(b)** Pots **(c)** Player's weight

Figure 3.B5: Scatter plot of share of winners and weight loss (%) per bet amount (weighted by the number of observations).



(a) Share of winners **(b)** Weight loss (%)

Figure 3.B6: Marginal effects of bet amount on probability of winning and weight loss.



(a) Probability of winning **(b)** Weight loss (%)

Descriptive statistics

Table 3.B1 shows that there are 21,077 games containing 912,737 players (user/game combinations).[37] Of these, 84% are female players[38] and 16% are male players. 44% are closed games and 56% are open games. 50.7% of all players were considered winners in their games. The average age of players is 35 years. On average, an individual participates in

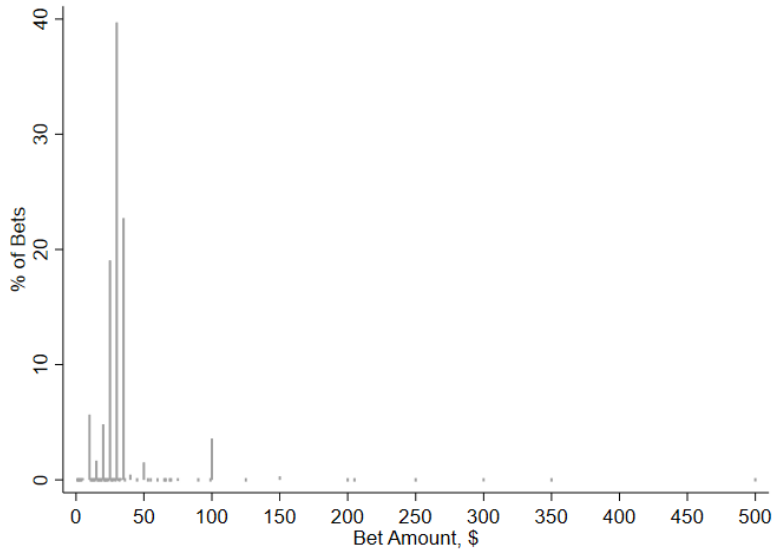
[37] The DietBet example shows that there is a demand for commitment: over five-year period almost 430,000 people used the program as a self-imposed commitment to lose weight.

[38] The fact that women represent the overwhelming majority of players is not surprising, e.g., Cawley and Price (2013) document that women were over-represented as participants in workplace weight loss programs.

Table 3.B1: Descriptive statistics I: Full sample.

	Obs	Mean	SD	Min	Max
			Game Characteristics		
Pot Amount, $	21077	1470.6	9757.8	2	442140
N Players in Games	21077	45.1	306.0	1	14448
N Winners in Games	21077	22.0	157.4	0	8777
Closed Game	21077	0.44	0.50	0	1
			User Characteristics		
N Games per User	426609	2.14	2.87	1	102
Male	426610	0.16	0.37	0	1
Age	233007	35.0	9.39	18	86.8
			Player Characteristics		
Bet Amount, $	912737	31.6	17.3	1	500
Amount Won, $	912737	32.5	37.6	0	1282.5
Share of Winners	912737	0.51	0.50	0	1
Start Weight, lb	912737	194.9	43.3	121.2	334.6
Final Weight, lb	912737	189.7	42.6	107.8	335
Weight Loss, lb	912737	5.20	4.61	-0.50	17.2
Weight Loss, %	912737	2.70	2.29	-0.41	11.9
Social Engagement	912737	0.93	10.0	0	261
N Weigh-ins in Game	912737	4.66	3.49	1	31
Male	912737	0.16	0.37	0	1
Completer	912737	0.64	0.48	0	1
			Player Characteristics: Winners		
Bet Amount, $	462791	33.5	19.8	1	500
Amount Won, $	462791	64.1	27.5	2	1282.5
Start Weight, lb	462791	193.3	42.6	121.2	334.6
Final Weight, lb	462791	184.2	40.8	107.8	329
Weight Loss, lb	462791	9.13	2.38	0.10	17.2
Weight Loss, %	462791	4.74	0.80	0.057	11.9
Social Engagement	462791	1.12	11.5	0	261
N Weigh-ins in Game	462791	5.85	3.73	1	31
Male	462791	0.19	0.40	0	1
			Player Characteristics: Nonwinners		
Bet Amount, $	449946	29.7	13.9	1	500
Start Weight, lb	449946	196.5	43.9	121.2	334.6
Final Weight, lb	449946	195.4	43.7	116.6	335
Weight Loss, lb	449946	1.16	2.25	-0.50	13.2
Weight Loss, %	449946	0.60	1.12	-0.41	3.99
Social Engagement	449946	0.73	8.27	0	261
N Weigh-ins in Game	449946	3.44	2.74	1	30
Male	449946	0.13	0.33	0	1

Figure 3.B7: Percent of bets by size



2.14 games. The number of players per game ranges significantly – while on average there are 45.1 players per game, there are small games (only 1 player) and large games (14,448 players).

The average bet amount is $31.6 (USD), with the lowest bet being $1 and highest bet being $500. Winners on average bet more than nonwinners ($33.5 vs $29.7). On average winners won $64.1, but the winning amounts range from $2 to $1,283.

The average initial weight of the player is 194.9 lb (88.4 kg). The final weight is 189.7 lb (86 kg). On average, players lost 5.2 lb (2.4 kg). Winners lost 9.13 lb (4.14. kg) on average, while nonwinners lost 1.16 lb (0.53 kg) on average.[39] A total of 64.5% of players completed an official weigh-in at the end of the game.

We examine whether player characteristics differ between the groups of completers and noncompleters. For this purpose, we use analysis of variance and chi-square tests. Table 3.B2 in Appendix B presents the results from the comparison of completers to noncompleters. Here we see that the groups differ significantly between each other on a number of characteristics: there are more males among completers, completers have a lower baseline weight and bet more money. Moreover, completers finish more weigh-ins and are more socially engaged.[40]

Since noncompleters did not submit their final weight, we make an assumption that people who did not submit the official final weigh-ins did not lose any weight. On average,

[39]Some players did not complete a final, verified weigh-in (we refer to them as "Noncompleters"). By not completing their final game weigh-in, noncompleters forfeit their wager.

[40]During the game, players have a possibility submit unofficial weigh-ins to track their progress and to share their participation on Facebook, share photos, add comments and likes.

players lost 2.7% (SD 2.3) of their initial weight, and the weight loss is statistically different between completers and noncompleters (4.2% vs 0.0% (P<0.00)).

We use both completers and noncompleters in our analysis, but to account for noncompletion and the assumption we make about players' weight loss in that case, we always control for it in our analysis by including a dummy variable equal to one if a person is a completer and zero otherwise.[41]

Table 3.B2: Completers vs noncompleters.

	Completers		Noncompleters		P value
Male, N (%)	88956	(10.0)	36258	(4.1)	<0.000
Start Weight lb, Mean (SD)	195.7	(47.8)	199.0	(49.1)	<0.000
Bet Amount $, Mean (SD)	32.4	(18.6)	30.2	(14.1)	<0.000
N Weigh-ins in Game, Mean (SD)	5.7	(3.6)	2.8	(2.3)	<0.000
Social Engagement, Mean (SD)	1.3	(22.5)	0.7	(14.3)	<0.000
Weight Loss (%), Mean (SD)	4.1	(1.6)	0.0	(0.1)	<0.000
N	617501		329748		

[41] We also repeat the analysis on the sample of completers only. The results stay qualitatively and quantitatively similar (available upon request).

Table 3.B3: Descriptive statistics II: Panel sample.

	Obs	Mean	SD	Min	Max
			Game Characteristics		
Pot Amount, $	17572	1675.2	10489.9	2	442140
N Players in Games	17572	51.2	329.4	1	14448
N Winners in Games	17572	25.0	169.2	0	8777
Closed Game	17572	0.40	0.49	0	1
			User Characteristics		
N Games per User	157788	4.08	4.04	2	102
Male	157788	0.15	0.35	0	1
Age	105648	34.7	9.00	18	86.3
			Player Characteristics		
Bet Amount, $	643916	32.5	18.5	1	500
Amount Won, $	643916	37.4	37.9	0	1050
Share of Winners	643916	0.58	0.49	0	1
Start Weight, lb	643916	195.3	42.8	121.2	334.6
Final Weight, lb	643916	189.5	42.0	108.7	335
Weight Loss, lb	643916	5.81	4.54	-0.50	17.2
Weight Loss, %	643916	3.00	2.23	-0.41	11.9
Social Engagement	643916	0.92	9.86	0	261
N Weigh-ins in Game	643916	5.03	3.68	1	31
Male	643916	0.16	0.36	0	1
Completer	643916	0.70	0.46	0	1
			Player Characteristics: Winners		
Bet Amount, $	375305	34.0	20.4	1	500
Amount Won, $	375305	64.2	27.4	3	1050
Start Weight, lb	375305	194.0	42.4	121.2	334.6
Final Weight, lb	375305	184.9	40.6	108.7	321
Weight Loss, lb	375305	9.12	2.35	0.10	17.2
Weight Loss, %	375305	4.71	0.77	0.057	11.9
Social Engagement	375305	1.04	10.8	0	261
N Weigh-ins in Game	375305	5.97	3.84	1	31
Male	375305	0.19	0.39	0	1
			Player Characteristics: Nonwinners		
Bet Amount, $	268611	30.4	15.1	1	500
Start Weight, lb	268611	197.1	43.2	121.2	334.6
Final Weight, lb	268611	195.9	43.0	116.6	335
Weight Loss, lb	268611	1.18	2.26	-0.50	13
Weight Loss, %	268611	0.60	1.12	-0.41	3.99
Social Engagement	268611	0.76	8.31	0	261
N Weigh-ins in Game	268611	3.71	2.96	1	30
Male	268611	0.11	0.31	0	1

3.10 Appendix C: Empirical Appendix

Table 3.C1: Effect of high-stake bets and marginal effects of bets on probability of winning and weight loss (%) – not controlling for completer status.

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High-Stake Bet	0.1068***	0.1052***	0.4297***	0.4246***
	(0.0028)	(0.0027)	(0.0133)	(0.0126)
	(1)	(2)	(3)	(4)
Bet Amount under $100	0.0028***	0.0029***	0.0103***	0.0117***
	(0.0001)	(0.0001)	(0.0005)	(0.0004)
Bet Amount over $100	0.0004***	-0.0001	0.00191***	-0.0005
	(0.0001)	(0.0001)	(0.0005)	(0.0004)
Add. controls	No	Yes	No	Yes
N	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788

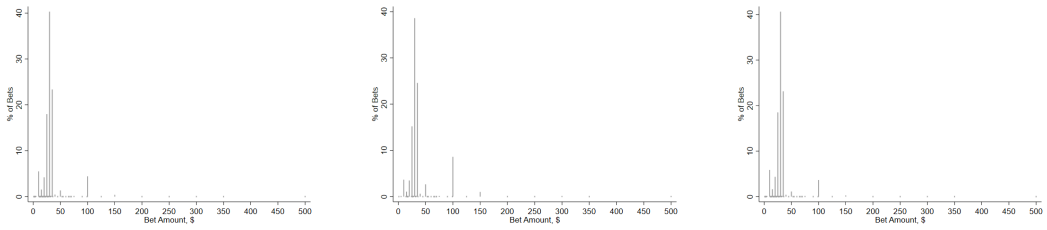
Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). All regressions include individual fixed effects. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$

Table 3.C2: Effect of high game pots on probability of winning and weight loss (%) – not controlling for completer status.

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High Pot Size	0.0126***	0.0189***	0.0350***	0.0752***
	(0.0013)	(0.0012)	(0.0059)	(0.0053)
Add. controls	No	Yes	No	Yes
N	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788

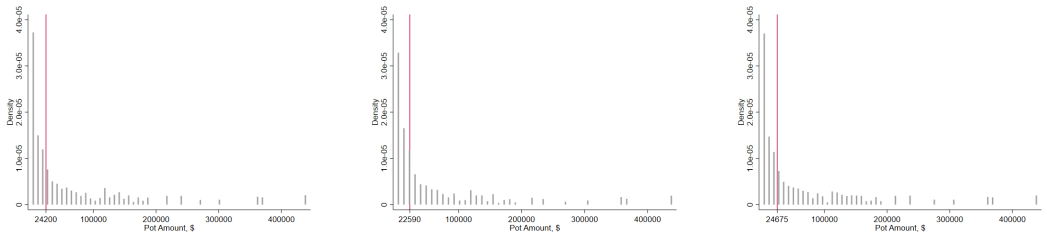
Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). High Pot Size is a dummy variable equal to one if pot size is larger than the sample median (24,200$), and zero otherwise. All regressions include individual fixed effects. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, bet and bet squared, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Figure 3.C1: Percent of bets by size.



(a) Bets **(b)** Pots **(c)** Player's weight

Figure 3.C2: Distributions of pot sizes with sample median.



(a) Panel sample **(b)** Male panel **(c)** Female panel

Figure 3.C3: Percent of bet sizes by month

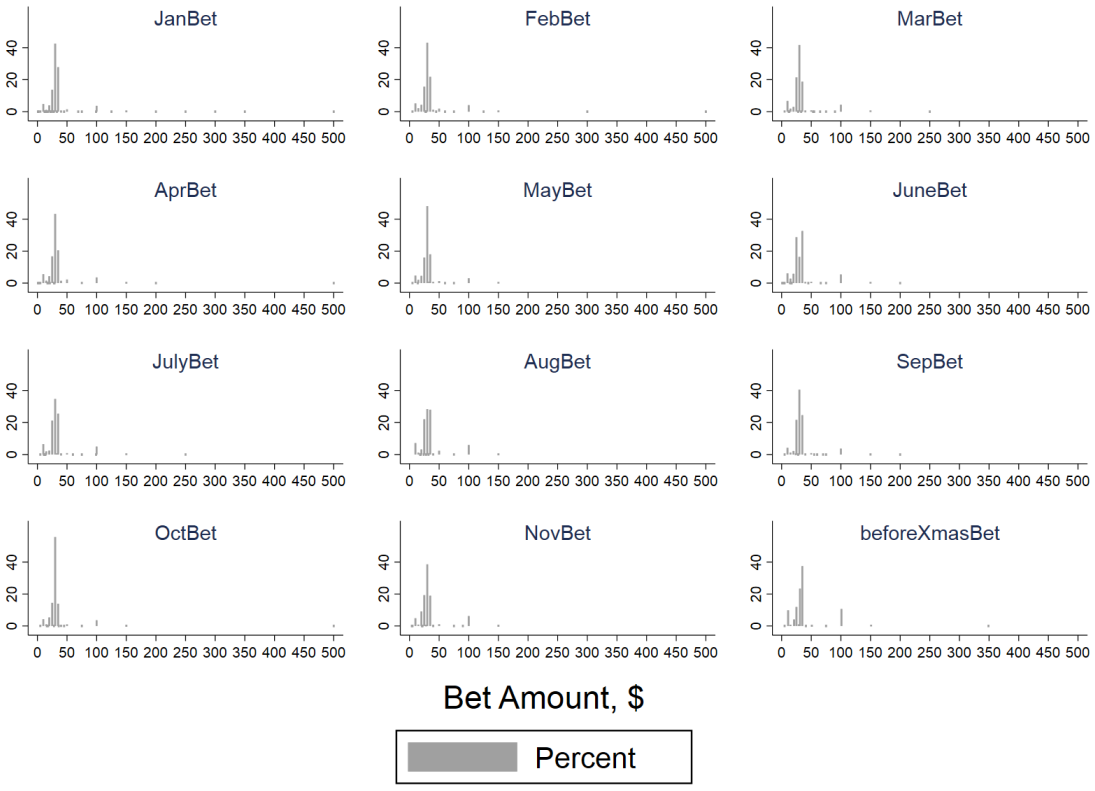


Figure 3.C4: Density of pots by month

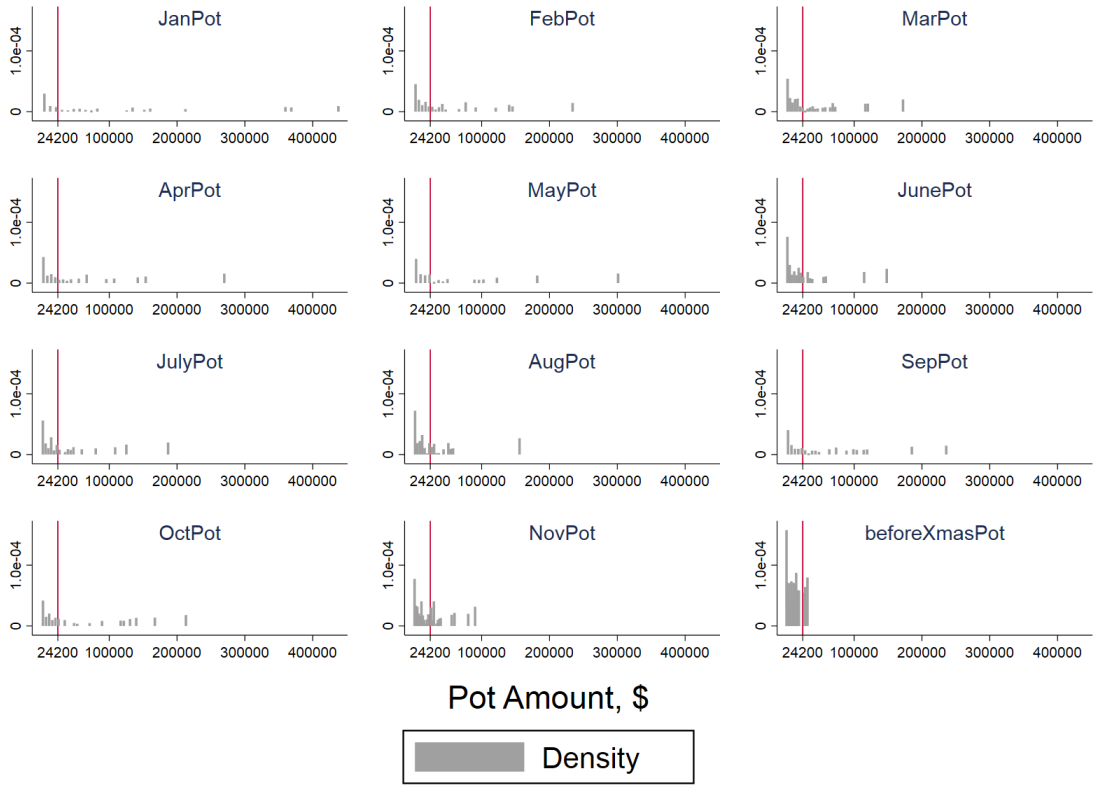


Table 3.C3: Placebo test: Probability of winning.

	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
High-Stake Bet	0.0527***	0.0530***	0.0517***	0.0530***	0.0519***	0.0533***	0.0529***	0.0535***	0.0535***	0.0524***
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)
High-Stake Bet	-0.0010	-0.0021	0.0158***	-0.0023	0.0159***	-0.0053	0.0015	-0.0059	-0.0079	0.0078
× X	(0.0046)	(0.0052)	(0.0054)	(0.0063)	(0.0059)	(0.0051)	(0.0052)	(0.0052)	(0.0056)	(0.0053)
	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Bet Amount	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***
under $100	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Bet Amount	0.0000	-0.0000	-0.0000	-0.0000	0.0000	-0.0000	-0.0000	0.0000	-0.0000	0.0000
over $100	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Bet Amount	-0.0002	-0.0003	-0.0001	0.0001	-0.0004	-0.0005**	-0.0008***	-0.0001	0.0002	-0.0001
under $100 × X	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Bet Amount	-0.0002	0.0001	-0.0001	-0.0000	-0.0005	0.0002	0.0002	-0.0007*	0.0002	-0.0005
over $100 × X	(0.0002)	(0.0002)	(0.0001)	(0.0004)	(0.0004)	(0.0002)	(0.0004)	(0.0004)	(0.0003)	(0.0005)
Add. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	643,916	643,916	643,916	643,916	643,916	643,916	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788	157,788	157,788	157,788	157,788	157,788	157,788

Dependent variable: probability of being a winner in the game (estimated by LPM). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 3.C4: Placebo test: Weight loss (%).

	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
High-Stake Bet	0.1337***	0.1393***	0.1360***	0.1389***	0.1364***	0.1418***	0.1370***	0.14175***	0.1395***	0.1391***
	(0.0066)	(0.0066)	(0.0065)	(0.0064)	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)
High-Stake Bet	0.0332*	-0.0101	0.0358*	-0.0052	0.0419*	-0.0396*	0.0250	-0.0358*	-0.0102	0.0015
× X	(0.0184)	(0.0194)	(0.0205)	(0.0230)	(0.0227)	(0.0202)	(0.0197)	(0.0206)	(0.0209)	(0.0205)
	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Bet Amount	0.0057***	0.0057***	0.0057***	0.0056***	0.0057***	0.0058***	0.0058***	0.00572***	0.0055***	0.0056***
under $100	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Bet Amount	0.0002	0.0000	0.0001	0.0001	0.0001	0.0001	0.0000	0.0001	0.0000	0.0001
over $100	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Bet Amount	-0.0001	-0.0017**	-0.0004	0.0001	-0.0009	-0.0022***	-0.0026***	-0.0014*	0.0016*	0.0007
under $100 × X	(0.0007)	(0.0008)	(0.0007)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0010)
Bet Amount	-0.0006	0.0012	-0.0004	-0.0002	-0.0008	0.0004	0.0024	-0.0024	0.0022*	-0.0008
over $100 × X	(0.0005)	(0.0011)	(0.0005)	(0.0013)	(0.0014)	(0.0009)	(0.0017)	(0.0018)	(0.0012)	(0.0017)
Add. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	643,916	643,916	643,916	643,916	643,916	643,916	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788	157,788	157,788	157,788	157,788	157,788	157,788

Dependent variable: weight loss in % (estimated by OLS). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 3.C5: Placebo test: Probability of winning and weight loss (%) for games with higher pot sizes.

	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
High Pot Size	0.0146***	0.0142***	0.0148***	0.0136***	0.0138***	0.0147***	0.0142***	0.0141***	0.0138***	0.0136***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
High Pot Size	-0.0086***	-0.0086***	-0.0113***	0.0006	-0.0035	-0.0130***	-0.0090***	-0.0070**	-0.0014	0.0014
× X	(0.0025)	(0.0028)	(0.0028)	(0.0029)	(0.0034)	(0.0029)	(0.0032)	(0.0028)	(0.0028)	(0.0032)
	Weight Loss (%)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
High Pot Size	0.0504***	0.0493***	0.0497***	0.0462***	0.0472***	0.0509***	0.0486***	0.0474***	0.0472***	0.0455***
	(0.0031)	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0030)
High Pot Size	-0.0363***	-0.0378***	-0.0297***	0.0025	-0.0157	-0.0611***	-0.0348***	-0.0187**	-0.0035	0.0108
× X	(0.0086)	(0.0096)	(0.0096)	(0.0102)	(0.0121)	(0.0102)	(0.0111)	(0.0095)	(0.0100)	(0.0113)
Add. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	643,916	643,916	643,916	643,916	643,916	643,916	643,916	643,916	643,916	643,916
N (clust)	157,788	157,788	157,788	157,788	157,788	157,788	157,788	157,788	157,788	157,788

Dependent variable: probability of winning and weight loss in % (estimated by OLS). High Pot Size is a dummy variable equal to one if pot size is larger than the sample median (24,200$), and zero otherwise. All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, gender, social engagement, N of weigh-ins, indicator whether the game is closed, bet and bet squared, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 3.C6: Probability of winning by year.

	2012	2013	2014	2015	2016	2017
	(1)	(2)	(3)	(4)	(5)	(6)
High-Stake Bet	0.1615	0.0591***	0.0699***	0.0415***	0.0418***	0.0185***
	(0.1218)	(0.0104)	(0.0053)	(0.0034)	(0.0026)	(0.0040)
	2012	2013	2014	2015	2016	2017
	(1)	(2)	(3)	(4)	(5)	(6)
Bet Amount under $100	0.0016**	0.0015***	0.0017***	0.0015***	0.0009***	0.0013***
	(0.0008)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0003)
Bet Amount over $100	-0.0012	0.0001	0.0000	-0.0003**	-0.0002	-0.0000
	(0.0008)	(0.0001)	(0.0003)	(0.0002)	(0.0002)	(0.0003)
Add. controls	Yes	Yes	Yes	Yes	Yes	Yes
N	2,856	65,697	87,311	140,154	183,860	49,435
N (clust)	1,094	20,161	28,459	42,452	54,197	18,300

Dependent variable: probability of being a winner in the game (estimated by LPM). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 3.C7: Weight loss (%) by year.

	2012	2013	2014	2015	2016	2017
	(1)	(2)	(3)	(4)	(5)	(6)
High-Stake Bet	0.1885	0.0907***	0.1712***	0.1153***	0.1117***	0.0651***
	(0.5177)	(0.0351)	(0.0189)	(0.0128)	(0.0010)	(0.0162)
	2012	2013	2014	2015	2016	2017
	(1)	(2)	(3)	(4)	(5)	(6)
Bet Amount under $100	0.0048	0.0044***	0.0062***	0.0047***	0.0025***	0.0032***
	(0.0030)	(0.0006)	(0.0007)	(0.0005)	(0.0004)	(0.0009)
Bet Amount over $100	0.0003	0.0001	0.0006	-0.0000	-0.0004	-0.0006
	(0.0036)	(0.0003)	(0.0011)	(0.0006)	(0.0006)	(0.0010)
Add. controls	Yes	Yes	Yes	Yes	Yes	Yes
N	2,856	65,697	87,311	140,154	183,860	49,435
N (clust)	1,094	20,161	28,459	42,452	54,197	18,300

Dependent variable: weight loss in % (estimated by OLS). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 3.C8: Probability of Winning and Weight loss (%) by year – non-parallel games.

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High-Stake Bet	0.0682***	0.0751***	0.1754***	0.2004***
	(0.0025)	(0.0025)	(0.0093)	(0.0093)
	(1)	(2)	(3)	(4)
Bet Amount under $100	0.0024***	0.0021***	0.0075***	0.0062***
	(0.0001)	(0.0001)	(0.0003)	(0.0003)
Bet Amount over $100	0.0003***	0.0001	0.0011***	0.0005
	(0.0001)	(0.0001)	(0.0004)	(0.0004)
Add. controls	No	Yes	No	Yes
N	455,428	455,428	455,428	455,428
N (clust)	140,347	140,347	140,347	140,347

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * $p<0.1$ ** $p<0.05$ *** $p<0.01$

Table 3.C9: Effect of high-stake bets on probability of winning and weight loss (%): Full sample.

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High-Stake Bet	0.0504***	0.0530***	0.1302***	0.1389***
	(0.0016)	(0.0016)	(0.0062)	(0.0063)
	(1)	(2)	(3)	(4)
Bet Amount under $100	0.0020***	0.0017***	0.0061***	0.0050***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Bet Amount over $100	0.0001*	-0.0000	0.0004**	0.0001
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Add. controls	No	Yes	No	Yes
N	912,737	912,737	912,737	912,737
N (clust)	426,609	426,609	426,609	426,609

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, pot size and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 3.C10: Effect of high game pots on probability of winning and weight loss (%): Full sample.

	Winner		Weight Loss (%)	
	(1)	(2)	(3)	(4)
High Pot Size	0.0169***	0.0138***	0.0571***	0.0471***
	(0.0008)	(0.0008)	(0.0029)	(0.0029)
Add. controls	No	Yes	No	Yes
N	912,737	912,737	912,737	912,737
N (clust)	426,609	426,609	426,609	426,609

Dependent variables: probability of being a winner in the game (estimated by LPM); and weight loss in % (estimated by OLS). High Pot Size is a dummy variable equal to one if pot size is larger than the sample median (22,855$), and zero otherwise. All regressions include individual fixed effects and an indicator for being a completer. Controls include players' starting weight, social engagement, N of weigh-ins, indicator whether the game is closed, bet and bet squared, and a categorical variable for game order. Standard errors in parentheses, clustered at the individual level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

4 Providers, Peers and Patients. How do Physicians' Practice Environments Affect Patient Outcomes?*

Abstract. We study how physicians practice environments affect their treatment decisions and quality of care. Using clinical registry data from Sweden, we compare stent choices of cardiologists moving across hospitals over time. To disentangle changes in practice styles attributable to hospital- and peer group-specific factors, we exploit quasi-random variation on cardiologists working together on the same days. We find that migrating cardiologists' stent choices rapidly adapt to their new practice environment after relocation and are equally driven by the hospital and peer environments. In contrast, while decision errors increase, treatment costs and adverse clinical events remain largely unchanged despite the altered practice styles.

4.1 Introduction

Individual behavior is to a large extent shaped through interactions with the environment. The social environment in particular, including peer networks, classmates and workplace colleagues, has been widely studied by labor economists to understand and quantify important drivers of productivity (see, e.g., Falk and Ichino, 2006; Mas and Moretti, 2009; Sacerdote, 2001). In healthcare, a growing literature in health economics has sought to understand causes of variations in physician practice styles and their consequences for care quality, healthcare utilization and overall health system efficiency (see, e.g., Currie *et al.*, 2016; Currie and MacLeod, 2020; Cutler *et al.*, 2019; Epstein and Nicholson, 2009; Epstein *et al.*, 2016; Fadlon and Van Parys, 2020; Grytten and Sørensen, 2003; Molitor, 2018;

*This chapter is written jointly with Daniel Avdic, Bo Lagerqvist and Ieva Sriubaite. A revised version of this article is published as *Avdic, D., Ivets, M., Lagerqvist, B., Sriubaite, I. (2023). Providers, peers and patients. How do physicians practice environments affect patient outcomes?. Journal of Health Economics, 89, 102741.* https://doi.org/10.1016/j.jhealeco.2023.102741.

We thank Sofia Amaral-Garcia, Amitabh Chandra, Anthony Harris, David Molitor, Jonathan Skinner and seminar participants in Essen, Copenhagen, 4th EuHEA student-supervisor conference in Lausanne, 27th European Workshop on Econometrics and Health Economics in Groningen, 6th Health Econometrics Workshop in Bergamo and 30th Annual EALE Conference in Lyon for valuable comments. All errors are our own.

Van Parys, 2016; Weng *et al.*, 2020).[1,2] However, despite calls for effective policies that seek to reduce inappropriate variations in healthcare utilization (see, e.g., Brownlee *et al.*, 2017; OECD, 2014a), the scientific evidence on the impact of healthcare practitioners' social environment on clinical practice behavior and patient outcomes has so far been scant.[3]

This paper seeks to add to the literature on the determinants of provider practice styles by studying how physicians' treatment choices are influenced by their practice environment and the consequences these choices have for their patients. To this end, we make two major contributions that so far have been largely overlooked in the literature. First, we propose a method to decompose the environmental effect into a physical and a social component, corresponding to a hospital-specific and a peer group-specific component. As we argue further below, this is an important distinction to make since the two components provide very different implications for policy. Second, by relating data on physicians' treatment choices to optimal management and associated patient outcomes, we are able to gauge and directly measure the impact of environmentally induced variation in physician treatment behavior on changes in the appropriateness, treatment costs, and the quality of care received by patients. This is in contrast to most existing studies on physician practice styles, which often rely solely on quantitative measures to evaluate practice heterogeneity.

To provide an empirical framework for identification and consistent estimation of causal effects, we apply and extend the physician migration approach used by Molitor (2018) in the important context of hospital treatment of patients suffering from coronary heart disease. Specifically, we focus on stent choice in coronary angioplasty; a medical procedure used to widen blocked or narrowed blood vessels in the heart.[4] We identify physicians who move (migrate) across hospitals and relate variation in the rate of use of a specific type of stent between the physician's pre-move (origin) and post-move (destination) hospitals to changes in the physician's own stent use across time in a difference-in-differences model.

To estimate the model, we use rich administrative data from the Swedish Coronary Angiography and Angioplasty Register (SCAAR) on all percutaneous coronary interventions (PCI) performed in Sweden between 2004 and 2013 and study how interventional cardi-

[1] Chandra *et al.* (2011) provide an overview of different explanations for why provider treatment decisions may vary across similar patients. Such reasons include (i) "defensive medicine", where providers perform unnecessary procedures to avoid complaints, bad reputation and possible lawsuits from patients; (ii) financial incentives associated with fee-for-service reimbursement models (McClellan, 2011); and (iii) unobserved heterogeneity across providers (Doyle *et al.*, 2010).

[2] Traditional demand factors, such as patient preferences and needs, are by and large unable to explain the substantial geographic variations in healthcare utilization observed in many countries (see, e.g., Chandra *et al.*, 2011; Finkelstein *et al.*, 2016; Skinner, 2011; Skinner *et al.*, 2011; Wennberg and Gittelsohn, 1973). For studies based on non-US data, see Bojke *et al.* (2013); Corallo *et al.* (2014); Godøy and Huitfeldt (2020); Kopetsch and Schmitz (2014); Moura *et al.* (2019); Phelps (2000); Prieto and Lago-Peñas (2012); Reich *et al.* (2012). Salm and Wübker (2020) provide an exception by finding that the vast majority of variation in ambulatory care use stems from demand factors, which they argue is due to supply-side constraints.

[3] An important exception is Chan (2016) who analyzes teamwork and moral hazard among physicians in an emergency department.

[4] Coronary angioplasty entails transporting a tiny deflated balloon to the blocked area using a catheter. Once in place, the balloon is inflated to unclog the blood vessel, thereby restoring blood flow to the heart. The angioplasty procedure usually also involves inserting a small wire mesh tube, a stent, into the artery. The stent is left in place to prop the artery open, decreasing the risk of reocclusion. This is called percutaneous coronary intervention (PCI). In this paper, we focus on stent choice among cardiologists. See also section 4.8 in Appendix B.

ologists' choices between the bare-metal stent (BMS) and the drug-eluting stent (DES) are determined by their environment.[5] Since the procedure, PCI, is identically performed irrespective of the type of stent used, this context provides an essentially ideal setting to study how the practice environment shapes physician preferences for treatment.

While empirical evidence on the extent to which physician practice styles are influenced by their work environment is informative, it does not per se convey much detail on *which* environmental factors are the drivers of such changes. Yet, such knowledge could be important. For example, physical, or hospital-specific, factors may be less informative about the malleability of physicians' underlying preferences if the possibility to operate in line with such preferences is restricted by factors beyond the individual physician's control, such as resource constraints, staff micromanagement, or hospital-specific guidelines. In contrast, social, or peer group-specific, factors are more directly related to the adjustment of physician beliefs for which much of the economic literature on physician practice styles lies at the heart of (see, e.g., Epstein and Nicholson, 2009).

To address this important question, we propose and implement a method to decompose the combined impact of the environment on physician treatment styles into a hospital-specific and a peer group-specific factor by exploiting quasi-random variation on physicians working together on given days. Specifically, given sufficient practice style variation among migrating physicians' coworkers (peers) *within* a hospital, the inclusion of hospital fixed effects in our econometric model will effectively purge all time-invariant hospital-specific variation in practice styles *across* hospitals from the analysis. Any remaining practice variation will consequently be derived from changes in the migrating physicians' coworker mix. Thus, resulting estimates of the environmental effect with and without hospital fixed effects gauge the relative magnitude of the adjustment in physician practice style arising from hospital- and peer group-specific factors, respectively.

Our estimation results show that cardiologists' use of DES in angioplasty treatments are strongly determined by the practice style of the hospital they currently work in. Migrating cardiologists rapidly adapt to their prevailing practice environment after relocation by changing their DES use with on average half a percentage point for each percentage point difference in DES utilization rates between the origin and destination hospitals. This result is robust to a set of alternative specifications and close to the corresponding estimate found by Molitor (2018). Furthermore, when decomposing the overall effect into a hospital-specific and a peer group-specific effect, we find that each component is responsible for roughly half of the practice style adjustment. To assess the extent of heterogeneity in response across cardiologists, we also provide results from a series of split sample regressions which reveal that our results are mainly derived from younger cardiologists and cardiologists moving to more DES-intensive environments.

In contrast, we find no empirical evidence to support the hypothesis that environmentally induced changes in migrating physicians' practice styles had important consequences

[5] The main trade-off in the choice of stent is based on the relative risk of two adverse clinical patient outcomes: restenosis (i.e., artery reocclusion) and stent thrombosis (i.e., blood clots caused by the stent itself). The former is mainly related to the use of BMS and the latter to DES. Due to uncertainty regarding the clinical evidence available at the time, these risks were not well-known, which left the decision of which stent to use largely up to the individual physician. See also section 4.8 in Appendix B.

for the quality of care received by patients. In addition to analyzing a set of adverse clinical events related to the medical procedure, we employ a machine learning algorithm to classify appropriate stent choices for each case based on out-of-sample predictions from academic hospitals and a rich set of patient and clinical characteristics. While our analysis does not reveal important systematic impacts on patient health as a result of changes in their physician's practice environment, we do find that migrating physicians are more likely to incorrectly apply DES after their move. This result suggests that the environmentally induced changes in physicians' practice behavior are mainly based on marginal "gray-zone" cases who run little risk of serious adverse medical events as a consequence of such choices. Moreover, a back-of-the-envelope calculation of the potential monetary savings from following the most efficient treatment approach suggests that the average migrating cardiologist incurred an additional cost of USD 1,200 per year from inappropriate stent choices, corresponding to roughly one-sixth of the price of a PCI.

One potential concern with our decomposition approach is that migrating physicians are non-randomly matched with peers after they move. This would introduce bias in our estimated parameters if migrants exert some control over whom they are working with and use this control to select coworkers with matching preferences. While this is unlikely to occur in practice, and would lead our estimates to be a lower bound on the true effect if it did, we nevertheless evaluate the robustness of our results to such endogeneity concerns by replacing our measure of practice environment with a *synthetic* environment. Based on the synthetic control method (see, e.g., Abadie and Gardeazabal, 2003; Abadie *et al.*, 2010, 2015), we construct an artificial matched comparison group using the sample of non-migrating cardiologists in our data. This method safeguards against estimation bias by comparing practice styles of migrating cardiologists with non-migrating cardiologists who were exposed to similar peer practice environments prior to the relocation. Reassuringly, we find that our estimates are largely robust to the definition of practice environment. In addition, we also estimate placebo models where we replace our main outcome with indicators for patient case-mix to study whether our peer effects are driven by patient, as opposed to stent, selection. The results from this analysis show, in line with the knowledge transfer hypothesis, that the type of patients that cardiologists are treating is unrelated to their peers' practice styles.

Our findings contribute to the scant literature on peer effects and social learning in healthcare. Social learning is broadly defined as the process of information transmission between economic agents when they observe and interact with each other within their social networks (see, e.g., Lin *et al.*, 1981). In line with our results, Huesch (2011) finds evidence for intragroup spillovers in the use of DES, suggesting that physicians are influenced by their peers. Furthermore, Nair *et al.* (2010) study peer effects in prescribing choices of physicians and find that such behavior is particularly influenced by research-active peers within physician groups. Heijmans *et al.* (2017) find similar results studying peer effects in cardiovascular risk management in networks with and without opinion leaders. On the other hand, Yang *et al.* (2014) document only small peer effects in prescription behavior for new drugs among physicians working in the same hospital at the same time. Silver (2021) studies the effect of working in a high-pressure peer group environment in an emergency department on physicians' treatment behavior and patient outcomes and find that physicians alter their treatment styles in response to their peer group, affecting quality of care.

Moreover, Epstein and Nicholson (2009) find that physician's treatment styles are responsive to changes in treatment styles of other physicians in the same hospital region in the context of Cesarean sections, but the effect dampens when accounting for common shocks at the hospital level. This is in line with our finding that both hospitals and peers are influential in altering practice styles of physicians. Finally, Burke *et al.* (2003) find that patients are more likely to receive certain procedures if an attending physician is in a group that performs these procedures more frequently, and Yuan *et al.* (2020) show that shared beliefs are crucial for successful implementation of new health technology within a peer network. Complementing these findings, the results from our split sample analyses show that effects are driven by younger cardiologists and cardiologists moving to more DES-intensive environments.

We also add contextual depth to the more general economic literature on peer effects.[6] A number of papers have investigated the influence of peers on academic performance, yielding mixed results. While some authors find significant peer effects (Sacerdote, 2001; Zimmerman, 2003), others find no effects at all (Foster, 2006; Lyle, 2007), or effects only for particular subgroups (Stinebrickner and Stinebrickner, 2006). In contrast, there exists strong evidence for positive social spillovers on task-oriented work behavior and productivity in non-academic settings. Mas and Moretti (2009) study peer effects at the workplace by analyzing the productivity of coworkers within the same team. They find evidence of positive productivity spillovers when working with highly productive peers, especially when they interact more frequently. Moreover, in an experimental setting, Falk and Ichino (2006) study individuals working on separate tasks within the sight of one another, finding that the productivity of workers is influenced by the productivity of their peers. These results motivate our approach to use physicians working on the same days as relevant peers in the analysis. Finally, Bandiera *et al.* (2010) study whether workers' behaviors are affected by the presence of peers that they are socially tied to, with the main finding that a worker's productivity is positively correlated with friends' abilities.

Furthermore, we contribute to the emerging health economics' literature on physician practice styles. Previous research on this topic has documented mixed results. Grytten and Sørensen (2003) find that primary care physicians' practice styles are largely stable, while for specialists Molitor (2018) and Weng *et al.* (2020) show that practice styles are malleable to their environment. Variation in physicians' treatment styles may also have a lasting influence on patient care (see, e.g., Currie *et al.*, 2016; Fadlon and Van Parys, 2020; Kwok, 2019). We complement these findings by providing empirical evidence that physicians equally strongly react to their social as well as their physical environments and relating how these altered practice styles affect quality of care.

Lastly, our results have broad implications for healthcare system efficiency. The finding that physicians' decisions are influenced by their social environment suggests a rationale for why practice styles cluster in certain areas. While such clustering may generate positive productivity and learning spillovers as in Chandra and Staiger (2007), it also implies that patients may receive suboptimal care depending on the dominating practice style at the admitting healthcare provider. In particular in supply-sensitive areas of healthcare, where

[6] See Herbst and Mas (2015) for a review of studies focusing on workers' response to speed and throughput of their peers.

the frequency of use of an activity is related to its local capacity, and where the choice of healthcare provider is subject to restrictions, such as place of residence, substantial allocation inefficiencies may exist. If quality of care is largely insensitive to such variations, as this paper shows in the context of cardiac catheterizations, a more integrated system where inappropriate practice variation can be mitigated through enhanced care coordination, monitoring, and follow-up based on evidence-based clinical guidelines could be vastly resource-saving (Wennberg, 2010). However, broadly defined uniform guidelines may not be the most efficient way to reduce inappropriate healthcare variations when patient populations are clinically diverse. For example, Chan *et al.* (2022) show that diagnosing pneumonia varies substantially across physicians with different skill levels where less skilled physicians are more likely to choose lower thresholds to reduce the risk of failing to correctly diagnose a patient.[7] Similarly, we find that younger and less experienced migrating physicians are more likely to inappropriately apply DES after their move. These findings suggest that investments in training of less experienced physicians to increase their skill may be a cost-efficient alternative to national guidelines to reduce unwarranted resource use.

4.2 Institutional Settings

The empirical analyses in this paper are based on inpatient medical records on percutaneous coronary interventions performed in Sweden between 2004 and 2013. Here, we provide a brief summary of the relevant characteristics of the Swedish healthcare system and clinical context we study. A detailed exposition of the institutional framework can be found in Appendix B.

Healthcare in Sweden

Healthcare in Sweden is mainly funded by direct income taxes raised by the three different levels of government: central, regional (21 county councils) and local (290 municipalities). The regional county councils are the main providers and financiers of healthcare in Sweden, being responsible for primary and specialized healthcare on both the in- and outpatient basis in their area of jurisdiction. Each council sets its own patient fees, although there is a national cap for the amount a patient has to pay out of pocket. Consequently, patient fees only account for around three percent of total healthcare spending. All Swedish citizens have strong financial protection from both direct healthcare costs as well as indirect income losses from temporary and permanent work disabilities.

Within a county, each hospital is responsible for providing care to residents within a given geographical catchment area. This means that place of residence largely determines which hospital a patient will be admitted to when seeking care. Furthermore, patients are generally assigned to an on-duty physician on the day of admission. Selection bias arising

[7] In line with this Chang *et al.* (2003) and Mehrotra *et al.* (2012) document that physician's experience affects costs and quality of care, with younger physicians accumulating higher costs without quality improvements.

from endogenous patient-physician sorting is therefore less likely to be a concern in the Swedish healthcare system compared to other contexts.

Coronary angioplasty

Coronary heart disease is generally treated by interventional cardiologists using a catheter-based treatment method called percutaneous coronary intervention (PCI), or coronary angioplasty. In a PCI, the cardiologist first inserts a catheter through either the femoral or radial artery, which is subsequently transported to the site of the blockage using a guide wire. Once the obstructed area is reached, a tiny balloon attached to the catheter is inflated, restoring blood flow by compressing atherosclerotic plaque against the artery wall. To keep the artery open after balloon dilation, the cardiologist may also place and leave a stent in the artery to reinforce the blood vessel's wall and to prevent it from reoccluding.

Two main types of stents are associated with performing a PCI: Bare-Metal Stents (BMS), first approved in 1994 and commonly referred to as first-generation stents, and the newer Drug-Eluting Stents (DES), first approved in Europe in 2002. In contrast to the BMS, the DES was developed to counteract reocclusion of the artery by being coated with drugs that inhibit cell proliferation, thus, significantly reducing the risk of restenosis, the gradual re-narrowing of a coronary artery after a blockage has been treated. Although the DES represents a major medical advance for angioplasty over the BMS, it has also been associated with the more severe side effect of stent thrombosis (ST) – the formation of blood clots in the blood vessels caused by the stent itself. Choosing between a BMS and a DES when performing angioplasty is therefore not trivial as the physician must weigh the risk and impact of one possible side effect against another. Furthermore, clinical guidelines for choosing the appropriate stent were unavailable for most of the time period we study in our empirical analysis, leaving the decision largely up to the individual physician's preferences.[8]

On the other hand, since the mode of treatment, PCI, is identical irrespective of the type of stent used, the choice of stent is not determined by external operational characteristics, such as treatment costs or the physical attributes of the clinician. Our context, therefore, suggests a close to ideal setting for studying how physicians' treatments preferences vary with their environment, since, in the absence of clear clinical guidance, observed choices are likely to be mainly a function of the physician's personal beliefs with respect to efficacy of the treatment.[9] These beliefs may be shaped and influenced by the degree and intensity of the physician's interactions with their workplace peers, which we explore in this paper.

[8] Dozens of medical journal articles were published on the topic each year, favoring one approach or the other (see, e.g., Daemen *et al.*, 2007; Daemen and Serruys, 2007; Kastrati *et al.*, 2007; Vlaar *et al.*, 2007), suggesting that the choice between BMS and DES belonged to the gray zone area of clinical practice during this time.

[9] This is a strength of our context insofar that in many other clinical settings where multiple treatment options are available, such as, for example, open heart surgery (CABG) vs PCI, other factors related to the mode of treatment play a crucial role for treatment choice, some of which a physician has very limited control over (e.g., physician education and training, physical attributes such as motorics and visual acuity, cost of treatment and the existence of an operating theater). Analyzing physician treatment choice in such scenarios would provide further limitations on what we could conclude about physician preferences.

4.3 Econometric Framework

In this section we describe our empirical approach for quantifying the effect of the environment on physician treatment styles. We first define how we measure physician exposure to their practice environment and how the environment can be partitioned into a hospital-specific and a peer group-specific component. We next describe our empirical model and how physicians' responses to a change in their practice environment can be identified and estimated using empirical variation from migrating cardiologists.

Definition of physician practice environment

The practice environment a physician is exposed to is a latent variable in the sense that it is not directly observable or quantifiable. A challenge is, therefore, to define and construct a variable that captures the relevant features of the practice environment for our purposes. Following the methodology taken from Molitor (2018) and adapted to our setting, we characterize cardiologist $j \in J$'s practice environment in hospital $h \in H$, where patient $i \in N_{ht}$ received a PCI in time period $t \in T$, as the ratio

$$E_{jht} = \frac{\sum_{i \in N_{kht}} 1(DES_i = 1)}{N_{kht}} \quad \forall\, k \neq j \in J, \tag{4.1}$$

where $N_{kht} \subset N_{ht}$ is the subset of patients *not* treated by cardiologist j. Hence, E_{jht} is cardiologist j's exposure to the practice environment with respect to the rate of DES use among eligible patients in hospital h and time t. Next, we define the *difference* in practice environments between a migrating cardiologist's origin (h_{O_j}) and destination (h_{D_j}) hospital at a given point in time as

$$\Delta_{jt} = E_{jh_{D_j}t} - E_{jh_{O_j}t}. \tag{4.2}$$

In other words, Δ_{jt} is the period-specific difference in DES leave-out shares between the hospital that cardiologist j practiced in before and after relocation, respectively. Note that this setup provides an intuitive framework for defining counterfactual treatment states of migrating physicians that we will use to motivate our empirical approach below.

Equations (4.1) and (4.2) constitute the basic framework for quantifying physicians' exposure to their practice environment over time and across hospitals. We now extend this framework by partitioning the overall practice environment into two separate dimensions: a physical (hospital-specific) and a social (peer group-specific) component, respectively. Conceptually, we can think of a physician's practice environment as a combination of physical (i.e, hospital infrastructure, technology, assets and resources) and social (i.e., peers, physician networks and coworkers) factors. The former component may be less relevant from a behavioral point of view, since physician responses to the availability of physical resources are not directly related to his or her preferences for a particular treatment.[10] On

[10] This is not to say that the hospital-specific environment does not include *any* preference-related factors, such as, for example, hospital management cultures. The argument here is that such factors are assumed to be fixed within the specific hospital in contrast to social factors that vary on the individual physician level.

the other hand, social interactions may be highly influential in forming and developing physician preferences for treatments and beliefs in their efficacy. Studying the net as well as the relative impact of these components in their capacity to alter physician practice styles is therefore important; theoretically, in terms of understanding the anatomy of physician decision-making; and in practice, to provide a basis for policy to enhance the effectiveness of healthcare delivery.

To empirically disentangle hospital- and peer group-specific components in physician practice environment, we postulate that cardiologists who are working in the same hospital on the same day form a relevant peer group from which we can draw inference.[11] Formally, let

$$P_{k_jht} = \frac{\sum_{i \in N_{k_jht}} 1(DES_i = 1)}{N_{k_jht}} \quad \forall\, k_j \neq j \in K_j \tag{4.3}$$

be the average DES share used by cardiologist j's *peers* k_j in hospital h and period t. Cardiologist j's peers are defined as all other K_j cardiologists who performed PCI on patients in the same hospital and at the same point in time as cardiologist j. We use this within-hospital variation to define and estimate physician j's *peer exposure* in time period t by the relation

$$E^P_{jht} = \sum_{k_j \in K_j} \sum_{d_t \in D_t} 1(d_{t_j} = 1, d_{t_{k_j}} = 1) \times P_{k_jht}, \tag{4.4}$$

where $d_t \in D_t$ is the specific calendar date *within* period t, and d_{t_j} and $d_{t_{k_j}}$ are indicator variables for whether physicians j and k_j were both treating patients on day d_t. In other words, E^P_{jht} is a weighted average of the overall practice environment of hospital h in time period t, with weights defined by the correspondence between cardiologist j and each of their peers with respect to the days they both performed PCI on admitted patients. Note that giving all K_j peers the same weight in Equation (4.3) would return E_{jht} from Equation (4.1).

The difference in peer practice environment between a migrating cardiologist's origin and destination hospitals, Δ^P_{jt}, is correspondingly defined by replacing E with E^P in Equation (4.2). The counterfactual practice environment (i.e., the environment in the hospital cardiologist j is not currently working in) is simply defined as the potential peer exposure derived from all cardiologists who worked in the counterfactual hospital over that period,

$$\Delta^P_{jt} = E^P_{jh_{D_j}t} - E^P_{jh_{O_j}t}. \tag{4.5}$$

The total variation in the hospital's practice environment is equal to the sum of the within- and the between-components, implying that we can decompose physician j's overall practice environment as

$$E_{jht} = E^P_{jht} + E^H_{ht}, \tag{4.6}$$

[11] This definition makes intuitive sense, as individuals who work together are able to observe and directly influence each other. It is also supported by the economic literature on peer effects in the workplace (see, e.g., Falk and Ichino, 2006; Mas and Moretti, 2009).

where E^H_{ht} is equal to the hospital-specific component, varying only across hospitals and time, and E^P_{jht} as the peer group-specific component, varying across cardiologists within hospitals over time. It follows that the total change in a migrating physician's practice environment can be decomposed as the sum of the *changes* in both environments,

$$\Delta_{jt} = \Delta^P_{jt} + \Delta^H_{jt}, \tag{4.7}$$

where $\Delta^H_{jt} = E^H_{h_{D_j}t} - E^H_{h_{O_j}t}$ is the equivalent of Δ^P_{jt} for hospitals (i.e., the change in the hospital-specific environment across a migrating cardiologist's origin and destination hospitals). Thus, the total impact of the change in environment of a migrating cardiologist at a given point in time consists of a physician-specific and a hospital-specific effect. Our approach to empirically disentangle these two effects is described in the following subsection.

Empirical model

The point of departure for our empirical modeling is based on the method in Molitor (2018) who uses longitudinal administrative data on cardiologists moving across hospitals to obtain empirical variation in physician practice environment. This variation is used to estimate causal effects of changes in the migrating physicians' practice environment on their own treatment choices in a difference-in-differences (DD) empirical design. The idea is simple yet intuitive: if physicians' practice styles are malleable to the environment they operate in, then we would expect to observe patients managed by migrating physicians to receive treatments more aligned with the practice environment in the destination hospital after, but not prior to, their relocation.

Formally, the patient-level DD model for patient $i \in N$, treated by cardiologist $j \in J$ at time $t \in T$ can be described by the equation

$$y_{ijt} = \alpha Post_t + \beta \Delta_{jt} + \gamma(\Delta_{jt} \times Post_t) + X'_{ijt}\Gamma + \lambda_j + \lambda_t + \epsilon_{ijt}. \tag{4.8}$$

The outcome y_{ijt} is defined by a dummy indicator variable equal to one if a patient undergoing PCI received a DES, and equal to zero if a BMS was used. Moreover, $Post_t = 1_{t>t_0}$ is a dummy variable which equals one for all time periods subsequent to cardiologist j's move to a new hospital at time t_0. The model also includes controls for cardiologist, λ_j, and time, λ_t, cluster-specific effects (i.e., $\sum_z \theta_z 1_{\lambda_{z'=z}}$ for $z = j, t$) and a vector of potentially time-varying observable patient, hospital and cardiologist characteristics, X_{ijt}, to adjust for observed and unobserved heterogeneity across patients, physicians and time. Finally, Δ_{jt}, defined in Equation (4.7), is a continuous variable with range $[-1, 1]$, characterized as the difference in physician j's practice environment between the *origin* (pre-migration) and *destination* (post-migration) hospitals with respect to the share of DES used in patients undergoing PCI at time t.

The main parameter of interest in Equation (4.8) is γ, which, under standard identifying assumptions of the DD estimator, captures the average physicians' response in their DES use to the difference in practice environments between the origin and destination hospitals after, relative to before, their relocation. Defining practice environment as the hospital's

risk-adjusted share of DES used on patients undergoing PCI, γ can be interpreted as the percentage change in physician j's *own* DES practice style for each percentage point difference in practice style *environment*. We refer to Equation (4.8) as our baseline model to provide a link and compare the results in Molitor (2018) to our decomposition approach described below.

To study the dynamic pattern of the migrating cardiologists' responses to their practice environment and test the common trend assumption, we extend our baseline model in Equation (4.8) by replacing $Post_t$ with a set of period-specific indicators

$$y_{ijt} = \beta\Delta_{jt} + \sum_{s|s\notin 0=-T'}^{T'} 1(s = t')\left(\alpha_{t'} + \gamma_{t'}\Delta_{jt'}\right) + X_{ijt}\Gamma + \lambda_j + \lambda_t + \epsilon_{ijt}, \tag{4.9}$$

where $t' = t - t_0 \in \{-T', \ldots, T'\}$ is the period-specific index recentered around the time of the cardiologist's move, t_0, where the latter is omitted from the analysis as indicated in Equation (4.9). This modification allows us to interpret the average period-specific cardiologist responses by time from their move on a common time index that can be plotted in an event-study fashion.[12]

Effect decomposition and quality of care

Our approach to identify physician responses to their practice environment relies on empirical variation derived from cardiologists moving across hospitals at different points in time. Whenever this happens, we maintain that they are exposed to a combined shift in practice environment arising from two sources: a hospital-specific, Δ^H_{jt}, and a peer group-specific, Δ^P_{jt}, component, as defined in Equation (4.7). To empirically disentangle these two effects, we make use of the fact that the former component is assumed to be constant within a hospital. Therefore, the additional inclusion of hospital-specific effects, λ_h, in Equations (4.8) and (4.9) will effectively purge the practice environment of the hospital-specific component and any remaining variation will hence be attributed to the peer effect, Δ^P_{jt}. Thus, we estimate Equations (4.8) and (4.9) with and without hospital fixed effects for our sample of movers and attribute the estimated γ without hospital fixed effects as the total impact of the practice environment (i.e., both hospital and peer). In contrast, the estimated effect with hospital fixed effects will be attributed to the peer group-specific effect component. Finally, the relative difference between these two effects as a share of the net effect is interpreted as the hospital-specific effect.[13]

So far our model framework has focused on changes in the practice styles of cardiologists induced by their practice environment. However, we are also interested in knowing whether any environmentally induced behavioral changes of physicians translate into

[12] Note that the event study also allows us to study the validity of the common trend assumption imposed on the DD model in Equation (4.8).

[13] It is possible that the hospital fixed effects do not capture the full range of dynamics in cardiologist responses since the estimated average response may conceal significant heterogeneity across migrants. Therefore, in Section 4.5 we also estimate split-sample models to study heterogeneity in cardiologist response by direction of the move (to more or less DES intensive hospitals) and by cardiologist experience (age in years).

changes in the quality of care received by patients who were treated by the migrating cardiologists. In particular, knowing *how* these behavioral changes affect the appropriateness of the treatment and patient health outcomes would provide useful information on whether and to which extent physician adaptation to their practice environment improved or worsened quality of healthcare services. To this end, we consider two additional sets of outcomes within our regression framework: physician decision errors and patient health outcomes. The latter category is based on a composite measure of relevant post-intervention adverse clinical events, including death, myocardial infarction and restenosis requiring a new intervention. The former outcome category is based on defining a measure of stent appropriateness using an auxiliary sample from which we employ a classification exercise based on machine learning techniques. We defer the details of this approach to the next section.

4.4 Data

We use data from the Swedish Coronary Angiography and Angioplasty Registry (SCAAR) for our empirical analyses.[14] Since 1998, SCAAR registers cardiac catheterization procedures performed in Swedish hospitals, including detailed clinical information on patient health status and comorbidities (e.g., diabetes mellitus, smoking status and BMI), angiography diagnostic results (e.g., location and severity of blockage by coronary artery segment) and relevant treatment outcomes (e.g., complications and adverse clinical events such as myocardial infarction or death). Importantly, the register also includes information on the treating hospital and responsible physician, performed procedure(s) and the time and dates of intervention, hospital admission and discharge.

Analysis sample

We initially sample all instances of PCI performed in Swedish hospitals and reported in SCAAR between 2004 and 2013.[15] To clearly identify our main outcome variable, the cardiologist's choice between using a DES and a BMS in the procedure, we drop patients who received multiple stents in the same treatment session from the sample. This restriction leaves us with a total of 51,381 PCI cases performed by 199 cardiologists in 28 hospitals.

The data include daily information on each cardiologist's angioplasty treatments and the hospital the activity takes place in. We use this information to define physician practice

[14] SCAAR is maintained by the Uppsala Clinical Research Center (UCR), sponsored by the Swedish Health Authorities and independent of commercial funding. Reporting in the SCAAR is Internet-based. The data are recorded online through a Web interface in the cardiac catheter laboratory, encrypted and sent to the UCR central server. Each hospital receives a feedback on the processes and quality of care measures. To monitor and maintain quality, a continuous screening process of the registry data is in place, operating by comparing 50 entered variables in 20 randomly selected interventions per hospital-year with the patients' hospital records. The overall correspondence in data during the study period is 95.2%.

[15] We restrict the starting year of our analysis to 2004 as this is the first year all hospital in Sweden that performed PCI contributed to the registry. The endpoint is chosen because the market for stents included additional options from 2013 onward due to the entry of a new second-generation DES and the corresponding sharp decline in the use of the BMS.

episodes by indicating the first and the last date a cardiologist practiced in a particular hospital. This method defines an origin and a destination hospital and a specific time-stamp for when the move took place. We only consider non-overlapping hospital episodes in our analysis. As a few cardiologists may occasionally practice in several hospitals, we classify physician practice episodes to hospitals where the cardiologist continuously treated patients over a period of at least six months. In addition, we exclude the quarter a physician transferred to the destination hospital from our analyses due to occasional overlap in practice episodes across hospitals.[16] With these restrictions we identify 51 migrating cardiologists treating 8,589 patients across 25 hospitals over the analysis period. Remaining cardiologists, who were based at the same hospital throughout the analysis period, are referred to as non-migrating cardiologists.

Columns (1) and (2) of Table 4.1 present means and standard deviations for our analysis sample of migrating cardiologists while columns (3) and (4) present corresponding figures for non-migrating physicians for comparison. The upper, middle and lower panels of the table partition this information into hospital-, cardiologist- and patient-specific characteristics, respectively. With respect to hospital characteristics, we observe no major differences across the two groups other than that non-moving cardiologists seem to work in moderately larger hospitals in terms of annual case volume. With respect to the characteristics of the cardiologists themselves, migrants tend to be somewhat younger and more likely to have a specialization in interventional cardiology (in contrast to, e.g., radiology or surgery).[17] Patient case-mix is remarkably similar in all aspects across the groups on average, although migrating cardiologists appear to be somewhat less prone to use DES. However, there are no differences in terms of patient health outcomes between migrants and non-migrants.

Decision errors and patient health outcomes

To study the impact of migrating cardiologists' changes in practice environment on quality of care, we replace our main outcome variable from Equations (4.8) and (4.9) with two sets of outcomes proxying for the appropriateness of the chosen treatment and for any adverse patient health consequence of such choices. We first define a dummy indicator variable for whether the treatment decision was the appropriate choice based on a risk-adjusted measure of treatment suitability and classified using a machine learning (ML) method for classification. To this end, we employ the Random Forest (RF) algorithm which has been demonstrated to have improved prediction accuracy in comparison with other supervised learning methods (Breiman, 2001; Svetnik *et al.*, 2003). Our method is essentially a simplified version of the method Mullainathan and Obermeyer (2022) use to identify and study physician diagnostic errors in heart attack treatments.

[16] We also exclude a few cases where a cardiologist continuously practices in several hospitals over an extended time period (e.g., Karolinska hospital in Solna and Huddinge in Stockholm county and Lund and Malmö hospital in Skåne county).

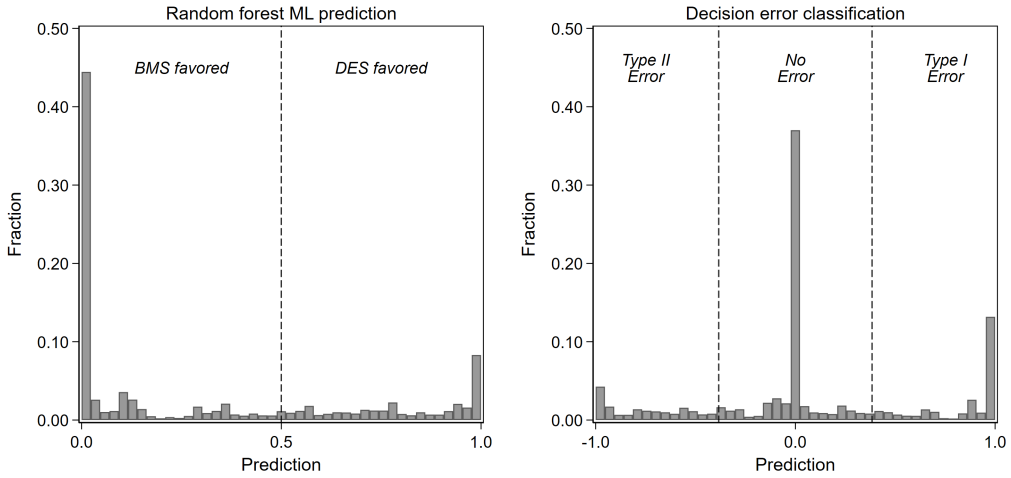
[17] All physicians in our sample are internists. Most physicians in our dataset also have a subspecialty in interventional cardiology, but there is also a small share who have a subspecialty in radiology and some who have not yet completed their cardiologist training.

Table 4.1: Descriptive sample statistics

	Moving cardiologists		Non-moving cardiologists	
	Mean	SD	Mean	SD
Hospital characteristics				
Teaching hospital	0.38	0.49	0.41	0.49
RiksHIA quality index	3.73	1.95	3.84	1.95
Case volume	7,861	7,349	8,912	7,468
Hospitals	25		28	
Cardiologist characteristics				
Male	0.93	0.25	0.90	0.30
Age	46.59	6.45	49.00	7.20
Patient volume	466.35	212.89	623.40	300.60
Specialization in interventional cardiology	0.85	0.35	0.70	0.46
Total error rate	0.40	0.05	0.39	0.07
Type I error rate	0.14	0.06	0.15	0.08
Type II error rate	0.26	0.08	0.24	0.10
Cardiologists	51		148	
Patient characteristics				
Risk factors				
Male	0.73	0.45	0.72	0.45
Age	65.81	10.94	66.00	11.11
Smoker	0.79	0.79	0.82	0.79
Diabetes	0.17	0.37	0.17	0.37
Chronic obstructive pulmonary disease	0.01	0.11	0.02	0.12
Peripheral vascular disease	0.00	0.05	0.00	0.07
Hypertension	0.49	0.50	0.50	0.50
Previous infarction	0.20	0.40	0.18	0.39
Previous CABG	0.09	0.28	0.08	0.27
Previous PCI	0.11	0.31	0.10	0.30
Outcomes				
DES treatment	0.36	0.48	0.42	0.49
Death (1 year)	0.04	0.19	0.04	0.19
MI (1 year)	0.07	0.26	0.07	0.26
TLR (1 year)	0.06	0.24	0.06	0.23
Total error rate	0.40	0.49	0.37	0.48
Type I error rate	0.19	0.39	0.20	0.40
Type II error rate	0.21	0.41	0.17	0.38
Cases	8,589		51,381	

SCAAR data for years 2004–2013. Means and standard deviations for samples of moving and non-moving cardiologists. Patient characteristics are missing for a subset of observations: gender (28 cases), smoking (4,893 cases), diabetes (680 cases), hypertension (1,535 cases), previous infarction (1,724 cases), previous CABG (158 cases), previous PCI (168 cases); and cardiologist characteristics: age (739 cases); specialization (692 cases); and hospital characteristics: RiksHIA quality index (693 cases). All observations with missing characteristics are included in the analysis by defining dummy variables for the missing categories.

Figure 4.1: Random forest ML prediction and decision error classification distributions in the analysis sample



NOTE.— SCAAR data for years 2004–2013. Left panel presents distribution of predictions of appropriate DES use in angioplasty treatments from estimation of the random forest (RF) machine learning algorithm explained in Section 4.4. Predictions are based on an auxiliary sample of non-moving cardiologists in university hospitals 2008-2011. See also Breiman (2001); Svetnik *et al.* (2003). Right panel shows corresponding decision errors by comparing migrating cardiologists' actual choices to RF classifications. Vertical lines correspond to thresholds for classification into Type I (false positive; application of DES when BMS is correct) and Type II (false negative; application of BMS when DES is correct) errors.

We assess the appropriateness of cardiologists' stent choices by relating actual physician choices to the predicted "appropriate" choices derived from the RF algorithm using auxiliary data based on angioplasty procedures performed by non-migrating cardiologists in Swedish academic hospitals. We predict the appropriate stent choice for each case in our analysis sample and define a dummy variable for overall error, equal to one whenever the observed choice does not match the predicted choice irrespective of the choice of stent. Figure 4.1 shows the distribution of predicted probabilities (left panel) and respective error rates (right panel).[18]

We furthermore decompose the overall decision error into Type I and Type II errors under the null hypothesis that the BMS is the appropriate treatment choice. To this end, a Type I error (i.e., a false positive) pertains to incorrectly inserting a DES when a BMS is suitable and a Type II error (i.e., a false negative) is defined by inserting a BMS when a DES was the correct option. This decomposition may provide additional insights into the consequences of inappropriate treatment choices since incorrect use of the DES potentially put patients at risk of more severe adverse events, such as ST, and higher treatment costs, since the DES is more expensive than the BMS (although the stent itself only constitutes a

[18] There exist many alternative ML algorithms which could be used to implement the method we use in this paper. For example, Ribers and Ullrich (2022) use the extreme gradient boosting algorithm (XGBoost) to study whether machine learning predictions can improve antibiotic prescribing efficiency and Mullainathan and Obermeyer (2022) use an ensemble method by combining LASSO and gradient boosting. The main difference between RF and gradient boosting is that the latter method builds decision trees one at a time while trees are fully independent in the former method. Gradient boosting methods are typically more accurate than RF methods, but also prone to overfitting if the data contain substantial statistical noise. See section 4.9 in Appendix C for further details and diagnostics on the RF algorithm we use in the analysis.

minor part of the total cost of treatment).[19] Table 4.2 presents a matrix of the cardiologists' treatment decisions in our sample and corresponding error rates.

Table 4.2: Treatment decision matrix and error rates

	Predicted: BMS	Predicted: DES	Error rate
Treated: BMS	3,875	1,755	0.31
Treated: DES	1,567	1,388	0.53

SCAAR data for years 2004-2013. Predicted treatments are classified according to predictions from estimation of the random forest (RF) machine learning algorithm explained in Section 4.4. Predictions are based on an auxiliary sample of non-moving cardiologists in university hospitals 2008-2011. Error rates are defined as the share of chosen non-recommended treatments among all treatments using the specific stent type. See also Figure 4.1.

Finally, we include a set of patient outcomes based on the prevalence of one-year post-intervention adverse clinical events, including patient death, myocardial infarction (MI), and total leison revascularization (TLR) to our regression model. The bottom panel of Table 4.1 shows the rates of these events in our analysis sample.

Estimation of physician practice environment

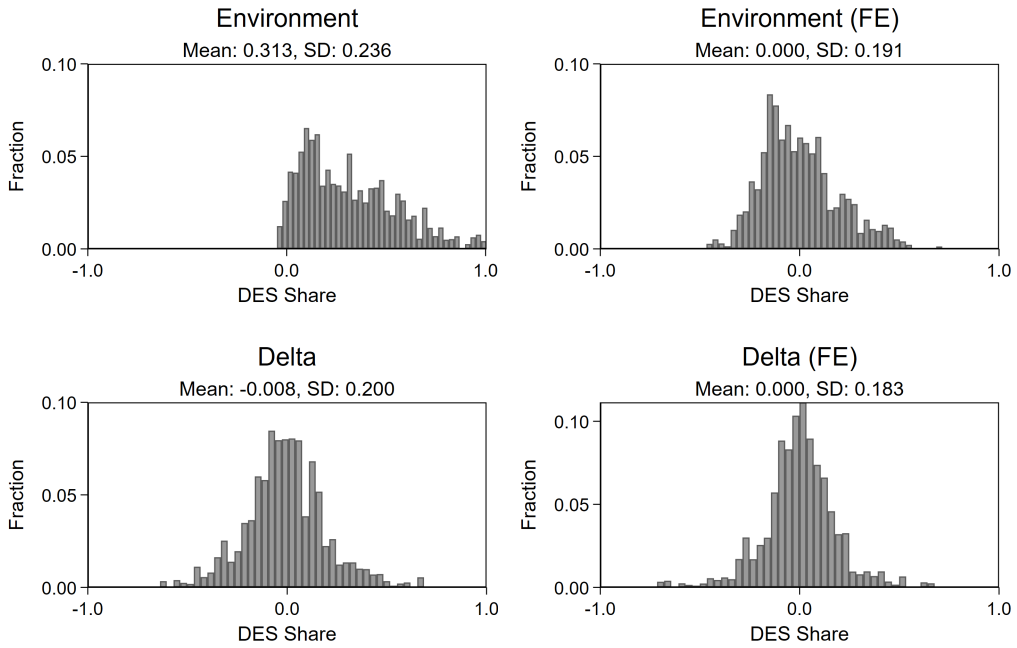
Since both the absolute number and the case-mix of patients treated by cardiologists may vary substantially, we modify each cardiologist's use of DES using the Empirical Bayes (EB) method. To this end, we estimate a mixed-effects model with both fixed (case-mix risk-adjustment) and random (shrinking imprecise physician DES shares to the population mean) elements to correct for potentially biased estimates of the physicians' practice environment as well as any existing risk selection between cardiologists and their patients (see, e.g., Rabe-Hesketh and Skrondal, 2008).

The distribution of the EB-adjusted practice environment across all migrating cardiologists and periods in our sample is shown in the upper left panel of Figure 4.2. The variation is large, covering almost the full range of the variable, and slightly skewed to the right with a mean of 0.31. The corresponding distribution after regression adjustment for hospital fixed effects (i.e., the within-hospital variation) is visualized in the upper right panel of the same figure. There is substantial variation remaining even after the hospital-specific component has been eliminated from the environment, suggesting that including hospital-specific effects is unlikely to generate problems of model overfitting.[20] The lower

[19] Another interesting analysis would be to study whether the decision errors of a migrating cardiologist's peers impact his or her own performance. This could be evaluated by simply replacing the environmental variable by the average decision error among peers in the origin and destination hospitals for each migrant and using the decision error dummy as outcome in the regression model. Unfortunately, while constructing this analysis we realized that the variation in decision errors across hospitals in Sweden is too small to provide reliable inference for answering this question empirically.

[20] The distribution of the risk-adjusted DES rates across the 21 county councils in Sweden is displayed in Figure 4.A1 of 4.7.

Figure 4.2: Distributions of migrating cardiologists' practice environments

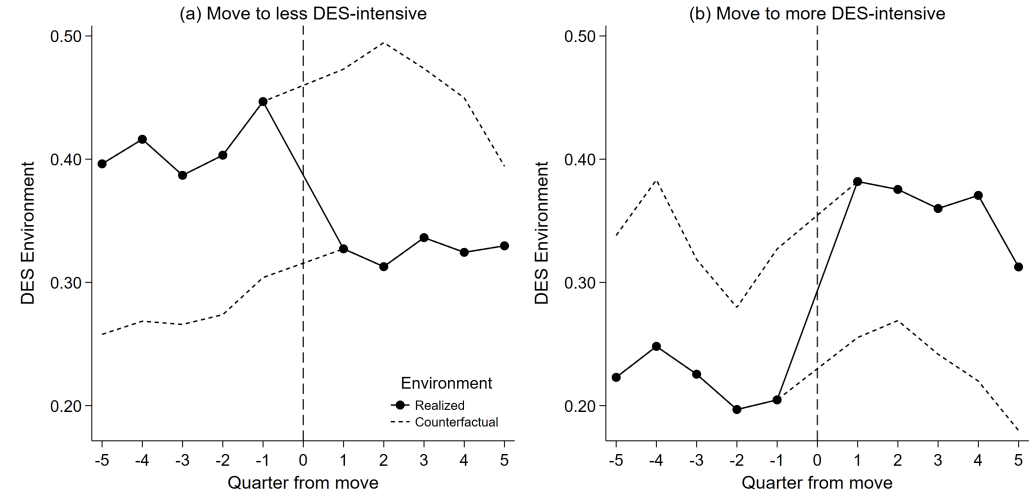


NOTE.— SCAAR data for years 2004–2013. Upper panels pertain to physicians' practice environment prior to relocation without (left panel) and with (right panel) adjustment for hospital fixed effects. Lower panels show corresponding distributions for the difference in practice environment between migrating cardiologists' origin and destination hospitals, Δ_{jt}

left and right panels of Figure 4.2 show corresponding distributions of Δ_{jt} with and without hospital-specific fixed effects, respectively. The change in practice environment among migrating cardiologists in our sample is symmetrically distributed across higher and lower shares of DES. Hence, our empirical approach is able to capture a wide range of treatment effects in both the positive and negative domains of changes in the physicians' practice environment.

Figure 4.3 provides a graphical illustration of the intuition behind the identification approach we use in our empirical analysis. The solid lines indicate the average practice style environment, measured by the average quarterly share of DES used among migrating cardiologists' peers, by time from their relocation. To avoid canceling out positive and negative changes in the practice environment, physicians moving from more to less DES-intensive environments and from less to more DES-intensive environments are plotted in the left and right panels of the figure, respectively. Moreover, the dashed lines show the corresponding estimated *counterfactual* environment in the hospitals associated with the migrating cardiologists: the destination hospital, prior to the relocation, and the origin hospital, after the relocation took place. At any point in time, the vertical difference between the two lines is computationally equivalent to the average difference in physician practice environments, Δ_{jt}, averaged over all J migrating cardiologists.

Figure 4.3: Average trends in migrating cardiologists' practice environments



NOTE.— SCAAR data for years 2004–2013. Practice environment defined as the share of DES used in angioplasty treatments in realized (solid lines) and counterfactual (dashed lines) hospitals by quarter from the cardiologist's move. Separate plots for cardiologists moving to hospitals with lower and higher intensity of DES use. Vertical dashed line indicates recentered quarter of physician relocation from the origin to the destination hospital. Quarter of move linearly interpolated.

The figure shows that there are significant jumps in the practice environment for both groups of migrating cardiologists at the time of relocation when the actual and the counterfactual environments are switched. The quarter of the move has been interpolated in the graph (and omitted from our analysis), since the cardiologist may treat patients in both the origin and destination hospitals during this period. The counterfactual environment can hence be interpreted as an estimate of the hypothetical environment that would have prevailed if the migrating physician would not have relocated. We can use this estimate to derive and evaluate the common trend assumption when estimating our DD model. In particular, if migrants react to the counterfactual environment prior to their move, we would conclude that our empirical approach is invalid. We study this in detail in the next section.[21]

4.5 Results

This section reports results from estimation of the econometric models described in Section 4.3 using our analysis sample explained in Section 4.4. We first provide main results obtained from estimation of our DD model on migrating cardiologists' responses to a change in their practice environment with respect to their use of DES when performing PCI. Next, we investigate the extent to which these responses improved or worsened the appropriateness of physicians' treatment choices and whether they were associated with significant changes in patient health outcomes and costs of treatment. Finally, we provide results from a set of robustness checks and heterogeneity analyses to evaluate the stability of our inference with respect to model specification, interpretation and variable definitions.

Do physicians adapt to their practice environment?

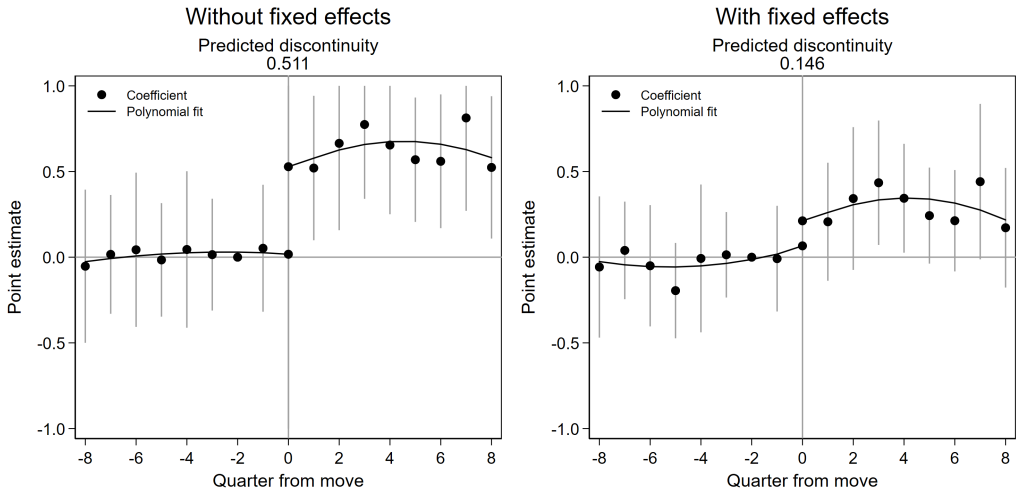
The left and right panels of Figure 4.4 display estimation results from the event study model in Equation (4.9) without and with hospital fixed effects. Each dot in the figure refers to an estimated $\gamma_{t'}$ parameter and the associated vertical spikes indicate corresponding 95 percent confidence bands. The solid vertical line in each panel pertains to the specific recentered year-quarter of cardiologists' move from the origin to the destination hospital. The quarter of relocation is omitted from the analysis and replaced with the predicted value based on a cubic polynomial, indicated by the solid line, and estimated separately for quarters before and after the move. The predicted discontinuity at the quarter of move is reported in the panel header. To ensure sufficient number of leads and lags while simultaneously keeping the panel of migrating cardiologists balanced, we follow the migrating cardiologist for eight quarters before and after the move. As the estimated parameters are only identified up to scale, we use the quarter prior to the move normalized to zero as baseline.

[21] We have also used an alternative approach to estimate the practice environment by applying a synthetic control method (see, e.g., Abadie and Gardeazabal, 2003; Abadie *et al.*, 2010, 2015) to gauge the stability of our results with respect to potential sorting of physicians and their peers. See Section 4.5 for details.

The estimated parameters prior to the physician's relocation are not significantly distinguishable from zero (i.e., the baseline period), suggesting that migrating physicians did not systematically respond to the counterfactual practice environment prior to their move. Moreover, for the model without hospital fixed effects, there is a visible sharp discontinuity occurring at the time of cardiologist relocation where the estimated $\gamma_{t'}$ coefficients become positive and highly significant. The estimated magnitude of this discontinuity is around 0.51. Interestingly, the cardiologists appear to rapidly and permanently adapt to the prevailing practice style at the destination hospital for the entire duration of the follow-up period.

The corresponding period-specific effect pattern in the right figure panel, where hospital fixed effects have been added to the model, describes a more muted, but still pronounced, change in the moving cardiologist's behavior at the time of relocation. In this case, we observe a somewhat more gradual adaptation to the destination hospital's practice environment over time and that the initial discontinuity at the time of relocation is somewhat smaller. We conclude from this analysis that cardiologists in our sample are partially malleable to their practice environment in terms of their own practice behavior, and that they are responsive to both their social and their physical environments.

Figure 4.4: Event study estimates of migrating cardiologists' changes in practice environment: Use of DES



NOTE.— SCAAR data for years 2004–2013. Dots correspond to coefficient estimates of $\gamma_{t'}$ from OLS estimation of Equation (4.9). Dependent variable is an indicator for whether a patient undergoing PCI received a DES. Covariates include hospital, cardiologist characteristics and patient risk factors reported in Table 4.1 and fixed effects for year-quarter, cardiologist, and hospital (right panel only). Vertical spikes pertain to robust 95 percent confidence intervals clustered by hospital.

Columns (1)–(4) of Table 4.3 report results from estimation of different models using our sample of migrating cardiologists. Column (1) provides corresponding coefficient estimates from the model used in Molitor (2018) to estimate the response of migrating cardiologists to changes in their practice environment. Our reported DD estimate of 0.72, interpreted as the average percentage point change in the physician's own practice style for each percentage point change in the practice environment between the origin and destination hospitals after relocation, is very close to the estimate of 0.67 found in Molitor (2018). Moreover, the coefficient of Δ_{jt}, interpreted as migrating physicians' average response to the destination hospital's practice environment *prior* to the move, is insignificant. This result supports our maintained common trend assumption that migrating cardiologists do not systematically change their own practice style in response to the destination hospital's practice environment before they relocate.

Next, Columns (2) and (3) show estimation results from our baseline DD model, defined in Equation (4.8), by successive inclusion of control variables. While the results from Column (2), in which only the control variables listed in Table 4.1 have been added, suggest a marginally significant response to Δ_{jt} prior to the move, this coefficient is once again insignificant after further adjustment for period-specific and cardiologist-specific effects in Column (3). The DD point estimates for these model specifications suggest a somewhat smaller physician response of between 0.49 and 0.52 and almost identical to the estimated discontinuity displayed in the left panel of Figure 4.4. In other words, about half of the migrating cardiologists' DES use can be attributed to their overall practice environment for our sample.

Finally, in Column (4) we decompose the overall effect from the change in practice environment by including hospital fixed effects in our regression model. Recall that migrating cardiologists face both a change in the hospital-specific and the peer group-specific practice environment when they move across hospitals. Assuming that the hospital-specific component is constant within a hospital, whereas the peer group-specific component varies within hospitals, we include hospital fixed effects to eliminate the impact of the former from the practice environment variable. This adjustment reduces the DD estimate by another 50 percent to 0.25. We interpret this result as that the peer group-specific effect is responsible for roughly half of the response in physician practice style. This suggests that physicians' reactions to their practice environment embody both the characteristics of the hospital itself, such as infrastructure, management and resources, as well as the social environment, captured by the physicians' workplace peers.

Table 4.4 reports split-sample results from estimation of our baseline DD model separately for cardiologists moving to hospitals with higher and lower shares of DES (Columns (1) and (2)), and for younger and older migrants, based on the median age of migrating cardiologists (Columns (3) and (4)). We focus on the peer environment by including hospital fixed effects in the model, corresponding to Column (4) of Table 4.3. The motive behind this analysis is to evaluate whether our main results are driven by specific subgroups. We anticipate that relatively younger physicians' practice styles are likely to be more malleable due to their lower practical experience and being in an earlier stage of their careers, consistent with the theory of champions, or opinion leaders, of clinical care (see, e.g., Shortell *et al.*, 2004). Furthermore, it is possible that migrating physicians are more susceptible to adopting treatment styles in practice environments where new medical technology is used

Table 4.3: Difference-in-Differences estimates of migrating cardiologists' changes in practice environment: Use of DES

	(1) DES	(2) DES	(3) DES	(4) DES
Post	-0.003	-0.030	0.014	0.003
	(0.022)	(0.034)	(0.020)	(0.023)
Δ_{jt}	-0.131	-0.253**	-0.164	0.013
	(0.085)	(0.126)	(0.105)	(0.087)
$Post \times \Delta_{jt}$	0.719***	0.485**	0.523***	0.247***
	(0.130)	(0.201)	(0.114)	(0.090)
Covariates		✓	✓	✓
Year FE	✓			
Origin hospital FE	✓			
Year-quarter FE			✓	✓
Cardiologist FE			✓	✓
Hospital FE				✓
Cardiologists	51	51	51	51
Observations	8,589	8,589	8,589	8,589

SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (4.8). Dependent variable is an indicator for whether a patient undergoing PCI received a DES. Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

more intensively, here characterized as a higher share of the relatively newer DES, due to the attractiveness of technology (see, e.g., Hofmann, 2015).

Our predictions align with the empirical evidence reported in Table 4.4 in that the estimated response to the change in practice environment is mainly driven by younger cardiologists and by cardiologists moving to more DES intensive environments. The first two columns suggest that the effect is positive for both positive and negative Δ_{jt}'s (albeit the latter coefficient is not statistically significant), while the last two columns indicate that more senior cardiologists do not respond at all to their peer practice environment when relocating. Thus, heterogeneity in the effect across both physicians and their environments seem to be important to understand clinicians' reactions to their practice environment.

Impact on quality of care

We next study the extent to which the environmentally induced changes in migrating cardiologists' DES use affected the appropriateness of physician treatment choice and their consequences for patients' health outcomes and the costs of treatment. To this end, we estimate versions of Equations (4.8) and (4.9) by replacing our outcome variable with the three indicators for major adverse cardiac events we consider: patient death, myocardial infarction (MI), and total lesion revascularization (TLR) within a year from the initial intervention. Moreover, we compare changes in physicians' rates of decision errors before and

Table 4.4: Difference-in-Differences estimates of migrating cardiologists' changes in practice environment: Heterogeneity analyses

	Environment ±		Physician age	
	(1) $\Delta_{jt} > 0$	(2) $\Delta_{jt} < 0$	(3) Below median	(4) Above median
Post	-0.021	-0.002	0.020	-0.059
	(0.051)	(0.043)	(0.025)	(0.038)
Δ_{jt}	-0.077	0.075	0.161	-0.032
	(0.129)	(0.146)	(0.142)	(0.106)
$Post \times \Delta_{jt}$	0.323**	0.184	0.292*	-0.080
	(0.154)	(0.187)	(0.159)	(0.121)
Covariates	✓	✓	✓	✓
Year-quarter FE	✓	✓	✓	✓
Cardiologist FE	✓	✓	✓	✓
Hospital FE	✓	✓	✓	✓
Cardiologists	24	27	23	28
Observations	3,776	4,813	4,429	4,160

SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (4.8). Dependent variable is an indicator for whether a patient undergoing PCI received a DES. Columns (1) and (2) split the sample into cardiologists moving to more and less DES-intensive hospitals. Columns (3) and (4) split the sample into younger and older cardiologists with median cardiologist age as threshold. Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

after their relocation using predictions from the RF machine learning algorithm to predict optimal treatment choice. Based on the results from these analyses, we conclude by providing a back-of-the-envelope calculation of the excess costs incurred from the inappropriate use of stents as a consequence of the change in practice environment.

Decision errors

Table 4.5 reports DD estimation results using decision errors, based on the correspondence between migrating cardiologists' choices and predictions from our RF machine learning algorithm, as outcomes. Columns (1), (2) and (3) show the estimates on the overall propensity to make inappropriate decisions, and for Type I and Type II errors, respectively. Recall that Type I errors (false positives) refer to the application of DES when BMS is the recommended treatment choice, and vice versa for Type II errors (false negatives). This distinction is relevant as it is possible that making errors of the former type may be subject to more severe risks for the patient due to the possibility of ST and higher medical expenses due to increased unit costs of DES. In contrast, the latter error type may be associated with higher total treatment costs in the form of a higher prevalence of restenosis and the consequential need for subsequent intervention.

The results from estimation show that the overall probability of making a treatment error is positive and significant after, relative to before, cardiologist relocation. The parameter estimate of 0.2 suggests that a migrating cardiologist inserts two percentage points stents incorrectly for each ten percentage points' change in peer practice environment between the origin and destination hospitals after relocation. Splitting the decision errors into Type I and Type II errors, we find that physicians are significantly more likely to make Type I errors after their change in practice environment. In contrast, the risk of committing Type II errors is slightly reduced, but not significantly so. Hence, this result suggests that migrating cardiologists are more likely to overuse DES when they move to a hospital with higher use of DES than they are to overuse BMS when moving to a hospital with lower use of DES.

Patient health outcomes

Columns (2)-(4) of Table 4.6 report results from estimation of Equation (4.8) for the three adverse patient health outcomes we consider: patient death, myocardial infarction (MI), and total lesion revascularization (TLR) within a year from the initial intervention. For comparison, the first column of the table reproduces the results from our preferred specification in Column (4) of Table 4.3. Each column corresponds to a specific outcome for our model with hospital fixed effects, implying that the reported point estimates refer to physician responses to the change in their peer environment. As before, the reported parameter estimates are interpreted as the rate of change in the outcome from a one percentage point change in the physicians' practice environment between the origin and destination hospitals. A negative sign implies that the risk of the event is less likely, whereas a positive coefficient indicates a higher risk.

Table 4.5: Difference-in-Differences estimates of migrating cardiologists' changes in practice environment: Decision errors

	(1) Error	(2) Type I	(3) Type II
Post	0.019	-0.008	0.028
	(0.030)	(0.019)	(0.025)
Δ_{jt}	-0.027	0.000	-0.027
	(0.108)	(0.077)	(0.087)
$Post \times \Delta_{jt}$	0.197*	0.242***	-0.045
	(0.099)	(0.080)	(0.079)
Covariates	✓	✓	✓
Year-quarter FE	✓	✓	✓
Cardiologist FE	✓	✓	✓
Hospital FE	✓	✓	✓
Cardiologists	51	51	51
Observations	8,589	8,589	8,589

SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (4.8). Dependent variables are indicators for whether a patient undergoing PCI received a non-recommended stent type. See Section 4.4 for details. Column (1) reports results for the propensity to commit any error while Columns (2) and (3) report error decomposition results for false positives and false negatives, respectively. Type I errors (false positives) refer to the application of DES when BMS is the recommended treatment choice, and Type II errors refer to the application of BMS when DES is the recommended treatment choice (false negatives). Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The reported parameter estimates suggest that rates of changes in patient outcomes are generally small and statistically indistinguishable from zero. The point estimate of 0.04 for MI is greatest in magnitude, but is only one-sixth of the response for the choice of stent. We interpret this finding as indicating that patient health outcomes are not systematically related to migrating physicians' adaptation to their peer practice environment. One possible reason for this result could be that the estimated changes in the cardiologists' use of DES after relocation were mainly based on low-risk patients for which the choice between a BMS and a DES was unlikely to put patients at serious health risks.

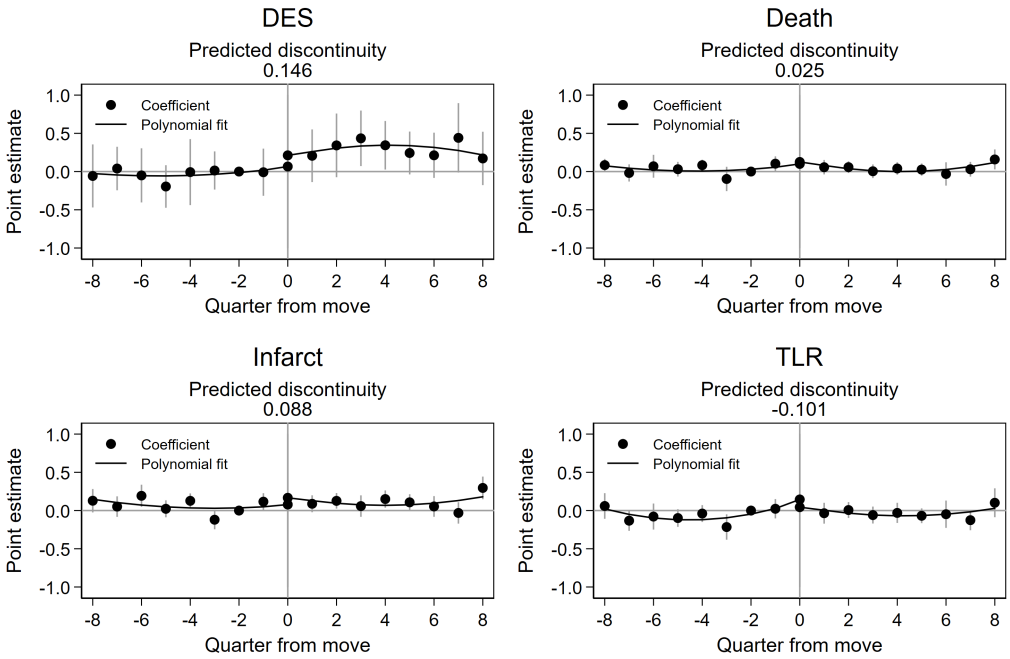
Table 4.6: Difference-in-Differences estimates of migrating cardiologists' changes in practice environment: Patient outcomes

	(1) DES	(2) Death	(3) Infarct	(4) TLR
Post	0.003	-0.009	0.001	-0.009
	(0.023)	(0.008)	(0.011)	(0.011)
Δ_{jt}	0.013	-0.047	-0.069*	-0.053
	(0.087)	(0.030)	(0.037)	(0.033)
$Post \times \Delta_{jt}$	0.247***	-0.011	0.041	0.028
	(0.090)	(0.027)	(0.042)	(0.033)
Covariates	✓	✓	✓	✓
Year-quarter FE	✓	✓	✓	✓
Cardiologist FE	✓	✓	✓	✓
Hospital FE	✓	✓	✓	✓
Cardiologists	51	51	51	51
Observations	8,589	8,589	8,589	8,589

SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (4.8). Dependent variables from left to right are indicators for whether a patient undergoing PCI received a DES and whether the patient died, suffered a myocardial infarction, or had another angioplasty within one year from the intervention, respectively. See Section 4.4 for details. Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 4.5 illustrates the corresponding event study graphs based on Equation (4.9) and the outcomes from Table 4.6. The four panels in the figure, separated by patient outcome, provide a similar pattern as above with no indications of important changes in patient health outcomes at any point over the two years before or after cardiologists' relocation. These results show that the changes in treatment behavior induced by variation in the migrating cardiologists' peer practice environment did not affect the quality of care in terms of patient outcomes to any important extent.

Figure 4.5: Event study estimates of migrating cardiologists' changes in practice environment: Patient outcomes



NOTE.— SCAAR data for years 2004–2013. Dots correspond to coefficient estimates of $\gamma_{t'}$ from OLS estimation of Equation (4.9). Dependent variables from top left to bottom right are indicators for whether a patient undergoing PCI received a DES and whether the patient died, suffered a MI, or was readmitted within one year, respectively. Covariates include hospital, cardiologist characteristics and patient risk factors reported in Table 4.1 and fixed effects for year-quarter, cardiologist, and hospital. Vertical spikes pertain to robust 95 percent confidence intervals clustered by hospital.

Costs of treatment

We have previously argued that the costs of using a DES and a BMS are comparable in terms of the direct and indirect costs of treatment. In particular, using figures from the Swedish agency for health technology assessment (SBU), the total expected cost of using a DES and a BMS for an average patient in Sweden in 2014 was SEK 66,901 and SEK 64,866, respectively. The lion's share of this cost (SEK 59,000) is derived from a fixed hospital reimbursement fee based on the PCI procedure and two nights stay at the hospital, according to figures used in the Nordic DRG patient classification system.[22] The remainder is the cost of the stent, modified by the expected number of stents inserted per intervention and the probability of a subsequent intervention. While the expected cost of a DES is significantly higher than the cost of a BMS, SEK 3,500 versus SEK 1,000, respectively, this is traded off against a lower risk of restenosis, 0.039 versus 0.074, while the expected number of inserted stents is the same for both stent types (Swedish Agency for Health Technology Assessment and Assessment of Social Services, 2014).

We use our previous estimation results in this section to calculate an estimate of the average excess cost that a migrating cardiologist incurred from adaptation to the new practice environment after relocation. Given that we do not find a difference in the propensity of

[22] See http://www.nordcase.org/eng/materials/manuals

revascularization for the migrating cardiologists, a back-of-the-envelope calculation of the increased cost burden from the additional Type I errors we estimate can be produced by multiplying the estimated number of inappropriately used DES by the difference in unit costs between the two stent types. Table 4.1 shows that the average absolute change in practice environment for the migrating cardiologists is 0.3 and the average annual number of PCIs per cardiologist is 65. The estimated increase in Type I errors is roughly 0.25 percentage points for each percentage point change in practice environment. Thus, on average, a migrating cardiologist inappropriately inserted $0.3 \times 0.25 = 0.075$ additional DES after relocation, amounting to an increase of around five stents per year. Multiplying this number with the cost difference between the BMS and the DES yields a cost increase of approximately SEK 12,500 (USD 1,200), or around one-sixth of the total cost of a PCI per migrating cardiologist. We conclude that this figure is rather small in the specific context of treatment of coronary heart disease.

Robustness checks

While our results are indicative of that peer effects are important determinants for cardiologists choice of stent in angioplasty treatments, there are several alternative explanations for our findings. First, peer effects are inherently difficult to identify due to the reflection problem, where involved agents simultaneously affect each other. While we acknowledge that this problem may also exist in our context of migrating physicians, we consider it unlikely that our results would be entirely driven by the moving physicians' influence over their peers. Some results also run contrary to such bias, such as the finding that effects are entirely driven by younger migrants. If migrants indeed affected their peers after relocation, we would rather expect more senior cardiologists to be the main drivers of the effect.

Moreover, peer effects can be modeled in different ways. In this paper, we opted for defining peers as same day coworkers, which seems a natural proxy for the intensity of social interactions between colleagues and, hence, a likely source of knowledge transfers. In theory, however, the exact mechanisms by which peers influence behavior may include more complex dynamics and learning-based non-linearities. The fact that our model is aggregated to the quarter-year level, that is average peer influence over a quarter, may provide some safeguarding against such specification errors.

Finally, our approach is based on the assumption that hospital-specific factors are likely to be fixed over time, and therefore eliminated when including hospital fixed effects in the model. Nevertheless, we cannot completely rule out confounding bias between hospital- and peer-specific effects in our estimation results. However, the fact that we are estimating event studies to explore changes in physician treatment behavior is likely to provide some merit to our approach since time-varying hospital factors are then only a concern if changes in the hospital environment are correlated with the timing of cardiologist migration.

In the remainder of this section, we report estimation results from a set of extensions to our main analysis to gauge the sensitivity of our findings to alternative model and sample specifications. We first check whether our results may be driven by patient, rather than stent, selection by replacing our outcome with different patient case-mix indicators. Next,

we analyze the stability of our results with respect to the definition of the practice environment by reestimating our main DD model using a synthetic environment and non-moving cardiologists to predict counterfactual states.

Peer effects and patient selection

A physician's peers may not only influence their treatment choices, but also which patients they treat on a given day. Even though patients are unable to choose their doctor, physicians may still have some influence over which patients they treat and, consequently, which observed treatment style they end up using (Chang and Obermeyer, 2020). Patients admitted to the hospital may be differently sorted to physicians depending on which peers are around at the time. If such sorting is related, directly or indirectly, to characteristics related to the appropriate choice of stent, our results may be driven by patient stent suitability rather than cardiologist preferences.

We conduct several tests to study whether patient selection, in contrast to stent selection, is a driver of our main results. First, we study the extent to which our estimates change when we omit patient characteristics from our model. Columns (1) and (2) of Table 4.7 report our main estimates from column (4) of Table 4.3 with and without the set of patient risk factors listed in Table 4.1. A dramatic change in the peer effect estimate would indicate that patient selection is an important explanation for our findings. However, as can be seen by comparing the two columns, estimates are very similar in magnitude and not statistically different from each other.

Next, column (3) of Table 4.7 shows estimates where we have replaced our main outcome variable with a "frailty" index, constructed using our RF algorithm, described in Appendix C, to predict one-year mortality for each patient in our sample. This analysis amounts to testing whether our peer effects are associated with treating more or less frail patients; a more comprehensive test of patient selection. Moreover, the results reported in column (4) instead use our RF-predicted DES appropriateness as an outcome, which more directly tests patient selection based on recommended stent type. Both specifications are placebo tests, insofar that we expect a null effect if patient selection is not an issue in our analysis. Reassuringly, the reported point estimates do not suggest that a migrating cardiologist's peers can predict neither patient mortality nor DES suitability. We conclude that patient selection is unlikely to be driving our peer effect results.

Synthetic environment

One empirical issue with the DD approach outlined so far is that migrating cardiologists are unlikely to randomly relocate between hospitals. This leads to two inferential problems with respect to the interpretation of our main findings. The first problem relates to the external validity of our estimated effects. Migrating physicians may constitute a selected group that is unrepresentative for the physician population at large. While Table 4.1 suggests some differences in observable characteristics between moving and non-moving physicians, such as age, the case-mix of patients they treat and the quality of care they provide is indistinguishable from those of non-moving cardiologists. We take this as evidence

Table 4.7: Difference-in-Differences estimates of migrating cardiologists' changes in practice environment: Patient selection

	(1) Main specification	(2) No patient controls	(3) Predicted mortality	(4) Predicted DES
Post	0.003	0.001	-0.011	0.018
	(0.023)	(0.024)	(0.008)	(0.020)
$\tilde{\Delta}_{jt}$	0.013	0.018	-0.014	-0.019
	(0.087)	(0.087)	(0.019)	(0.066)
$Post \times \tilde{\Delta}_{jt}$	0.247***	0.229**	-0.019	-0.033
	(0.090)	(0.092)	(0.056)	(0.064)
Covariates	✓		✓	✓
Year-quarter FE	✓	✓	✓	✓
Cardiologist FE	✓	✓	✓	✓
Hospital FE	✓	✓	✓	✓
Cardiologists	51	51	51	51
Observations	8,589	8,589	8,589	8,589

SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (4.8). Dependent variables from left to right are indicators for whether a patient undergoing PCI received a DES estimated with and without patient characteristics as controls (Columns 1 and 2, respectively), predicted mortality and predicted DES appropriateness using the Random Forest ML algorithm described in Appendix C (Columns 3 and 4, respectively). Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

supporting the notion that the subpopulation of cardiologists moving across hospitals is not widely different from non-moving cardiologists with respect to relevant characteristics.

The second problem relates to the internal validity of our estimates and is potentially more severe as it may invalidate our approach altogether. Specifically, if physicians generally move to hospitals based on their preferences for using DES, the associations we estimate and interpret as caused by changes in practice environment cannot be empirically distinguished from the sorting of physicians to hospitals with practice environments based on their clinical preferences. A similar argument can be raised with respect to the specific peers that the physicians are working together with within a hospital. Although the results from Figure 4.3 and Table 4.3 are reassuring in the sense that the common trend assumption is not rejected, we may still be concerned that the counterfactual practice environment is estimated with bias.[23] To test whether our approach is robust to alternative definitions of practice environments, we propose to extend our analysis by using a synthetic control method derived from a different source of variation to estimate the counterfactual practice environment.

To find a suitable control group that can serve to identify the counterfactual state of migrating cardiologists should they not have moved, we define a synthetic practice style environment from the pool of non-migrating cardiologists (see, e.g., Abadie and Gardeazabal, 2003; Abadie *et al.*, 2010, 2015).[24] For each migrating cardiologist $j \in J$, we define $\tilde{\Delta}_{jt} = \sum_c w_c \Delta_{ct}$ as the counterfactual environment based on non-migrating cardiologists, $c \in C \notin J$. The weights, w_c, are chosen to minimize functions of pre-migration DES share levels ($\sum_{s \in t < t_0} \Delta_{js} - \tilde{\Delta}_{js}$) and slopes ($\sum_{s \in t < t_0} \partial \Delta_{js}/\partial s - \partial \tilde{\Delta}_{js}/\partial s$) based on a constrained quadratic optimization routine (see, e.g., Botosaru and Ferman, 2019; Ferman and Pinto, 2021). A corresponding approach is applied to estimate the counterfactual environment in the pre-migration period using post-migration DES share levels and slopes. Finally, the resulting counterfactual estimates are applied to versions of the event study model in Equation (4.9) where the original practice style environment, Δ_{jt}, is replaced with its synthetic equivalent, $\tilde{\Delta}_{jt}$.

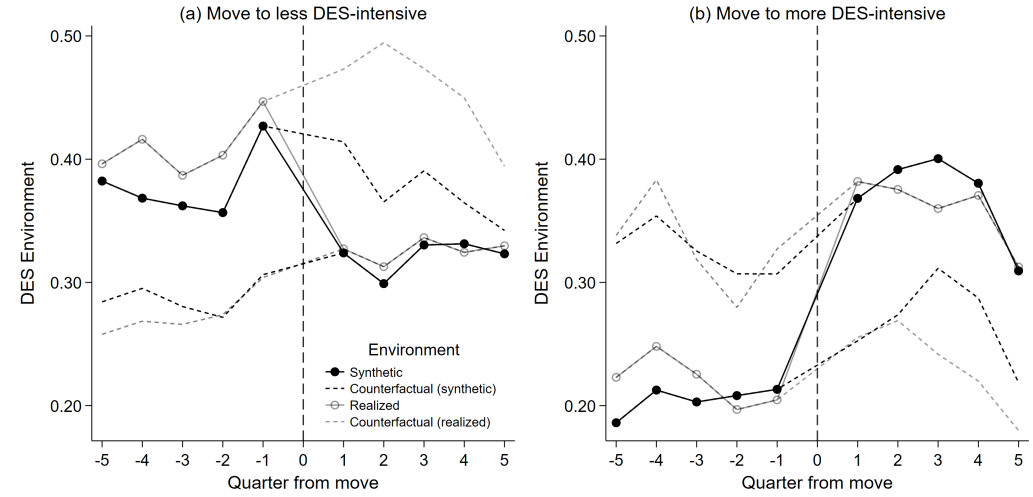
Figure 4.6 illustrates the synthetic environment approach (darker lines) and how it relates to the previous approach by overlaying the corresponding trends in practice environment from Figure 4.3 (brighter lines). The two definitions mostly overlap, with the exception of the post-migration counterfactual environment among cardiologists moving to

[23] It is a priori unlikely that physician sorting based on preferences for individual treatments occurs because they do not possess the individual freedom to schedule their work hours in such detail. Moreover, such sorting would most likely generate conservative bias in our estimates since estimated changes in both the practice environment and the responses therefrom would be based on matching of physicians with similar preferences. In such cases, these changes would thus be smaller than if they were truly random.

[24] Although the synthetic control method was originally developed for a single treated unit, the framework can easily accommodate estimation with multiple treated units by fitting separate synthetic controls for each of the treated units (see, e.g., Abadie, 2021). While there is no important conceptual difference in the contexts of one versus multiple treated units, practice issues relating to the non-uniqueness of the solution to the minimization problem when selecting weights for the synthetic controls are exacerbated in the latter. To address this issue, Abadie and Lhour (2021) propose a synthetic control estimator that incorporates a penalty for pairwise matching discrepancies between the treated units and each of the units that contributes to their synthetic controls.

less DES-intensive hospitals that is somewhat lower than the corresponding environment using the original approach. This suggests that, while the two types of counterfactual environments are partially based on the same empirical variation, there are also some notable differences between them.

Figure 4.6: Average trends in migrating cardiologists' realized and synthetic practice environments



NOTE.— SCAAR data for years 2004–2013. Practice environment defined as the share of DES used in angioplasty treatments in realized (solid lines) and counterfactual (dashed lines) hospitals by quarter from cardiologist move. Brighter lines pertain to estimates of Δ_{jt} while darker lines pertain to the estimated synthetic practice environment, $\tilde{\Delta}_{jt}$. See Section 4.5 for details on the construction of this variable. Separate plots for cardiologists moving to hospitals with higher and lower intensity of DES use. Vertical dashed line indicates recentered quarter of physician relocation from the origin to the destination hospital. Quarter of move linearly interpolated.

Finally, we study whether our main estimation results are sensitive to the definition of practice environment. Table 4.8 reports estimation results from our main DD model where we replace Δ_{jt} with $\tilde{\Delta}_{jt}$ in the analysis. Reassuringly, the results are close to our main estimation from Table 4.6: a change in DES use of migrating cardiologists is of around 0.31 percentage points for each percentage point change difference in synthetic practice environment between origin and destination hospitals, but has no corresponding impacts on adverse patient outcomes. We conclude from this analysis that our main results are robust to the definition of practice environment with respect to whether it is derived from the hospital or from the pool of non-migrating cardiologists.

4.6 Conclusion

This paper empirically analyzes how physicians' treatment decisions are influenced by their practice environment and how such decisions affect the quality of care received by patients. We study these questions in the context of the choice between using bare metal stents (BMS) or drug-eluting stents (DES) among interventional cardiologists in Sweden performing percutaneous coronary interventions (PCI) on patients diagnosed with coro-

Table 4.8: Difference-in-Differences estimates of migrating cardiologists' changes in synthetic practice environment: Patient outcomes

	(1) DES	(2) Death	(3) Infarct	(4) TLR
Post	-0.022	-0.009	0.005	-0.011
	(0.023)	(0.008)	(0.012)	(0.011)
$\tilde{\Delta}_{jt}$	0.122	-0.060	-0.019	-0.047
	(0.139)	(0.036)	(0.025)	(0.043)
$Post \times \tilde{\Delta}_{jt}$	0.312**	0.019	0.006	0.056
	(0.128)	(0.028)	(0.038)	(0.053)
Covariates	✓	✓	✓	✓
Year-quarter FE	✓	✓	✓	✓
Cardiologist FE	✓	✓	✓	✓
Hospital FE	✓	✓	✓	✓
Cardiologists	51	51	51	51
Observations	8,589	8,589	8,589	8,589

SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (4.8) using the estimated synthetic practice environment, $\tilde{\Delta}_{jt}$ in place of Δ_{jt}. See Section 4.5 for details on the construction of this variable. Dependent variables from left to right are indicators for whether a patient undergoing PCI received a DES and whether the patient died, suffered a myocardial infarction, or had another angioplasty within one year from the intervention, respectively. See Section 4.4 for details. Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

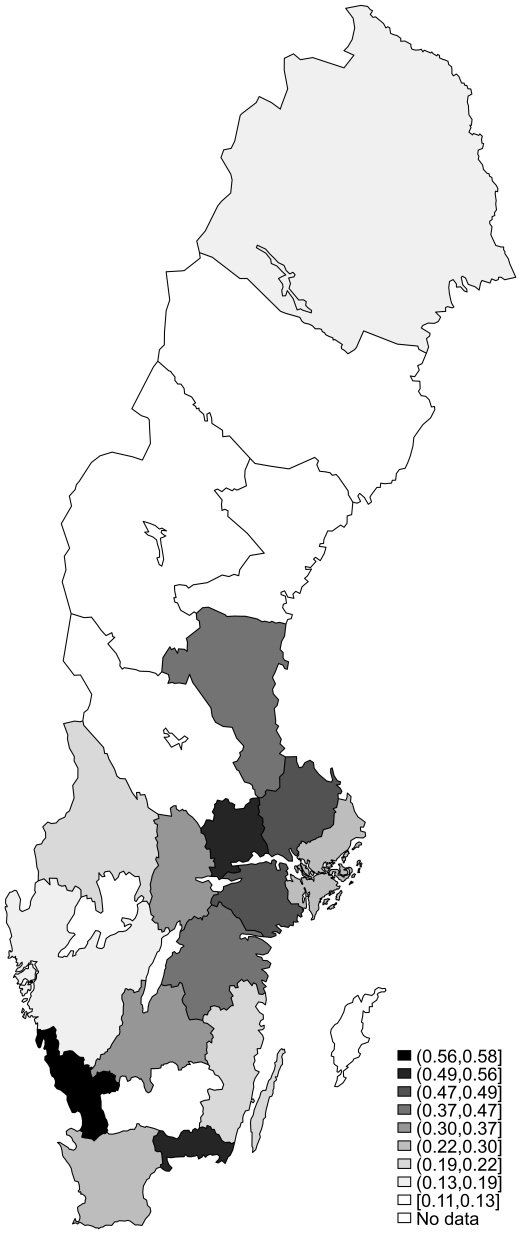
nary artery disease. To obtain empirical variation in a physician's practice environment, we identify cardiologists who moved between hospitals and relate changes in their own treatment behavior and subsequent patient outcomes to differences in the hospital's practice environment before and after they relocated. The overall physician response to their environment is then decomposed into a physical (hospital-specific) and a social (peer group-specific) component by exploiting quasi-random variation in the practice behavior of migrating physicians' coworkers within a hospital. Finally, we relate the environmentally induced changes in practice environment to variations in physicians' rate of decision errors and patient adverse clinical events to gauge whether the practice style changes led to important changes in quality and costs of care provision.

Similar to the results reported in Molitor (2018), we find that migrating cardiologists rapidly, but not fully, adapt to the prevailing practice environment in their use of DES after relocating. Our estimates suggest that cardiologists change their use of DES with around 0.5 percentage points for each percentage point difference in practice environment between the origin and destination hospitals. Decomposing the overall effect into a hospital-specific and a peer group-specific component we find that around half of the response is driven by the latter effect, suggesting that a physician's peer group is as influential as the physical work environment in altering treatment styles. Furthermore, while we observe that physician decision errors, measured using a Random Forest (RF) machine learning algorithm, increased after relocation, we find no evidence that neither major adverse cardiac events, such as heart attacks or patient death, nor treatment costs, were strongly associated with changes in the migrating physicians' treatment styles. This could potentially be explained by that medical decisions were still made within prevailing medical guidelines and did not lead to significantly increased health risks for cardiac patients. Finally, estimation results from a set of split-sample heterogeneity analyses show that our main effects are primarily driven by younger cardiologists and by cardiologists who move to more DES intensive environments, suggesting that both environmental as well as individual physician characteristics appear to be important for the magnitude of physician response.

In conclusion, the results obtained in this paper have important bearing on current health policy with respect to the causes and consequences of unwarranted regional variations in healthcare use (see, e.g., Corallo *et al.*, 2014). Recent evidence on the extent to which regional variations are driven by healthcare providers, such as clinics or hospitals, or individual clinicians have emphasized the role of the latter (see, e.g., Gutacker *et al.*, 2018). That physicians strongly respond and adapt to their prevailing practice environment, and that such conforming arises from both the hospital itself and from the workplace peers, suggest a rationale for why physician treatment styles may cluster in specific geographic areas. The absence of an impact on patient outcomes from such adjustments also provides an explanation for the conundrum of a weak observable correlation between regional variations in the costs and the quality of healthcare provision (see, e.g., Fisher *et al.*, 2003a,b). Although concrete policy advice may require more substantiated evidence, which we leave for further work, we believe that our results show that information campaigns aimed at harmonizing treatment choice among healthcare professionals, such as clinical guidelines, may need to be complemented with alternative measures, such as additional physician training, to significantly reduce unwarranted variations in healthcare use.

4.7 Appendix A: Additional Tables and Figures

Figure 4.A1: Distribution of raw DES rates across hospital regions in Sweden, 2004–2013



NOTE.— SCAAR data for years 2004–2013. Regional administrative map of the 21 county councils in Sweden. Intensity of shaded areas reflect average shares of DES use among patients undergoing angioplasty treatment across all years.

4.8 Appendix B: Institutional Setting

The empirical analyses in this paper are based on inpatient medical records on all percutaneous coronary interventions performed in Sweden between 2004 and 2013. In this section, we first provide relevant background information on the Swedish healthcare system. This is followed by a brief description of the general treatment of coronary heart disease and the specific medical procedure we study.

Healthcare in Sweden[25]

Healthcare in Sweden is mainly funded by direct income taxes raised by the three different levels of government: central, regional (21 county councils) and local (290 municipalities). Responsibilities for health and medical care are shared between the governments according to a scheme stipulated in the Swedish Health and Medical Service Act from 1982. Within each government tier, principals (i.e., elected politicians and bureaucrats) have substantial discretion in designing the system in their area of administration subject to a few general principles, such as that all citizens are entitled to accessible and high-quality healthcare services based on their individual needs. Both county councils and municipality executive boards are political bodies that consist of representatives elected by residents every four years coinciding with the national election.

The main responsibilities of the central government are to set goals for national health policy, coordinate and provide advice to health and medical care providers and to regulate prices and approval of new medical services and products. Municipalities are mainly responsible for organizing long-term care for the elderly in their home or in aged care facilities and to accommodate the needs of residents with physical or psychological disabilities. Finally, the county councils are the main providers and financiers of healthcare in Sweden being responsible for primary and specialized healthcare on both the in- and outpatient basis in their respective geographical area. Since the end of the 1990's, both municipality and county healthcare boards are allowed to contract out services to private providers in purchaser-provider split models. While the outsourcing of healthcare services to private agents have become increasingly commonplace within the primary, outpatient and long-term care sectors over time, virtually all inpatient care is still operated by public providers.

The vast majority of healthcare spending in Sweden is paid for by county and municipal-level direct income taxes raised from area residents. Contributions from the central government are relatively small and mainly consist of provider pay-for-performance incentive schemes and redistribution between regions. Each county council sets its own patient fees, although there is a national limit for the amount a patient has to pay out of pocket (approximately USD 130 per annum as of 2020). Consequently, patient fees only account for around three percent of total spending on healthcare. Both employed and unemployed Swedish citizens are also covered by a national statutory sickness and disability insurance, replacing up to eighty percent of lost earnings and financed through employer social contributions. This insurance can be further topped up for employees covered by collective

[25] www.kliniskastudier.se/english/sweden-research-country/swedish-healthcare-system.html provides a concise summary of the main features of the Swedish healthcare system in English.

agreements or complementary private insurance schemes. Hence, virtually all Swedish citizens have strong financial protection from both direct healthcare costs as well as indirect income losses from temporary and permanent work disabilities.

One important feature of the Swedish inpatient healthcare system that is relevant for our empirical strategy is that recipients of healthcare are constrained in their choices of hospital service provider and treating physician. Specifically, each hospital is responsible for providing care to all residents within a geographical catchment area. This means that place of residence determines which hospital a patient will be admitted to when seeking care. Furthermore, hospitals are not obliged to accommodate patient requests for a specific treating physician. As a general rule, a patient will be assigned to an on-duty physician on the day of admission. This implies that patients are quasi-randomly allocated to physicians and that selection bias arising from endogenous patient-physician sorting is unlikely to be a concern.

Treatment of coronary heart disease

Coronary arteries supply oxygen and blood to the heart. When cholesterol and other fatty plaque build up inside these arteries, the wall of the blood vessel thickens, narrowing the channel within the artery and reduces blood flow to the heart. This process, called atherosclerosis, starves the heart muscle of oxygen and may cause heart tissue damage, known as Myocardial Infarction (MI) or, more commonly, a heart attack. Worldwide, about 15.9 million myocardial infarctions occurred in 2015 (Vos *et al.*, 2016).

Coronary heart disease is generally treated by interventional cardiologists using a catheter-based treatment method called percutaneous coronary intervention (PCI), or coronary angioplasty.[26] In a PCI, the cardiologist first inserts a catheter through either the femoral or radial artery, which is subsequently transported to the site of the blockage using a guide wire. Once the obstructed area is reached, a tiny balloon attached to the catheter is inflated, compressing the atherosclerotic plaque against the artery wall and thereby restoring blood flow. To keep the artery open at the site of the blockage after balloon dilation, the cardiologist may also place and leave a stent (an expandable small metal mesh tube) in the artery to reinforce the blood vessel's wall and prevent it from reoccluding.

Prior to invasive treatment, a diagnostic technique, angiography, is used to determine the size, severity and location of the suspected artery blockage(s). To this end, a catheter is guided into one of the major coronary arteries to inject a contrast dye into the blood passing through the heart. The diagnosing physician, the *angiographer*, can then determine the locations with restricted blood flow from a series of images (angiograms) taken by an X-ray machine. Sometimes, when considered suitable by the responsible physician, the angiography is directly followed by a PCI in the same treatment session, a procedure known as *ad-hoc* PCI.

[26] PCI began as percutaneous transluminal coronary angioplasty (PTCA), a term still found in the literature. It now encompasses balloons, stents, and other modifications to the catheter tip, including devices that cut out plaque to open narrowed arteries.

Bare-Metal and Drug-Eluting Stents

Two main types of stents are associated with performing a PCI: Bare-Metal Stents (BMS), first approved in 1994 and commonly referred to as first-generation stents, and the newer Drug-Eluting Stents (DES), first approved in Europe in 2002. The principal difference between the BMS and the DES is that the latter is coated with a drug that reduces the incidence of restenosis, the medical term for the gradual re-narrowing of a coronary artery after a blockage has been treated with angioplasty. Because the process of compressing, or "crushing", the atherosclerotic plaque often causes trauma to the artery wall, the body will attempt to heal itself by repairing the tissue damage caused by the intervention by proliferation of endothelial cells (a layer on the surface of blood vessels). Restenosis occurs from excessive tissue growth as a consequence of such healing processes, which reoccludes the blood vessel at the site of the stent. In contrast to the BMS, the DES was developed to counteract reocclusion of the artery by being coated with drugs that inhibit cell proliferation, thus significantly reducing the risk of restenosis.

Although the DES represents a major medical advance for angioplasty over the BMS, it has also been associated with the more severe side effect of stent thrombosis (ST); the formation of blood clots in the blood vessels caused by the stent itself.[27] As the drugs coated on the DES inhibit the body's natural healing process (i.e., the formation of an endothelial layer), they simultaneously expose the body to an increased risk of thrombus formation (blood clots). Thus, the DES has been linked with an increased risk of ST occurring up to several years after the initial intervention. So-called Dual Anti-Platelet Therapy (DAPT), most commonly involving acetylsalicylic acid (aspirin) and clopidogrel, is considered crucial to reduce the risk of ST. Early cessation of these drugs after angioplasty using DES significantly increases the risk of both ST and MI.

The above discussion suggests that the choice between a BMS and a DES when performing angioplasty is not trivial. Although clearer guidelines exist today as to which type of stent should be used in each case, this choice belonged to the "gray zone" of medical decision-making, where guidance from clinical evidence is inadequate in providing clear indications for use,[28] during the time period we study in this paper. In addition, the choice between a BMS and a DES does not involve significant differences in other categories of use, such as prices[29] (e.g., costs of equipment necessary for the procedure), mode of treatment (e.g., minimally invasive versus highly invasive), or physical attributes of the clinician (e.g., visual acuity or motor skills). This context provides us with a close to ideal

[27] While this is true for the first generation of DES (Taxus and Cypher), the second generation DES has been associated with significantly less ST than its predecessor (Chitkara and Gershlick, 2010). However, the latter stent type only began gaining popularity at the end of our analysis period.

[28] See, for example, Naylor (1995).

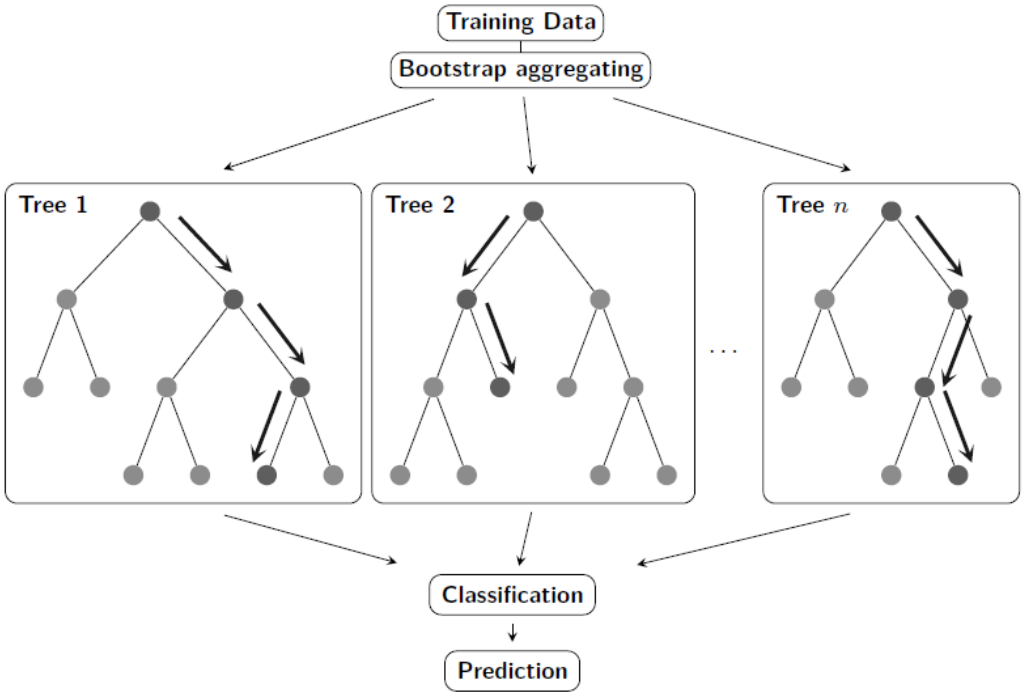
[29] See, e.g., Ekman *et al.* (2006) who estimate that the expected one-year cost of a PCI with a Taxus DES in 2004 amounted to SEK 72,000 (USD 8,500) versus SEK 67,000 (USD 7,900) for a BMS. In 2014, the corresponding figures were SEK 67,000 and SEK 65,000, respectively (Swedish Agency for Health Technology Assessment and Assessment of Social Services, 2014). Both direct and indirect (i.e., repeat revascularization) treatment costs are included as Swedish hospitals are typically paid prospectively on a capitation basis with global budgets. This contrasts, for example, with much larger cost differences in the US (see, e.g., Karaca-Mandic *et al.*, 2017). In addition, we can rule out large incentives for adoption from lobbying by the medical device industry as this is much more muted in the Swedish centralized healthcare system compared to more market-based systems.

setting for studying how physician preferences for treatments vary with their environment, since observed choices are likely to be mainly a function of the physician's personal preferences with respect to the relative efficacy of each treatment option.

4.9 Appendix C: Random Forest Algorithm

Random Forest (RF) is a supervised machine learning method for classification based on the construction of decision trees. The computational steps of the RF algorithm are illustrated in Figure 4.C1. A decision tree splits the data into a set of subsamples defined by a classification rule represented by a tree branch. Each branch could either lead to another subtree or have a leaf/terminal node with an assigned class. The most frequently classified outcome among all individual decision trees performed defines the terminal prediction (class) of the RF. Application of this data splitting method can be further pruned by setting constraints on model parameters to boost the accuracy on the out-of-sample predictions and stability of the tree.

Figure 4.C1: Random Forest machine learning algorithm



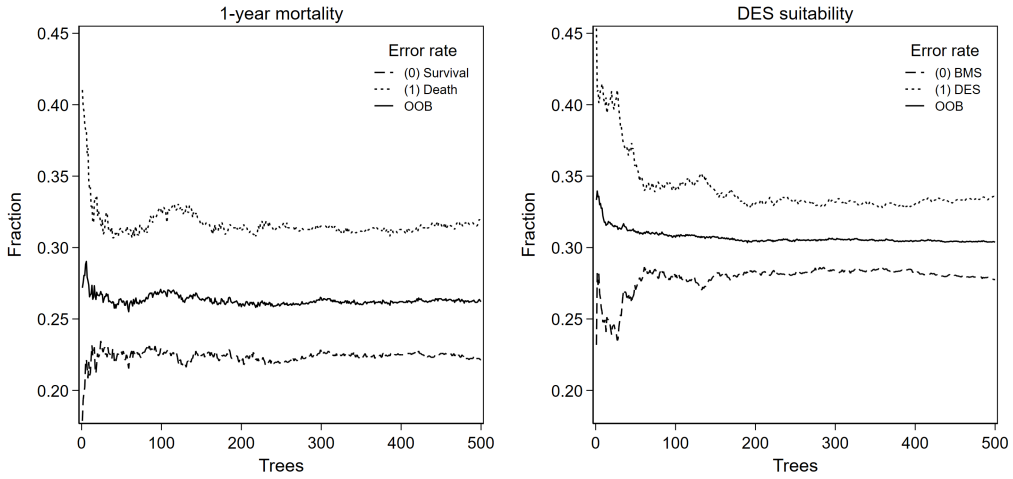
We employ the RF algorithm to classify two binary outcomes: DES suitability and one-year mortality. To classify suitability for DES, we selected a sample of patients treated by non-moving cardiologists in academic hospitals 2008-2011 over which period the use of both stent types were of approximately the same proportions. We randomly split this auxiliary data sample into two parts: a training sample used to fit the RF algorithm and a validation sample used to analyze the algorithm's performance. This resampling procedure is based on a 70:30 split. The analysis sample included only a small share of patients who died after 1 year (4.2 percent). The class imbalance in binary outcome classification leads to reduced classification performance for many classification algorithms (Japkowicz and Stephen, 2002). For this reason, we balanced the sample by downsampling the majority class (classified as survived).

To maximize model's performance and minimize loss, we perform hyper-parameter tuning to fit the RF algorithm for each classified outcome. We grow 500 individual decision trees to reduce and stabilize the out-of-bag (OOB) error and randomly sampled 100 variables at each split out of total 198 predictors that include a rich set of pre-determined clinical factors and patient characteristics. Because of the different sizes of the training samples, we set each tree's terminal node to have at least 2000 observations for classifying DES suitability and at least 100 observations for classifying one-year mortality. The tuning of all parameters is based on the performance evaluation on the validation sample.

The best performing RF algorithm classified DES suitability and one-year mortality with the approximate overall OOB error of 30 and 26 percent, respectively. In the classification of DES suitability, the rates of false positive and negatives are 32 and 28 percent, while in the classification of one-year mortality the corresponding rates are equal to 31 and 22, respectively. Changes in error rates by the cumulative number of decision trees are presented in Figure 4.C2.

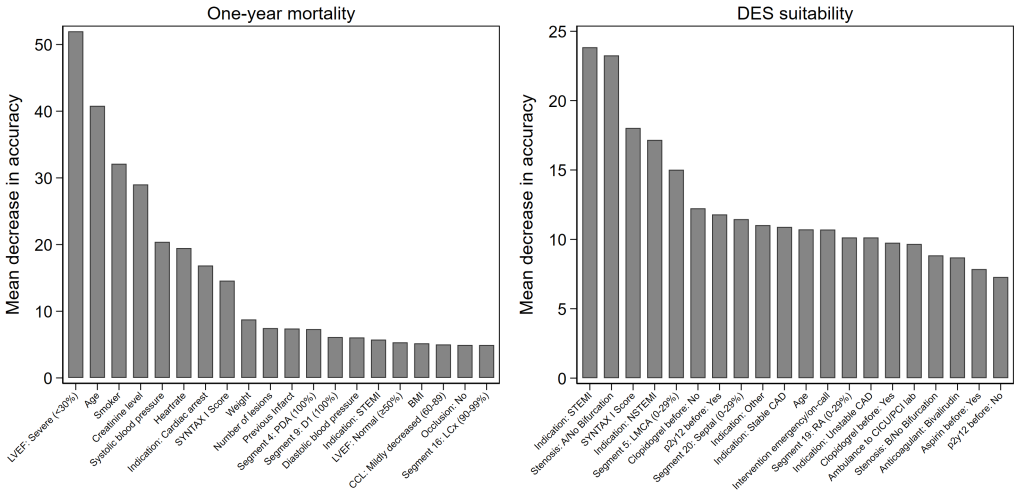
Finally, Figure 4.C3 presents the importance of variables used in the classification of each outcome. The highest loss in the classification accuracy are attributed to the pumping capacity of the patient's heart (LVEF), patient age, smoking status and kidney functioning when classifying one-year mortality. In addition, the algorithm also selected variables that describe underlying medical conditions associated with a higher risk of mortality such as hyperlipidemia status, creatinine test result, and previous infarct or thrombosis. For the classification of DES stability, the algorithm selected a number of variables that characterize the lesion and the location and severity of the occlusion such as coronary artery segment. The strongest predictors for the suitability of DES are indications for ST-Elevated Myocardial Infarction (STEMI), stenosis classification and SYNTAX 1 score.

Figure 4.C2: Error rates in Random Forest prediction by number of regression trees



NOTE.— SCAAR data for years 2008-2011. Shares refer to the fraction of incorrectly classified choices in the RF validation sample. Variables included in the prediction are patient gender, age, BMI, reason for hospitalization, smoking status, blood pressure, diabetes, COPD, peripheral vascular disease, hypertension, hyperlipidemia, previous infarction, previous CABG, previous PCI, previous stroke, creatinine clearance level (CCL), hemoglobin test, occlusion, angiography results by segment including degree of stenosis severity and duration, left ventricular ejection fraction (LVEF), location of lesions, 3-vessel and/or LM lesion, number of treated segments, SYNTAX 1 score, and primary diagnosis according to the ICD-10 classification.

Figure 4.C3: Variable importance weights in Random Forest prediction



NOTE.— SCAAR data for years 2008-2011. Higher values indicate greater importance of variable in predicting outcomes. Variables included in the prediction are patient gender, age, BMI, reason for hospitalization, smoking status, blood pressure, diabetes, COPD, peripheral vascular disease, hypertension, hyperlipidemia, previous infarction, previous CABG, previous PCI, previous stroke, creatinine clearance level (CCL), hemoglobin test, occlusion, angiography results by segment including degree of stenosis severity and duration, left ventricular ejection fraction (LVEF), location of lesions, 3-vessel and/or LM lesion, number of treated segments, SYNTAX 1 score, and primary diagnosis according to the ICD-10 classification.

5 Early Childcare and Children's Non-Cognitive Development: An Interdisciplinary Approach*

Abstract. Interdisciplinary perspective can improve our understanding of the subject, when the evidence is mixed. I connect the findings from non-economic literature on developmental neuroscience, child psychology and development with economic literature evaluating ECC programs and policies. I highlight the importance of considering the *cumulative lifetime quantity* of childcare – the combination of the ECC starting age and mean weekly hours of care. By focusing on the ECC quantity, a consistent theory and evidence-supported hypothesis emerges: early and extensive childcare before the age of one and a half for girls and two for boys is harmful for children's socioemotional and behavioral development. The adverse effects of ECC before age two are especially pronounced among boys and cohorts with more extensive ECC access at earlier ages. I test this hypothesis by exploring the 2005 ECC reform in Germany that substantially expanded supply of childcare places to children under age three. I apply the Difference-in-Differences approach to the data from the German Survey of Youth and Adolescents (KiGGS) and find supporting evidence.

5.1 Introduction

.

Every year more children in high-income countries attend early childcare (ECC).[1] Despite extensive evidence from non-economic and economic fields about ECC, there is still a debate about how it affects child development. Additionally, the findings produced by these fields have been largely disconnected. I use theory and evidence from the fields of developmental neuroscience, child psychology, and child development and juxtapose this information with the evidence from ECC program and policy evaluation produced by the

*This chapter is single-authored and I am thankful to Robert Koch Institute for providing the data. I am grateful to participants at the CINCH Mondays Seminar Series (Essen), Katharina Blankart, Ansgar Wübker, Christiane Wuckel and Stephan Sommer for the discussion, valuable feedback and suggestions. All remaining errors are my own.

[1] In the Organization for Economic Cooperation and Development (OECD) countries, ECC enrollment rates under age three have increased from 22% in 2005 to 36% in 2019. In France, the Netherlands, Belgium, Luxembourg, Norway, Denmark, Iceland, Israel and Korea more than half of children are enrolled (OECD, 2021).

economic literature. Consequently, this study suggests new testable insights on how ECC could affect children's development.

There are two opposing views supported by empirical evidence on the ECC effects on child development. One perspective suggests that high-quality ECC can improve and promote social and cognitive skills, supported by evidence from experimental programs targeting disadvantaged children (see Currie, 2001). The other view raises concerns about the potential harmful effects of extensive childcare initiated early in life and continuing until school entry (see Belsky, 2001) – the type usually experienced by children in economically advanced countries. A large part of this debate surrounds the role of ECC quality and quantity. ECC proponents suggest that high-quality ECC could be beneficial regardless of quantity of childcare – combination of ECC starting age and mean weekly hours of care – attributing negative findings to low ECC quality. The other view suggests that quality, though important, is not the only factor that matters. It suggests that cumulative life course ECC quantity is an important predictor of a child's socioemotional and behavioral problems.

To settle this debate, the US National Institute of Child Health and Human Development (NICHD) initiated a large-scale study in the 1990's. Scientific reports resulting from this effort found that *quality* is important for cognitive-linguistic development, but care *quantity* matters for socioemotional and behavioral development. Specifically, higher quality of care fosters intellectual functioning, while greater quantity of care initiated early in life hinders socioemotional functioning and predicts more problem behavior (NICHD, 2000, 2002, 2003a, 2005). The findings indicate that ECC quality and quantity matter for different aspects of child development.

The neuroscientific research sheds light on the reasons behind the NICHD Study's findings. It establishes that the first years of life are an important period for a sequentially-developing human brain. Specifically, in the first one and a half/two years of life there is a right-brain spurt – responsible for emotional development, followed by the left-brain spurt – responsible for cognitive and motor development (Chiron *et al.*, 1997; Schore, 2017). These growth spurts represent sensitive/critical periods that lay a foundation for future cognitive and emotional functioning. Therefore, the differential findings for ECC quality and quantity could be due to differential effects from children's exposure to them during specific brain development periods.

The research also suggests gender differences in ECC response due to the differences in brain maturation (Schore, 2017). Specifically, boys' emotional right brain develops slower than girls', making them more vulnerable to various stressors in their social environment over a longer time during critical brain development stages due to their immature capacity to self-regulate. This can negatively influence boys' right brain development and lead to long-term deficits in their socioemotional functions and behaviors. Simultaneously, girls develop socioemotional and language skills faster than boys, and these differences in development can lead to better early self-regulation of their behavior. This suggests that girls and boys have different sensitivity to socially stressful situations.

The non-economic literature distinguishes two main concurrent mechanisms how ECC can be stressful for children and how it can influence their ability to self-regulate. The first mechanism suggests that a child in daycare might be affected physiologically. Specifically,

on the days when children attend daycare, their cortisol[2] levels do not follow a normal diurnal circadian rhythm (Tout *et al.*, 1998) and are elevated, signifying greater stress compared to the days when they are at home (Gunnar *et al.*, 2010). *Age* appears to be the most important moderator of this relationship: for children under 36 months, the effect of daycare attendance and cortisol elevation is the most pronounced (Vermeer and van IJzendoorn, 2006). This chronic stress during the critical periods of brain development can have permanent influence and lead to long-term adverse consequences in their socioemotional development and behavior (Lupien *et al.*, 2009; Meaney and Szyf, 2022; Scientific Council, 2014).

The second mechanism is based on the Bowlby's attachment and Schore's regulation theories. Attachment theory suggests that early and extensive childcare could disrupt a child's attachment security or, more generally, lead to less positive mother-child interaction (Bowlby, 1969). The regulation theory suggests that the attachment disruption at specific early critical stages of brain development can directly influence developing right brain structure responsible for the child's future socioemotional functioning (Schore, 2000, 2001). It underlines the important role of mother-child interaction quality on a child's emotional and stress regulation during the right-brain growth spurt.

The NICHD Study finds evidence supporting this mechanism: ECC quantity is associated with the increased risk of insecure attachment in combination with low maternal sensitivity (NICHD, 1997, 2001). It also shows that parenting changes in response to ECC quantity: more hours of care across the first 36 months of life associate with less maternal sensitivity (Jaeger, 1999; NICHD, 2003b). This suggests that early and extensive childcare affects not only a child's, but also a mother's side of the relationship: by worsening parental ability to regulate a child's emotions. Thus, in addition to being affected physiologically by the chronic stress during the period when children rely on quality caregiver-child interaction to regulate their emotions, they experience less of such adult inputs not only in daycare settings, but also at home (see Ahnert *et al.*, 2000).

Therefore, early, extensive, and continuous childcare could lead to worse non-cognitive skills and harm children's development by undermining their future human capital base since non-cognitive skills promote the formation of cognitive skills (Cunha and Heckman, 2008; Heckman *et al.*, 2006), and are important predictors of educational and labor market success (see, e.g., Barón and Cobb-Clark, 2010; Brunello and Schlotter, 2011; Caliendo *et al.*, 2015; Heckman *et al.*, 2013; Prevoo and ter Weel, 2015). Moreover, Layard *et al.* (2014) show that children's emotional health and social behavior are the most (and intellectual development is the least) important predictors of adult life satisfaction. This should be concerning policymakers developing ECC policies.

Meanwhile, despite this evidence many countries are introducing publicly subsidized ECC policies. Germany conducted an ECC reform in 2005 that substantially increased access to early and extensive childcare for children under age three. I use this policy to test my theory- and evidence-supported hypothesis that integrates the findings from non-economic and economic fields about how ECC affects children's development.

[2]Cortisol is a stress-sensitive hypothalamic-pituitary-adrenocortical (HPA) axis hormone that is involved in regulation of stress and emotions. The maturation of the HPA axis continues into the child's third year, and this maturation corresponds with a child's self-regulatory abilities (Watamura *et al.*, 2004).

I use a staggered expansion of new ECC slots across counties and time – some counties expanded relatively fast, while others were slower, in part because of a complicated and lengthy administrative process – as the basis for my Difference-in-Differences (DiD) identification strategy. I utilize data from the German Health Interview and Examination Survey for Children and Adolescents (KiGGS).

I contribute to non-economic and economic literature by connecting the findings from the two fields and creating a unique interdisciplinary perspective that extends our understanding of ECC effects. I contribute to non-economic literature by providing *causal* estimates of greater ECC access on children's outcomes, while this literature usually provides *associations* only. Additionally, this literature documents that early and extensive childcare has an adverse, but *modest* impact on children's non-cognitive development compared to family background factors – whilst controlling for parenting in the analysis – suggesting that family matters more than childcare quantity (see, e.g., Belsky *et al.*, 2007; NICHD, 2003a). Since parenting changes in response to early and extensive daycare use, this could bias estimates and lead to problematic conclusions that understate ECC's overall impact. I capture the ECC policy's ecological perspective on children's development that allows policymakers to see a greater picture of the overall policy effect.

I contribute to economic literature by looking beyond the initial period of the reform-initiated ECC expansion (done previously by Felfe and Lalive, 2018), by covering cohorts that were affected by later expansion stages. This is important, because more full-time ECC slots became available for children affected by the later expansion when they were younger, exposing them to more extensive care during earlier developmental stages. This influences their non-cognitive outcomes. Consequently, policymakers could draw different conclusions about policy effects depending on the ECC expansion period under consideration. I also suggest mechanisms that could contribute to the uncovered effect.

I find that receiving a greater access to early and extensive childcare led to a significant decrease in pro-social skills of around 74% of a standard deviation in children from earlier affected cohorts for whom extensive ECC became available around age two. The effect is concentrated among boys (94.71% of a standard deviation). For boys from later affected cohorts for whom extensive ECC became available closer to age one and a half, greater ECC access had more pronounced effects and led to higher hyperactivity/inattention, peer problems, and lower pro-social behavior (126.39%, 170.83%, and 156.20% of a standard deviation, respectively). For girls, especially from earlier affected cohorts, I find indications of skill improvements.

The remainder of the paper is organized as follows. Section 5.2 reviews previous economic research. Section 5.3 provides details on the policy. Section 5.4 describes the data and key data elements. Sections 5.5 and 5.6 present the empirical method and results, respectively. Section 5.7 contains a discussion; and Section 5.8 concludes.

5.2 Previous Economic Research on ECC

Evaluation of ECC policies and programs that publicly subsidize extensive and early non-parental care presents a great opportunity to estimate its causal effect on children's de-

velopment, and compliments the findings from the non-economic literature. I review the economic literature against the framework suggested by the non-economic literature. I focus on overall findings for cognitive and non-cognitive skills, and gender heterogeneity.

I review the evidence on the programs similar to the German ECC program along these dimensions: universal;[3] targeted at children under age three;[4] with maternal care or some other type of informal care being the main alternative care arrangement;[5] programs with comparable quality and structure (hence, I look at Canada (Quebec)[6] and Europe). Finally, I review studies that measure children's outcomes in preschool and primary school ages since my outcomes are measured in these ages.

Research on non-parental care for children under age three at first sight appears to be mixed. However, when reviewing the evidence focusing on lifetime ECC quantity – the combination of children's age when starting ECC program and mean weekly hours of care – a more consistent picture emerges.

Evidence from the Quebec policy shows that it had no positive effect on children's cognitive skills in the short run (Baker *et al.*, 2008; Haeck *et al.*, 2015); and a negative effect on children's non-cognitive skills in the short (Baker *et al.*, 2008) and medium run (Baker *et al.*, 2019). The negative effects are mostly concentrated in boys (Kottelenberg and Lehrer, 2018) and among children who gained access to daycare before age two, while children who gained access to daycare from age three benefited (Kottelenberg and Lehrer, 2014).[7]

Additional evidence comes from European countries. For Oslo, Norway, Drange and Havnes (2019) find positive effects of ECC enrollment at ages one and two on cognitive development compared to children who were delayed for 12 months or more, with stronger effect for underperforming groups. In line with this, Peter *et al.* (2016) for the UK show that starting daycare after the age two and a half compared to earlier ages[8] has a negative effect for children's socioemotional skills. The effect is mostly concentrated in boys from disadvantaged backgrounds. Another UK study finds sizable negative short-term effects of non-parental care utilization during the child's first two years of life on cognitive test scores and motor skills (Herbst, 2013).

For Italy, Carta and Rizzica (2018) find no detrimental effects on children's cognitive skills from extending access to two-year old children; while Fort *et al.* (2020) examine the effects of attending childcare before age three (i.e., from four months to up to 36 months) in the city of Bologna and find negative effects on cognitive and non-cognitive skills, con-

[3] Enrollment in universal programs is open to everyone and covers economically and socially diverse groups of children.

[4] In light of the evidence from the non-economic research it is important to consider whether the ECC program substantially extends access to the very young (children under age two), or to children two to three years of age, since we can expect different effects for these groups.

[5] A shift from parental care to formal ECC could affect children differently than a shift from informal care, which is likely to be of inferior quality.

[6] Baker *et al.* (2008, 2019) indicate that the Quebec program was designed with affinity to ECC programs in European countries and has similar quality to ECC programs in many other high income countries.

[7] These findings are in line with Loeb *et al.* (2007) for the US, who find that starting center-based care before age two is related to more negative social behavior. They also find that children who started center care between ages two and three enjoyed the strongest cognitive benefits.

[8] The majority of these children started childcare between ages one and a half and four.

centrated among girls from more affluent families. However, it is not clear at what ages most of these children received ECC access.

For Germany, Gathmann and Sass (2012) look at the reform in Thuringia that led to a reduction in ECC use by disadvantaged households and document negative effects in cognitive and non-cognitive skills for one- and two-year old girls. Felfe and Lalive (2018) look at first cohorts affected by the 2005 ECC policy and document heterogeneous effects for children based on their propensity to attend ECC. They find that children who were most likely to attend ECC benefited in terms of their motor skill development, while children who were least likely to attend benefited with respect to their socioemotional skill development, implying that ECC benefited relatively more boys and relatively disadvantaged populations. However, since they only look at the initial ECC expansion stage – where children were receiving less extensive ECC access and at older ages (between age two and two and a half), compared to the population of children affected by the reform's later stages – we could expect different results. I contribute by looking at later affected cohorts, for whom more full-time slots become available at younger ages.

Summarizing the economic literature, I note that it generally finds negative effects on both cognitive and non-cognitive skills for children exposed to extensive ECC programs before age two, especially for boys. However, girls and boys (especially from disadvantaged families) start to benefit after around age one and a half and two, respectively. This underlines the importance of timing of ECC exposure and counterfactual care quality.

This is in line with the non-economic literature that suggests that children starting ECC after age one and a half/two could benefit from stimulating social and cognitive environments because their time in ECC would coincide with the development of their left brain hemisphere responsible for linguistic and motor functions, while these children also would have more mature self-regulatory abilities. The difference in quality of care between home and ECC environment with respect to caregiver's education and the availability of cognitively, socially and motor stimulating activities is important during this time.[9] Girls'/boys' cognitive and non-cognitive skills benefit depending on the differences in quality between ECC and home environments if they start ECC after the age of one and a half/two. Meanwhile, girls/boys who are exposed to extensive ECC before the age of one and a half/two are negatively affected due ECC exposure coinciding with their right brain hemisphere development and their immature self-regulatory ability.

The above evidence allows me to formulate the following hypothesis: greater access to early (prior to age one and a half (for girls)/two (for boys)) and more extensive childcare could have a negative effect on children's socioemotional development and behavior. The adverse effect of ECC before age two should be concentrated among boys and be more pronounced in cohorts that receive greater and more extensive ECC access at earlier ages.

[9] If during this time period home environment is more stimulating than ECC (as could be in the case of higher educated families) these children might not benefit from their exposure to ECC; while children from lower educated families might benefit, since ECC environment during this developmental stage could be more stimulating than their home environment.

5.3 German Childcare System

The childcare system consists of two tiers: preschool kindergarten is available to children three to six, while children under age three attend ECC. States (*Bundesländer*) are responsible for childcare regulation, while counties (*Kreise*) are responsible for organization and implementation.[10]

ECC is provided in two different modes: care centers or family daycare (mostly occurs in caregivers' homes)[11] – and is heavily subsidized. The fees are related progressively to family income and care hours. Parents pay on average 30% of the total cost and fees are independent of daycare mode. Large families and families with low income can have further reductions. Thus, the reform provided low-cost care for children under age three across all German counties.

Prior to the 2005 reform, ECC supply in Western Germany was low, while there was a large excess demand. In 2005 the stated demand for an ECC place among Western German parents with children under age three was around 35%, but only 5% of children in this age group were offered one (Bien *et al.*, 2007). To deal with this excess demand the centers were using waiting lists. Families who signed up early, single parents, and those who already have an older sibling enrolled in ECC were given priority.

Reforms expanding ECC facilities

In 2005 the German federal government introduced a reform that significantly expanded the number of ECC places for children under age three in Western Germany.[12,13] In December 2008 the child support act (*Kinderförderungsgesetz*) was introduced, establishing a legal claim right to an ECC slot for all children aged one or older from August 2013 onward.[14] This was a commitment by the federal government to sufficiently expand ECC supply in all counties to satisfy the legal claim for a daycare slot for all eligible children by August 2013.[15]

Figure 5.1 provides a graphical representation of the differences in ECC supply across Western German counties over the years and allows us to assess the scale of the expansion. It conveys that there is a substantial heterogeneity across counties in the timing of the expansion. ECC coverage expanded strongly over time: in 2002 it was 2.2%; in 2006 –

[10] States, for example, regulate opening hours, staff qualifications, group size, and staff-to-child ratios, while counties monitor childcare centers' compliance with these rules.

[11] The center quality is regulated with respect to staff per child ratios, staff qualification, opening hours and group sizes. Family daycare is characterized by lower number of children (but the same staff-child ratio) and lower staff qualifications.

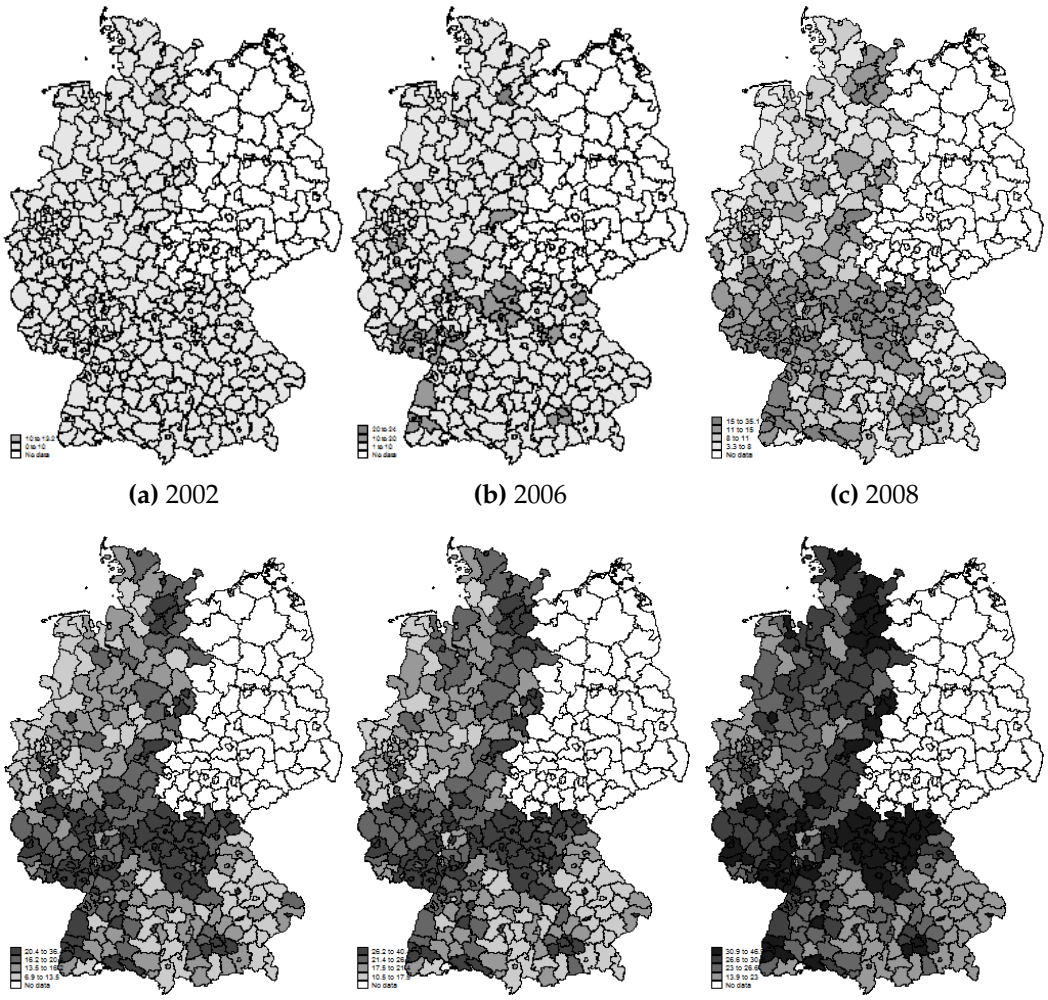
[12] Historically, ECC coverage in Eastern Germany was relatively higher and it experienced ECC expansion to a much lesser degree.

[13] According to the daycare expansion law (*Tagesbetreuungsausbaugesetz*) the plan was to add additional 230,000 places for children under age three by 2010, while adhering to the quality standards.

[14] It was decided in 2007 during the summit (*Krippengipfel*) of the three federal levels (federal, state, and county) that by 2013 a nationwide coverage rate should be 35%.

[15] In case of a municipality's inability to satisfy legal claim for eligible children, parents could sue for reimbursement of private childcare costs due to the lack of sufficient public provision; also for compensation of foregone earnings in cases when the lack of public ECC prevented parental employment.

Figure 5.1: ECC coverage (slots per 100 children under age 3) over the years (*Kinder- und Jugendhilfestatistik*).



(a) 2002 **(b)** 2006 **(c)** 2008

(d) 2010 **(e)** 2012 **(f)** 2014

7.3%; in 2008 – 11.7%; in 2010 – 17%; in 2012 – 22.1%; and in 2014 – 27%.[16] The expansion also led to a large regional variation in ECC coverage between counties. Overall, these reforms constituted a substantial positive supply shock to subsidized ECC and led to a great expansion in ECC slots availability over the years.

It is important to understand where this variation in the ECC expansion was coming from. This has been previously discussed elsewhere, thus, the following paragraphs draw on the information found in Bauernschuster *et al.* (2016) and Felfe and Lalive (2018).

[16]There are no administrative data on public childcare provision available for the years prior to 2002 and from 2003 to 2005.

The ECC expansion was financed with funds from federal, state, and local governments.[17] Thus, the process of the new ECC slots provision involved complex and intertwined decisions from authorities on municipal, county and state levels. The authorities on municipal and county levels were responsible for assessing demand on the local level given demographic and economic factors. The authorities on the state level were responsible for approving proposals for opening new ECC centers by the non-profit organizations.

This administrative process was prone to various issues that differed substantially between the counties, including the differences in routines and knowledge about the complicated funding system (with subsidized funding coming from the federal, state and municipality levels); various regulations for building and opening ECC; shortages in construction grounds and of qualified staff; delays in application approvals or rejections due to non-compliance with the regulations (Hüsken, 2011).

This process created a geographic variation in the expansion's timing: some counties were able to fulfill the necessary requirements and receive the funding relatively fast, while it took others some time to organize – those counties received funding later. Thus, ECC expansion timing differences across counties were not only due to variations in demographic and economic factors predicting demand, but also due to local supply shocks in the creation of new ECC slots resulting from complicated and lengthy administrative processes (Felfe and Lalive, 2018). The latter component is arguably not related to the expected changes in children's non-cognitive development and, therefore, I rely on it in my identification strategy.

5.4 Data and Descriptive Analysis

Data sources

For my analysis of ECC on children's development, I draw on data from German Health Interview and Examination Survey for Children and Adolescents (KiGGS) from the Robert Koch Institute (RKI). KiGGS is a nationally representative survey that contains comprehensive data on health and development of children and adolescents living in Germany. The sampling takes place in 167 cities and municipalities (*Gemeinde*) in Germany (111 of them are in Western Germany) which are located in 167 counties (*Kreise*). There are currently three waves of the survey: Baseline study (Wave 0), Wave 1 and 2. The initial cohort includes children aged zero to 17, which were followed up longitudinally across all three waves. Additionally, a new cohort consisting of children aged zero to 17 was added to Wave 2. While KiGGS has a longitudinal component I use it cross-sectionally.

Data for the KiGGS baseline study (Wave 0), collected in 2003-2006, include 17,641 children and adolescents aged zero to 17. Data for KiGGS Wave 1, collected in 2009-2012, follow up on children, adolescents and young adults who had participated in Wave 0 (now aged six to 25). KiGGS Wave 2, collected in 2014-2017, continues to follow children, adolescents and young adults from previous waves. In addition, new participants (aged zero

[17] The reform did not crowd out the public funding of other family-related programs (child allowance, child benefit or parental allowance).

to 17) were randomly selected and invited to the study. Thus, Wave 2 includes 23,000 participants aged zero to 31.[18]

The content of KiGGS is rich. The data provide information on out-of-family care and parental evaluation of children's development. The sample size averages around 1,000 children per each age, although some counties and ages are oversampled. I use the provided weights in all the results presented in the paper.

In KiGGS a primary caregiver completes a questionnaire and provides information on child and family background; whether the child has ever been in out-of-family care, and if yes, from what age. I use this information to impute whether a child was in out-of-family care before age three. In my analysis I use the sample of children up to age 10 living in 111 Western German counties from all three waves of the survey.[19] I restrict the analysis to Western Germany, since Eastern Germany already had relatively high ECC coverage levels as a legacy of the former GDR.

I utilize Strength and Difficulties Questionnaire (SDQ) as the main measure of a child's non-cognitive skills.[20] The SDQ index consists of categories that cover questions about child's inattention/hyperactivity, emotional, peer, behavioral problems and pro-social behavior. The scores are built up from a menu of questions. Each question solicits a response on a three-point scale, and the score is constructed as the sum of responses on all questions for a given behavior or skill.[21] The questions are asked starting from age three. Thus, my sample includes children aged three to 10. I standardize each SDQ category and pro-social score to have a mean of zero and a standard deviation of one.

I also collect data on the counties' ECC coverage rates from Statistical Offices of the German Laender (*Statistische Landesämter*), available for the year 2002 and annually from 2006 to 2017. I combine the individual KiGGS data with the administrative data on the ECC coverage on the county level. This constitutes my main analysis dataset.

In supplementary analyses I also utilize the *Indikatoren und Karten zur Raum- und Stadtentwinklung* (INKAR) database for regional-level information such as county fertility and female employment rates, share of population zero to three, GDP per capita, net migration, and share of foreigners.[22] Finally, I also utilize county-level information on center quality from *Statistik der Kinder- und Jugendhilfe* from Statistical Offices of the German Laender.

[18] For more information on the data please refer to https://www.kiggs-studie.de/english/survey/kiggs-overview.html.

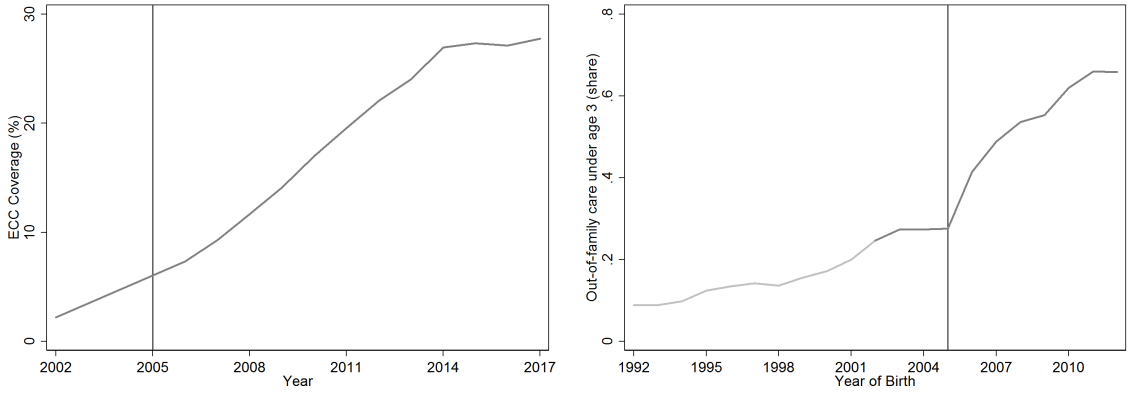
[19] I limit my sample to children up to age 10 because this is the oldest age children affected by the reform would reach by the time of the last survey (Wave 2), i.e., children born in 2006 (first affected cohort) would be 10 in 2017 (the last year of the survey).

[20] It is a standard behavioral measure in the child development literature, see Goodman (1997) for a description.

[21] See Appendix 5.9 for a list of questions used to construct SDQ categories and the pro-social score.

[22] The INKAR database is published by the Federal Institute for Research on Building, Urban Affairs and Spatial Development and provides regional information on infrastructure and socioeconomic characteristics.

Figure 5.2: ECC coverage (slots per 100 children under age 3) (*Kinder- und Jugendhilfestatistik*) and out-of-family care for children under age 3 over the cohorts (KiGGS).



(a) ECC Coverage (%)

(b) Out-of-Family Care under Age 3

Descriptive evidence

I begin by graphing side-by-side the share of ECC slots and out-of-family care under age three over the years and different cohorts, respectively. The left panel of Figure 5.2 shows the ECC coverage rate from administrative data from 2002 to 2017. We see that in 2002 the coverage in Western Germany was around 2.2%, but has been increasing since – especially rapidly from 2006 – and reached around 27% by 2014. The right panel of Figure 5.2 shows the share of children in out-of-family care under age three per each cohort in KiGGS.[23] This share was gradually increasing until 2005, and thereafter experienced a rapid rise, coinciding with the expansion in ECC supply. This implies that more children in out-of-family care, born after 2005, were attending publicly provided ECC.

Fast- vs slow-expanding counties

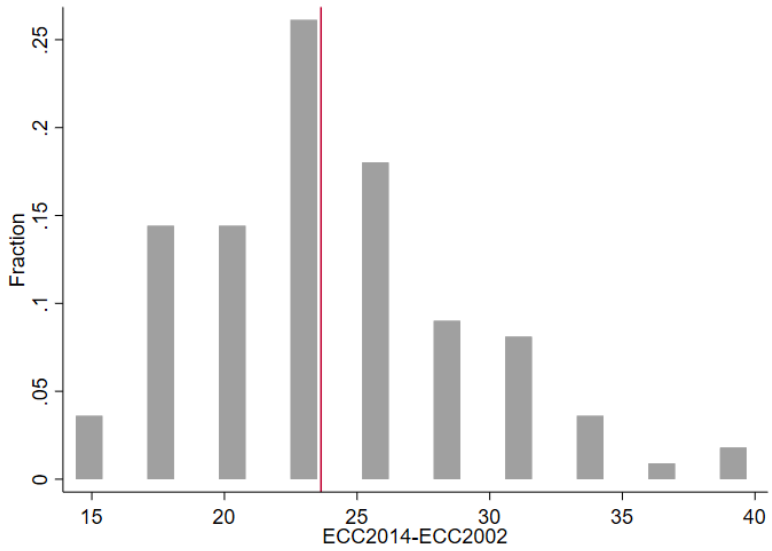
I use the fact that ECC supply expanded unequally across space and over time and define fast- and slow-expanding counties by generating a binary treatment variable in terms of the ECC expansion speed. I achieve it by ordering Western German counties with respect to the absolute difference in their public ECC coverage from 2002 to 2014 (Figure 5.3).

I choose 2002 as a baseline year because it is the first pre-reform year for which administrative data on the ECC coverage are available and because during this time ECC expansion was not yet at the center of a political discussion and, thus, had not yet been affected by any political decision-making to expand ECC. I choose 2014 as the post-reform year since the expansion happens until 2014 and the counties do not converge afterwards (as seen in the left panels of Figures 5.2 and 5.4).

[23] It is based on an answer in KiGGS if parents report that a child has been in out-of-family care before age three. I create a binary variable that is equal to one if they report out-of-family care before age three, and zero otherwise. I then calculate a share of children per birth year that has experienced out-of-family care before age three.

Fast-expanding counties are defined as counties that experienced above median expansion over the years. Respectively, the counties that expanded below median are defined as slow-expanding counties.

Figure 5.3: ECC expansion density plot (*Kinder- und Jugendhilfestatistik*).



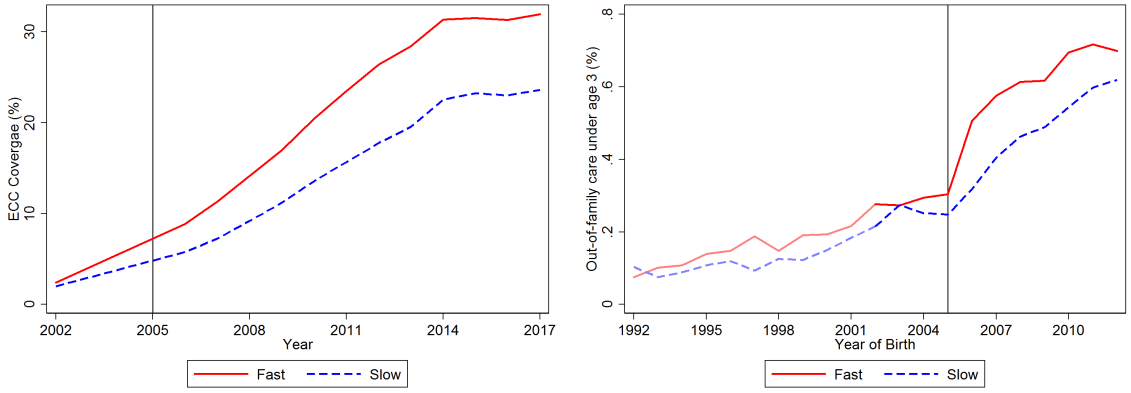
The left panel of Figure 5.4 shows the ECC coverage from the administrative data for the fast- and slow-expanding counties. The right panel shows the difference in out-of-family care under age three from KiGGS data between fast- and slow-expanding counties. Both types of counties were expanding the ECC supply, but this expansion was greater in the fast-expanding counties.

Table 5.A1 in Appendix shows means and standard deviations of my dependent and control variables divided into fast- and slow-expanding counties before and after the reform for earlier (born in 2006-2011) and later affected cohorts (born in 2008-2014) measured in primary school (6-10) and preschool (3-5) ages, respectively. I note a few differences between the fast- and slow-expanding counties before the reform: fast-expanding counties had a larger share of children under age three receiving out-of-family care, a larger share of educated, higher socioeconomic status (SES), non-immigrant, and single-parent households, a larger share of households living outside big urban areas. To account for these differences, I control for child and family characteristics in my analysis.

5.5 Empirical Method

To estimate a causal effect of ECC on children's development, I exploit geographic variation in the timing of childcare expansion emanating from the positive ECC supply shock

Figure 5.4: ECC coverage (slots per 100 children under age 3) (*Kinder- und Jugendhilfestatistik*) and out-of-family care for children under age 3 for fast- and slow-expanding counties over the cohorts (KiGGS).



(a) ECC Coverage (%)

(b) Out-of-Family Care under Age 3

discussed in Section 5.3. I utilize the Difference-in-Differences (DiD) method similar to Havnes and Mogstad (2011) and Baker *et al.* (2008, 2019). According to this methodology, I compare the outcomes of children who live in fast- and slow-expanding counties (treatment and comparison groups, respectively) and who were and were not affected by the reform (born after and before the reform, respectively).

For my main analysis, I specify the following linear DiD model for outcome variables such as out-of-family care before age three and various children's non-cognitive development measures:

$$Y_{ict} = \alpha + \beta PolicyExposure_{ct} + \delta_t + \pi_c + \lambda_w + X_{ict}\Gamma + \epsilon_{ict} \tag{5.1}$$

Y_{ict} is the outcome of interest for a child i in county c in observation period[24] $t \in \{0, 1, 2\}$, where 0 stands for the pre-reform period that includes children born from 1992 to 1999 observed in Wave 0; 1 stands for the pre-reform period that includes children born from 2000 to 2005 observed in Waves 1 and 2; and 2 stands for the post-reform period that includes children born from 2006 to 2014 observed in Wave 2.

The main variable of interest is $PolicyExposure_{ct}$, which equals 1 if children live in fast-expanding counties and are born in 2006 or later and observed in Wave 2 (exposed to the policy), and 0 otherwise. δ_t are observation period fixed effects. I also include wave and county dummies – λ_w and π_c, respectively – to eliminate unobservable time-invariant heterogeneity across counties and temporal shocks that affect all children in the same wave simultaneously.

I also include a set of controls X_{ict} for parents' and children's characteristics: parental education level, household SES, parents' age and immigration status, indicator for residency

[24] Since some waves contain both, affected and unaffected cohorts, I define an observation period variable that separates them into different periods.

in a large city, number of younger and older siblings, and a child's age and gender. I estimate the model both with and without controls. ϵ_{ict} is the error term, which is assumed to be normal. I cluster standard errors at the county and wave level to allow for any arbitrary serial correlation within county and wave.

The coefficient of interest is β. It captures an intention-to-treat (ITT) effect since I estimate the reduced form impact on all children that were affected by the reform and live in fast-expanding counties after the reform rather than only on the children that attend ECC. ITT effect identifies the whole impact of the reform including its direct and indirect effects, i.e., any peer effects on children who were not attending ECC or change in the family environment.

Since I estimate a linear DiD model and to give β a causal interpretation I need to ensure that the core identifying assumption holds. This assumption is called common time trend and states that in the absence of the reform, children's outcomes in fast- and slow-expanding counties would have followed a common time trend. Figures 5.B1 and 5.B2 in Appendix provide graphical evidence supporting this assumption in the time period before the reform.[25] Moreover, balancing test (Table 5.B1) shows that demographic observables are unrelated to the speed of counties' ECC expansion instigated by the reform (except for the number of siblings in the later affected cohorts' sample).[26]

To test my hypothesis I estimate Equation 5.1 on the samples of earlier and later affected cohorts. For them the share of full-time ECC slots was increasing over time and disproportionately more in fast-expanding counties, while the average starting age of out-of-family care gradually decreased (Figure 5.C1 in Appendix). Children from earlier affected cohorts had fewer full-time slots and were attending ECC at older ages due to the initial supply constraint; while children from the later affected cohorts had more full-time slots available to them and were attending ECC at younger ages due to a more relaxed supply constraint. Finally, since according to my hypothesis I also expect different effects among boys and girls I then estimate Equation 5.1 on the sub-samples divided by gender.

5.6 Results

I first present the results from the DiD specification with and without controls for earlier affected cohorts and then for later affected cohorts (Tables 5.1 and 5.2, respectively). Note, higher scores for hyperactivity, emotional, peer and behavioral problems are associated with worse outcomes, while the opposite is true for pro-social behavior.

From Table 5.1 we see that out-of-family care has increased significantly in fast-expanding districts after the reform. There is some evidence of a negative effect for these children: they exhibit statistically significant lower pro-social behavior, while the coefficients on hyperactivity/inattention score and emotional problems indicate skill improvements, but for peer and behavioral problems the coefficients indicate skill deterioration,

[25] Only emotional problems in earlier affected cohorts depart from a common trend in observation period 1 at 10% significance level. The direction of the failure indicates that I would underestimate the causal policy effect for this outcome.

[26] For more details on this, please refer to section 5.10 in Appendix.

albeit imprecisely estimated. The effect size for decline is pro-social score (-0.1136) constitutes 7% (-0.1136/1.627*100) of the standard deviation.

The estimated effect appears to be small relative to the mean or standard deviation. This estimate is an ITT effect. To arrive at the effect of treatment on the treated we need to divide it by the probability of treatment. If we define treatment as the increase in the ECC use, then 9.5% of children are treated. When we scale the mean effect by this number, the effect appears quite large. For example, the decline in pro-social behavior relative to the mean constitutes only 1.39% (-0.1136/8.151*100=1.39%), but if we divide it by the share of treated children we get 14.6% (1.39/9.5*100=14.6%), which is 73.7% (7/9.5*100) of a standard deviation (SD).

Table 5.1: Impact of ECC policy exposure on measures of non-cognitive outcomes for earlier affected cohorts (aged 6 to 10).

	(1) In Care	(2) Hyperactivity	(3) Emotional	(4) Peer	(5) Behavioral	(6) Prosocial
Panel A: Ages 6-10: No Controls						
Policy Exposure	0.1040***	-0.0709	-0.0316	0.0344	0.0349	-0.1156**
	(0.0246)	(0.0594)	(0.0603)	(0.0551)	(0.0533)	(0.0571)
N	7,571	7,580	7,584	7,587	7,586	7,593
Mean	0.230	3.379	1.909	1.350	1.969	8.151
Panel B: Ages 6-10: With Controls						
Policy Exposure	0.0949***	-0.0652	-0.0250	0.0209	0.0483	-0.1136**
	(0.0232)	(0.0574)	(0.0587)	(0.0534)	(0.0529)	(0.0560)
N	7,571	7,580	7,584	7,587	7,586	7,593
Mean	0.230	3.379	1.909	1.350	1.969	8.151

Dependent variables: In Care (under age 3), Hyperactivity, Emotional, Peer, Behavioral problems, Prosocial behavior are standardized. For each dependent variable I show the pre-policy mean for fast-expanding counties, and the coefficient from a regression on the Policy Exposure dummy. All regressions include county, wave, and observation period fixed effects. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, child's gender, and an indicator for residence in a large city (>100K+). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 5.2 shows that the average effect is not statistically significant (except for marginally significant increase in hyperactivity), even though these children were more likely to be in out-of-family care before age three. All other coefficients also indicate towards skill and behavior deterioration, albeit not precisely estimated. However, behind this mean effect might be hiding important gender heterogeneity of the ECC policy impact. I investigate it in the next step.

Table 5.2: Impact of ECC policy exposure on measures of non-cognitive outcomes for later affected cohorts (aged 3 to 5).

	(1) In Care	(2) Hyperactivity	(3) Emotional	(4) Peer	(5) Behavioral	(6) Prosocial
Panel A: Ages 3-5: No Controls						
Policy Exposure	0.1054***	0.0765	0.0422	0.0270	-0.0212	-0.0761
	(0.0246)	(0.0548)	(0.0580)	(0.0542)	(0.0577)	(0.0673)
N	3,376	3,452	3,451	3,451	3,454	3,461
Mean	0.222	3.334	1.657	1.436	2.109	7.640
Panel B: Ages 3-5: With Controls						
Policy Exposure	0.1093***	0.1005*	0.0506	0.0456	0.0069	-0.1071
	(0.0233)	(0.0562)	(0.0560)	(0.0528)	(0.0540)	(0.0683)
N	3,376	3,452	3,451	3,451	3,454	3,461
Mean	0.222	3.334	1.657	1.436	2.109	7.640

Dependent variables: In Care (under age 3), Hyperactivity, Emotional, Peer, Behavioral Problems, Pro-social behavior are standardized. For each dependent variable I show the pre-policy mean for fast-expanding counties, and the coefficient from a regression on the Policy Exposure dummy. All regressions include county, wave, and observation period fixed effects. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, child's gender, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Gender differences

I hypothesize that boys and girls respond differently to early non-parental care, with boys exposed to early and extensive childcare before age two experiencing worse outcomes. To test this, I estimate Equation 5.1 separately for boys and girls from earlier and later exposed cohorts.

When looking at the ECC policy effects in earlier exposed cohorts separated by gender (Table 5.3), we see that the policy had more pronounced negative effect on boys: boys exposed to greater ECC access exhibit less pro-social behavior (94.71% SD) and marginally more behavioral problems (73.48% SD). For girls the estimates are smaller (though not precisely estimated) – e.g., 26.18% SD for pro-social behavior – and show a mixed picture: with improvements in some skills and deterioration in others.

Table 5.3: Gender heterogeneity of ECC policy exposure on measures of non-cognitive outcomes for earlier affected cohorts (aged 6-10).

	(1) In Care	(2) Hyperactivity	(3) Emotional	(4) Peer	(5) Behavioral	(6) Prosocial
Panel A: Boys						
Policy Exposure	0.1136***	-0.1148	-0.0152	0.0197	0.1344*	-0.1815**
	(0.0331)	(0.0829)	(0.0789)	(0.0777)	(0.0750)	(0.0773)
N	3,860	3,864	3,864	3,865	3,865	3,871
Mean	0.205	3.838	1.870	1.475	2.172	7.866
Panel B: Girls						
Policy Exposure	0.0785***	0.0027	-0.0304	0.0188	-0.0416	-0.0312
	(0.0292)	(0.0716)	(0.0830)	(0.0640)	(0.0687)	(0.0699)
N	3,711	3,716	3,720	3,722	3,721	3,722
Mean	0.257	2.888	1.951	1.217	1.753	8.454

Dependent variables: In Care (under age 3), Hyperactivity, Emotional, Peer, Behavioral Problems, Pro-social behavior are standardized. For each dependent variable I show the pre-policy non-standardized mean for fast-expanding counties, and the coefficient from a regression on the Policy Exposure dummy. All regressions include county, wave fixed, and observation period effects, and controls. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age and gender, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

In the next step I repeat the analysis on sub-samples of boys and girls from later affected cohorts (Table 5.4) and find that boys with greater ECC access experience statistically significant higher hyperactivity/inattention and peer problems, while also exhibiting less pro-social behavior (126.39%, 170.83%, and 156.20% SD, respectively). The results for boys show that policy-initiated greater ECC access places them on at-risk levels – defined as an increase of 1 SD above the mean (NICHD, 2003a) – of problem behavior. The results for girls are, again, smaller (20.93% SD for pro-social behavior), mixed and not statistically significant.

Table 5.4: Gender heterogeneity of ECC policy exposure on measures of non-cognitive outcomes for later affected cohorts (aged 3-5).

	(1) In Care	(2) Hyperactivity	(3) Emotional	(4) Peer	(5) Behavioral	(6) Prosocial
Panel A: Boys						
Policy Exposure	0.0744**	0.1995**	0.1067	0.1937***	0.0563	-0.1977**
	(0.0336)	(0.0819)	(0.0815)	(0.0722)	(0.0873)	(0.1003)
N	1,705	1,740	1,740	1,741	1,743	1,746
Mean	0.213	3.462	1.609	1.517	2.179	7.350
Panel B: Girls						
Policy Exposure	0.1388***	0.0211	0.0414	-0.1184	-0.0795	-0.0459
	(0.0348)	(0.0786)	(0.0773)	(0.0790)	(0.0906)	(0.0996)
N	1,671	1,712	1,711	1,710	1,711	1,715
Mean	0.230	3.202	1.707	1.351	2.036	7.938

Dependent variables: In Care (under age 3), Hyperactivity, Emotional, Peer, Behavioral Problems, Prosocial behavior are standardized. For each dependent variable I show the pre-policy non-standardized mean for fast-expanding counties, and the coefficient from a regression on the Policy Exposure dummy. All regressions include county, wave, and observation period fixed effects, and controls. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Robustness checks

I first check the robustness of my findings by adding to the controls in my model a variable for policy lead (*PolicyLead*) of the binary variable *PolicyExposure*. The *PolicyLead* variable checks whether the policy already had an effect before it actually took place – in observation period 1 – and serves as a placebo test to check whether there is an effect where we expect none. Table 5.5 presents the results.

The coefficient on the policy lead is not statistically significant except for emotional problems in observation period 1 in the earlier affected cohorts' sample. This was also evident from the common time trend graph (Figure 5.B1) and the balancing table (Table 5.B1), where the trends between the fast- and slow-expanding counties diverge for this outcome in the observation period 1. For other child development outcomes the coefficient for the policy lead is not statistically significant.

Table 5.5: Placebo test with lead on the impact of ECC policy exposure on measures of non-cognitive outcomes for earlier (aged 6-10) and later (aged 3-5) affected cohorts.

	(1) ECC	(2) Hyperactivity	(3) Emotional	(4) Peer	(5) Behavioral	(6) Prosocial
Panel A: Ages 6-10						
Policy Exposure	0.0997***	-0.0511	-0.0659	-0.0010	0.0376	-0.1197**
	(0.0226)	(0.0547)	(0.0596)	(0.0520)	(0.0516)	(0.0553)
Policy Lead	0.0137	0.0436	-0.1265*	-0.0679	-0.0331	-0.0189
	(0.0259)	(0.0607)	(0.0672)	(0.0608)	(0.0617)	(0.0579)
N	7,571	7,580	7,584	7,587	7,586	7,593
Mean	0.230	3.379	1.909	1.350	1.969	8.151
Panel B: Ages 3-5						
Policy Exposure	0.1167***	0.1086	-0.0137	0.0455	-0.0202	-0.0708
	(0.0287)	(0.0771)	(0.0716)	(0.0696)	(0.0757)	(0.0902)
Policy Lead	0.0136	0.0151	-0.1340	-0.0004	-0.0551	0.0745
	(0.0426)	(0.1018)	(0.1037)	(0.0997)	(0.1042)	(0.1133)
N	3,376	3,452	3,451	3,451	3,454	3,461
Mean	0.222	3.334	1.657	1.436	2.109	7.640

Dependent variables: In Care (under age 3), Hyperactivity, Emotional, Peer, Behavioral Problems, and Pro-social behavior are standardized. For each dependent variable I show the number of observations, the pre-policy non-standardized mean for fast-expanding counties, and the coefficient from a regression on the Policy Exposure and Policy Lead dummies. All regressions include county, wave, and observation period fixed effects, and controls. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age and gender, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

For the next robustness check I choose a different expansion period instead of 2002 to 2014 and repeat my analysis with 2002 to 2015 and 2002 to 2016 expansion periods. The results stay qualitatively similar. Tables 5.D1, 5.D2, 5.D3, and 5.D4 in Appendix present the results for the 2002 to 2016 expansion period.[27]

Another concern relates to the choice of my treatment variable where I chose to split the counties by their expansion speed based on the median, which could seem arbitrary. I repeat my analysis and let my treatment variable to vary in its treatment intensity, allowing for continuity in treatment. In this generalized model, I am able to exploit the full variation in local ECC coverage. I re-estimate Equation 5.1 where the $PolicyExposure_{ct}$ is now continuous and captures the ECC coverage rate for county c in year t. It equals 0 if a child was born before 2006 and for $t \in [2006; 2014]$ $PolicyExposure_{ct} = \frac{\sum_{j=0}^{2} ECCcoverage_{ct+j}}{3}$. Thus, for children born in county c after 2005 $PolicyExposure_{ct}$ represents an average ECC coverage rate for the next three years. Tables 5.D5, 5.D6 and 5.D7 in Appendix present the estimation results. The results for boys are qualitatively similar. The results for girls indicate skill improvements (albeit smaller and not precisely estimated for later cohorts).

[27] The results for the 2002 to 2015 expansion period are available upon request.

However, girls from earlier affected cohorts show statistically significant improvements in their non-cognitive skills.

A threat to identification under DiD would be if another reform affected my treatment and control groups differently. There was another childcare reform in 1996 that expanded kindergarten access to three to six year olds. This reform had neutral or positive effect on child development (see, e.g., Bach *et al.*, 2019; Cornelissen *et al.*, 2018; Dustmann *et al.*, 2013; Kuehnle and Oberfichtner, 2020). If, for example, counties that expanded faster during the 1996 reform also expanded faster during the 2005 reform, then we could expect children's development in these counties to follow different time trends. However, Table 5.B1 in Appendix shows that my child development measures do not differ in their pre-trends.

The German government also introduced a parental leave reform in 2007 that changed financial incentives for childbearing by shifting the means-tested parental leave benefit that mainly targeted low-income families to a benefit that is tied to women's pre-birth earnings, thus, benefiting higher-earning women, with a minimum benefit being offered to all mothers. This reform was enacted by the federal government and applied to all German counties. Wave fixed effects should absorb the differences in effects due to the reform.

However, one might suggest that the 2007 reform might have affected fast- and slow-expanding counties systematically differently, i.e., the reform had a larger impact on birth rates and disproportionately changed parental composition in fast-expanding counties. For example, Raute (2019) finds that this reform had a positive effect on fertility that was primarily driven by higher-educated, higher-earning women. Thus, there might be a concern that the children born after the 2007 reform in fast-expanding counties are more likely to be born to more advantaged parents, which can influence their developmental outcomes. Also, if the fertility response was stronger (e.g., on intensive margin) in fast-expanding counties, then children in those counties might be more likely to have more siblings and will have to share their parents' attention and resources. However, if I regress parental and family composition variables (household SES, education, and the number of older and younger siblings) on the interaction of a post-2007-reform dummy with my dummy for fast- and slow-expanding counties, I find no indication that there is a change in those variables in my samples (Table 5.D8 in Appendix). Thus, I do not expect that the 2007 reform affected fast- and slow-expanding counties differently with respect to family and parental composition.

In my analysis I control for county-specific fixed effects, therefore, it is not necessary that the ECC expansion was unrelated to time-invariant county-specific characteristics. However, it is useful to understand the determinants of the expansion across counties. Additionally, there is still a concern that fast- and slow-expanding counties might have developed differently over time, i.e., expansion of ECC supply could be accompanied by changes in counties' socioeconomic composition. Thus, it is important to look into other sources that may be responsible for the different trajectories of the treatment and control groups. I check this by looking at the time trends in counties' socioeconomic composition over time using county-level INKAR data.

Figure 5.D1 in Appendix plots county socioeconomic (female employment, fertility, net migration, foreign share, GDP, and population share under age three) characteristics over

time. I also conduct a formal test (Tables 5.D9 and 5.D10 in Appendix). Here we see that fast-expanding counties had a marginally more negative pre-trend only for net migration. There were no changes in county composition post reform, except for fertility rate.

The latter finding is in accord with Bauernschuster *et al.* (2016), who also find that fertility has increased in response to the 2005 reform.[28] They document that the response was stronger at the intensive than extensive margin. Notably, this increase in fertility is not reflected in the number of siblings that the cohorts of children affected by the reform have (see Table 5.B2). Thus, it is unlikely that any changes in county characteristics are responsible for the uncovered effect.

5.7 Discussion

Alternative explanations

My results represent reduced form findings and are subject to various interpretations. One such explanation is that the findings reflect a transitional effect of different ECC expansion stages and are related to initial lower quality of care in fast-expanding counties. To check this I use data from the Statistics of the Child and Youth Services (*Statistik der Kinder- und Jugendhilfe*) that include structural daycare quality measures (number of children per caregiver and caregivers' educational qualifications). Figure 5.D2 and Tables 5.D9, and 5.D10 in Appendix show that the staff per child ratios and the share of staff with childhood education degrees are very similar in both types of counties, indicating that it is not a change in ECC quality that drives my results.

Another potential concern relates to a change in parental reporting. Specifically, parental assessment of their child's development could be associated with the time a parent spends with the child: parents who spend more time with children at home and can observe them throughout the day might become less critical; the opposite could be true if parents send children to ECC. Additionally, parents can report more negative outcomes if childcare provider bring up certain behaviors (not noticed previously by the parents) to their attention. Therefore, one might suggest that the ECC policy might have induced a change in parental assessment resulting from a change in daycare arrangement and not actually from the reduction in children's skills.

I argue that this concern does not apply to this case since the parental assessments are conducted at future points in time as the treatment – when children are aged three to 10 – and, thus, are out of ECC. Furthermore, at the time of assessment most children have already been exposed to some type of out-of-family childcare other than ECC, i.e., preschool,[29] and all children start primary school at age six. Therefore, parents' reference points when evaluating children's behavior are other children in preschool or primary school. Additionally, my findings are consistent with the results from studies that use other reporting measures, e.g., teachers' or caregivers' reports (as in the NICHD Study), providing confidence that it is not due to parental reporting bias.

[28] The 2005 and 2007 reforms by the German government were explicitly targeting fertility levels increase.
[29] 85% of three- to five-year olds in my sample have already been in some type of out-of-family care.

Mechanisms and future research

The non-economic literature suggests two concurrent mechanisms for the uncovered ECC effect. One of them is physiological, suggesting that children in daycare are exposed to higher levels of stress. Two potential sources of stress for children in ECC are maternal separation stress and stress from peer interactions (see Schore, 2017; Vermeer and van IJzendoorn, 2006). The evidence on this mechanism is still missing from the economic literature.

The second mechanism is worse parent-child interaction since ECC reform not only affects children by incentivizing out-of-family care use, but also affects parents by increasing maternal labor supply that could contribute to greater parental and family stress. The economic literature provides supporting evidence: the Quebec reform significantly increased maternal labor supply and led to worse parenting (aversive, hostile, ineffective and inconsistent) and changes in parental inputs which could negatively contribute to child development (Baker *et al.*, 2008). Home environment and parental investments changed disproportionately for boys than girls, potentially contributing to different outcomes (Kottelenberg and Lehrer, 2018). In line with this, Baker and Milligan (2016) show that boys and girls receive different parental home inputs and Bertrand and Pan (2013) document that non-cognitive returns to parental inputs differ markedly by gender: with boys' non-cognitive development, unlike girls', being extremely responsive to the change in such inputs.

Emerging studies also suggest that sensitive parenting plays a role in children's physiological regulation of stress (Hostinar *et al.*, 2014), implying that both mechanisms are likely to work in tandem creating a vicious cycle with worse child behavior negatively influencing parenting and worse parenting negatively influencing child behavior. Thus, both mechanisms indicate important avenues for future research because they can improve our understanding of physiological pathways in ECC effects on child development and the role of parenting and parental investments in human capital formation in early childhood, to design better ECC and family leave policies.

Limitations and discussion

The current study's limitation is that the outcomes for earlier and later affected cohorts are measured during different ages. Therefore, the findings could include a fading age effect, where the effect is more pronounced in preschool age and fades out when children reach primary school age. Notably, the non-economic and economic literature documents the effect persistence in primary school ages and even in adolescence (see, e.g., Baker *et al.*, 2019; NICHD, 2005; Vandell *et al.*, 2010), indicating – if any – a limited fade-out effect.

My results are in line with the economic literature that finds negative effects of universal ECC policies on children's non-cognitive skills (see, e.g., Baker *et al.*, 2008, 2019; Kottelenberg and Lehrer, 2014, 2018). The difference in findings between Quebec policy – where research finds more pronounced skill declines – and German reform is likely to be due to the fact that children in Germany were exposed to less extensive ECC and at older ages, since parental leave is paid (at least partially) up to two years, and it provides 36 months of job-protection for a mother, while parental leave in Quebec is shorter and the reform

incentivized full-day care use and increased mothers' full-time employment (Lefebvre and Merrigan, 2008).[30]

My results differ from Felfe and Lalive (2018), who find positive effects on socioemotional skills for children who were least likely to attend ECC, especially for boys and children from disadvantaged families. They study the initial cohorts affected by the policy expansion (born between July 2005 and June 2008): these cohorts had fewer full-time slots and the age at which daycare became available to them was older (between ages two and two and a half) than for later cohorts (see Figure 5.C1 in Appendix).[31]

The economic literature indicates that children who start childcare after age one and a half (girls)/two (boys) might benefit from stimulating social and cognitive environmental depending on the the quality difference between home and ECC environments. It also shows that girls start to benefit earlier (between age one and a half and two), while boys benefit more later (between age two and two and a half). Provided that the affected cohorts in Felfe and Lalive (2018) were starting daycare between ages two and two and a half and if boys from disadvantaged families were the most treatment-resistant, this explains why they find relatively larger positive effect on their socioemotional development.

Drawing conclusions about the ECC policy based on the effects found from the initial expansion could be misleading since the population of children eventually affected will be different with respect to ECC starting age and hours in care. Also, more children now receive access to extensive ECC even at younger age (since 2013 there is a legal claim (*Rechtsanspruch*) to an ECC place once a child turns one). This implies that my results might be an underestimate of the effects for yet later cohorts, who receive greater access to extensive ECC closer to age one.

Policy implications

My findings indicate that initiation of extensive ECC before age two could be harmful to boys' non-cognitive development. This is disconcerting since non-cognitive skills are more predictive of later life success than cognitive skills, especially for boys (Heckman *et al.*, 2013).

Policymakers can implement a policy similar to the academic red-shirting based on a child's individual disposition, where age-eligible children postpone ECC entrance to allow more time for their socioemotional development. Also greater support should be provided to families with children under age two, i.e., building awareness and providing parents with skills to support positive parenting. Thus, parents should be informed about potential harmful effects of early and extensive childcare, which allows them to make better-informed decisions about whether, when and for how long to send their children to day-

[30] Paid (unpaid) parental leaves in Quebec can be taken for up to 35(65) weeks (Statistics Canada, 2021; Quebec Government, 2022). 83% of mothers return to work within 27 to 52 weeks after getting their child (Statistics Canada, 2021), while in Germany employment rate of women with children under age three was around 54% in 2014 (OECD, 2014b).

[31] The daycare access expanded gradually with older children receiving access first. This is supported by the observation that twice as many parents with children aged two to three demand daycare compared to parents of children aged one to two (Bien *et al.*, 2007).

care. Additionally, providing parents with information on good parenting practices and reducing parental stress could ensure that parenting is not negatively affected.

The current findings together with previous evidence also provide support for more targeted use of pubic funds – with respect to which places to subsidize (i.e., less subsidies to full-time places for very young children and more to ages two and older; and more subsidies to children from specific groups who benefit more from such policies). For example, the evidence shows that children from disadvantaged families start to benefit more around the age one and a half (girls)/two (boys) if their out-of-family ECC environment is more stimulating than their family environment (see Drange and Havnes, 2019; Felfe and Lalive, 2018; Gathmann and Sass, 2012; Peter *et al.*, 2016); while children from advantaged families might benefit less or even be negatively affected (see Felfe and Lalive, 2018; Fort *et al.*, 2020), implying that their home environment is more stimulating than out-of-family ECC environment.

Additionally, the paper adds to the quality-quantity discussion. Specifically, what type of quality-quantity matters and when. While the literature suggests, that it is the care quality with respect to caregiver's sensitivity and responsiveness that is more important during the first years and a half or two years of child's life during the socioemotional brain development stage; it is more cognitively, motor and linguistically stimulating care environment that matters more after that age during the consequent brain development stage that starts around the age of one and a half or two (see, e.g., Luby *et al.*, 2022). Thus, this raises a question *who* can provide the best type of care at each of these developmental stages, i.e., parents, informal, or formal caregivers. The answer to this question can guide public policy.

Finally, declining fertility and aging population underline the importance of making greater investment in human capital development through the life cycle – especially during the earliest ages due to the greater rate of return on such investments (see Heckman, 2006) – to increase future workers' productivity and to ensure sustainable economic growth. Thus, when introducing public policies aiming to boost human capital development, it is crucial that policymakers understand what kind of investment (parental vs out-of-family ECC), at what age and for whom will be the most beneficial to avoid unintended dis-investments in human capital formation during the earliest ages (implying a negative mirror effect in skills' formation).

5.8 Conclusion

In recent years more children in high-income countries start to attend extensive childcare early. This is due in part to the introduction of universal subsidized care for children under age three. To a notable extent, these policies are being introduced in a manner that does not fully account for existing evidence from economic and non-economic literature regarding ECC effects on children's development. To policymakers' defense, the accumulated extensive evidence from these fields has been largely disconnected, providing limited guidance when creating ECC policies.

The current investigation fills this gap and applies interdisciplinary perspective by connecting the findings from non-economic and economic fields to improve our understanding of ECC effects on children's development. It highlights the importance of considering the cumulative lifetime quantity of childcare (combination of ECC starting age and mean weekly care hours). By focusing on the ECC quantity, a consistently theory- and evidence-supported hypothesis emerges: early, extensive and continuous childcare before age one and a half (for girls)/two (for boys) is harmful for children's socioemotional and behavioral functions; the adverse effects of ECC before age two are especially pronounced among boys. This is disconcerting given the popularity of ECC policies that affect large numbers of children and promote early and extensive non-parental childcare during this age. This is especially apparent in consideration of the research regarding the importance of non-cognitive skills for prediction of adult life satisfaction and labor market success.

I contribute to non-economic literature by providing causal estimates of the ECC's impact on children's development using Difference-in-Differences design in the quasi-experimental setting of the 2005 ECC reform in Germany applied to KiGGS data. I provide an ecological perspective of the ECC policy's effect on children's development. I also contribute to economic literature by looking beyond the period of initial ECC expansion, since the cohorts of children ultimately affected by later expansion receive greater access to extensive care at younger ages.

I find evidence of negative effects on non-cognitive skills that are more pronounced in later affected cohorts that were exposed to more extensive care and at earlier ages, concentrated among boys. Specifically, in later affected cohorts, boys face a statistically significant reduction in pro-social behavior and increased hyperactivity/inattention and peer problems. While in earlier affected cohorts that were treated less extensively and at older ages, the negative effects are less pronounced. These results suggest that the policy-induced ECC access expansion has a negative effect on boys' non-cognitive development. Girls – especially from earlier affected cohorts (exposed to extensive ECC around age two) – show indication of skill improvements.

To conclude, non-economic and economic literature indicates that ECC programs that increase access to early and extensive childcare might have negative effects on children's non-cognitive development, especially if exposing a child to out-of-family care too much too early in a child's life. The negative effects of ECC before age two are concentrated among boys. These findings could consistently be detected across skill assessment questionnaires, programs and countries. Thus, an important question for policymakers is how to use these lessons and implement public policies that minimize these negative impacts.

5.9 Appendix A

SDQ questionnaire

The following presents the list of questions used in construction of each category.[32]

Hyperactivity/inattention problems:

- Restless, overactive, cannot stay still for long
- Constantly fidgeting or squirming
- Easily distracted, concentration wanders
- Good attention span, sees work through to the end
- Thinks things out before acting

Emotional problems:

- Often complains of headaches, stomach-aches or sickness
- Many worries or often seems worried
- Often unhappy, depressed or tearful
- Nervous or clingy in new situations, easily loses confidence
- Many fears, easily scared

Peer relationship problems:

- Rather solitary, prefers to play alone
- Has at least one good friend
- Picked on or bullied by other children
- Gets along better with adults than with other children
- Generally liked by other children

Conduct/Behavior problems:

- Often fights with other children or bullies them
- Steals from home, school or elsewhere
- Often lies or cheats
- Often loses temper
- Generally well behaved, usually does what adults request

Pro-social behavior:

- Shares readily with other children, for example toys, treats, pencils
- Helpful if someone is hurt, upset or feeling ill

[32]The answers are either “not true”, “somewhat true”, or “certainly true”.

- Kind to younger children
- Considerate of other people's feelings
- Often offers help to others (parents, teachers, other children)

The maximum score for each category is 10, minimum zero. Higher scores on the first four subscales and lower score on the pro-social behavior subscale indicate greater problems.[33]

[33]See www.sdqinfo.com for more information.

Descriptive statistics

Table 5.A1: Descriptive statistics: earlier and later affected cohorts (aged 6-10 and 3-5).

-	Earlier Cohorts (Aged 6-10)				Later Cohorts (Aged 3-5)			
	Fast		Slow		Fast		Slow	
	Before	After	Before	After	Before	After	Before	After
Panel A: Child development outcomes								
Hyperactivity problems score	3.379	3.134	3.284	3.237	3.334	3.379	3.543	3.370
	(2.391)	(2.321)	(2.310)	(2.419)	(2.178)	(2.229)	(2.157)	(2.219)
Emotional problems score	1.909	1.687	1.921	1.765	1.657	1.458	1.598	1.331
	(1.869)	(1.853)	(1.853)	(1.896)	(1.612)	(1.594)	(1.623)	(1.483)
Peer problems score	1.350	1.173	1.408	1.199	1.436	1.306	1.420	1.228
	(1.627)	(1.593)	(1.584)	(1.537)	(1.506)	(1.535)	(1.551)	(1.396)
Behavioral problems score	1.969	1.788	2.032	1.822	2.109	1.904	2.221	2.015
	(1.548)	(1.592)	(1.557)	(1.506)	(1.333)	(1.404)	(1.496)	(1.582)
Prosocial behavior score	8.151	8.229	8.090	8.371	7.640	7.657	7.520	7.685
	(1.627)	(1.639)	(1.643)	(1.652)	(1.668)	(1.738)	(1.715)	(1.781)
Panel B: Out-of-family care (under the age of 3)								
In Care	0.230	0.546	0.201	0.409	0.222	0.718	0.164	0.545
	(0.421)	(0.498)	(0.401)	(0.492)	(0.416)	(0.450)	(0.371)	(0.498)
Panel C: Child characteristics								
Boy	0.517	0.507	0.513	0.519	0.508	0.498	0.512	0.525
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Child's age	7.924	7.523	7.923	7.548	4.022	3.969	4.008	4.017
	(1.406)	(1.245)	(1.366)	(1.211)	(0.814)	(0.818)	(0.816)	(0.825)
Panel D: Family characteristics								
HH tertiary education	0.469	0.486	0.380	0.399	0.439	0.509	0.370	0.434
	(0.499)	(0.500)	(0.485)	(0.490)	(0.497)	(0.500)	(0.483)	(0.496)
High HH SES	0.251	0.272	0.179	0.170	0.258	0.270	0.179	0.190
	(0.433)	(0.445)	(0.383)	(0.376)	(0.438)	(0.444)	(0.383)	(0.393)
HH immigrant	0.165	0.180	0.195	0.207	0.186	0.178	0.237	0.205
	(0.371)	(0.384)	(0.396)	(0.405)	(0.389)	(0.382)	(0.426)	(0.404)
Mother's age	38.153	38.745	37.873	38.148	34.133	35.256	33.998	34.875
	(5.190)	(5.458)	(5.151)	(5.615)	(5.256)	(5.340)	(5.131)	(5.490)
Father's age	41.098	42.139	40.923	41.626	37.607	39.091	36.999	38.478
	(5.995)	(5.933)	(6.103)	(6.044)	(5.780)	(6.345)	(5.618)	(6.222)
HH two parents	0.865	0.896	0.896	0.868	0.907	0.904	0.907	0.891
	(0.342)	(0.306)	(0.305)	(0.339)	(0.291)	(0.295)	(0.290)	(0.311)
Number of older siblings	1.067	0.756	1.080	0.765	1.158	0.716	1.210	0.751
	(0.824)	(0.766)	(0.819)	(0.744)	(0.776)	(0.760)	(0.768)	(0.787)
Number of younger siblings	0.892	0.581	0.886	0.568	0.769	0.425	0.771	0.494
	(0.819)	(0.698)	(0.830)	(0.684)	(0.808)	(0.586)	(0.825)	(0.635)
Resides in city ($> 100K+$)	0.208	0.258	0.272	0.311	0.212	0.243	0.282	0.330
	(0.406)	(0.438)	(0.445)	(0.463)	(0.409)	(0.429)	(0.450)	(0.471)

Displayed are the means of each variable with the standard deviation beneath in parenthesis. The data come from KiGGS. The sample contains 8,291 children from earlier cohorts (aged 6-10) and 3,579 children from later cohorts (aged 3-5). The earlier cohorts' sample is split by fast- and slow-expanding counties for Before (including cohorts born 1992-2005 and observed in Wave 0 (2004-2006), Wave 1 (2009-2012) and Wave 2 (2014-2017)) and After periods (including cohorts born 2006-2011 and observed in Wave 2 (2014-2017)). The later cohorts' sample is split by fast- and slow-expanding counties for Before (including cohorts born 1997-2003 and observed in Wave 0 (2004-2006)) and After periods (including cohorts born 2008-2014 and observed in Wave 2 (2014-2017)). In earlier cohorts' sample family characteristics are missing for a subset of observations: HH education (666 cases), HH immigrant (666 cases), mother's age (725 cases), father's age (1,466 cases), HH with two parents (611 cases), residence (576 cases). In later cohorts' sample family characteristics are missing for a subset of observations: HH education (50 cases), HH immigrant (39 cases), mother's age (61 cases), father's age (280 cases), HH with two parents (17 cases). All observations with missing characteristics are included in the analysis by defining dummy variables for the missing categories.

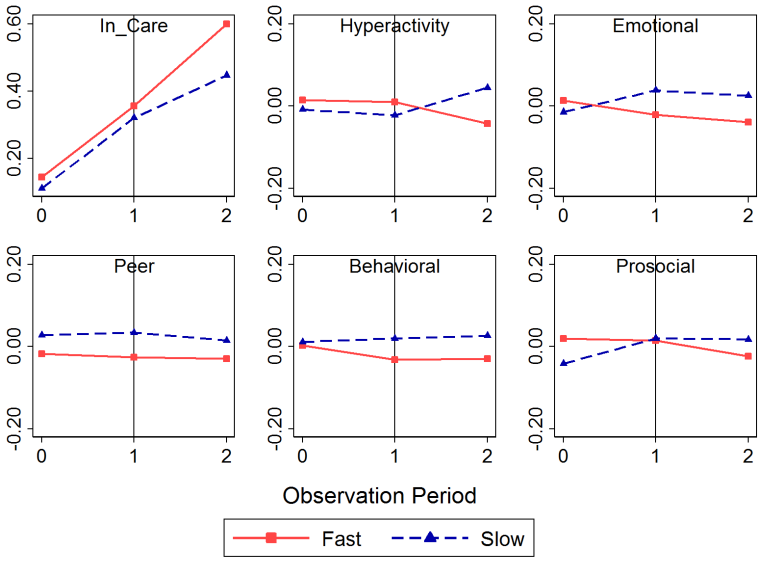
5.10 Appendix B

Common time trend

To causally interpret β a common trend assumption needs to hold. It states that absent of the reform developmental outcomes of children in fast- and slow-expanding counties would follow similar time trends. I visually inspect the slopes of child development outcomes before the reform took effect.

Figures 5.B1 and 5.B2 present visual evidence in the samples of children from earlier (aged 6-10) and later (aged 3-5) affected cohorts. I observe the cohorts of children that were not affected by the reform in observation periods 0 and 1[34,35] and see that child outcomes evolve in parallel between treatment and comparison counties. Only for emotional problems the common trend deviates in the pre-reform period.

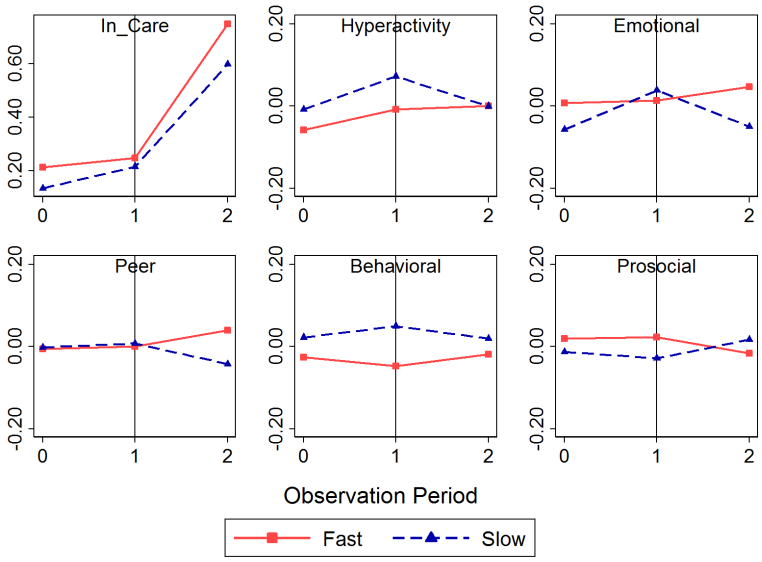
Figure 5.B1: Common time trend for outcomes (standardized) for children from earlier affected cohorts (aged 6-10) (KiGGS).



[34]Earlier affected cohorts: observation period 0 includes children born from 1992 to June 1998 and observed in Wave 0. Observation period 1 includes children born from June 1998 to 2005 and observed in Waves 1 and 2. Observation period 2 includes children born from 2006 to 2011 and observed in Wave 2.

[35]Later affected cohorts: observation period 0 includes children born from 1997 to June 2000 and observed in Wave 0. Observation period 1 includes children born from July 2000 to 2003 and observed in Wave 2. Observation period 2 includes children born from 2008 to 2014 and observed in Wave 2.

Figure 5.B2: Common time trend for outcomes (standardized) for children from later affected cohorts (aged 3-5) (KiGGS).



Balance test

To further test the common time trend assumption I perform balancing tests for child outcomes as well as baseline child and family characteristics. Important concerns are whether the timing of the ECC expansion relates to pre-reform trends in child development outcomes; and whether the speed of ECC expansion can be predicted by pre-reform trends in several baseline variables.

Table 5.B1 presents the results on the pre-reform differences in levels and trends between fast- and slow-expanding counties. The absence of difference in levels is not required for the DiD identification. More important is the lack of difference in pre-reform trends. I observe a few differences in pre-reform levels with respect to family characteristics: fast-expanding counties had larger shares of higher educated and high SES households, more single-parent households (in earlier affected cohorts' sample) and older fathers (in later affected cohorts' sample). However, the pre-reform trends for these variables are not significant.

Importantly, there are no significant differences in trends for child development outcomes between fast- and slow-expanding counties in the pre-reform period, except for a marginally significant negative trend for emotional problems in earlier affected cohorts' sample. In sum, the evidence provided in Table 5.B1 supports the main identification strategy.

Table 5.B1: Balancing table: Pre-reform differences between fast- and slow-expanding counties for earlier and later affected cohorts (aged 6-10 and 3-5).

	Fast vs Slow			
	Earlier Cohorts (Aged 6-10)		Later Cohorts (Aged 3-5)	
	Level	Trend	Level	Trend
Panel A: Child development outcomes				
Hyperactivity problems score	0.0409	0.0444	-0.0949*	-0.0480
	(0.0321)	(0.0719)	(0.0564)	(0.0438)
Emotional problems score	-0.0078	-0.1341*	0.0360	-0.0436
	(0.0346)	(0.0761)	(0.0511)	(0.0451)
Peer problems score	-0.0408	-0.0701	0.0138	-0.0213
	(0.0328)	(0.0738)	(0.0514)	(0.0447)
Behavioral problems score	-0.0419	-0.0384	-0.0761	-0.0284
	(0.0317)	(0.0701)	(0.0529)	(0.0432)
Prosocial behavior score	0.0341	-0.0314	0.0686	0.0168
	(0.0336)	(0.0665)	(0.0610)	(0.0441)
Panel B: Out-of-family care (under the age of 3)				
In care	0.0292*	0.0029	0.0552**	-0.0032
	(0.0152)	(0.0349)	(0.0273)	(0.0180)
Panel C: Child characteristics				
Boy	0.0044	-0.0030	-0.0017	0.0060
	(0.0082)	(0.0176)	(0.0166)	(0.0172)
Child's age	-0.0266	0.0165	0.0740	-0.0243
	(0.0236)	(0.0455)	(0.0723)	(0.0337)
Panel D: Family characteristics				
HH tertiary education	0.0883***	0.0148	0.0734**	0.0091
	(0.0186)	(0.0397)	(0.0287)	(0.0224)
High HH SES	0.0723***	0.0142	0.0803***	-0.0130
	(0.0157)	(0.0322)	(0.0271)	(0.0183)
HH immigrant	-0.0296	0.0081	-0.0517	-0.0274
	(0.0193)	(0.0348)	(0.0337)	(0.0207)
Mother's age	0.2572	0.3843	0.1791	0.0234
	(0.1726)	(0.3634)	(0.3132)	(0.2394)
Father's age	0.1267	0.0436	0.6695**	-0.1109
	(0.2071)	(0.4389)	(0.3192)	(0.2542)
HH two parents	-0.0315***	0.0211	-0.0005	-0.0079
	(0.0116)	(0.0239)	(0.0159)	(0.0114)
Number of older siblings	-0.0093	-0.0002	-0.0516	-0.0710**
	(0.0250)	(0.0521)	(0.0390)	(0.0306)
Number of younger siblings	0.0102	-0.0366	-0.0013	-0.0808**
	(0.0285)	(0.0591)	(0.0471)	(0.0378)
Resides in city ($> 100K+$)	-0.0641	0.0568	-0.0651	-0.0399
	(0.0604)	(0.1062)	(0.0875)	(0.0458)

Each row represents a different dependent variable. For the outcome variable in each row in the Level column I present the results from the following equation: $Y_{ic} = \beta_0 + \beta_1 Fast_c + \beta_2 D_t + \lambda_w + \epsilon_{ic}$. For the outcome variable in each row in Trend column I present the test of the common time trend assumption from the following equation: $Y_{ict} = \beta_0 + \beta_1 D_t * Fast_c + \beta_2 Fast_c + \beta_3 D_t + \lambda_w + \epsilon_{ict}$. $Fast_c$ is an indicator whether the county is a fast-expanding county and 0 otherwise. D_t is the observation period indicator (earlier cohorts' sample: $= 0$ for children born from 1992 to 1999 and observed in Wave 0; and $= 1$ for children born from 2000 to 2005 and observed in Waves 1 and 2; later cohorts' sample: $= 0$ for children born from 1997 to June 2000 and observed in Waves 0; and $= 1$ for children born from July 2000 to 2003 and observed in Waves 0). λ_w is wave fixed effects. Using data from the period before the policy implementation I report the coefficients β_1 for each corresponding outcome. Standard errors are clustered at the county and wave level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Additionally, I estimate a version of Equation 5.1 with demographic characteristics as dependent variables to ensure that there are no shifts in these characteristics between fast- and slow-expanding counties over time that might confound the observed changes in children's outcomes. Table 5.B2 presents the results. I see that almost all variables do not have a significant link to residing in fast-expanding counties post policy. The only exceptions are the increase in the share of children from two-parent households in fast-expanding counties post policy in earlier affected cohorts' sample and slightly younger children in later affected cohorts' sample. These variables are not usually linked to developmental concerns. However, this underlines the importance of controlling for child and family characteristics in the analysis.

Table 5.B2: Differences between fast- and slow-expanding counties post reform in demographic variables for earlier and later affected cohorts (aged 6-10 and 3-5).

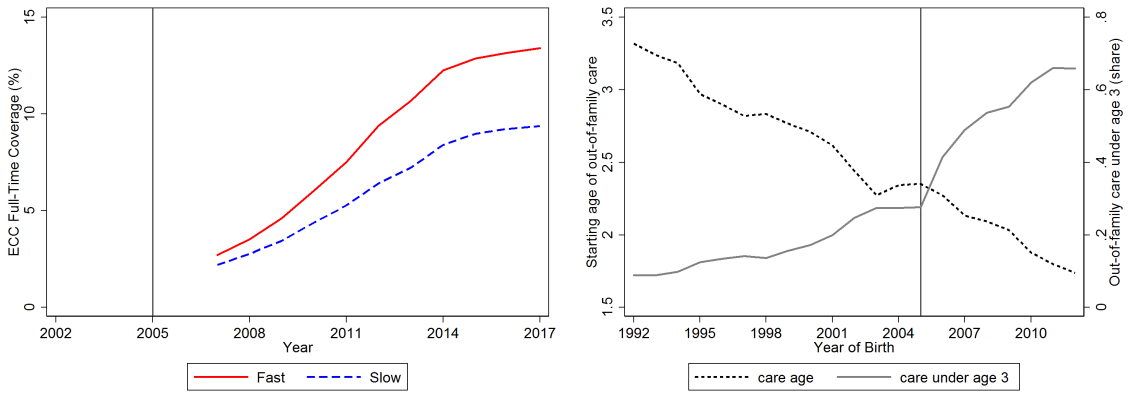
	Policy Exposure	
	Earlier Cohorts (Aged 6-10)	Later Cohorts (Aged 3-5)
Panel A: Child characteristics		
Boy	-0.0180	-0.0269
	(0.0212)	(0.0203)
Age	0.0053	-0.0590*
	(0.0574)	(0.0308)
Panel B: Family characteristics		
HH tertiary education	-0.0050	0.0026
	(0.0249)	(0.0314)
High HH SES	0.0242	-0.0005
	(0.0178)	(0.0231)
HH immigrant	0.0154	0.0328
	(0.0228)	(0.0281)
Mother's age	0.3062	0.3179
	(0.3007)	(0.2845)
Father's age	0.4271	0.1186
	(0.3444)	(0.3534)
HH two parents	0.0599***	0.0146
	(0.0168)	(0.0201)
Number of older siblings	0.0113	0.0332
	(0.0453)	(0.0450)
Number of younger siblings	0.0166	-0.0540
	(0.0375)	(0.0442)
Resides in city ($> 100K+$)	0.0171	0.0196
	(0.0149)	(0.0140)

Each row represents a different dependent variable. Regression includes county, observation period, and wave fixed effects. Standard errors are clustered at the county and wave level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

5.11 Appendix C

Early childcare coverage

Figure 5.C1: Full-time ECC coverage rate in fast- vs slow-expanding counties (*Kinder- und Jugendhilfe Statistik*) and average starting age of out-of-family care and out-of-family care for children under age 3 (%) (KiGGS).



(a) Full-time: fast vs slow

(b) ECC starting age

5.12 Appendix D

Robustness checks

2016-2002 expansion period

Table 5.D1: Impact of ECC policy exposure on measures of non-cognitive outcomes for children from earlier affected cohorts (aged 6-10).

	(1) Hyperactivity	(2) Emotional	(3) Peer	(4) Behavioral	(5) Prosocial
Panel A: No Controls					
Policy Exposure	-0.0198	-0.0008	0.0503	0.0054	-0.1389**
	(0.0600)	(0.0604)	(0.0551)	(0.0535)	(0.0569)
N	7,580	7,584	7,587	7,586	7,593
Mean	3.227	1.836	1.304	1.949	8.127
Panel B: With Controls					
Policy Exposure	-0.0330	-0.0132	0.0252	0.0008	-0.1247**
	(0.0575)	(0.0592)	(0.0538)	(0.0527)	(0.0561)
N	7,580	7,584	7,587	7,586	7,593
Mean	3.227	1.836	1.304	1.949	8.127

Dependent variables: In Care (under age 3), Hyperactivity, Emotional, Peer, Behavioral Problems, Pro-social behavior are standardized. For each dependent variable I show the pre-policy mean for fast-expanding counties, and the coefficient from a regression on the Policy Exposure dummy. All regressions include county, wave, and observation period fixed effects. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, child's gender, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 5.D2: Gender heterogeneity of ECC policy exposure on measures of non-cognitive outcomes for children from earlier affected cohorts (aged 6-10).

	(1) Hyperactivity	(2) Emotional	(3) Peer	(4) Behavioral	(5) Prosocial
Panel A: Boys					
Policy Exposure	0.0241	0.0194	0.0794	0.1201	-0.1630**
	(0.0833)	(0.0782)	(0.0772)	(0.0752)	(0.0771)
N	3,864	3,864	3,865	3,865	3,871
Mean	3.838	1.870	1.475	2.172	7.866
Panel B: Girls					
Policy Exposure	-0.0577	-0.0356	-0.0187	-0.0965	-0.0889
	(0.0711)	(0.0842)	(0.0644)	(0.0685)	(0.0707)
N	3,716	3,720	3,722	3,721	3,722
Mean	2.888	1.951	1.217	1.753	8.454

Dependent variables: In Care (under age 3), Hyperactivity, Emotional, Peer, Behavioral Problems, Pro-social behavior are standardized. For each dependent variable I show the pre-policy non-standardized mean for fast-expanding counties, and the coefficient from a regression on the Policy Exposure dummy. All regressions include county, wave, and observation period fixed effects, and controls. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 5.D3: Impact of ECC policy exposure on measures of non-cognitive outcomes for children from later affected cohorts (aged 3-5).

	(1) Hyperactivity	(2) Emotional	(3) Peer	(4) Behavioral	(5) Prosocial
Panel A: No Controls					
Policy Exposure	0.1483***	-0.0006	0.1504***	0.0552	-0.1709**
	(0.0541)	(0.0581)	(0.0536)	(0.0580)	(0.0664)
N	3,452	3,451	3,451	3,454	3,461
Mean	3.166	1.717	1.313	2.034	7.690
Panel B: With Controls					
Policy Exposure	0.1553***	0.0278	0.1273**	0.0678	-0.1616**
	(0.0558)	(0.0555)	(0.0536)	(0.0541)	(0.0686)
N	3,452	3,451	3,451	3,454	3,461
Mean	3.166	1.717	1.313	2.034	7.690

Dependent variables: In Care (under age 3), Hyperactivity, Emotional, Peer, Behavioral Problems, Pro-social behavior are standardized. For each dependent variable I show the pre-policy mean for fast-expanding counties, and the coefficient from a regression on the Policy Exposure dummy. All regressions include county, wave, and observation period fixed effects. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, child's gender, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 5.D4: Gender heterogeneity of ECC policy exposure on measures of non-cognitive outcomes for children from later affected cohorts (aged 3-5).

	(1) Hyperactivity	(2) Emotional	(3) Peer	(4) Behavioral	(5) Prosocial
Panel A: Boys					
Policy Exposure	0.2773***	0.0844	0.2627***	0.1818**	-0.3372***
	(0.0808)	(0.0827)	(0.0726)	(0.0905)	(0.0949)
N	1,740	1,740	1,741	1,743	1,746
Mean	3.333	1.696	1.391	2.086	7.402
Panel B: Girls					
Policy Exposure	0.0522	0.0226	-0.0077	-0.0674	-0.0199
	(0.0786)	(0.0758)	(0.0792)	(0.0902)	(0.1005)
N	1,712	1,711	1,710	1,711	1,715
Mean	3.002	1.737	1.235	1.983	7.974

Dependent variables: In Care (under age 3), Hyperactivity, Emotional, Peer, Behavioral Problems, Pro-social behavior are standardized. For each dependent variable I show the pre-policy non-standardized mean for fast-expanding counties, and the coefficient from a regression on the Policy Exposure dummy. All regressions include county, wave, and observation period fixed effects, and controls. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Continuous treatment

Table 5.D5: Impact of ECC policy exposure on measures of non-cognitive outcomes for children from earlier and later affected cohorts (aged 6-10 and 3-5).

	(1) Hyperactivity	(2) Emotional	(3) Peer	(4) Behavioral	(5) Prosocial
Panel A: Ages 6-10					
Policy Exposure	-0.0036	-0.0100**	-0.0037	-0.0046	-0.0052
	(0.0046)	(0.0047)	(0.0045)	(0.0045)	(0.0045)
N	7,580	7,584	7,587	7,586	7,593
Mean	3.203	1.842	1.320	1.963	8.115
Panel B: Ages 3-5					
Policy Exposure	0.0030	-0.0028	0.0057	-0.0013	-0.0128**
	(0.0046)	(0.0053)	(0.0047)	(0.0041)	(0.0058)
N	3,452	3,451	3,451	3,454	3,461
Mean	3.338	1.636	1.373	2.135	7.608

Dependent variables: Hyperactivity, Emotional, Peer, Behavioral Problems, Pro-social behavior are standardized. For each dependent variable I show the pre-policy mean, and the coefficient from a regression on the continuous Policy Exposure variable. All regressions include county, wave, and observation period fixed effects, and controls. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, child's gender, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 5.D6: Gender heterogeneity of ECC policy exposure on measures of non-cognitive outcomes for children from earlier affected cohorts (aged 6-10).

	(1) Hyperactivity	(2) Emotional	(3) Peer	(4) Behavioral	(5) Prosocial
Panel A: Boys					
Policy Exposure	0.0053	-0.0036	0.0014	0.0059	-0.0159**
	(0.0065)	(0.0063)	(0.0070)	(0.0066)	(0.0062)
N	3,864	3,864	3,865	3,865	3,871
Mean	3.615	1.804	1.456	2.139	7.824
Panel B: Girls					
Policy Exposure	-0.0118**	-0.0160**	-0.0076	-0.0145**	0.0045
	(0.0059)	(0.0068)	(0.0053)	(0.0057)	(0.0056)
N	3,716	3,720	3,722	3,721	3,722
Mean	2.768	1.881	1.177	1.776	8.422

Dependent variables: Hyperactivity, Emotional, Peer, Behavioral Problems, Pro-social behavior are standardized. For each dependent variable I show the pre-policy non-standardized mean, and the coefficient from a regression on the continuous Policy Exposure variable. All regressions include county, wave, and observation period fixed effects, and controls. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 5.D7: Gender heterogeneity of ECC policy exposure on measures of non-cognitive outcomes for children from later affected cohorts (aged 3-5).

	(1) Hyperactivity	(2) Emotional	(3) Peer	(4) Behavioral	(5) Prosocial
Panel A: Boys					
Policy Exposure	0.0124*	-0.0005	0.0154**	-0.0005	-0.0258***
	(0.0064)	(0.0072)	(0.0065)	(0.0062)	(0.0084)
N	1,740	1,740	1,741	1,743	1,746
Mean	3.392	1.598	1.425	2.143	7.344
Panel B: Girls					
Policy Exposure	-0.0076	-0.0032	-0.0080	-0.0056	-0.0032
	(0.0064)	(0.0063)	(0.0060)	(0.0067)	(0.0070)
N	1,712	1,711	1,710	1,711	1,715
Mean	3.143	1.712	1.304	2.019	7.934

Dependent variables: Hyperactivity, Emotional, Peer, Behavioral Problems, Pro-social behavior are standardized. For each dependent variable I show the pre-policy non-standardized mean, and the coefficient from a regression on the continuous Policy Exposure variable. All regressions include county, wave, and observation period fixed effects, and controls. Standard errors in parentheses, clustered at the county and wave level. Additional control variables include child's age, mother's and father's age, the number of older and younger siblings, parents' migrant status, household's education and SES, an indicator for two-parent family, and an indicator for residence in a large city ($> 100K+$). The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

2007 parental leave reform

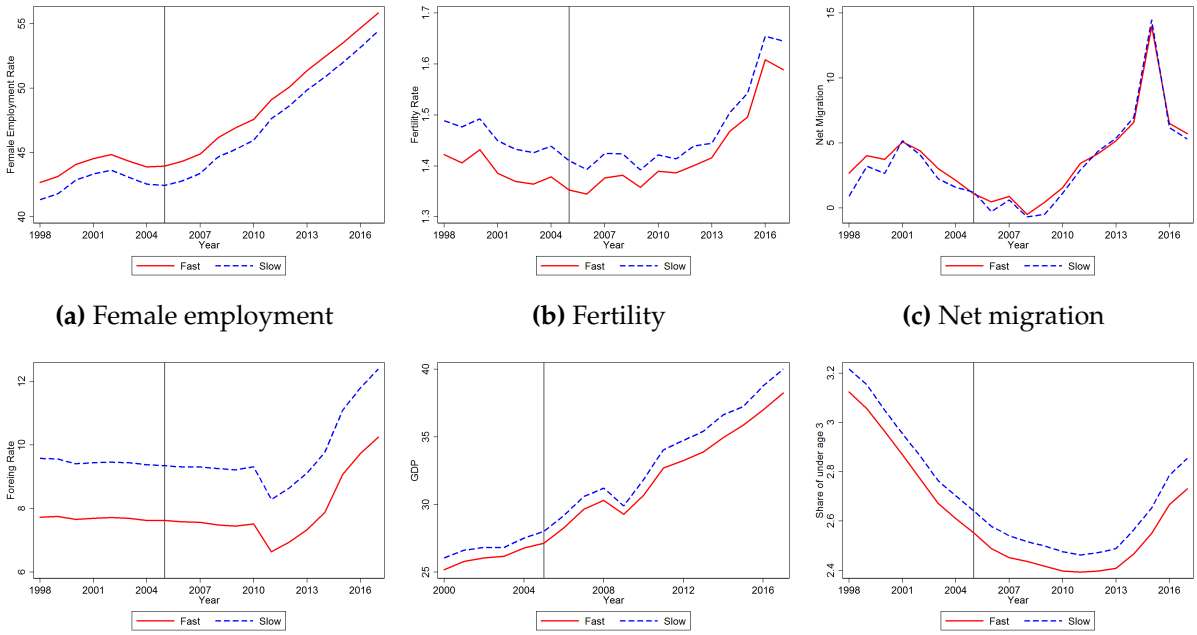
Table 5.D8: Differences between fast- and slow-expanding counties post 2007 reform in family characteristics for children from earlier (aged 6-10) and later (aged 3-5) affected cohorts.

	2007 Policy Exposure	
	Earlier Cohorts (Aged 6-10)	Later Cohorts (Aged 3-5)
Family characteristics		
HH tertiary education	0.0037	0.0004
	(0.0277)	(0.0313)
High HH SES	0.0290	-0.0007
	(0.0202)	(0.0232)
Number of older siblings	0.0145	-0.0530
	(0.0430)	(0.0444)
Number of younger siblings	0.0257	0.0343
	(0.0503)	(0.0449)

Each row represents a different dependent variable. Regression includes county, wave, and observation period fixed effects. Standard errors are clustered at the county and wave level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

County and childcare characteristics

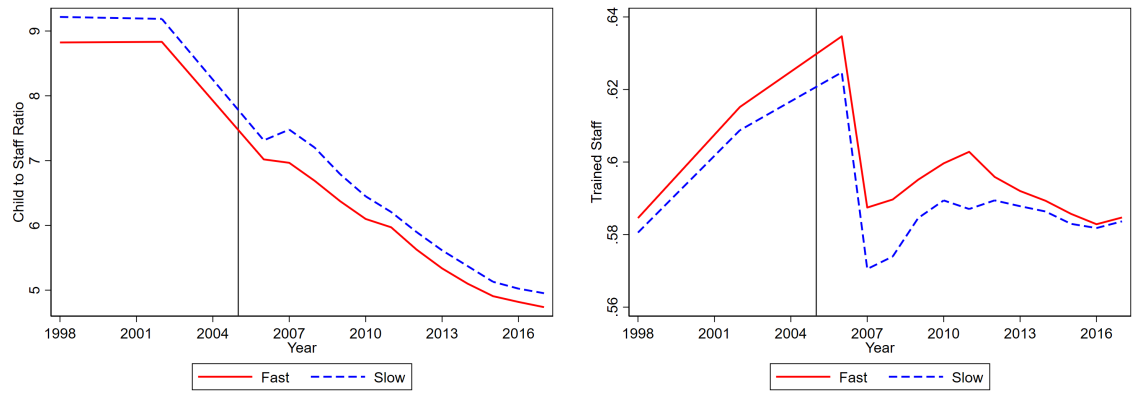
Figure 5.D1: County characteristics for fast- vs slow-expanding counties (*INKAR*).



(a) Female employment **(b)** Fertility **(c)** Net migration

(d) Foreign share **(e)** GDP **(f)** Population share under age 3

Figure 5.D2: Childcare characteristics for fast- vs slow-expanding counties (*Statistik der Kinder- und Jugendhilfe*).



(a) Staff to child ratio **(b)** Share of staff with chidhood education

Table 5.D9: Balancing table: Pre-reform differences between fast- and slow-expanding counties' characteristics.

	Fast vs Slow	
	Level	Trend
County characteristics		
Fertility	-0.0637***	0.0012
	(0.0132)	(0.0013)
Female employment	1.2722***	-0.0081
	(0.3766)	(0.0221)
GDP	-0.7636	0.0460
	(1.2514)	(0.0542)
Foreign rate	-1.7716***	0.0143
	(0.4738)	(0.0133)
Net migration	0.7505**	-0.1622*
	(0.3577)	(0.0881)
Population share under age 3	-0.0913***	0.0002
	(0.0284)	(0.0038)
Childcare characteristics		
Staff per child ratio	-0.4921***	-0.0137
	(0.1417)	(0.0189)
Share of trained staff	0.0052	0.0006
	(0.0092)	(0.0010)

Each row represents a different dependent variable. For the outcome variable in each row in the Level column I present the results from the following equation: $Y_{ct} = \beta_0 + \beta_1 Fast_c + \pi_t + \epsilon_{ct}$. For the outcome variable in each row in Trend column I present the test of the common time trend assumption from the following equation: $Y_{ct} = \beta_0 + \beta_1 Year_t * Fast_c + \lambda_c + \pi_t + \epsilon_{ct}$. $Fast_c$ is an indicator whether the county is a fast-expanding county, 0 otherwise. π_t and λ_c are year and county fixed effects, respectively. $Year_t$ is a continuous time variable. Using data from the period before (for years 1998-2004) the policy implementation I report the coefficients β_1 for each corresponding outcome. Standard errors are clustered at the county level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

Table 5.D10: Post-reform differences between fast- and slow-expanding counties' characteristics.

	Policy Exposure
County characteristics	
Fertility	0.0220***
	(0.0068)
Female employment	0.2465
	(0.1663)
GDP	-0.4969
	(0.4464)
Foreign rate	-0.0574
	(0.0911)
Net migration	-0.5680
	(0.3500)
Share under age 3	0.0008
	(0.0242)
Childcare characteristics	
Staff to child ratio	0.0733
	(0.1046)
Share of trained staff	0.0034
	(0.0043)

Each row represents a different dependent variable. For the outcome variable in each row I present the results from the following equation: $Y_{ct} = \beta_0 + \beta_1 Post_t * Fast_c + \lambda_c + \pi_t + \epsilon_{ct}$. $Fast_c$ is an indicator whether the county is a fast-expanding county, 0 otherwise. π_t and λ_c are year and county fixed effects, respectively. $Post_t$ is an indicator variable equal 0 in pre-reform period (year 1998-2004) and 0 – post reform (years 2005-2017). Standard errors are clustered at the county level. The stars represent significance at the following p-values: * p<0.1 ** p<0.05 *** p<0.01

6 Conclusion

This thesis focuses on the topic of health, environment, and behavior. All the studies are united by practical value that aims to support evidence-based decision-making for individuals and policymakers in real-world settings.

Chapter 2 explores fertility effects after the 1918-19 influenza pandemic in Sweden and finds that the pandemic had a long-lasting negative effect on fertility. The findings can help navigate fertility consequences of the COVID-19 pandemic and shape policies to counteract potential negative effects. The finding is especially important for the low fertility context: falling fertility rates can further accelerate population decline and drive rapid population aging and slower labor force growth, placing the sustainability of social security systems into question and contributing to slower productivity and economic growth in economies around the world.

Chapter 3 finds that some people fail to commit successfully to lose weight due to the overconfidence bias in their self-control abilities, and that these people could be helped by making greater investments. This finding can help individuals and policymakers in determining working weight loss strategies and understanding why they might not work for everyone. It also underlines the importance of using the opportunities to target certain people who are more likely to be overconfident about their self-control abilities and to encourage commitment during specific periods when people seek positive behavioral lifestyle change. Furthermore, the study has implications for assessing digital health technologies aiming at positive behavioral change. The results should be used together with addressing the upstream environmental drivers of overweight and obesity epidemic that contribute heavily to non-communicable disease burden.

The Chapter 4 finding that physicians equally adapt their treatment styles to their hospital and peer environments indicates that physicians' behavior is an important determinant of the geographical variation in treatment choices, and that this behavior is subject to environmental influences. This can inform policymakers interested in equitable and efficient healthcare use, and those targeting uneven distribution of healthcare services across geographic regions.

Finally, the Chapter 5 finding that early and extensive childcare could have a negative impact on children's development, especially for boys and children exposed at earlier ages, can influence the decisions of parents about whether, when, and for how long to send their children to daycare, and the decisions of policymakers on how to efficiently design and subsidize ECC and other family policies to mitigate these negative effects.

In their small manner, each paper seeks to align itself with UN's goals of sustainable development of our planet. In a time of growing social and financial pressures on social security systems, exacerbated by the COVID-19 pandemic, greater attention must be paid

not only by nations and institutions, but also by individuals seeking to do their part. The actions we take now as individuals and as societies will in many ways determine the well-being of our and future generations for decades to come. Thus, it is important to choose a plan of action that relies on a solid evidence base. This thesis aims to add to an evidence base used for better-informed individual and public decision-making to ensure economic growth, reduce inequalities, and provide equitable and efficient healthcare and early child-care.

Bibliography

AASSVE, A., CAVALLI, N., MENCARINI, L., PLACH, S. and BACCI, M. L. (2020). The COVID-19 pandemic and human fertility. *Science*, **369** (6502), 370–371.

—, —, —, — and SANDERS, S. (2021). Early assessment of the relationship between the COVID-19 pandemic and births in high-income countries. *Proceedings of the National Academy of Sciences*, **118** (36), e2105709118.

ABADIE, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, **59** (2), 391–425.

—, DIAMOND, A. and HAINMUELLER, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of Californias tobacco control program. *Journal of the American Statistical Association*, **105** (490), 493–505.

—, — and — (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, **59** (2), 495–510.

— and GARDEAZABAL, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, **93** (1), 112–132.

— and LHOUR, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, **116** (536), 1817–1834.

ABEL, A. B. (2002). An exploration of the effects of pessimism and doubt on asset returns. *Journal of Economic Dynamics and Control*, **26** (7-8), 1075–1092.

ABRAMITZKY, R., DELAVANDE, A. and VASCONCELOS, L. (2011). Marrying up: The role of sex ratio in assortative matching. *American Economic Journal: Applied Economics*, **3** (3), 124–157.

AGADJANIAN, V. and PRATA, N. (2002). War, peace, and fertility in Angola. *Demography*, **39** (2), 215–231.

AHNERT, L., RICKERT, H. and LAMB, M. E. (2000). Shared caregiving: Comparisons between home and child-care settings. *Developmental Psychology*, **36** (3), 339.

ALAM, S. A. and PÖRTNER, C. C. (2018). Income shocks, contraceptive use, and timing of fertility. *Journal of Development Economics*, **131**, 96–103.

ALLEN, L. (2017). Are we facing a noncommunicable disease pandemic? *Journal of Epidemiology and Global Health*, **7** (1), 5–9.

ALMOND, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 U.S. population. *Journal of Political Economy*, **114**, 672–712.

—, CURRIE, J. and DUQUE, V. (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature*, **56** (4), 1360–1446.

— and MAZUMDER, B. (2013). Fetal origins and parental responses. *Annual Review of Economics*, **5** (1), 37–56.

ÅMAN, M. (1990). *Spanska sjukan: den svenska epidemin 1918-1920 och dess internationella bakgrund*. Uppsala: Acta Universitatis Upsaliensis.

ANDREASEN, V., VIBOUD, C. and SIMONSEN, L. (2008). Epidemiologic characterization of the 1918 influenza pandemic summer wave in Copenhagen: Implications for pandemic control strategies. *The Journal of Infectious Diseases*, **197** (2), 270–278.

ANGRIST, J. (2002). How Do Sex Ratios Affect Marriage and Labor Markets: Evidence from America's Second Generation. *Quarterly Journal of Economics*, **117** (3), 997–1038.

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

ARNI, P., DRAGONE, D., GOETTE, L. and ZIEBARTH, N. R. (2021). Biased health perceptions and risky health behaviorstheory and evidence. *Journal of Health Economics*, **76**, 102425.

ASHRAF, N., KARLAN, D. and YIN, W. (2006). Deposit collectors. *Advances in Economic Analysis & Policy*, **5** (2).

AUGER, N., FRASER, W. D., SMARGIASSI, A., BILODEAU-BERTRAND, M. and KOSATSKY, T. (2017). Elevated outdoor temperatures and risk of stillbirth. *International Journal of Epidemiology*, **46** (1), 200–208.

AUGURZKY, B., BAUER, T. K., REICHERT, A. R., SCHMIDT, C. M. and TAUCHMANN, H. (2012). Does money burn fat? Evidence from a randomized experiment. *Ruhr Economic Paper*, (368).

BACH, M., KOEBE, J. and PETER, F. (2019). Long-run effects of universal childcare on personality traits. *DIW Berlin Discussion Paper*, (1815).

BAKER, M., GRUBER, J. and MILLIGAN, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, **116** (4), 709–745.

—, — and — (2019). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy*, **11** (3), 1–26.

— and MILLIGAN, K. (2016). Boy-girl differences in parental time investments: Evidence from three countries. *Journal of Human Capital*, **10** (4), 399–441.

BANDIERA, O., BARANKAY, I. and RASUL, I. (2010). Social incentives in the workplace. *The Review of Economic Studies*, **77** (2), 417–458.

BARBER, B. M. and ODEAN, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, **116** (1), 261–292.

BARKER, D. J. (1990). The fetal and infant origins of adult disease. *British Medical Journal*, **301** (6761), 1111.

BARNETT, A. G. and DOBSON, A. J. (2010). *Analysing seasonal health data*, vol. 30. Springer.

BARÓN, J. D. and COBB-CLARK, D. A. (2010). Are young people's educational outcomes linked to their sense of control? *Available at SSRN 1594792*.

BARRO, R. J., URSÚA, J. F. and WENG, J. (2020). *The Coronavirus and the Great Influenza Pandemic: Lessons from the "Spanish Flu" for the Coronavirus's potential effects on mortality and economic activity*. Tech. rep., National Bureau of Economic Research.

BARRY, J. M. (2004). *The Great Influenza: The Epic Story of the Greatest Plague in History*. Viking Penguin, New York.

BAUERNSCHUSTER, S., HENER, T. and RAINER, H. (2016). Children of a (policy) revolution: The introduction of universal child care and its effect on fertility. *Journal of the European Economic Association*, **14** (4), 975–1005.

BAVEL, J. (2001). Malthusian sinners: Illegitimate fertility and economic crisis. A case study in Leuven, 1846-1856. *Belg Tijdschr Nieuwste Geschied*, **31** (3-4), 371–401.

BEACH, B., CLAY, K. and SAAVEDRA, M. (2022). The 1918 influenza pandemic and its lessons for COVID-19. *Journal of Economic Literature*, **60** (1), 41–84.

—, FERRIE, J. P. and SAAVEDRA, M. H. (2018). *Fetal shock or selection? The 1918 influenza pandemic and human capital development*. Tech. rep., National Bureau of Economic Research.

BECKER, G. S. (1960). An economic analysis of fertility. In *Demographic and economic change in developed countries*, Columbia University Press, pp. 209–240.

— (1973). A theory of marriage: Part I. *Journal of Political Economy*, **81** (4), 813–46.

— (1974). A theory of marriage: Part II. *Journal of Political Economy*, **82** (2), S11–26.

BELSKY, J. (2001). Emanuel Miller Lecture: Developmental risks (still) associated with early child care. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, **42** (7), 845–859.

—, VANDELL, D. L., BURCHINAL, M., CLARKE-STEWART, K. A., MCCARTNEY, K., OWEN, M. T. and NETWORK, N. E. C. C. R. (2007). Are there long-term effects of early child care? *Child Development*, **78** (2), 681–701.

BENARTZI, S. and THALER, R. (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic perspectives*, **21** (3), 81–104.

BENGTSSON, T., CAMPBELL, C., LEE, J. Z. *et al.* (2004). *Life under pressure: Mortality and living standards in Europe and Asia, 1700-1900*. MIT Press.

— and DRIBE, M. (2014). The historical fertility transition at the micro level: Southern Sweden 1815-1939. *Demographic Research*, **30**, 493–533.

BERTRAND, M. and PAN, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*, **5** (1), 32–64.

BEYER, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, **59** (5), 960–970.

— and BOWDEN, E. M. (1997). Gender differences in seff-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, **23** (2), 157–172.

BHALOTRA, S., KARLSSON, M. and NILSSON, T. (2017). Infant health and longevity: Evidence from a historical intervention in Sweden. *Journal of the European Economic Association*, **15** (5), 1101–1157.

BIEN, W., RAUSCHENBACH, T. and RIEDEL, B. (2007). Wer betreut Deutschlands Kinder. *DJI-Kinderbetreuungsstudie. Berlin*.

BLAND, P. B. (1919). Influenza in its relation to pregnancy and labor. *American Journal of Obstetrics*, **79** (184), 97.

BLAND, S. H., O'LEARY, E. S., FARINARO, E., JOSSA, F. and TREVISAN, M. (1996). Long-term psychological effects of natural disasters. *Psychosomatic Medicine*, **58** (1), 18–24.

BLOOM-FESHBACH, K., SIMONSEN, L., VIBOUD, C., MØLBAK, K., MILLER, M., GOTTFREDSSON, M. and ANDREASEN, V. (2012). Natality decline and miscarriages associated with the 1918 influenza pandemic: The Scandinavian and United States experiences. *Journal of Infectious Diseases*, **204** (8), 1157–1164.

BOJKE, C., CASTELLI, A., STREET, A., WARD, P. and LAUDICELLA, M. (2013). Regional variation in the productivity of the English National Health Service. *Health Economics*, **22** (2), 194–211.

BOLTON, D., O'RYAN, D., UDWIN, O., BOYLE, S. and YULE, W. (2000). The long-term psychological effects of a disaster experienced in adolescence: II: General psychopathology. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, **41** (4), 513–523.

BONANNO, G. A., HO, S. M., CHAN, J. C., KWONG, R. S., CHEUNG, C. K., WONG, C. P. and WONG, V. C. (2008). Psychological resilience and dysfunction among hospitalized survivors of the SARS epidemic in Hong Kong: A latent class approach. *Health Psychology*, **27** (5), 659.

BONGAARTS, J. (1978). A framework for analyzing the proximate determinants of fertility. *Population and Development Review*, pp. 105–132.

BOTOSARU, I. and FERMAN, B. (2019). On the role of covariates in the synthetic control method. *The Econometrics Journal*, **22** (2), 117–130.

BOUCEKKINE, R., DESBORDES, R. and LATZER, H. (2009). How do epidemics induce behavioral changes? *Journal of Economic Growth*, **14** (3), 233–264.

BOWLBY, J. (1969). Attachment and loss v. 3 (vol. 1). *Random House. Furman, W., & Buhrmester, D.(2009). Methods and measures: The network of relationships inventory: Behavioral systems version. International Journal of Behavioral Development*, **33**, 470–478.

— (1973). *Attachment and loss, vol. II: Separation*, vol. 2. Basic Books New York.

— (1988). A secure base: Clinical applications of attachment theory (collected papers). *London: Tavistock*.

BREIMAN, L. (2001). Random forests. *Machine Learning*, **45** (1), 5–32.

BROWN, R. and THOMAS, D. (2018). On the long-term effecs of the 1918 U.S. influenza pandemic, unpublished Manuscript.

BROWNLEE, S., CHALKIDOU, K., DOUST, J., ELSHAUG, A. G., GLASZIOU, P., HEATH, I., NAGPAL, S., SAINI, V., SRIVASTAVA, D., CHALMERS, K. *et al.* (2017). Evidence for overuse of medical services around the world. *The Lancet*, **390** (10090), 156–168.

BRUCKNER, T. A., MODIN, B. and VÅGERÖ, D. (2014). Cold ambient temperature in utero and birth outcomes in Uppsala, Sweden, 1915–1929. *Annals of Epidemiology*, **24** (2), 116–121.

BRUNELLO, G. and SCHLOTTER, M. (2011). Non-cognitive skills and personality traits: Labour market relevance and their development in education & training systems. *IZA Discussion paper*.

BURKE, M. A., FOURNIER, G. and PRASAD, K. (2003). *Physician social networks and geographical variation in medical care*. Center on Social and Economic Dynamics Washington, DC.

BYRNES, J. P., MILLER, D. C. and SCHAFER, W. D. (1999). Gender differences in risk taking: a meta-analysis. *Psychological Bulletin*, **125** (3), 367–383.

CALIENDO, M., COBB-CLARK, D. A. and UHLENDORFF, A. (2015). Locus of control and job search strategies. *Review of Economics and Statistics*, **97** (1), 88–103.

CAMERER, C. and LOVALLO, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, **89** (1), 306–318.

CARNEIRO, P. M. and HECKMAN, J. J. (2003). Human capital policy.

CARTA, F. and RIZZICA, L. (2018). Early kindergarten, maternal labor supply and children's outcomes: Evidence from Italy. *Journal of Public Economics*, **158**, 79–102.

CASTRO, R., BEHRMAN, J. R. and KOHLER, H.-P. (2015). Perception of HIV risk and the quantity and quality of children: the case of rural Malawi. *Journal of Population Economics*, **28** (1), 113–132.

CAWLEY, J. and MEYERHOEFER, C. (2012). The medical care costs of obesity: An instrumental variables approach. *Journal of Health Economics*, **31** (1), 219–230.

— and PRICE, J. A. (2011). Outcomes in a program that offers financial rewards for weight loss. *Economic Aspects of Obesity*, p. 91.

— and — (2013). A case study of a workplace wellness program that offers financial incentives for weight loss. *Journal of Health Economics*, **32** (5), 794–803.

CDC (2022). Early brain development and health. https://www.cdc.gov/ncbddd/childdevelopment/early-brain-development.html, Last Accessed: 2022-07-23.

CHAN, D. C. (2016). Teamwork and moral hazard: Evidence from the emergency department. *Journal of Political Economy*, **124** (3), 734–770.

—, GENTZKOW, M. and YU, C. (2022). Selection with variation in diagnostic skill: Evidence from radiologists. *The Quarterly Journal of Economics*, **137** (2), 729–783.

CHANDRA, A., CUTLER, D. and SONG, Z. (2011). Who ordered that? The economics of treatment choices in medical care. *Handbook of Health Economics*, **2**, 397–432.

— and STAIGER, D. O. (2007). Productivity spillovers in health care: Evidence from the treatment of heart attacks. *Journal of Political Economy*, **115** (1), 103–140.

CHANDRA, S., CHRISTENSEN, J., MAMELUND, S.-E. and PANETH, N. (2018). Short-term birth sequelae of the 1918–1920 influenza pandemic in the United States: State-level analysis. *American Journal of Epidemiology*, **187** (12), 2585–2595.

— and YU, Y. (2015a). The 1918 influenza pandemic and subsequent birth deficit in Japan. *Demographic Research*, **33**, 313–326.

— and — (2015b). Fertility decline and the 1918 influenza pandemic in Taiwan. *Biodemography and Social Biology*, **61** (3), 266–72.

CHANG, C. Y. and OBERMEYER, Z. (2020). Association of clinical characteristics with variation in emergency physician preferences for patients. *JAMA Network Open*, **3** (1), e1919607–e1919607.

CHANG, W.-C., LI, T.-C. and LIN, C.-C. (2003). The effect of physician experience on costs and clinical outcomes of laparoscopic-assisted vaginal hysterectomy: A multivariate analysis. *The Journal of the American Association of Gynecologic Laparoscopists*, **10** (3), 356–359.

CHESNAIS, J.-C. (1992). *The demographic transition: Stages, patterns, and economic implications. A Longitudinal Study of Sixty-Seven Countries Covering the Period 1720-1984*. 9780198286592, Oxford University Press.

CHIRON, C., JAMBAQUE, I., NABBOUT, R., LOUNES, R., SYROTA, A. and DULAC, O. (1997). The right brain hemisphere is dominant in human infants. *Brain*, **120** (6), 1057–1065.

CHITKARA, K. and GERSHLICK, A. (2010). Second versus first-generation drug-eluting stents. *Journal of Interventional Cardiology*, **5** (1), 23–26.

CICCONE, S. J. (2011). Investor optimism, false hopes and the January effect. *Journal of Behavioral Finance*, **12** (3), 158–168.

CLARK, G. (2014). *The son also rises: Surnames and the history of social mobility*. Princeton University Press.

CLAY, K., LEWIS, J. and SEVERNINI, E. (2019). What explains cross-city variation in mortality during the 1918 influenza pandemic? Evidence from 438 US cities. *Economics & Human Biology*, **35**, 42–50.

COHAN, C. L. and COLE, S. W. (2002). Life course transitions and natural disaster: Marriage, birth, and divorce following Hurricane Hugo. *Journal of Family Psychology*, **16** (1), 14.

COHEN, P. N. (2021). Baby bust: Falling fertility in US counties is associated with COVID-19 prevalence and mobility reductions.

COMOLLI, C. L. (2017). The fertility response to the Great Recession in Europe and the United States: Structural economic conditions and perceived economic uncertainty. *Demographic research*, **36**, 1549–1600.

CONNING, J. and UDRY, C. (2007). Rural financial markets in developing countries. *Handbook of Agricultural Economics*, **3**, 2857–2908.

CORALLO, A. N., CROXFORD, R., GOODMAN, D. C., BRYAN, E. L., SRIVASTAVA, D. and STUKEL, T. A. (2014). A systematic review of medical practice variation in OECD countries. *Health Policy*, **114** (1), 5–14.

CORNELISSEN, T., DUSTMANN, C., RAUTE, A. and SCHÖNBERG, U. (2018). Who benefits from universal child care? Estimating marginal returns to early child care attendance. *Journal of Political Economy*, **126** (6), 2356–2409.

CORREIA, S., LUCK, S. and VERNER, E. (2022). Pandemics depress the economy, public health interventions do not: Evidence from the 1918 flu. *The Journal of Economic History*, **82** (4), 917–957.

CUNHA, F. and HECKMAN, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, **43** (4), 738–782.

CURRIE, J. (2001). Early childhood education programs. *Journal of Economic Perspectives*, **15** (2), 213–238.

—, MACLEOD, W. B. and VAN PARYS, J. (2016). Provider practice style and patient health outcomes: The case of heart attacks. *Journal of Health Economics*, **47**, 64–80.

CURRIE, J. M. and MACLEOD, W. B. (2020). Understanding doctor decision making: The case of depression treatment. *Econometrica*, **88** (3), 847–878.

CUTLER, D., SKINNER, J. S., STERN, A. D. and WENNBERG, D. (2019). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *American Economic Journal: Economic Policy*, **11** (1), 192–221.

DAEMEN, J. and SERRUYS, P. W. (2007). Drug-eluting stent update 2007: Part I: A survey of current and future generation drug-eluting stents: Meaningful advances or more of the same? *Circulation*, **116** (3), 316–328.

—, TANIMOTO, S., GARCÍA-GARCÍA, H. M. *et al.* (2007). Comparison of three-year clinical outcome of sirolimus-and paclitaxel-eluting stents versus bare metal stents in patients with ST-segment elevation myocardial infarction (from the RESEARCH and T-SEARCH Registries). *The American Journal of Cardiology*, **99** (8), 1027–1032.

DAI, H., MILKMAN, K. L. and RIIS, J. (2014). The fresh start effect: Temporal landmarks motivate aspirational behavior. *Management Science*, **60** (10), 2563–2582.

DEAUX, K. and FARRIS, E. (1977). Attributing causes for one's own performance: The effects of sex, norms, and outcome. *Journal of Research in Personality*, **11** (1), 59–72.

DEE, A., KEARNS, K., O'NEILL, C., SHARP, L., STAINES, A., O'DWYER, V., FITZGERALD, S. and PERRY, I. J. (2014). The direct and indirect costs of both overweight and obesity: a systematic review. *BMC Research Notes*, **7**, 1–9.

DEL BONO, E., WEBER, A. and WINTER-EBMER, R. (2015). Fertility and economic instability: The role of unemployment and job displacement. *Journal of Population Economics*, **28** (2), 463–478.

DELAHOY, M. J. (2020). Characteristics and maternal and birth outcomes of hospitalized pregnant women with laboratory-confirmed COVID-19 – COVID-NET, 13 States, March 1–August 22, 2020. *MMWR. Morbidity and Mortality Weekly Report*, **69**.

DELLA VIGNA, S. and MALMENDIER, U. (2006). Paying not to go to the gym. *The American Economic Review*, **96** (3), 694 – 719.

DÍAZ-ZAVALA, R. G., CASTRO-CANTÚ, M. F., VALENCIA, M. E., ÁLVAREZ-HERNÁNDEZ, G., HABY, M. M. and ESPARZA-ROMERO, J. (2017). Effect of the holiday season on weight gain: A narrative review. *Journal of Obesity*, **2017**.

DIGRANDE, L., NERIA, Y., BRACKBILL, R. M., PULLIAM, P. and GALEA, S. (2010). Long-term posttraumatic stress symptoms among 3,271 civilian survivors of the September 11, 2001, terrorist attacks on the World Trade Center. *American Journal of Epidemiology*, **173** (3), 271–281.

DONALDSON, D. and KENISTON, D. (2015). How positive was the positive check? Investment and fertility in the aftermath of the 1918 influenza in India, unpublished Manuscript.

DOYLE, J. J., EWER, S. M. and WAGNER, T. H. (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of Health Economics*, **29** (6), 866–882.

DRANGE, N. and HAVNES, T. (2019). Early childcare and cognitive development: Evidence from an assignment lottery. *Journal of Labor Economics*, **37** (2), 581–620.

DRIBE, M. and LUNDH, C. (2005). Finding the right partner: Rural homogamy in nineteenth-century Sweden. *International Review of Social History*, **50** (S13), 149–177.

—, — and NYSTED, P. (2007). Widowhood strategies in preindustrial society. *Journal of Interdisciplinary History*, **38** (2), 207–232.

DUSTMANN, C., RAUTE, A. and SCHÖNBERG, U. (2013). Does universal child care matter? Evidence from a large expansion in pre-school education. *Unpublished Manuscript*.

D'UVA, T. B., O'DONNELL, O. and VAN DOORSLAER, E. (2017). Who can predict their own demise? Heterogeneity in the accuracy and value of longevity expectations. *The Journal of the Economics of Ageing*, p. 100135.

ECKEL, C. C. and GROSSMAN, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results*, **1**, 1061–1073.

EDVINSSON, S. (2011). *A Country Doctor: Health Care in a Mid-Nineteenth-Century Swedish Remote Area*, Pickering and Chatto, p. 93114.

EKMAN, M., SJÖGREN, I. and JAMES, S. (2006). Cost-effectiveness of the Taxus paclitaxel-eluting stent in the Swedish healthcare system. *Scandinavian Cardiovascular Journal*, **40** (1), 17–24.

EPSTEIN, A. J. and NICHOLSON, S. (2009). The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics*, **28** (6), 1126–1140.

—, — and ASCH, D. A. (2016). The production of and market for new physicians skill. *American Journal of Health Economics*, **2** (1), 41–65.

ERIKSSON, A. W. and FELLMAN, J. (2000). Seasonal variation of livebirths, stillbirths, extramarital births and twin maternities in Switzerland. *Twin Research and Human Genetics*, **3** (4), 189–201.

FADLON, I. and VAN PARYS, J. (2020). Primary care physician practice styles and patient care: Evidence from physician exits in Medicare. *Journal of Health Economics*, **71**, 102304.

FALK, A. and ICHINO, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, **24** (1), 39–57.

FELFE, C. and LALIVE, R. (2018). Does early child care affect children's development? *Journal of Public Economics*, **159**, 33–53.

FENOLL, A. A. and GROSSBARD, S. (2020). Intergenerational residence patterns and COVID-19 fatalities in the EU and the US. *Economics & Human Biology*, **39**, 100934.

FERMAN, B. and PINTO, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, **12** (4), 1197–1221.

FIELD, A. E., COAKLEY, E. H., MUST, A., SPADANO, J. L., LAIRD, N., DIETZ, W. H., RIMM, E. and COLDITZ, G. A. (2001). Impact of overweight on the risk of developing common chronic diseases during a 10-year period. *Archives of Internal Medicine*, **161** (13), 1581–1586.

FIELDING-MILLER, R. K., SUNDARAM, M. E. and BROUWER, K. (2020). Social determinants of COVID-19 mortality at the county level. *PloS One*, **15** (10), e0240151.

FINKELSTEIN, A., GENTZKOW, M. and WILLIAMS, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics*, **131** (4), 1681–1726.

FINKELSTEIN, E. A., LINNAN, L. A., TATE, D. F. and BIRKEN, B. E. (2007). A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees. *Journal of Occupational and Environmental Medicine*, **49** (9), 981–989.

FINLAY, J. (2009). Fertility response to natural disasters: The case of three high mortality earthquakes, World Bank Policy Research Working Paper Series.

FISHER, E. S., WENNBERG, D. E., STUKEL, T. A., GOTTLIEB, D. J., LUCAS, F. L. and PINDER, E. L. (2003a). The implications of regional variations in Medicare spending. Part 1: The content, quality, and accessibility of care. *Annals of Internal Medicine*, **138** (4), 273–287.

—, —, —, —, — and — (2003b). The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care. *Annals of Internal Medicine*, **138** (4), 288–298.

FLEGAL, K. M., CARROLL, M. D., OGDEN, C. L. and CURTIN, L. R. (2010). Prevalence and trends in obesity among US adults, 1999-2008. *JAMA*, **303** (3), 235–241.

FLETCHER, J. M. (2018). The effects of in utero exposure to the 1918 influenza pandemic on family formation. *Economics & Human Biology*, **30**, 59–68.

FORT, M., ICHINO, A. and ZANELLA, G. (2020). Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families. *Journal of Political Economy*, **128** (1), 158–205.

FOSTER, G. (2006). It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics*, **90** (8-9), 1455–1475.

GÄCHTER, S., JOHNSON, E. J. and HERRMANN, A. (2022). Individual-level loss aversion in riskless and risky choices. *Theory and Decision*, **92** (3), 599–624.

GALLUP (2011). Personal Weight Situation. https://news.gallup.com/poll/7264/personal-weight-situation.aspx Last Accessed: 2020-11-16.

— (2014). Personal Weight Situation. https://news.gallup.com/poll/7264/personal-weight-situation.aspx Last Accessed: 2020-11-16.

GALOR, O. and WEIL, D. N. (1996). The gender gap, fertility, and growth. *American Economic Review*, **86** (3), 374–387.

GATHMANN, C. and SASS, B. (2012). Taxing childcare: Effects on family labor supply and children.

GEERTZ, C. (1963). *Agricultural involution: the process of ecological change in Indonesia*, vol. 11. University of California Press.

GINÉ, X., KARLAN, D. and ZINMAN, J. (2010). Put your money where your butt is: A commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, **2** (4), 213–35.

GLASER, M., NÖTH, M. and WEBER, M. (2004). Behavioral finance. *Blackwell Handbook of Judgment and Decision Making*, pp. 527–546.

GODØY, A. and HUITFELDT, I. (2020). Regional variation in health care utilization and mortality. *Journal of Health Economics*, **71**, 102254.

GOLDMANN, E. and GALEA, S. (2014). Mental health consequences of disasters. *Annual Review of Public Health*, **35**, 169–183.

GOLDSTEIN, J. R., KREYENFELD, M., JASILIONIENE, A. and ÖRSAL, D. K. (2013). Fertility reactions to the "Great Recession" in Europe: Recent evidence from order-specific data. *Demographic Research*, **29**, 85–104.

GOODMAN, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, **38** (5), 581–586.

GOUVEIA, S. O. and CLARKE, V. (2001). Optimistic bias for negative and positive events. *Health Education*.

GRIFFITHS, M. and WOOD, R. (2001). The psychology of lottery gambling. *International Gambling Studies*, **1** (1), 27–45.

GRIMARD, F. (1993). Household consumption smoothing through ethnic ties: Evidence from Cote d'Ivoire. *Journal of Development Economics*, **53** (2), 391–422.

GRYTTEN, J. and SØRENSEN, R. (2003). Practice variation and physician-specific effects. *Journal of Health Economics*, **22** (3), 403–418.

GUIMBEAU, A., MENON, N. and MUSACCHIO, A. (2020). The Brazilian bombshell? The long-term impact of the 1918 influenza pandemic the South American way. *National Bureau of Economic Research*.

GUL, F. and PESENDORFER, W. (2001). Temptation and self-control. *Econometrica*, **69** (6), 1403–1435.

GUNNAR, M. R., KRYZER, E., VAN RYZIN, M. J. and PHILLIPS, D. A. (2010). The rise in cortisol in family day care: Associations with aspects of care quality, child behavior, and child sex. *Child Development*, **81** (3), 851–869.

GUTACKER, N., BLOOR, K., BOJKE, C. and WALSHE, K. (2018). Should interventions to reduce variation in care quality target doctors or hospitals? *Health Policy*, **122** (6), 660–666.

HAECK, C., LEFEBVRE, P. and MERRIGAN, P. (2015). Canadian evidence on ten years of universal preschool policies: The good and the bad. *Labour Economics*, **36**, 137–157.

HALPERN, S. D., FRENCH, B., SMALL, D. S., SAULSGIVER, K., HARHAY, M. O., AUDRAIN-MCGOVERN, J., LOEWENSTEIN, G., BRENNAN, T. A., ASCH, D. A. and VOLPP, K. G. (2015). Randomized trial of four financial-incentive programs for smoking cessation. *New England Journal Medicine*, **372**, 2108–2117.

HARRIS, M. C. (2017). Imperfect information on physical activity and caloric intake. *Economics & Human Biology*, **26**, 112–125.

HAVNES, T. and MOGSTAD, M. (2011). No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy*, **3** (2), 97–129.

HAZAN, C. and SHAVER, P. R. (1994). Attachment as an organizational framework for research on close relationships. *Psychological Inquiry*, **5** (1), 1–22.

HECKMAN, J., PINTO, R. and SAVELYEV, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, **103** (6), 2052–86.

HECKMAN, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, **312** (5782), 1900–1902.

—, STIXRUD, J. and URZUA, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, **24** (3), 411–482.

HEDIN, L.-E. (1967). Some notes on the financing of the Swedish railroads 1860-1914. *Economy and History*, **10**, 3–37.

HEIJMANS, N., VAN LIESHOUT, J. and WENSING, M. (2017). Information exchange networks of health care providers and evidence-based cardiovascular risk management: An observational study. *Implementation Science*, **12** (7), 1–12.

HELANDER, E. E., WANSINK, B. and CHIEH, A. (2016). Weight gain over the holidays in three countries. *New England Journal of Medicine*, **375** (12), 1200–1202.

HELGERTZ, J. and BENGTSSON, T. (2019). The long-lasting influenza: The impact of fetal stress during the 1918 influenza pandemic on socioeconomic attainment and health in Sweden, 1968–2012. *Demography*, **56** (4), 1389–1425.

HERBST, C. M. (2013). The impact of non-parental child care on child development: Evidence from the summer participation dip. *Journal of Public Economics*, **105**, 86–105.

HERBST, D. and MAS, A. (2015). Peer effects on worker output in the laboratory generalize to the field. *Science*, **350** (6260), 545–549.

HERLIHY, D. (1997). *The Black Death and the transformation of the West*. Harvard University Press.

HILL, R. and HANSEN, D. A. (1962). *Man and society in disaster.*, New York: Basic Books, chap. Families in disaster., pp. 185–221.

HIRT-SCHIERBAUM, L. and IVETS, M. (2021). *Incentivizing motivation and self-control preferences*. Tech. rep., Ruhr Economic Papers.

HOFMANN, B. M. (2015). Too much technology. *British Medical Journal*, **350** (7996), h705.

HÖGBERG, U., WALL, S. and BROSTRÖM, G. (1986). The impact of early medical technology on maternal mortality in late 19th century Sweden. *International Journal of Gynecology and Obstetrics*, **24** (4), 251–261.

HOLTENIUS, J. and GILLMAN, A. (2014). The Spanish flu in Uppsala, clinical and epidemiological impact of the influenza pandemic 1918-1919 on a Swedish county. *Infection Ecology and Epidemiology*, **4** (1), 21528.

HOSTINAR, C. E., SULLIVAN, R. M. and GUNNAR, M. R. (2014). Psychobiological mechanisms underlying the social buffering of the hypothalamic–pituitary–adrenocortical axis: A review of animal models and human studies across development. *Psychological bulletin*, **140** (1), 256.

HSIANG, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *PNAS*, **107** (35), 15367–15372.

HUESCH, M. D. (2011). Is blood thicker than water? Peer effects in stent utilization among Floridian cardiologists. *Social Science & Medicine*, **73** (12), 1756–1765.

HÜSKEN, K. (2011). *Kita vor Ort. Betreuungsatlas auf Ebene der Jugendamtsbezirke 2010.* Deutsches Jugendinstitut.

INFLUENSABYRÅN (1919). *Koncept. Registratur 1918–1919*. Stockholm: Kungl. Medicinalstyrelsen.

INFLUENSAKOMMITTÉN (1924). Influensakommitténs rapport. *Svenska Läkartidningen*, **10**, 1–32.

JAEGER, E. (1999). Child care and mother-child interaction in the first 3 years of life. *Developmental Psychology*, **35** (6), 1399–413.

JALLOH, M. F., LI, W., BUNNELL, R. E., ETHIER, K. A., O'LEARY, A., HAGEMAN, K. M., SENGEH, P., JALLOH, M. B., MORGAN, O., HERSEY, S. *et al.* (2018). Impact of Ebola experiences and risk perceptions on mental health in Sierra Leone. *BMJ Global Health*, **3** (2), e000471.

JAPKOWICZ, N. and STEPHEN, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, **6** (5), 429–449.

JEFFERY, R. W. (2012). Financial incentives and weight control. *Preventive Medicine*, **55**, S61–S67.

—, BJORNSON-BENSON, W. M., ROSENTHAL, B. S., KURTH, C. L. and DUNN, M. M. (1984). Effectiveness of monetary contracts with two repayment schedules on weight reduction in men and women from self-referred and population samples. *Behavior Therapy*, **15** (3), 273–279.

—, GERBER, W. M., ROSENTHAL, B. S. and LINDQUIST, R. A. (1983). Monetary contracts in weight control: Effectiveness of group and individual contracts of varying size. *Journal of Consulting and Clinical Psychology*, **51** (2), 242.

JOHN, L. K., LOEWENSTEIN, G., TROXEL, A. B., NORTON, L., FASSBENDER, J. E. and VOLPP, K. G. (2011). Financial incentives for extended weight loss: A randomized, controlled trial. *Journal of General Internal Medicine*, **26** (6), 621–626.

JONSSON, E. (2009). History of health technology assessment in Sweden. *International Journal of Technology Assessment in Health Care*, **25** (1), 42–52.

JÖRBERG, L. and KRANTZ, O. (1978). *Ekonomisk och social politik i Sverige, 1850-1939*. Ekonomisk-historiska institutionen, Lunds Universitet.

KAHNEMAN, D. and TVERSKY, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47** (2), 263–292.

— and — (1984). Choices, values, and frames. *American Psychologist*, **39** (4), 341.

KAHNEMANN, D. and TVERSKY, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–291.

KARACA-MANDIC, P., TOWN, R. J. and WILCOCK, A. (2017). The effect of physician and hospital market structure on medical technology diffusion. *Health Services Research*, **52** (2), 579–598.

KARAFILLAKIS, E., JALLOH, M. F., NURIDDIN, A., LARSON, H. J., WHITWORTH, J., LEES, S., HAGEMAN, K. M., SENGEH, P., JALLOH, M. B., BUNNELL, R. *et al.* (2016). 'Once there is life, there is hope' Ebola survivors' experiences, behaviours and attitudes in Sierra Leone, 2015. *BMJ Global Health*, **1** (3), e000108.

KARLSSON, M., NILSSON, T. and PICHLER, S. (2014). The impact of the 1918 Spanish flu epidemic on economic performance in Sweden: An investigation into the consequences of an extraordinary mortality shock. *Journal of Health Economics*, **36**, 1–19.

KASTRATI, A., DIBRA, A., SPAULDING, C., LAARMAN, G. J. *et al.* (2007). Meta-analysis of randomized trials on drug-eluting stents vs. bare-metal stents in patients with acute myocardial infarction. *European Heart Journal*, **28** (22), 2706–2713.

KHALIL, A., VON DADELSZEN, P., DRAYCOTT, T., UGWUMADU, A., OBRIEN, P. and MAGEE, L. (2020). Change in the incidence of stillbirth and preterm delivery during the COVID-19 pandemic. *JAMA*, **324** (7), 705–706.

KINARI, Y. (2016). Properties of expectation biases: Optimism and overconfidence. *Journal of Behavioral and Experimental Finance*, **10**, 32–49.

KOBASA, D., JONES, S. M., SHINYA, K., KASH, J. C., COPPS, J., EBIHARA, H., HATTA, Y., KIM, J. H., HALFMANN, P., HATTA, M., FELDMANN, F., ALIMONTI, J. B., FERNANDO, L., LI, Y.,

KATZE, M. G., FELDMANN, H. and KAWAOKA, Y. (2007). Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. *Nature*, **445** (7125), 319–323.

KOPELMAN, P. (2007). Health risks associated with overweight and obesity. *Obesity Reviews*, **8**, 13–17.

KOPETSCH, T. and SCHMITZ, H. (2014). Regional variation in the utilisation of ambulatory services in Germany. *Health Economics*, **23** (12), 1481–1492.

KOTTELENBERG, M. J. and LEHRER, S. F. (2014). Do the perils of universal childcare depend on the childs age? *CESifo Economic Studies*, **60** (2), 338–365.

— and — (2018). Does Quebec's subsidized child care policy give boys and girls an equal start? *Canadian Journal of Economics/Revue canadienne d'économique*, **51** (2), 627–659.

KUEHNLE, D. and OBERFICHTNER, M. (2020). Does starting universal childcare earlier influence childrens skill development? *Demography*, **57** (1), 61–98.

KWOK, J. (2019). How do primary care physicians influence healthcare? Evidence on practice styles and switching costs from Medicare. *Unpublished Manuscript*.

LANZIERI, G. (2014). Towards a baby recession' in Europe? Differential fertility trends during the economic crisis. Eurostat, Statistics in Focus 13/2013.

LAYARD, R., CLARK, A. E., CORNAGLIA, F., POWDTHAVEE, N. and VERNOIT, J. (2014). What predicts a successful life? A life-course model of well-being. *The Economic Journal*, **124** (580), F720–F738.

LEE, R. (1981). Short-term variation: Vital rates, prices and weather. in, The population history of England, 1541-1871. E.A. Wrigley and R.S. Schofield, eds.

— (1990). The demographic response to economic crisis in historical and contemporary populations. *Population Bulletin of the United Nations*, (29), 1.

LEFEBVRE, P. and MERRIGAN, P. (2008). Child-care policy and the labor supply of mothers with young children: A natural experiment from Canada. *Journal of Labor Economics*, **26** (3), 519–548.

LEGRAND, T., KOPPENHAVER, T., MONDAIN, N. and RANDALL, S. (2003). Reassessing the insurance effect: A qualitative analysis of fertility behavior in Senegal and Zimbabwe. *Population and Development Review*, **29** (3), 375–403.

LICHTENSTEIN, S., FISCHHOFF, B. and PHILLIPS, L. (1982). Calibration of probabilities: The state of the art to 1980. D. Kahneman, P. Slovic, and A. Tverski (Eds.) Judgement under uncertainty: Heuristics and biases.

LIN, N., ENSEL, W. M. and VAUGHN, J. C. (1981). Social resources and strength of ties: Structural factors in occupational status attainment. *American Sociological Review*, **46** (4), 393–405.

LINDSTROM, D. P. and BERHANU, B. (1999). The impact of war, famine, and economic decline on marital fertility in Ethiopia. *Demography*, **36** (2), 247–261.

LOEB, S., BRIDGES, M., BASSOK, D., FULLER, B. and RUMBERGER, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, **26** (1), 52–66.

LUBY, J. L., ROGERS, C. and MCLAUGHLIN, K. A. (2022). Environmental conditions to promote healthy childhood brain/behavioral development: Informing early preventive interventions for delivery in routine care. *Biological Psychiatry Global Open Science*, **2** (3), 233–241.

LUNDEBERG, M. A., FOX, P. W. and PUŃĆOHAŔ, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, **86** (1), 114.

LUNDGREN, Å. (1989). *Spanskans år*. Bokförlaget Forum.

LUNDGREN, Å. (1991). *Storspovens sång*. Bokförlaget Forum.

LUNDH, C. (2007). Remarriage, gender and social: A longitudinal study of remarriage in Southern Sweden, 1766-1894. *Continuity and Change*, **22** (3), 373–406.

LUPIEN, S. J., MCEWEN, B. S., GUNNAR, M. R. and HEIM, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, **10** (6), 434–445.

LUPPI, F., ARPINO, B. and ROSINA, A. (2020). The impact of COVID-19 on fertility plans in Italy, Germany, France, Spain, and the United Kingdom. *Demographic Research*, **43**, 1399–1412.

LUSHER, L. (2016). College better: Parimutuel betting markets as a commitment device and monetary incentive. *Natural Field Experiments*, **561**.

LYLE, D. S. (2007). Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point. *The Review of Economics and Statistics*, **89** (2), 289–299.

LYNN, M., FLYNN, S. M. and HELION, C. (2013). Do consumers prefer round prices? Evidence from pay-what-you-want decisions and self-pumped gasoline purchases. *Journal of Economic Psychology*, **36**, 96–102.

MADDEN, K. M. (2017). The seasonal periodicity of healthy contemplations about exercise and weight loss: Ecological correlational study. *JMIR Public Health and Surveillance*, **3** (4), e92.

MAGNUSSON, L. (1996). *Sveriges ekonomiska historia.* Raben Prisma.

MAMELUND, S.-E. (2003). Spanish influenza mortality of ethnic minorities in Norway 1918–1919. *European Journal of Population*, **19** (1), 83–102.

— (2004). Can the Spanish influenza pandemic of 1918 explain the baby boom of 1920 in neutral Norway? *Population*, **59** (2), 229–260.

— (2011). Geography may explain adult mortality from the 1918–20 influenza pandemic. *Epidemics*, **3** (1), 46–60.

—, HANEBERG, B. and MJAALAND, S. (2016). A missed summer wave of the 1918–1919 influenza pandemic: Evidence from household surveys in the United States and Norway. *Open Forum Infectious Diseases*, **3** (1), ofw040.

MANSOUR, S. B., JOUINI, E. and NAPP, C. (2006). Is there a pessimistic bias in individual beliefs? Evidence from a simple survey. *Theory and Decision*, **61** (4), 345–362.

MARKOWITZ, S., NESSON, E. and ROBINSON, J. J. (2019). The effects of employment on influenza rates. *Economics & Human Biology*, **34**, 286–295.

MAS, A. and MORETTI, E. (2009). Peers at work. *American Economic Review*, **99** (1), 112–45.

MATOVIC, M. R. (1986). The Stockholm marriage: Extra-legal family formation in Stockholm 1860-1890. *Continuity and Change*, **1** (2), 385–413.

MATYSIAK, A., SOBOTKA, T. and VIGNOLI, D. (2021). The great recession and fertility in europe: A sub-national analysis. *European Journal of Population*, **37** (1), 29–64.

MCCLELLAN, M. (2011). Reforming payments to healthcare providers: The key to slowing healthcare cost growth while improving quality? *The Journal of Economic Perspectives*, **25** (2), 69–92.

MEANEY, M. J. and SZYF, M. (2022). Environmental programming of stress responses through DNA methylation: Life at the interface between a dynamic environment and a fixed genome. *Dialogues in Clinical Neuroscience*.

MEHROTRA, A., REID, R. O., ADAMS, J. L., FRIEDBERG, M. W., MCGLYNN, E. A. and HUSSEY, P. S. (2012). Physicians with the least experience have higher cost profiles than do physicians with the most experience. *Health Affairs*, **31** (11), 2453–2463.

MENKEN, J., TRUSSELL, J. and WATKINS, S. (1981). The nutrition fertility link: An evaluation of the evidence. *The Journal of Interdisciplinary History*, **11** (3), 425–441.

MERKLE, C. and WEBER, M. (2011). True overconfidence: The inability of rational information processing to account for apparent overconfidence. *Organizational Behavior and Human Decision Processes*, **116** (2), 262–271.

MOLITOR, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy*, **10** (1), 326–56.

MORENS, D. M. and FAUCI, A. S. (2007). The 1918 influenza pandemic: Insights for the 21st century. *Journal of Infectious Diseases*, **195** (7), 1018–1028.

MOURA, A., SALM, M., DOUVEN, R. and REMMERSWAAL, M. (2019). Causes of regional variation in Dutch healthcare expenditures: Evidence from movers. *Health Economics*, **28** (9), 1088–1098.

MULLAINATHAN, S. and OBERMEYER, Z. (2022). Diagnosing physician error: A machine learning approach to low-value health care. *The Quarterly Journal of Economics*, **137** (2), 679–727.

NAIR, H. S., MANCHANDA, P. and BHATIA, T. (2010). Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research*, **47** (5), 883–895.

NATIONAL RESEARCH COUNCIL (1998). *From death to birth: Mortality decline and reproductive change*. National Academies Press.

NAYLOR, C. D. (1995). Grey zones of clinical practice: Some limits to evidence-based medicine. *The Lancet*, **345** (8953), 840–842.

NEELS, K. (2010). Temporal variation in unemployment rates and their association with tempo and quantum of fertility: Some evidence for Belgium, France and the Netherlands. In *Annual meeting of the Population association of America, Dallas*, vol. 17.

NICHD (1997). The effects of infant child care on infant-mother attachment security: Results of the NICHD study of early child care. *Child Development*, pp. 860–879.

— (2000). The relation of child care to cognitive and language development. *Child Development*, pp. 960–980.

— (2001). Child-care and family predictors of preschool attachment and stability from infancy. *Developmental Psychology*, **37** (6), 847–62.

— (2002). Early child care and childrens development prior to school entry: Results from the NICHD study of early child care. *American Educational Research Journal*, **39** (1), 133–164.

— (2003a). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, **74** (4), 976–1005.

— (2003b). Early child care and mother–child interaction from 36 months through first grade. *Infant Behavior and Development*, **26** (3), 345–370.

— (2005). Early child care and childrens development in the primary grades: Follow-up results from the NICHD study of early child care. *American Educational Research Journal*, **42** (3), 537–570.

NIEDERLE, M. and VESTERLUND, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, **122** (3), 1067–1101.

NOBLES, J., FRANKENBERG, E. and THOMAS, D. (2015). The effects of mortality on fertility: Population dynamics after a natural disaster. *Demography*, **52** (1), 15–38.

NYSTRÖM, H. (1994). *Hungerupproret 1917*. Zelos.

O'BRIEN, K., VENN, B. J., PERRY, T., GREEN, T. J., AITKEN, W., BRADSHAW, A. *et al.* (2007). Reasons for wanting to lose weight: Different strokes for different folks. *Eating Behaviors*, **8** (1), 132–135.

OECD (2014a). Geographic variations in health care: What do we know and what can be done to improve health system performance? OECD Health Policy Studies, OECD Publishing.

— (2014b). OECD Family Database. OECD Publishing, Paris, http://www.oecd.org/els/family/database.htm Last Accessed: 2021-02-02.

— (2019). The heavy burden of obesity: The economics of prevention. https://www.oecd.org/health/the-heavy-burden-of-obesity-67450d67-en.htm, Last Accessed: 2022-11-22.

— (2021). OECD Family Database. OECD Publishing, Paris, http://www.oecd.org/els/family/database.htm Last Accessed: 2022-01-15.

— (2022a). Health spending (indicator). https://data.oecd.org/healthres/health-spending.htm, Last Accessed: 2022-07-23.

— (2022b). Interim Report September 2022: Paying the Price of War.

— (2022c). Life expectancy at birth (indicator). https://data.oecd.org/healthstat/life-expectancy-at-birth.htm, Lat Accessed: 2022-07-23.

— (2022d). OECD Family Database (Fertility Rates). https://www.oecd.org/els/family/database.htm, Last Accessed: 2022-07-23.

O'LEARY, A., JALLOH, M. F. and NERIA, Y. (2018). Fear and culture: Contextualising mental health impact of the 2014–2016 Ebola epidemic in West Africa.

OLOFSSON, C. (1921). *Sveriges järnvägar.* Stockholm: Utg., 1st edn., accessed from historiskt.nu.

ÖSTLIND, A. (1945). Svensk samhällsekonomi 1914-1922 (The Swedish Economy, 1914-1922). *Stockholm: Svenska Bankföreningen*.

PALLONI, A. and RAFALIMANANA, H. (1999). The effects of infant mortality on fertility revisited: New evidence from Latin America. *Demography*, **36** (1), 41–58.

PARMAN, J. (2015). Childhood health and sibling outcomes: Nurture reinforcing nature during the 1918 influenza pandemic. *Explorations in Economic History*, **58**, 22–43.

PAUL-EBHOHIMHEN, V. and AVENELL, A. (2008). Systematic review of the use of financial incentives in treatments for obesity and overweight. *Obesity Reviews*, **9** (4), 355–367.

PERCHESKI, C. and KIMBRO, R. (2014). How did the Great Recession affect fertility. *Focus*, **30** (2), 26–30.

PETER, F. H., SCHOBER, P. S. and SPIESS, K. C. (2016). Early birds in day care: The social gradient in starting day care and childrens non-cognitive skills. *CESifo Economic Studies*, **62** (4), 725–751.

PHELAN, S., WING, R. R., RAYNOR, H. A., DIBELLO, J., NEDEAU, K. and PENG, W. (2008). Holiday weight management by successful weight losers and normal weight individuals. *Journal of Consulting and Clinical Psychology*, **76** (3), 442.

PHELPS, C. E. (2000). Information diffusion and best practice adoption. In *Handbook of Health Economics*, vol. 1, Elsevier, pp. 223–264.

POLIVY, J. (2001). The false hope syndrome: Unrealistic expectations of self-change. *International Journal of Obesity*, **25** (1), S80–S84.

— and HERMAN, C. P. (2000). The false-hope syndrome: Unfulfilled expectations of self-change. *Current Directions in Psychological Science*, **9** (4), 128–131.

PÖRTNER, C. (2008). *Gone with the wind? Hurricane risk, fertility and education*. Working Papers UWEC-2006-19-R, University of Washington, Department of Economics.

PRESTON, S. H. (1978). *The effects of infant and child mortality on fertility*. New York: Academic Press.

PREVOO, T. and TER WEEL, B. (2015). The importance of early conscientiousness for socio-economic outcomes: Evidence from the British Cohort Study. *Oxford Economic Papers*, **67** (4), 918–948.

PRIETO, D. C. and LAGO-PEÑAS, S. (2012). Decomposing the determinants of health care expenditure: The case of Spain. *The European Journal of Health Economics*, **13** (1), 19–27.

QUEBEC GOVERNMENT (2022). Becoming a parent. https://www.quebec.ca/en/family-and-support-for-individuals/pregnancy-and-parenthood/parent, Last Accessed: 2022-05-01.

RABE-HESKETH, S. and SKRONDAL, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.

RABELO, I., LEE, V., FALLAH, M. P., MASSAQUOI, M., EVLAMPIDOU, I., CRESTANI, R., DECROO, T., VAN DEN BERGH, R. and SEVERY, N. (2016). Psychological distress among Ebola survivors

discharged from an Ebola treatment unit in Monrovia, Liberia–a qualitative study. *Frontiers in Public Health*, **4**, 142.

RAUTE, A. (2019). Can financial incentives reduce the baby gap? Evidence from a reform in maternity leave benefits. *Journal of Public Economics*, **169**, 203–222.

REICH, O., WEINS, C., SCHUSTERSCHITZ, C. and THÖNI, M. (2012). Exploring the disparities of regional health care expenditures in Switzerland: Some empirical evidence. *The European Journal of Health Economics*, **13** (2), 193–202.

RELTON, C., STRONG, M. and LI, J. (2011). The Pounds for Pounds weight loss financial incentive scheme: An evaluation of a pilot in NHS Eastern and Coastal Kent. *Journal of Public Health*, **33** (4), 536–542.

RIBERS, M. A. and ULLRICH, H. (2022). Machine predictions and human decisions with variation in payoff and skills: The case of antibiotic prescribing, unpublished manuscript.

RICHTER, A. and ROBLING, P. O. (2013). Multigenerational effects of the 1918-19 influenza pandemic in Sweden, Swedish Institute for Social Research, Working Paper Series.

RIEGER, M. O., WANG, M. and HENS, T. (2011). Prospect theory around the world. *NHH Dept. of Finance & Management Science Discussion Paper*, (2011/19).

ROBERTSON, A. (2014). *Obesity and inequities. Guidance for addressing inequities in overweight and obesity*. World Health Organization.

ROMLID, C. (1997). Swedish midwives and their instruments in the eighteenth and nineteenth centuries. In H. Marland and A. M. Rafferty (eds.), *Midwives, society and childbirth: Debates and controversies in the modern period*, London, England: Routledge, pp. 38–60.

ROSENZWEIG, M. R. and SCHULTZ, T. P. (1983). Consumer demand and household production: The relationship between fertility and child mortality. *American Economic Review*, **73** (2), 38–42.

SACERDOTE, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, **116** (2), 681–704.

SALM, M. and WÜBKER, A. (2020). Sources of regional variation in healthcare utilization in Germany. *Journal of Health Economics*, **69**, 102271.

SANDSTRÖM, G. (2011). Time–space trends in Swedish divorce behaviour 1911–1974. *Scandinavian Journal of History*, **36** (1), 65–90.

SCHMIDT, U. and TRAUB, S. (2002). An experimental test of loss aversion. *Journal of Risk and Uncertainty*, **25** (3), 233–249.

SCHOELLER, D. A. (2014). The effect of holiday weight gain on body weight. *Physiology & Behavior*, **134**, 66–69.

SCHÖN, L. (2010). *Sweden's road to modernity: An economic history*. SNS förlag.

SCHORE, A. N. (2000). Attachment and the regulation of the right brain. *Attachment & Human Development*, **2** (1), 23–47.

— (2001). Effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. *Infant Mental Health Journal: Official Publication of The World Association for Infant Mental Health*, **22** (1-2), 7–66.

— (2017). All our sons: The developmental neurobiology and neuroendocrinology of boys at risk. *Infant Mental Health Journal*, **38** (1), 15–52.

SCIENTIFIC COUNCIL, N. (2014). Excessive stress disrupts the development of brain architecture. *Journal of Childrens Services*, **9** (2), 143–153.

SHEPPERD, J. A., CARROLL, P., GRACE, J. and TERRY, M. (2002). Exploring the causes of comparative optimism. *Psychologica Belgica*, **42** (1/2), 65–98.

SHORTELL, S. M., MARSTELLER, J. A., LIN, M., PEARSON, M. L., WU, S.-Y., MENDEL, P., CRETIN, S. and ROSEN, M. (2004). The role of perceived team effectiveness in improving chronic illness care. *Medical Care*, **42** (11), 1040–1048.

SILVER, D. (2021). Haste or waste? Peer pressure and productivity in the emergency department. *The Review of Economic Studies*, **88** (3), 1385–1417.

SKINNER, J. (2011). Causes and consequences of regional variations in health care. In *Handbook of Health Economics*, vol. 2, Elsevier, pp. 45–93.

—, GOTTLIEB, D. J. and CARMICHAEL, D. (2011). A new series of Medicare expenditure measures by hospital referral region: 2003-2008. *Dartmouth Atlas Project*.

SOBOTKA, T., JASILIONIENE, A., GALARZA, A. A., ZEMAN, K., NEMETH, L. and JDANOV, D. (2021). Baby bust in the wake of the COVID-19 pandemic? First results from the new STFF data series.

—, SKIRBEKK, V. and PHILIPOV, D. (2011). Economic recession and fertility in the developed world. *Population and Development Review*, **37** (2), 267–306.

STATISTICS CANADA (2021). Study: Family matters: Parental leave in Canada. https://www150.statcan.gc.ca/n1/en/daily-quotidien/210210/dq210210a-eng.pdf?st=S3RGAcP-, Last Accessed: 2022-05-01.

STATISTICS SWEDEN (1917). *Fattigvården*. Stockholm: Kungl. Statistiska Centralbyrån.

STATISTICS SWEDEN (1920). Årsbok för sveriges kommuner. årg. 03(1920). Stockholm.

STATISTICS SWEDEN (1929). Befolkningsrörelsen. översikt för åren. 1911-1920. Stockholm.

STATISTICS SWEDEN (1999). Befolkningsutvecklingen under 250 år.

STEVENSON, J., KRISHNAN, S., STONER, M., GOKTAS, Z. and COOPER, J. (2013). Effects of exercise during the holiday season on changes in body weight, body composition and blood pressure. *European Journal of Clinical Nutrition*, **67** (9), 944–949.

STINEBRICKNER, R. and STINEBRICKNER, T. R. (2006). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics*, **90** (8-9), 1435–1454.

STRAND, L. B., BARNETT, A. G. and TONG, S. (2012). Maternal exposure to ambient temperature and the risks of preterm birth and stillbirth in Brisbane, Australia. *American Journal of Epidemiology*, **175** (2), 99–107.

STRULIK, H. and VOLLMER, S. (2015). The fertility transition around the world. *Journal of Population Economics*, **28** (1), 31–44.

SVETNIK, V., LIAW, A., TONG, C., CULBERSON, J. C., SHERIDAN, R. P. and FEUSTON, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, **43** (6), 1947–1958.

SWEDISH AGENCY FOR HEALTH TECHNOLOGY ASSESSMENT AND ASSESSMENT OF SOCIAL SERVICES (2014). *Läkemedelsavgivande stentar i hjärtats kransartärer*. Tech. rep., SBU Alert-rapport nr 2014-06.

SWINBURN, B., EGGER, G. and RAZA, F. (1999). Dissecting obesogenic environments: The development and application of a framework for identifying and prioritizing environmental interventions for obesity. *Preventive Medicine*, **29** (6), 563–570.

SWINBURN, B. A., SACKS, G., HALL, K. D., MCPHERSON, K., FINEGOOD, D. T., MOODIE, M. L. and GORTMAKER, S. L. (2011). The global obesity pandemic: Shaped by global drivers and local environments. *The Lancet*, **378** (9793), 804–814.

TAUBENBERGER, J. K. and MORENS, D. M. (2006). 1918 influenza: The mother of all pandemics. *Emerging Infectious Diseases*, **12** (1), 15–22.

TEITELBAUM, M. S. (2014). *The British fertility decline: Demographic transition in the crucible of the industrial revolution*. Princeton University Press.

THALER, R. H. and BENARTZI, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, **112** (S1), S164–S187.

THOMPSON, D., EDELSBERG, J., COLDITZ, G. A., BIRD, A. P. and OSTER, G. (1999). Lifetime health and economic consequences of obesity. *Archives of Internal Medicine*, **159** (18), 2177–2183.

THOMSON, A. (1997). European midwifery in the inter-war years. In H. Marland and A. M. Rafferty (eds.), *Midwives, society and childbirth: debates and controversies in the modern period*, London, England: Routledge, pp. 20–37.

TOUT, K., DE HAAN, M., CAMPBELL, E. K. and GUNNAR, M. R. (1998). Social behavior correlates of cortisol activity in child care: Gender differences and time-of-day effects. *Child Development*, **69** (5), 1247–1262.

TOWNSEND, R. M. (1994). Risk and insurance in village India. *Econometrica: Journal of the Econometric Society*, **62** (3), 539–591.

TROGDON, J. G., FINKELSTEIN, E. A., HYLANDS, T., DELLEA, P. S. and KAMAL-BAHL, S. (2008). Indirect costs of obesity: A review of the current literature. *Obesity Reviews*, **9** (5), 489–500.

TVERSKY, A. and KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185** (4157), 1124–1131.

TZANNATOS, Z. and SYMONS, J. (1989). An economic approach to fertility in Britain since 1860. *Journal of Population Economics*, **2** (2), 121–138.

UNITED NATIONS (2015). Resolution adopted by the General Assembly on 25 September 2015. Transforming our world: The 2030 Agenda for Sustainable Development. *General Assembly*, **70/1**.

VAN PARYS, J. (2016). Variation in physician practice styles within and across emergency departments. *PloS One*, **11** (8), e0159882.

VANDELL, D. L., BELSKY, J., BURCHINAL, M., STEINBERG, L., VANDERGRIFT, N. and NETWORK, N. E. C. C. R. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. *Child Development*, **81** (3), 737–756.

VEENENDAAL, M. V., PAINTER, R. C., ROOIJ, S., BOSSUYT, P. M., POST, J., GLUCKMAN, P. D., HANSON, M. A. and ROSEBOOM, T. J. (2013). Transgenerational effects of prenatal exposure to the 1944–45 Dutch famine. *BJOG: An International Journal of Obstetrics and Gynaecology*, **120** (5), 548–554.

VERMEER, H. J. and VAN IJZENDOORN, M. H. (2006). Children's elevated cortisol levels at daycare: A review and meta-analysis. *Early Childhood Research Quarterly*, **21** (3), 390–401.

VLAAR, P. J., DE SMET, B. J. G. L. and ZIJLSTRA, F. (2007). DES or BMS in acute myocardial infarction? *European Heart Journal*, **28** (22), 2693–2694.

VOLPP, K. G., JOHN, L. K., TROXEL, A. B., NORTON, L., FASSBENDER, J. and LOEWENSTEIN, G. (2008). Financial incentive–based approaches for weight loss: A randomized trial. *JAMA*, **300** (22), 2631–2637.

VOS, T., ALLEN, C., ARORA, M., BARBER, R. M., BHUTTA, Z. A., BROWN, A., CARTER, A., CASEY, D. C., CHARLSON, F. J., CHEN, A. Z. *et al.* (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, **388** (10053), 1545–1602.

WATAMURA, S. E., DONZELLA, B., KERTES, D. A. and GUNNAR, M. R. (2004). Developmental changes in baseline cortisol activity in early childhood: Relations with napping and effortful control. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, **45** (3), 125–133.

WAYBETTER (since 2010). Waybetter inc. about. http://www.waybetter.com/about.

WEINSTEIN, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, **39** (5), 806–820.

WENG, W., VAN PARYS, J., LIPNER, R. S., SKINNER, J. S. and SIROVICH, B. E. (2020). Association of regional practice environment intensity and the ability of internists to practice high-value care after residency. *JAMA Network Open*, **3** (4), e202494–e202494.

WENNBERG, J. and GITTELSOHN, A. (1973). Small area variations in health care delivery: A population-based health information system can guide planning and regulatory decision-making. *Science*, **182** (4117), 1102–1108.

WENNBERG, J. E. (2010). *Tracking medicine: A researcher's quest to understand health care*. Oxford University Press.

WHO (2016). Smoking prevalence. https://data.worldbank.org/indicator/SH.PRV.SMOK.MA Last Accessed: 2019-30-08.

— (2021). Obesity and overweight. http://www.who.int/mediacentre/factsheets/fs311/en/ Last Accessed: 2022-11-30.

WICKSELL, S. (1922). *Bilaga till Statistiksakkunnigas betänkande: redogörelse för huvuddragen av den officiella statistikens utveckling och nuvarande organisation i Sverige jämte kort översikt av dess organisation i vissa främmande länder samt i internationellt hänseende*. Stockholm: Statens offentliga utredningar.

WIWANITKIT, V. (2010). Influenza, swine flu, sperm quality and infertility: A story. *Journal of Human Reproductive Sciences*, **3** (2), 116.

WOERNER, A. (2021). Overcoming time inconsistency with a matched bet: Theory and evidence from exercising.

YANG, M., LIEN, H.-M. and CHOU, S.-Y. (2014). Is there a physician peer effect? Evidence from new drug prescriptions. *Economic Inquiry*, **52** (1), 116–137.

YANOVSKI, J. A., YANOVSKI, S. Z., SOVIK, K. N., NGUYEN, T. T., O'NEIL, P. M. and SEBRING, N. G. (2000). A prospective study of holiday weight gain. *New England Journal Medicine*, **342**, 861–867.

YOUNG, A. (2005). The gift of the dying: The tragedy of AIDS and the welfare of future African generations. *The Quaterly Journal of Economics*, **120** (2), 423–466.

YUAN, C. T., NEMBHARD, I. M. and KANE, G. C. (2020). The influence of peer beliefs on nurses' use of new health information technology: A social network analysis. *Social Science & Medicine*, **255**, 113002.

ZIMMERMAN, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, **85** (1), 9–23.