Hungarian DOI: 10.14754/CEU.2023.04

Essays in Labor and Health Economics

by
Anna Bárdits

Submitted to
Central European University
Department of Economics and Business

In partial fulfillment of the requirements for the degree
Doctor of Philosophy in Economics

Supervisor: Ádám Szeidl

Budapest - Vienna,

2023

Hungarian DOI: 10.14754/CEU.2023.04

Copyright notice

Author: Anna Bárdits
Title: Essays in Labor and Health Economics
Degree: Ph.D.
Dated: September, 2023

Hungarian DOI: 10.14754/CEU.2023.04

Co-author contribution

Chapter 1: Precautionary Fertility: Conceptions, Births, and Abortions around Employment Schocks

Joint work with Anna Adamecz, Márta Bisztray, Ágnes Szabó-Morvai, and Andrea Weber

All authors contributed to the idea and focus of the paper, reviewing the literature, and the development of the empirical strategy and identification. The theoretical derivations are mostly the work of Márta Bisztray (in Appendix A.3), the other co-authors contributed with ideas. I wrote most of the codes for data cleaning and analysis, some parts of the code were written by Ágnes Szabó-Morvai. I wrote a large part of the first manuscript, which was then improved and modified by all authors, in particular by Ágnes Szabó-Morvai, Andrea Weber and me.

Chapter 2: The Gains from Family Foster Care: Evidence from Hungary

Joint work with Gábor Kertesi

The paper was developed in close cooperation with Gábor Kertesi. The original idea for the paper came from me but the empirical strategy and the focus of the paper substantially changed since Gábor joined, as we worked out these together. We discussed and collected relevant literature and details of the Hungarian institutional surroundings together. I did the coding for the data cleaning and analysis, and I wrote the text of this chapter.

Hungarian DOI: 10.14754/CEU.2023.04

Abstracts

The thesis consists of one single-authored and two jointly-authored chapters, that are linked by their focus on investigating causal questions with high policy relevance in the fields of labor and health economics. All chapters use large Hungarian datasets for the empirical analyses. The individual chapters are summarized in the following abstracts.

Chapter 1: Precautionary Fertility: Conceptions, Births, and Abortions around Employment Shocks[1]

Joint work with Anna Adamecz, Márta Bisztray, Ágnes Szabó-Morvai, and Andrea Weber

This chapter studies fertility responses to employment shocks. Using unique Hungarian administrative data that allow linking firm-level mass layoff and closure events to individual-level records on births and abortions, we show that the main response happens in anticipation of the shock. Responses differ by the availability of dismissal protection. While pregnancies increase in anticipation of all events, births only rise in case of mass layoffs when pregnant women are protected from layoffs. If the firm closes protection is lost and we find an increase in abortions. We interpret these results as evidence for precautionary fertility behavior. Women threatened by job displacement bring births forward to exploit dismissal protection, a strategy that breaks down if the firm closes permanently.

Chapter 2: The Gains from Family Foster Care: Evidence from Hungary

Joint work with Gábor Kertesi

In this chapter, we investigate how the type of home environment – family foster care or residential care – affects the adult outcomes of individuals who were raised in state care during adolescence. While it is established in the literature that living in institutional care is detrimental for babies, the effect of living in different types of care as an older child is underexplored. We use Hungarian individual-level administrative panel data and follow the children from 6$^{\text{th}}$ grade (around age 13) until age 19. We show that the adult outcomes of children growing up in family foster care are substantially better even after controlling for a rich set of variables, including indicators of cognitive and non-cognitive skills, and psychological problems observed in 6$^{\text{th}}$ grade. Young adults who grew up in family foster care are more likely to finish secondary education, they spend less time without either working or studying, and they are less likely to use tranquilizers than comparable youth raised in residential care. For girls, teenage pregnancy is less likely. IV estimations using local foster mother capacity as an instrument also point to a beneficial effect of foster care.

Chapter 3: The Effect of Physical Education Time on Student's Body Composition

In this chapter, I analyze the effect of physical education (PE) time on the body composition of students. Previous literature has established that spending more time at PE classes is largely

[1]A previous version of the chapter was published as a IZA discussion paper (Bárdits et al., 2023a) and as a KRTK KTI working paper (Bárdits et al., 2023b)

ineffective in reducing children's body mass index (BMI). However, BMI can increase or can stay unchanged even when body composition becomes healthier, as physical activity can reduce fat mass and build muscle mass at the same time. Therefore, relying solely on BMI as an outcome can be misleading in assessing the true potential of PE time on students' health. In my analysis I use a new outcome, body fat percentage, to overcome this problem. To identify the effect of PE time, I use the introduction of daily PE classes in Hungarian primary and secondary schools in 2012. The staggered implementation of the policy created a large variation in time ever spent with PE between multiple subsequent age cohorts, making it possible to compare cohorts with different past PE times while controlling for year fixed effects. Using data on the whole population of Hungarian 5^{th} to 12^{th} grade students I estimate OLS models on the following outcomes: mean BMI, mean body fat percentage, and obesity defined by a BMI threshold and by a body fat percentage threshold. The estimated effects of past PE time are statistically insignificant and close to zero both for BMI and body fat percentage outcomes. Heterogeneity analysis reveals that increased PE time left the outcomes unchanged in specific subgroups of students (by age, income, gender, and school infrastructure) as well. The findings indicate that increasing PE time alone is not sufficient to tackle childhood obesity.

Hungarian DOI: 10.14754/CEU.2023.04

Acknowledgements

I wish to express my heartfelt thanks to those who have supported me throughout my doctoral journey. First of all, I would like to thank my advisor, Ádám Szeidl for his guidance. His ideas, insights, and questions had a profound impact on this thesis.

I appreciate the valuable and thorough feedback provided by the examiners of the thesis, Aline Bütikofer and Tímea Molnár.

I am also grateful to my co-authors, Anna Adamecz, Márta Bisztray, Ágnes Szabó-Morvai, Andrea Weber, and Gábor Kertesi. I learned a lot through our collaboration.

I thank other members of the CEU faculty, who provided much help during various research seminars. I especially thank Róbert Lieli who contributed a lot during these seminars and also taught my favorite classes. I also thank Gábor Kézdi without whom I wouldn't have even started a PhD.

I am grateful to the participants of the seminars of KTI's Health and Population research group for the thought-provoking discussions about my research, and particularly to Anikó Bíró, for organizing these events. The help of KTI's Data Bank, in particular, Tir Melinda's extensive knowledge of the data was essential in the data work for this thesis.

I would also like to express my gratitude to Kati Szimler and her colleagues for their help with administrative issues.

I also want to acknowledge my PhD peers, especially Lili Márk, Boldizsár Juhász, and Ceyda Üstün, who made me feel like I am part of a community during my studies. I am grateful to my friends outside CEU for regularly updating me about what is going on in the real world.

I thank my children for being amazing, and for completely changing my perception of time. Now I know that five free minutes is a lot. I am also grateful to my parents for their invaluable help with the kids, and my husband for everything, especially for calming my nerves.

Contents

List of Tables

 Hungarian DOI: 10.14754/CEU.2023.04

List of Figures

Hungarian DOI: 10.14754/CEU.2023.04

1 Chapter 1: Precautionary Fertility: Conceptions, Births, and Abortions around Employment Shocks

Joint work with Anna Adamecz, Márta Bisztray, Ágnes Szabó-Morvai, and Andrea Weber

1.1 Introduction

In modern labor markets with high female labor force participation rates, fertility decisions are increasingly determined by the compatibility of career and family goals (Doepke et al., 2022). Women no longer decide between a career or a family, but their aim is to have it all. Family policies support these goals by guaranteeing mothers' access to equal opportunity and equal treatment in the workplace. Besides providing maternity leave regulations, many countries also protect mothers from dismissal during pregnancy and maternity leave, and guarantee them the right to return to their previous job (ILO, 2010).

Maternity policies might also play an important role in a situation where careers are especially vulnerable: around a job loss. It is well established that job displacement is related to large and persistent earnings and employment losses (Bertheau et al., 2022; Jacobson et al., 1993a). These losses are generally found to be larger for women who are more likely to end up in part-time employment or in unstable jobs than men (Illing et al., 2021). As a consequence, women reduce their fertility after a job loss with the aim of getting their career back on track (Del Bono et al., 2012; Huttunen and Kellokumpu, 2016). Less is known about how fertility responds in anticipation of a job loss, however. In this chapter, we argue that in an environment with maternal dismissal protection, pregnancies can be used as a precautionary strategy to avoid job displacement. The idea is that a woman who is aware of economic problems and anticipates a potential mass layoff at her workplace chooses to become pregnant to protect herself against the layoff risk and wait out the crisis during the maternity leave period. This strategy will be successful as long as the firm survives the temporary crisis. In case of a firm closure, the precautionary mechanism breaks down and the woman might choose to terminate the pregnancy.

We study Hungary, a country that has adopted the latest ILO Maternity Protection Convention according to which pregnant women are protected from dismissal.[2] Hungarian family policy offers generous leave benefits for employed mothers and lower benefits if they are unemployed, creating an additional incentive for women to avoid job loss during pregnancy. We make use of unique and rich administrative matched employer-employee data which allows us to identify mass layoff events and plant closures, and which can be linked to health records with individual information on births and abortions. These data offer an ideal setting to study fertility responses around job loss.

We start by documenting that large layoff events are on average preceded by indicators of economic problems at the firm. While employment stays relatively stable, we show that orders decline significantly in the 6 months leading up to the layoff event. Second, we compare employment and earnings outcomes of women employed in firms with a mass layoff or a closure with a comparison group of similar women employed in firms with no layoff event. In line with the previous findings, we show that women affected by a layoff event at their workplace experience economically large losses after the event. The magnitude of the losses is similar in both types of layoff events. Third, we study the development of conceptions, births, and abortions around the layoff event. In the year preceding the layoff event, we find an increase in conceptions of women employed in firms

[2]Convention 183, Article 8

with closures relative to the comparison group (10 out of 1000 women increase, $p < 0.05$). We also observe a large increase (5 out of 1000 women) in the case of mass layoffs, but the finding is statistically insignificant ($p > 0.1$). The large positive coefficient estimates are in line with the precautionary motive: women who anticipate the layoff respond by becoming pregnant. Birth and abortion outcomes of pregnancies conceived in the year preceding the event differ by event type, however. While births increase in firms with a mass layoff, abortions increase in firms that are closing. Effect sizes are of the same magnitude in absolute terms: in case of a mass layoff event births increase by 8 out of 1000 women ($p < 0.01$), and abortions increase by 7 out of 1000 women ($p < 0.05$) before a closure event. This finding is evidence of the riskiness of the precautionary strategy. A pregnancy cannot protect a woman's job or career if the firm ceases to exist. She must find a job with a new employer and loses the high maternity benefit if she becomes unemployed before giving birth.

We perform heterogeneity analysis to test the robustness of these findings. First, we identify groups with relatively high pregnancy rates who should be more flexible in timing their fertility in response to the threat of a layoff. We show that effects are indeed driven by young women and women with a high probability of getting pregnant. Second, we identify groups with high abortion rates conditional on getting pregnant. These women might be more likely to use abortions as a form of contraception. Our results show that women with high abortion probability are driving the increase in abortions in the closure sample. However, there is no difference in fertility responses between women with high and low probabilities of abortion in case of a mass layoff. These findings suggest that our results are due to strategic fertility decisions rather than responses to unplanned pregnancies.

Our research contributes to several strands of the literature on the effects of economic shocks on fertility and abortions. First, a large literature has studied the cyclicality of fertility in various settings (Adsera, 2005; Dehejia and Lleras-Muney, 2004). But relatively few studies address the effects of economic conditions on abortions. The primary objective of these studies is to test whether in times of economic hardship abortions are increasingly used to terminate unplanned pregnancies. Several studies confirm this hypothesis and document that lower unemployment or increased generosity of income support programs tend to reduce abortion rates (Blank et al., 1996; González and Trommlerová, 2021; Herbst, 2011). Abortion rates in Hungary are generally high compared to Western European countries, like Germany, and closer to rates in the UK and the U.S., which makes our findings relevant to this literature. Our results reveal an interesting time pattern. We show a large abortion response only in anticipation of the initial employment shock, but in the years after the shock, the effects on abortion rates are smaller and insignificant. This suggests that abortions are less important in dealing with income losses in the longer run.

Second, studies investigating the effects of job displacements at the individual level have – due to the lack of data on abortions – focused on fertility responses *after* the loss of a job and studied total fertility effects by looking at medium to long run outcomes (Del Bono et al., 2012; Huttunen and Kellokumpu, 2016). Our medium-term results in the first three years after displacement show a slight decline in the number of births which confirms the previous literature. We contribute a new result on abortions and find no significant change in abortions relative to the comparison group in the years after displacement.

Third, we also contribute to the literature studying the anticipation of job loss. Survey evidence confirms that individuals have some prior knowledge about a future job loss (Hendren, 2017; Mueller and Spinnewijn, 2022). But it has been hard to deal with anticipation in a setup studying employment effects of mass layoffs and plant closures, as affected individuals are by construction required to remain employed until the shock occurs (Schwerdt, 2011). Halla et al. (2020) conclude that wives of displaced husbands adjust their job search intensity only after the shock has occurred. Our fertility results draw a more nuanced picture indicating that women anticipate their own job loss.

Lastly, our results also contribute to the large literature studying the effects of family policies. We show how maternity policy can affect fertility decisions when women face a high risk of job loss. Women who remain employed and thus eligible for high maternity benefits choose to bring forward their planned fertility to the period of uncertainty and thereby potentially rescue their careers. But women who lose their jobs and their access to high maternity benefits are more likely to terminate their pregnancies. This result implies that there is still scope to improve protection.

In the next section, we discuss the trends in births, abortions, and the relevant institutional background. We present a simple model of the anticipatory fertility decisions in Section 1.3. Section 1.4 describes the data, and the empirical strategy is introduced in Section 1.5. We present our main results and the related robustness checks in Sections 1.6 and 1.7. We conclude in Section 1.8.

1.2 Fertility trends and institutional background

1.2.1 Births and abortions

Hungary is a small developed country with low fertility, wide access to abortions, and a generous state-financed maternity benefits system. To put the Hungarian institutional and fertility landscape in context, we present it along with data on other developed countries.

The number of births per 1000 women of reproductive age (15 to 49 years) was around 40 in Hungary in our period of interest (2009-2017). This birth rate is close to the EU average of 43 to 44 and lower than the birth rate above 60 in the US in this period (Figure 1.1a).

In Hungary, the number of abortions has been steadily declining since the 1990s, but in 2016 it was still 33% of the number of births. The abortion rate, i.e abortions per 1000 women of age 15 to 49 (15 to 44 in the US) was 13.3 in 2016, slightly higher compared to the US (11.6) and the UK (10.4), and significantly higher compared to Germany (4.4) (Figure 1.1b).

Most births and abortions in Hungary take place in public healthcare institutions. Deliveries are financed by the National Health Insurance Fund which covers every citizen during the observation period. Abortion is not covered by this fund, but the price is low, about USD 90 to 100 in the period of our study (37 to 41 percent of the local minimum wage in 2010), and it can be further decreased if the woman proves financial difficulties. According to the categorization of the Guttmacher Institute, access to abortion is very easy in Hungary, similar to most developed countries Singh et al., 2018. Abortions can be legally carried out on request before the 12^{th} week of pregnancy, after having two consultations with the staff of the Family Protection Services.[3] All legal abortions are carried out surgically, as abortion pills are not authorized.[4]

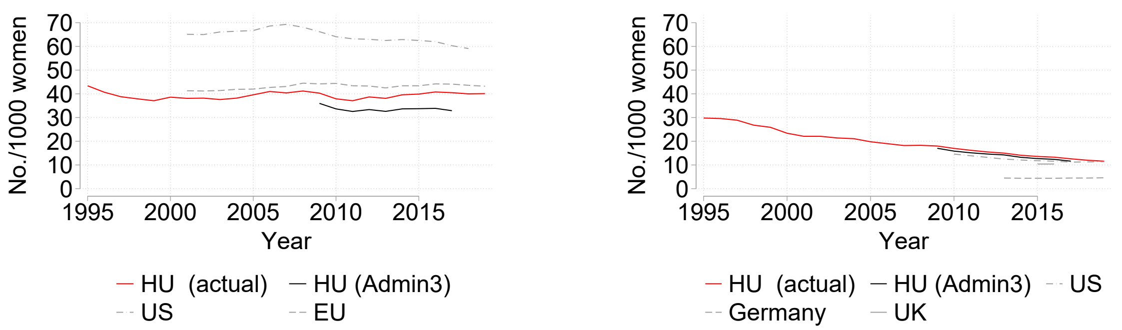
Other forms of contraception are largely available for Hungarian women: according to survey estimates more than 90 percent of women use some type of modern contraception. However, in the European context, Hungary is one of the worst-performing countries in terms of access to contraceptives. The main reason behind this is that contraceptive pills require a prescription, and

[3] Law 1992/79.

[4] As a minor exception, abortion pills were used by a private medical institution in Hungary between 2010 and 2012. Index, 2012

Figure 1.1: Birth rates and abortion rates (1995-2020)

(a) Birth rates (b) Abortion rates



Data source: US: CDC and Guttmacher Institute; EU: Eurostat birth data and Eurostat abortion data; HU: KSH
Note 1: US figures refer to women of age 15 to 44. As fertility is lower at the ages of 45 to 49, this in itself could lead to higher birth rates in the US even if the age-specific fertility was the same. But the difference is not substantial. In Hungary, the birth rate in the 15-44 year age group is 38.2, compared to the 37.1 birth rate in the 15-49 years age group.
Note 2: The difference between the official Hungarian live birth statistics and our estimation data (Admin 3) stems from omitting births in private hospitals and births at home.
Note 3: The EU average of abortion rates is not available due to missing data for some countries. Instead, we report selected country-level data.

the national healthcare system does not cover the price, not even for vulnerable groups.[5] Regardless of this more limited access to contraceptives, the rate of unplanned pregnancies (24 percent of all pregnancies) is close to the European mean (25 percent) based on the estimates of Bearak et al. (2022). This finding suggests that the relatively higher abortion rate in Hungary is not the result of less careful contraception but rather comes from the fact that a larger share of unplanned pregnancies gets aborted.

1.2.2 Family policy

Hungary provides a generous system of maternity benefits, especially for employed women OECD, 2022. Child-related benefits (Appendix Table A.2) are linked to previous employment and wages, and women are generally eligible for benefits until the 2nd birthday of the child. Specifically, women who have been employed for at least 12 months in the two years preceding childbirth and are employed until 42 days before childbirth, are eligible for a baby-care allowance until the child is 6 months old, and a childcare benefit from 7 to 24 months of age of the child. Both the baby-care allowance and the childcare benefit pay 70 percent of the previous wage, but while the baby-care allowance is uncapped, the childcare benefit is maximized at a fairly high level (1.4 times the minimum wage). If a woman becomes unemployed during pregnancy, she will be entitled to a 50 to 70 percent lower amount.

Dismissal protection laws prohibit firms from laying off a pregnant employee, once she has informed the employer about the pregnancy, except if she seriously neglects her duties. Also, she

[5] Source: Contraception Policy Atlas, https://www.epfweb.org/sites/default/files/2023-02/Contraception_Policy_Atlas_Europe2023.pdf, downloaded (2023.07.18)

has a guaranteed right to return to her previous job at the end of maternity leave. In our data, 41 percent of non-pregnant women get displaced in the mass layoff sample, while the same share for pregnant employees is only 20 percent, showing that pregnancy substantially decreases the layoff risk.[6] Similarly strong dismissal protection policies are implemented in many European countries (e.g. Austria, Belgium, France, Germany, Italy, etc.). In other countries (e.g. USA, UK, Canada), dismissal protection is weaker and is restricted to protection from discriminatory dismissal (ILO, 2022).

1.3 Fertility decisions around job displacement

The empirical literature shows that fertility decisions are shaped by the institutional framework with parental leave regulations and dismissal protection laws (see e.g. De Paola et al., 2021; Fitzenberger and Seidlitz, 2023; Lalive and Zweimüller, 2009) as well as by economic conditions. In this section, we outline a short theoretical framework that explains how institutions and economic shocks might interact in determining fertility decisions. We follow the spirit of dynamic models of fertility Hotz et al. (1997) where a woman decides on the optimal timing of birth. This framework will be useful to motivate our interpretation of fertility responses in anticipation of the two types of employment shocks considered in our setup, mass layoffs, and firm closures.

In the Hungarian context, the level of dismissal protection and leave benefits for pregnant women differs substantially between job displacements from firm closures and mass layoffs. This is due to two features of family policy. First, dismissal protection for pregnant women is only available as long as the firm exists but is lost when the firm closes. Second, high maternity benefits and the option to return to the previous job after the leave are only available for employed women. But a woman who loses her job from firm closure during pregnancy falls to the low benefit level and has no job to return to.

We consider an employed woman who decides whether or not to get pregnant. She derives income from employment while working and parental leave benefits after giving birth. We assume that her income increases with her job tenure. After giving birth the mother takes a period of parental leave, receives the benefit, and subsequently returns to her previous job. A layoff is associated with the loss of firm-specific capital and the need to restart the career with a new employer, which puts her at a lower position in the tenure profile. Figure 1.2 schematically summarizes income flows around the birth of a child in Panel (1.2a) and in case of a job loss in Panel (1.2b).

Next, we consider how these career interruptions interact to determine the timing of fertility decisions for women who anticipate a job loss. Panel (1.2c) shows how a precautionary pregnancy helps avoid income losses from job displacement in a mass layoff. If the woman starts her pregnancy right before the displacement, she is protected from layoff. Instead of having to restart her career at a new firm she collects maternity benefits and thereby waits out the crisis at her firm and then, re-enters the firm after the leave period. Compared to the dashed line which denotes the income profile without pregnancy, precautionary fertility timing can avoid large income losses.

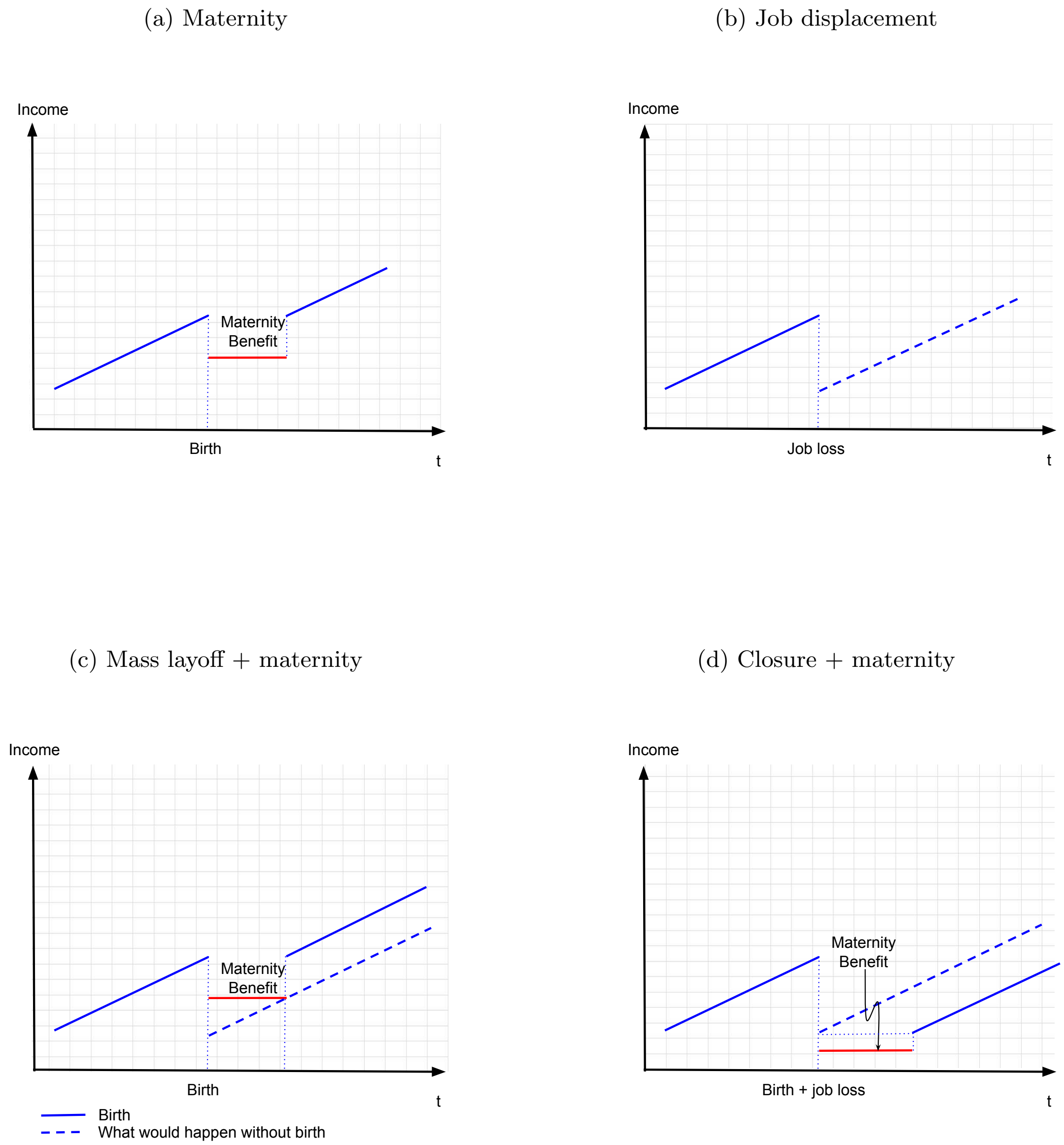
In case of a firm closure, depicted in Panel (1.2d), things work out worse than that. As the firm stops existing, the woman loses her job. If the firm closes while she is pregnant, she receives

[6] Even if dismissal protection was perfect, it would be possible that some women are displaced in our data while they are pregnant, first because we include voluntary separations from the firm as well, second because pregnant women can be dismissed if they do not fulfill work requirements, and third because not every woman announces pregnancy to the employer, and dismissal protection can be only enforced in this case. In addition to these, anecdotal evidence shows that some employers try to trick the laws to be able to dismiss pregnant employees, e.g. pressuring the pregnant woman informally to leave the job "voluntarily".

low maternity leave benefits unless she manages to find a new job shortly after the firm closure. But finding a job while pregnant is difficult: evidence from the literature supports that displaced pregnant women suffer relatively high losses in employment and working hours Meekes and Hassink, 2020. In any case, the woman has no option to return to the pre-displacement job with high earnings but she has to restart her career after the maternity leave period. Her income loss from giving birth around job displacement is thus larger than the income loss without birth, which can be seen from the comparison of the dashed and the solid lines.

The figures illustrate the risk involved in a precautionary pregnancy for women who do not yet know the exact type of employment shock when they get pregnant. In case of a mass layoff, the precautionary pregnancy helps avoid any income loss from displacement. But if the firm closes, the combined income loss from maternity and job displacement is the largest and can only be avoided by terminating the pregnancy. In Section A.3 in the Appendix, we present the formal derivations of the model.

Figure 1.2: Income flows in case of four states of the world



(a) Maternity

(b) Job displacement

(c) Mass layoff + maternity

(d) Closure + maternity

1.4 Data and sample

1.4.1 Data

We use administrative individual-level monthly panel data [7]. The data are hosted by the Databank of the Centre for Economic and Regional Studies and link administrative records of the National Health Insurance Fund Administration, the Hungarian State Treasury, the National Tax and Customs Administration, the Ministry of Finance and the Educational Authority, based on anonymized social security numbers. For a more detailed description of data compilation and cleaning, see Sebők (2019). The data contain information on 5.17 million people, a random 50 percent sample of the Hungarian population[8] drawn in 2003 and followed until 2017. We observe gender, age, county of residence, employment, occupation, wages, state transfers, registered unemployment, and employer identifiers each month. The employer identifiers are linked to a yearly database covering firm-level information on firm size, sector, foreign ownership, and revenues.

We use daily healthcare records to measure fertility outcomes. This part of the dataset contains the International Statistical Classification of Diseases (ICD) codes and dates of each person's public hospital visits. These data are only available for the years between 2009 and 2017. Based on these records, we can identify births (ICD codes O6, O7, and O8) and surgical abortions (ICD code O04) at public hospitals. These records cover the majority of the relevant events: we observe 88-93 percent of births and 95-98 percent of abortions reported in the official summary statistics (see Table A.1). Some of the childbirth records could be missing because of children born in private institutions, at home, or abroad, while the missing abortions are due to abortions in private institutions.

After aggregating birth and abortion data to the monthly level, we link them to individuals at the estimated date of conception. Throughout the analysis, we use the conception date instead of the date of the actual childbirth or abortion. This means that when we compare abortion and birth frequencies, we talk about pregnancies conceived at the same time. As we do not observe the date of conception, we pin down the conception dates 9 months before childbirth and 2 months before an abortion. Although these are crude approximations, they are very close to the actual conception date in the majority of the cases: administrative birth records[9] show that 90.9 percent of children were born about 9 months after conception (37^{th} to 41^{st} obstetric weeks), and 83.1 percent of abortions were carried out about 2 months after conception (7^{th} to 11^{th} obstetric weeks) in the period between 2003 and 2020 in Hungary. (See Appendix Figure A.1)

In our analysis an important outcome is the number of pregnancies, which we calculate as the sum of births and abortions, omitting miscarriages. Miscarriages amount to about 10% of all pregnancies according to the official records and their number is rather stable over time. The reason for not using miscarriage data in this study is its weaker reliability. Only less than 10% of miscarriages reported in the official summary statistics can be identified in our administrative dataset. Appendix A.2 discusses the potential impact of measurement error on the estimated effect

[7] The administrative database used in this paper is a property of the National Health Insurance Fund Administration, the Central Administration of National Pension Insurance, the National Tax and Customs Administration, the National Employment Service, and the Educational Authority of Hungary. The data was processed by the Databank of the Centre for Economic and Regional Studies. The present study has been produced using the corporate financial statement and performance statistics data files, and the live birth, pregnancy termination, and miscarriage data files of the Hungarian Central Statistical Office. The calculations and the conclusions within the document are the intellectual product of the authors.

[8] Foreigners are not included in the data. However, the share of foreigners is small: the Hungarian Statistical Office reported less than 2 percent of the population was non-Hungarian in 2009.

[9] Hungarian Central Statistical Office, Live birth database

of job loss on observed pregnancies.

1.4.2 Sample

1.4.2.1 Firm closures and mass layoffs To form our treatment sample, we first identify closures and mass layoffs of private for-profit firms in the data and restrict our attention to those that happened between 2010 and 2014. This way, for each woman we observe at least 1 year of abortion and birth history (and 8 years of employment and earnings history) before the shock, and at least 3 years after that.

We define the date of a firm closure as the month when the number of employees drops to 0 and stays 0 for two consecutive years. We take multiple cautionary steps to avoid including "false firm deaths" Kuhn, 2002 when instead of real closure, a firm ID disappears for some other reason (e.g. ID change due to a new legal form, or a merger). First, we require firms to exist for at least 2 years preceding the closure. Second, similar to other papers in the literature using firm closures for identification (e.g. Eliason and Storrie, 2006), we only include firms if the number of employees is at least 10 at least once in our observed period, based on yearly firm records. We also require that the number of employees present in the data is at least 5 in the month before closure.[10] Third, we exclude firms if more than 30 percent of employees transferred to the same new firm after the month of closure, and if at one receiving firm, at least five people and 30 percent of the new entrants to the firm came from this same sending firm.

The date of a mass layoff is pinned down at the month when the number of employees decreases by at least 20 percent and does not increase for 12 months following the decrease. If there are multiple mass layoffs at one firm, we include all of them. We drop those few firms which experience a mass layoff and a closure as well. We use the same criteria of firm size and age for downsizing firms and closures. Again, to avoid false layoffs, we exclude firms from the sample if more than 10 percent of previous employees move to the same new firm after the layoff.

1.4.2.2 Definition and description of the treatment and control groups We define two treatment groups: women affected by closures, and women affected by mass layoffs.[11] We include all women in the sample who are employed at firms when a layoff event occurs, even if they are not getting displaced. As a result, in the closure treatment sample every woman loses her job, whereas, in the mass layoff treatment sample, only a fraction (41 percent) is displaced (see Figure A.2).

Women in the treatment groups are required to satisfy the following selection criteria: they have to be of reproductive age (15-49 years), work at the firms in the quarter preceding the layoff event, and have at least 12 months of tenure at the time of the event.[12]

We follow the approach of Del Bono et al. (2012) and include not only women who stay at the firm until the last month before the layoff event but also those who leave two or three months before that. The reason is that workers who stay until the very end are a selected sample, and including early leavers mitigates this selection. However, we exclude those leaving even earlier than three months. As employment of young women of reproductive age is unstable, and we do not observe the reason for leaving a firm, these very early separations are more likely to be voluntary.

[10]As in our data 50 percent of the Hungarian population is included, requiring 5 employees in the individual-level data means that the firm's actual size before the month of closure/mass layoff is required to be at least 10 on average.

[11]Women affected by multiple closures or mass layoffs are excluded from the sample. 87 percent is affected by only 1 event.

[12]Employment is defined by the employment status, and working at a firm with a valid firm id variable in the administrative data. We use uninterrupted employment spells at the same firm to define tenure.

Table 1.1: Number of observations in different steps of the sample selection

Sample		Closure		Mass layoff	
		Treated	Control	Treated	Control
Reproductive age, min 12 months of tenure, firm min 2 years old		4282	18506491	13838	18506491
Exact match on wage history, county, age, calendar month		3628	230438	11558	637384
Matched on the propensity score	No caliper*	2707	19154	8299	60751
	Caliper (main)	**2496**	**16860**	**4068**	**19736**
	Strict caliper*	2276	14359	3388	14155

Note: In cases when a woman can serve as a potential control for multiple time periods, she is counted as a separate observation for each applicable period. The sample with bold text is used in the main analysis. We run robustness checks on the samples denoted with *.

Requiring 12 months of tenure ensures that, in case of giving birth, the woman would be eligible for the high child benefits linked to previous employment, had the firm not closed. It also makes our results comparable to previous studies, using the same tenure criterion e.g., Del Bono et al., 2012; Huttunen and Kellokumpu, 2016.

To form the control groups, we employ a combination of exact matching and propensity score matching based on individual and firm characteristics. Our strategy aims to ensure that the treated and control groups are as comparable as possible in terms of observable features. Given the limited number of pre-treatment years for births and abortions and our focus on employed women, the incidence of births in these years is exceedingly low. Hence, even if we observe a similar number of births pre-treatment, it does not necessarily imply that we should expect these groups to exhibit the same number of births in the absence of shocks. Therefore, we conduct careful matching on multiple variables to ensure that the women are genuinely similar and do not merely appear similar based on the mechanically low pre-treatment birth rates. Our dataset offers an exceptionally large pool of potential control women, even when applying stringent matching criteria. However, the main constraint lies in the limited number of treated women, leading to a loss of some of them when striving for closely matched controls. We present the number of observations during the selection process in Table 1.1. There is a trade-off between achieving similarity and maintaining an adequate number of observations. To address this issue, we conduct robustness checks using both larger and smaller control groups. In the following, we describe the details of the matching.

The reference month, in which the matching is done, is set to the last month before the closure or mass layoff generally.[13] First, for every treatment woman, we find a pool of possible control women who work at non-closing and non-downsizing firms in the reference month and satisfy the other selection criteria used for treatment women (i.e are of reproductive age, and have a minimum of 12 months of tenure at their firm). From this pool of control women, we match exactly on age group (15-19, 20-24, etc), county of residence, and yearly wage category history (0-50000 HUF; 50000-10000 HUF; etc.) from the 4^{th} year to the 1^{st} year before the reference month. Note that we do not use the wage in the year of the closure, as these wages might already be affected by the coming shock in the treatment group. The exact matching ensures that the treatment and control

[13] For those who leave the firm 2 or 3 months before the closure or mass layoff, the reference month is set to the last month when they still work at the firm

women are comparable in the aspects we find most important. They are the same age and from the same region with the same wage history at the time of matching. In addition, matching control women in a specific month automatically pins down the date of the pseudo-event for them.

Then, from the exact matches we select the (maximum) 10 nearest neighbors within a caliper based on the propensity score.[14]. The propensity score is estimated using a probit model:

$$P(T_i = 1|X) = \Phi(X_i'\beta_i), \tag{1.1}$$

where T_i is a binary variable equal to 1 for treated women, and X_i denotes a large set of independent variables, including individual and firm characteristics[15]. The following variables in X are measured right before the event: the woman's age (in years), occupation (9 categories), an indicator of having a young child (based on previous child transfers received by the woman), tenure (in months), and experience (in months). We also include longer histories of wages, and months spent employed, from year -5 to year -1. In addition, X_i includes firm characteristics: size, revenue, foreign ownership, and sector measured one year before the reference month. Note that we do not use firm characteristics in the year right before the shock. Closing and downsizing firms already experience some distress before the actual shock happens, and we want to avoid matching on characteristics already affected by the coming events.

For the closure sample, the caliper is set to 0.09, and for the mass layoff sample to 0.001. In choosing the caliper there is a trade-off: with a small caliper we end up with very similar control women but lose both treatment and control observations if there are no close-enough matches, while a large caliper (or no caliper at all) allows for keeping many observations but at the expense of reducing similarity.

[14]In the analysis, we use sample weights to account for the fact that for some treated women fewer than 10 controls are matched, and that some controls are matched to multiple treated women. The weight of a treated observation is always 1. The weight of a control observation depends on the number of treated observations she is matched to and reversely depends on the number of other controls in the same exact match set. The weights are the default weights calculated by the psmatch2 package of Stata by Leuven and Sianesi, 2003. The weighting is important on the descriptive graphs, as it ensures that we do not overweight observations in cases when we find more matches to treated women. However, for the regression results, entirely omitting the weighting would leave our results practically unchanged, as we use exact match fixed effects in our main models.

[15]The matching is implemented using the Stata package psmatch2 Leuven and Sianesi, 2003.

Table 1.2: Means in the treatment and control groups

		Time of measurement	Closure			Mass Layoff		
			Control	Treated	Diff.	Control	Treated	Diff.
Age		Year 0	36.2	36.2	-0.014	38.2	38.2	0.005
Receives child benefits		Year 0	0.049	0.038	-0.011**	0.028	0.025	-0.003
Tenure (months)		Year 0	46.6	43.3	-3.326***	58.7	61.4	2.704***
Experience (months)		Year 0	81.6	82.2	0.665*	89.8	91.0	1.199***
White collar		Year 0	0.48	0.49	0.001	0.38	0.33	-0.054**
Wage (10000 HUF)		Year 0	13.53	13.38	-0.15*	14.17	14.38	0.22***
Percent losing job		Month 0	3.08	100.00	96.92***	2.89	40.71	37.82***
Firm characteristics								
Size	Small (-49)	Year -1	0.48	0.64	0.156***	0.35	0.30	-0.044***
	Medium (50-249)	Year -1	0.31	0.21	-0.100***	0.29	0.30	0.008
	Large (250-)	Year -1	0.21	0.15	-0.056***	0.36	0.39	0.036***
Log revenue (Mn HUF)		Year -1	6.603	5.678	-0.925***	7.331	7.475	0.143**
Avg. wage (10000 HUF)		Year -1	14.89	13.80	-1.09***	15.44	14.79	-0.64***
Foreign owned		Year -1	0.20	0.14	-0.056***	0.33	0.35	0.024**
Firm age		Year -1	7.39	6.20	-1.185***	7.72	7.71	-0.009
Women (15-49) share		Year -1	0.46	0.48	0.029***	0.46	0.46	-0.003
Fertility variables								
Pregnancies		Year(-3)-(-1)	0.014	0.012	-0.003	0.010	0.010	0.001
		Year 0	0.034	0.041	0.008*	0.029	0.033	0.004
Births		Year(-3)-(-1)	0.002	0.003	0.001	0.001	0.001	0.000
		Year 0	0.024	0.026	0.003	0.019	0.027	0.008***
Abortions		Year(-3)-(-1)	0.012	0.009	-0.003**	0.009	0.010	0.001
		Year 0	0.010	0.015	0.005*	0.011	0.007	-0.004*
Calendar year								
2010		Month 0	0.17	0.17	0	0.20	0.20	0
2011		Month 0	0.18	0.18	0	0.23	0.23	0
2012		Month 0	0.23	0.23	0	0.31	0.31	0
2013		Month 0	0.28	0.28	0	0.14	0.14	0
2014		Month 0	0.15	0.15	0	0.11	0.11	0
Number of observations			16860	2496		19736	4068	

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Receiving child benefits includes all benefits available up to the 3rd birthday of the child.

Hungarian DOI: 10.14754/CEU.2023.04

We choose calipers in a way to achieve a balanced sample in the sense that none of the independent variables of interest are substantially different in magnitude in the treatment and the control group. While our identification of parallel trends allows for different treated and control samples, we thrive for treatment and control samples that are as similar as possible, as pre-treatment fertility data is limited in time, and births are mechanically low. After the end of the sample selection, we have 2496 treated and 16860 women in the closure sample, and 4068 treated and 19736 control women in the mass layoff sample.

Table 2.3 presents the descriptive differences for some individual and firm-level variables. The table reveals that the treatment and control samples are similar, but not the same within the same event. For example, the age, share who receives child benefits, tenure, experience, occupation, and wages are reasonably close between treatment and control samples for both types of events. While the share of women of reproductive age in the workforce is around 46-48 percent for all samples, there is a larger difference in terms of other firm characteristics, especially for closures: treatment firms are generally smaller with lower revenues, wages, and foreign ownership. For closures, the firms are also statistically significantly younger.

We present the means of the fertility outcomes in the pre-treatment period separately for the year just before the events (Year 0), and the (maximum) 3 years preceding it (Year(-3)-(-1)). The number of conceived pregnancies is around 1 per 100 women from year -3 to year -1 for closures and mass layoffs as well, both for treated and control women. In this period most of the pregnancies get aborted, and the births are close to 0 in these years, which is a consequence of including only employed women in our sample. In the case of closures, abortions are significantly less likely in the treatment group in this period (by 3 abortions per 1000 women). In year 0, there are larger differences between the fertility variables, which, as described above, can already be the consequence of the anticipated shock.

In addition to differences within the samples, our samples are somewhat different between the two types of events as well. Women working in closing firms are on average younger, with lower tenure and experience than those working in firms with mass layoffs. Women in closing firms are more likely to work in white-collar occupations but they have on average lower wages than women in firms with mass layoffs. Closing firms tend to be smaller than firms with mass layoffs, they are less likely to be foreign-owned, and they have lower revenue in the year before the event. Note that while all women in the closing firms lose their jobs, only 41% of women working in the mass layoff group are displaced. There are some differences in the calendar year in which we match: for mass layoffs, a higher share of women is affected by events before 2013, but both samples include better and worse years in terms of GDP growth. In most specifications, we run separate regressions on the two samples for different events, but we also show pooled results to make sure that the differential responses to the shocks are not driven by the differences between the two samples.[16].

1.4.3 Firm outcomes around the layoff event

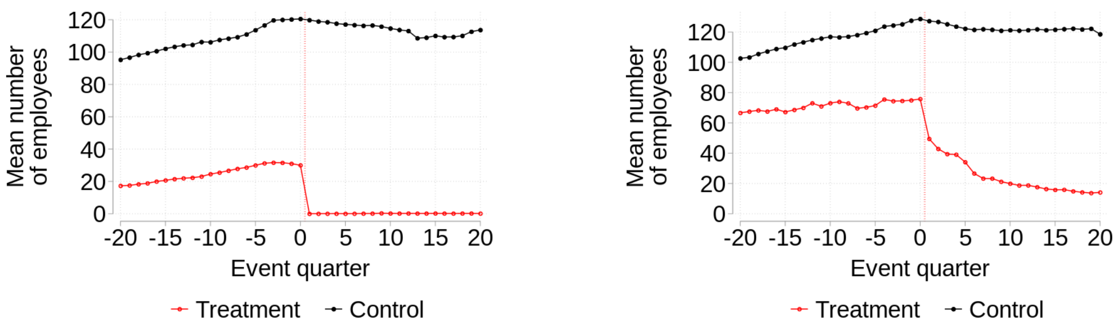
In this section, we discuss firm dynamics around the layoff event, looking at variables that might trigger the anticipation of layoff events among employees. First, Figure 1.3 shows that the evolution of the number of employees follows similar dynamics in treated and control firms. We do not see large numbers of employees exiting prior to the layoff events, but firm growth appears to be somewhat slower in treated than in control firms in the years leading up to the event. Annual firm revenues show a similar pattern. Log revenues grow a bit slower in closing firms in pre-treatment years, with

[16]2014 was the best year in the period for GDP growth (4.2 percent), and 2012 was the worst (-1.3 percent).

a decrease in revenues in the year of the closure. Growth in firm revenues also stops in the year of the mass layoff, and there is a substantial drop in the next two years (Figure A.3)

Figure 1.3: Firm size around the layoff event

(a) Closure (b) Mass layoff



The last month of Quarter 0 is the month of the layoff event. For control firms, the date of the pseudo-event is set to the month when the most control women are matched

On the other hand, data on new orders in the manufacturing sector shown in Figure 1.4 demonstrates that orders start decreasing significantly on average 6 to 12 months before the layoff event. This pattern indicates that treated firms suffer negative shocks leading to the layoff event at the end of year 0. We assume that the negative shock can be observed either by the women themselves or by their colleagues who pass on the information. Survey evidence also confirms that individuals have some prior knowledge about a future job loss (Hendren, 2017 Mueller and Spinnewijn, 2022).

Thus, it appears that the firms in the treated group typically start to experience problems during the year before the layoff event. It is plausible that employees can sense these problems, and update their expectations about the probability of a future job loss, already before the layoffs happen.

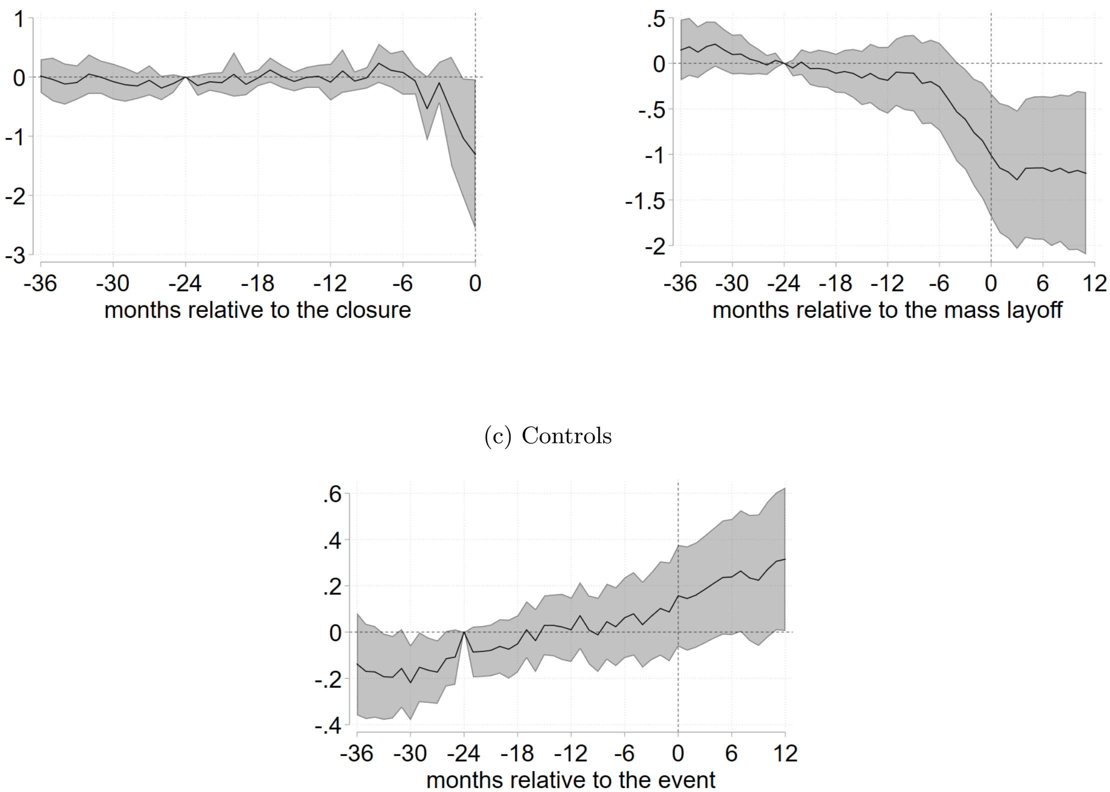
Even though women can perceive economic problems at the firms, it might be hard to predict the actual outcomes. This idea is supported by the interview we conducted with a liquidation commissioner (who supervises liquidation procedures at firms). In general, when a firm starts to face problems, rumors start to spread around among the employees. After that, the firm can recover and go on with the business, there can be mass layoffs, or the firm can close altogether. But when the problems start, no one knows for sure how the troubles are going to end. Probably everyone assigns different probabilities for each outcome. The initial expectations are updated later when more information is revealed about the type of shock, and the behavior of the employees adjusts accordingly.

1.5 Empirical strategy and identification

1.5.1 Empirical strategy

In our empirical strategy, we estimate event study and difference-in-differences models on the sample defined in Section 1.4. First, we run the following event study regression:

Figure 1.4: Firm orders



(a) Closure

(b) Mass layoff

(c) Controls

Sample: matched manufacturing firms with closure (a) or mass layoff (b) or with control women (c).
Regression: $logNewOrders_{it} = \sum_{k=-36}^{12} \beta_k EventMonth_{it}^k + \alpha_t + \alpha_i + \epsilon_{it}$ where i is firm, t is month, α_t is calendar month fixed effect and α_i is firm fixed effects and ϵ_{it} is the error term. Figures present the estimated β_k with the 95% confidence interval. We assign a single month of the event to control firms based on the highest number of women working there and being used as controls with that specific event time, considering closure and mass layoff events jointly.

$$Y_{it} = \alpha + \beta T_i + \lambda_t + \sum_{\substack{k=-5 \\ k\neq-3}}^{k=5} [\delta_k (T_i \times \mathbf{1}_{k=t})] + \gamma \hat{P}_i + \tau \mu_{g(i)} + u_{it} \tag{1.2}$$

where Y_{it} denotes the outcome variables: average wages, employment indicators, number of births, abortions, and pregnancies measured at the time of conception for woman i in event year t. The layoff events (or pseudo-events for control women) take place between the last month of event year 0 and the first month of event year 1.[17] T_i is the treatment assignment indicator, with value 1

[17]For treated women, the last month of event year 0 denotes the last month when they still work at the closing

if woman i worked at a firm with a layoff event in the three months preceding the event. Note that T_i is 1 for individuals working at downsizing firms even if they are not displaced. Event year fixed effects (λ_t) are also included. The coefficients δ_t are of main interest, showing the treatment-control difference in the outcome in event year t relative to the difference in the baseline event year (year -3).

$\mu_{g(i)}$ denote exact match dummies in match set g. To control for remaining differences in the pre-treatment characteristics of women (see Table 2.3) we include the propensity score ($\hat{P}_i$) estimated in equation 2.4. Calendar year fixed effects are not included in the equation, because the matching is done in a given month, so including exact match dummies controls for calendar time.

To get robust standard errors accounting also for the fact that the regression is run after matching, we cluster the standard errors by exact match sets. Abadie and Spiess, 2020 show that standard errors clustered like this are valid in regressions run after matching even if the regression equation is misspecified with regard to the population regression equation. Their results apply to non-parametric nearest neighbor matching without replacement, while we match on the propensity score within the exact match sets, and allow for replacement. As we are not aware of analytical results for the correctly specified standard errors with this extra detail in the matching, in addition to clustered standard errors, we also calculate standard errors by bootstrapping for the main coefficients of interest.

After estimating yearly effects, we pool event years into three separate time periods, and run three-period DiD regressions for the same outcome variables, using the following equation:

$$\begin{aligned} Y_{it} = \alpha + \beta T_i + \gamma_1 Year_t^0 + \gamma_2 Year_t^{1,2,3} + \delta_1 (T_i \times Year_t^0) + \delta_2 (T_i \times Year_t^{1,2,3}) + \\ + \gamma \hat{P}_i + \tau \mu_{g(i)} + u_{it} \end{aligned}, \tag{1.3}$$

where $Year_t^0$ is a dummy equal to 1 in event year 0 (the year just before the event), and $Year_t^{1,2,3}$ is a dummy equal to 1 in event years 1 to 3. The reference time period is all event years available before year 0. Using these three stacked time periods is motivated by the theoretical results suggesting that women already react to the coming layoff event before it actually happens. We interpret δ_1 - the treatment-control difference in the outcomes in event year 0 relative to the difference in the reference time period - as the effect of anticipating the coming closure or mass layoff. The coefficient δ_2 shows the average yearly effect of the shock in the following three years.

1.5.2 Identification

The identifying assumption of equations 1.2 and 1.3 is parallel trends conditional on observables. I.e. had the shock of the layoff event not affected the treatment group, their fertility would have changed the same way as that of the control group.

We took multiple steps to support this assumption. First, we ensured by the matching that controls are similar to treated women on many observables. Along with variables measured right before the shock, the matching also includes 4-year histories of wages and employment: this makes it more likely that women in the treatment and control group are not only similar right before the shock, but they are also on similar paths in their careers.

or downsizing firm, or in case of those women who end up not leaving a downsizing firm, it denotes the last month before the mass layoff. For control women, the last month of event year 0 is the month when they are matched to treatment women.

Second, we restricted our sample to women with at least 12 months of tenure and matched on firm characteristics one year before the shock. The average tenure in our treatment and control sample is almost 4 years in case of closures and around 5 years in case of mass layoffs. This increases the probability that the estimated fertility effects are not driven by some underlying variable correlated with firm and fertility choice. We argued in Section 1.4 that problems at the firm might already be present and perceived by employees in the year preceding the closure or layoff (one sign of these problems being the decrease in new orders of a firm). Women who start working at the firm during the year before the layoff might also perceive that the firm they are getting hired at is in trouble, and more risk-tolerant women might accept jobs at these types of firms with a higher probability. These women might be more prone to other risky behavior, and by including them, we might overestimate abortion effects because of this. By including women with long tenures, and by matching on firm characteristics, we minimize the probability that women know that they are getting employed at a firm that is about to have a layoff event in the future, at the time when they are hired. In the descriptive statistics, for pre-treatment years (-3) to (-1) we see that treated women are less likely to have an abortion in the case of closures than control women, while for mass layoffs we do not see a difference in abortions for treated and control women in these years. This supports that it is unlikely that women with higher abortion probabilities select to work at firms that are more likely to be affected by a layoff event in the future.

Third, we not only include women who stay until the last month of closure or mass layoff but include also those who leave the firm earlier to mitigate selection over the downsizing period.

1.6 Results

1.6.1 Event study estimates

In this subsection, we present the raw yearly means of the outcome variables and the yearly event study estimates of Equation 1.2. These results provide a general picture of the yearly evolution of the outcome variables and the dynamics of the effects.

1.6.1.1 Labor market outcomes: employment, wages First, we present evidence that women suffer large and persistent economic losses after closures and mass layoffs. Figures 1.5 and 1.6 show the raw means and the estimated yearly treatment effects (δ_t-s in Equation 1.2) for two outcomes: an indicator for being employed throughout the given year, and the mean yearly wage.

The career of treatment and control women evolves similarly before the shocks: employment and wages steadily increase for both.

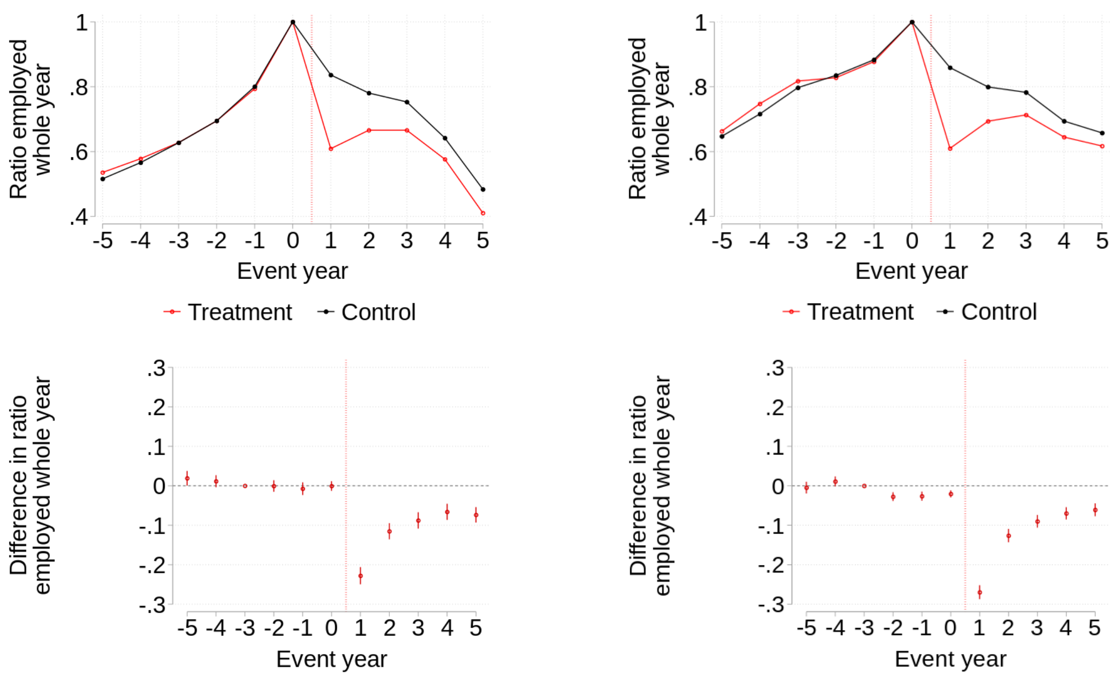
The share of women working throughout the year before the shock is 1 – a consequence of our criterion of 12 months of tenure. Closures and mass layoffs decrease the employment share by 23 and 27 percentage points in the first post-treatment year. The gap between treatment and control employment shrinks but persists in the following years (by 8 to 12 percentage points in years 2 to 5). These effect sizes are similar in magnitude to the results of Ichino et al. (2017) which find that plant closures in Austria decrease employment probability by 27% in the first two years by 10 to 14% in years 3 to 10.

The course of treated and control wages also diverges from event year 1, starting from a HUF 20,000 or a 10-14% difference, and persisting until event year 5 at a similar level. The average effects of the two types of shocks on labor market outcomes are similar, which supports the idea that these are comparable shocks. These effect sizes are also comparable to previous estimates: two seminal papers find on US data that earnings losses of displaced workers are 25% per year (Jacobson et al., 1993b) and 9% per year (Stevens, 1997).

Other labor market variables show a similar pattern, such as the number of months spent working during the year (Figure A.5), registered unemployment (Figure A.6), and the wages of employed women (Figure A.7).

Figure 1.5: Employment in the treatment and control group before and after the shocks: raw means and regression estimates

(a) Closure (b) Mass layoff



The last month of event year 0 is the time of matching. The number of observations by event years is shown in Figure A.4

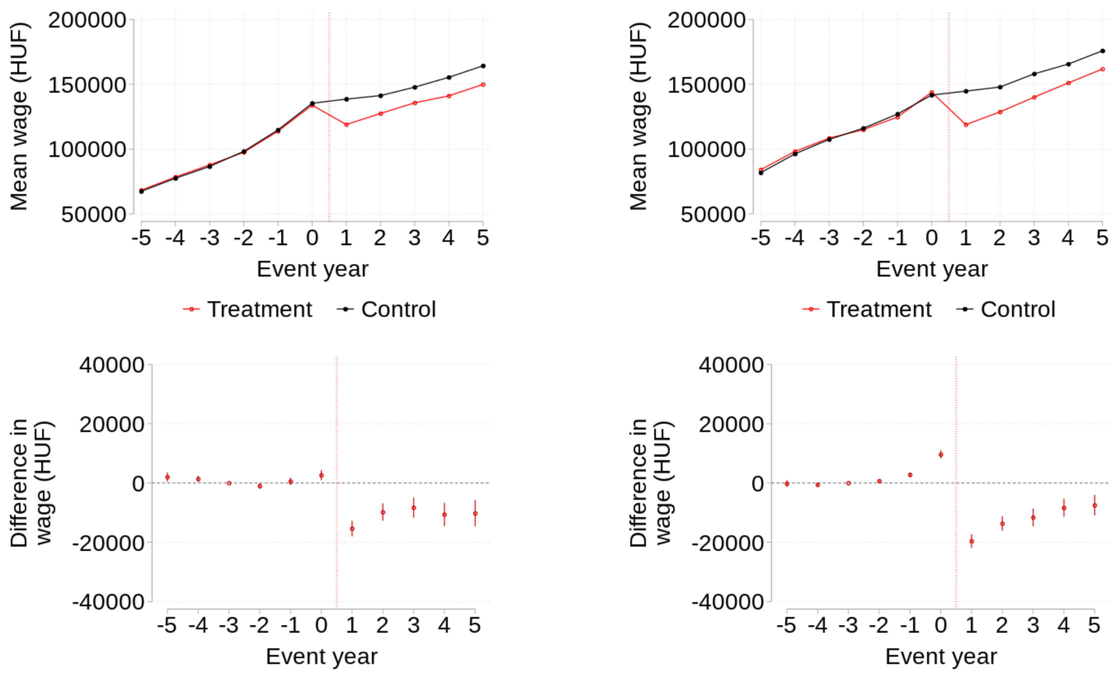
1.6.1.2 Main outcomes: pregnancies, births, and abortions After establishing the negative effect on labor market outcomes, we turn to the main variables of interest: pregnancies (Figure 1.7), births (Figure 1.8), and abortions (Figure 1.9), measured at the estimated time of conception.

A defining feature of the fertility graphs is the appearance of treatment-control differences already in event year 0, the year before the shocks. Wages and employment are still the same this year, thus, these effects cannot be reactions to the current economic situation of women. Rather, we interpret these as women anticipating the coming shocks and the threat of job loss, and reacting by strategically adjusting their fertility.

The graphs of fertility variables support the idea of precautionary pregnancies: pregnancies increase in the year before closures and mass layoffs as well. In line with the strategy being

Figure 1.6: Wages in the treatment and control group before and after the shocks: raw means and regression estimates

(a) Closure (b) Mass layoff



The red vertical line indicates the time of the layoff event. The last month of event year 0 is the time of matching. The number of observations by event years is shown in Figure A.4.

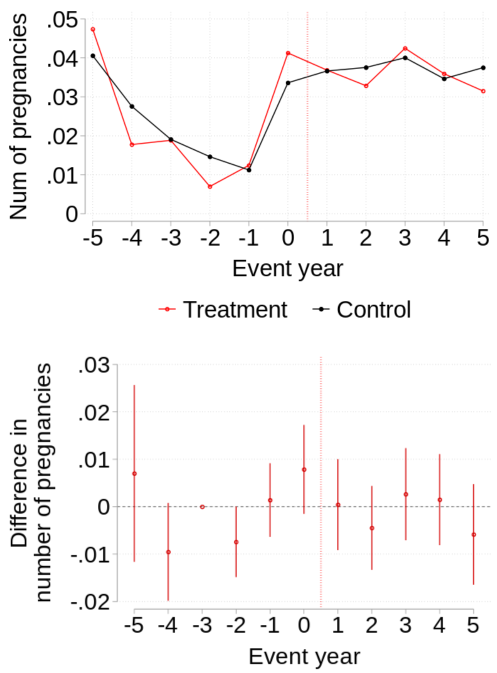
successful only if the firm survives, the resolution of the pregnancies is markedly different in year 0 for the two types of layoff events. Births increase in case of mass layoffs, and abortions increase in case of closures. The effects in the post-treatment years appear to be more moderate than the initial responses.

To get an even more detailed picture of fertility outcomes around the events, we also plot the descriptive graphs and estimate the regressions at a quarterly level on Figures A.8 - A.10. From these figures, we see that the largest increase in pregnancies during the year before the layoff event happens in event quarter -2 (7 to 9 months before the layoff event), and pregnancies are also somewhat higher in event quarter 0, both for closures and mass layoffs. For closures, pregnancies ending with abortions drive this increase, and for mass layoffs pregnancies ending with birth. Both pregnancies and abortions increase the most in even quarter -2, suggesting that the same women drive the increase in pregnancies and abortions. In addition, we see that the somewhat higher levels of abortions in the case of closures in the year after the event are driven by increased abortions in the quarter just after the shock. One explanation for this is that for some women the exact time of the shock is unexpected, and they only learn that the firm closes after the event already happened.

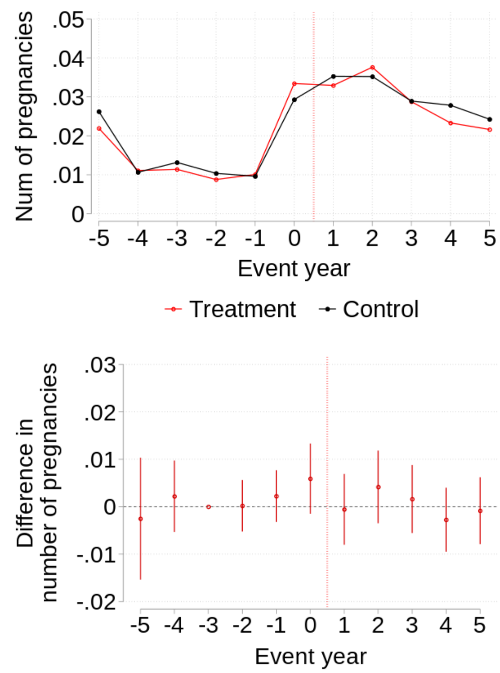
But, as pregnancies, births, and abortions are rare events, quarterly and yearly estimates for fertility outcomes are noisy, and even large effects can be statistically insignificant in these specifications. In the next section, we adopt a difference-in-differences specification to increase power and explicitly account for the potentially different responses before and after the shocks.

Figure 1.7: Pregnancies: raw means in the treatment and the control group and regression estimates

(a) Closure



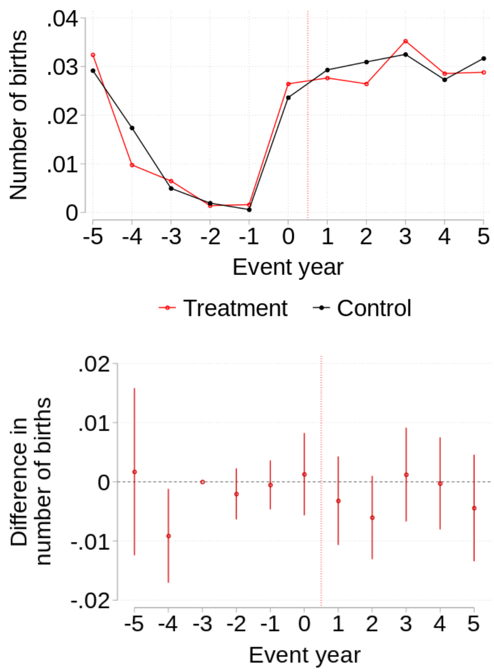
(b) Mass layoff



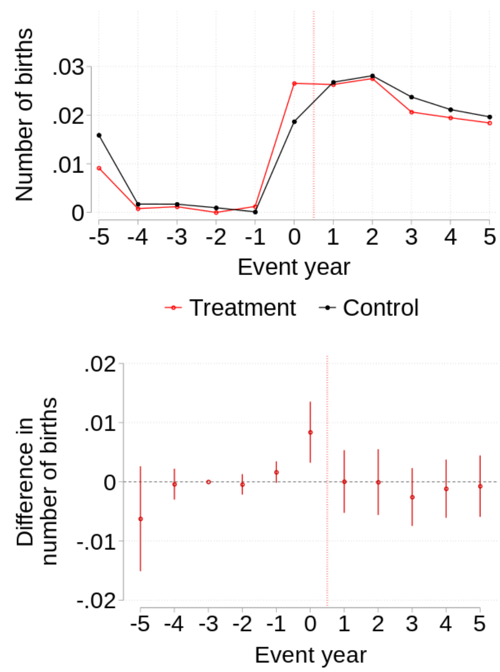
The red vertical line indicates the time of the layoff event. The last month of event year 0 is the time of matching. The number of observations by event years is shown in Figure A.4. The pregnancies, births, and abortions are counted in the year of conception.

Figure 1.8: Births: raw means in the treatment and the control group and regression estimates

(a) Closure

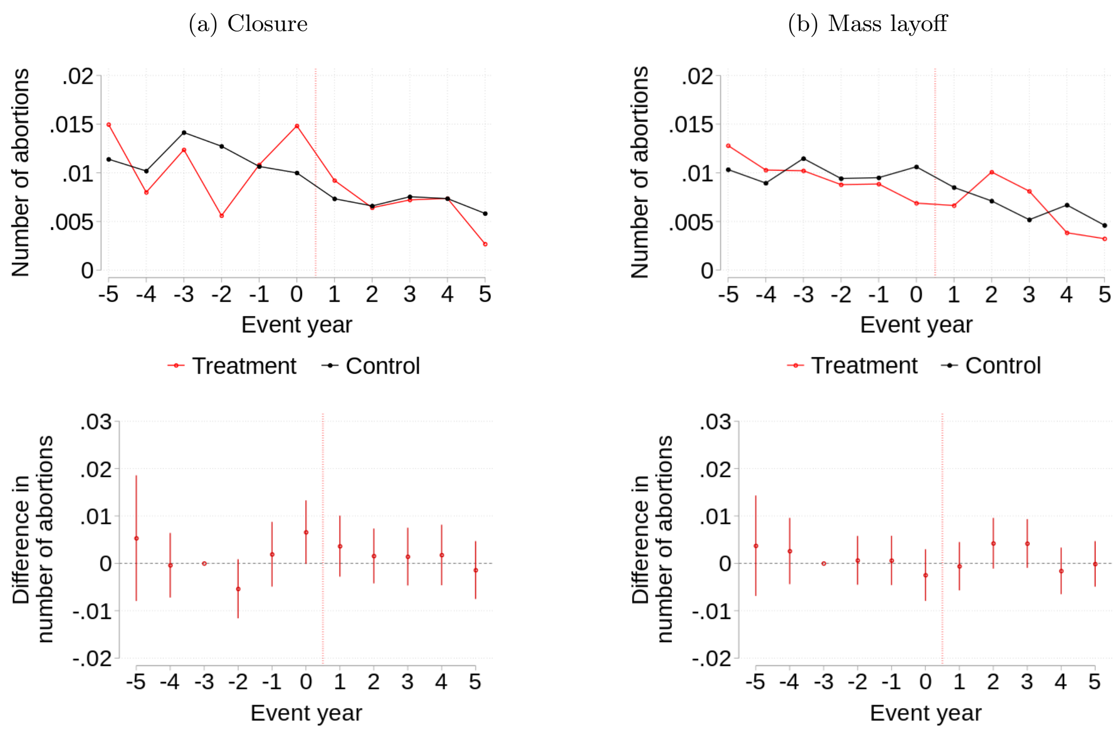


(b) Mass layoff



The red vertical line indicates the time of the layoff event. The last month of event year 0 is the time of matching. The number of observations by event years is shown in Figure A.4. The pregnancies, births, and abortions are counted in the year of conception.

Figure 1.9: Abortions: raw means in the treatment and the control group and regression estimates



The red vertical line indicates the time of the layoff event. The last month of event year 0 is the time of matching. The number of observations by event years is shown in Figure A.4. The pregnancies, births, and abortions are counted in the year of conception.

1.6.2 DiD estimates

In this subsection, we further study women's fertility responses using the difference-in-differences equation 1.3. Years -1 and before are pooled and serve as the baseline category, and we estimate the response separately in the anticipation period ($Year^0$) and in years 1 to 3 after the shock ($Year^{1,2,3}$) to explicitly model anticipatory and post-shock effects. We use three post-treatment years because these years are observed for the whole sample. We also estimate the regressions for the labor market outcomes and present the results in Table A.3.

The first row in Table 1.3 shows that pre-treatment fertility is not statistically significantly different between the treatment and control groups. The coefficients on Year 0 indicate that in the absence of the shock, births are expected to increase in the year before, by a similar magnitude for closures (0.018, $p < 0.01$) and mass layoffs (0.017, $p < 0.01$.). The jump in births is a consequence of our sample being employed for some time and of reproductive age, but these estimates further show that control women act similarly regardless of the event type. Abortions are not significantly different from zero for either group in year 0, meaning that pregnancies increase by around the level of births. In the 3 years following the shock, the control group continues to have similar levels of births (0.025, $p < 0.01$ for closures, and 0.024 $p < 0.01$ for mass layoffs), and a decrease in abortions (-0.005 $p < 0.01$ for closures and -0.003 $p < 0.01$ for mass layoffs). The negative coefficients on abortions can both be the consequence of the decreasing trend in Hungary in abortions over time and also the general decline in abortions after age 35.[18]

Next, we analyze our main coefficients of interest. First, we study the effects in the year preceding the layoff events. The coefficient on Treated X Year 0 in Table 1.3, Column (1) shows that for closures, pregnancies increase by 10 per 1000 women in the anticipation period. This is a large and statistically significant estimate[19]. The number of counterfactual pregnancies - number of pregnancies we would expect in the absence of the treatment[20] - is 29 per 1000 women. Compared to this number the coefficient of 0.010 translates into a 35 percent increase. In the case of mass layoffs (Col. (4)), the point estimate is also large (0.005, or a 19% increase compared to the counterfactual) but insignificantly different from zero. The pregnancy increase in both types of events supports that in anticipation of the shock, women react by increasing pregnancies to avoid being displaced and time childbirth to the crisis of the firm.

The resolution of the pregnancies is different for the two types of layoff events. Women working at firms about to have a mass layoff, increase births by 8 per 1000 women ($p < 0.01$) in anticipation of the coming events (Col. (5) Table 1.3). This is a large, 44% increase, compared to the counterfactual number of 18 births per 1000 women. We can put this effect size into a larger context by comparing it to the national level of 40 births per 1000 women in a year. On the other hand, the coefficient estimate on the number of births in the closure sample is insignificant and smaller.

Columns (3) and (6) in Table 1.3 report the estimates for abortions. Closures increase abortions by 7 per 1000 women (are at an 88% higher level compared to the counterfactual) in year 0. This is a large effect and considering that there are 15 abortions per 1000 women of reproductive age per year in Hungary, it is even more noteworthy. For mass layoffs, we estimate a relatively large reduction in abortions in year 0, but it is insignificant (-0.003, -43%). The estimates are in line with the notion that timing births to the shocks is only optimal if the firm survives the crisis. If

[18] See Figure 1.10 in the next section.

[19] At the 1% level with clustered robust standard errors, and at the 5% with bootstrapped standard errors (see the p-values in the lower panel of the table)

[20] Calculated as pre-treatment mean in the control group (0.015) + coefficient on Treated (-0.002) + coefficient on Year 0(0.016)

women learn that the firm closes, and they lose employment and high maternity benefits, the cost of childbirth substantially increases, and some pregnant women decide to have an abortion.

Next, we turn to the longer-term effects. The estimated yearly effects on births in the 3 post-treatment years are negative (-0.001, or -4% compared to the counterfactual) but insignificant in both samples. These point estimates are between the effect sizes of Huttunen and Kellokumpu (2016) (about 3% significant effect) looking at job displacement events in Finland, and Del Bono et al. (2012) (5-10% significant effect) who analyze Austrian data.

Abortions are estimated to increase in the 3 post-shock years, (by 0.003 for closures and by 0.002 for mass layoffs). These estimates are insignificantly different from zero ($p > 0.01$) but are large compared to the counterfactual level of abortions (60% larger for closures, and 25% larger for mass layoffs). For comparison, González and Trommlerová (2021) estimate the effect of a negative income shock in Spain on abortions and find a significant 13.5% increase, thus our point estimates seem to indicate an even larger abortion response in the Hungarian context. The negative birth and the positive abortion post-shock effects may indicate that births are reduced in the post-treatment period by using abortions. For closures, even these large point estimates on abortions are smaller than the effect in the anticipation period. This suggests that abortions play a more important role in responding to immediate shocks rather than dealing with long-term economic hardship. However, given that our post-treatment estimates are not significant, these interpretations should be treated with caution.

It is worth noting that while the birth and abortion responses for the two types of events are substantially different in year 0, they are very similar in the post-shock years. This is in line with our finding that the economic consequences in the two groups are also similar.

To calculate the net effect of the shocks, we estimate a difference-in-differences equation pooling year 0 and the 3 post-treatment years (Table A.4). The regression estimates reveal that neither closures nor mass layoffs change the overall number of births in the 4-year period surrounding the shocks statistically significantly. Although we do not observe completed fertility in our data, this pattern suggests that mass layoffs do not increase lifetime fertility. Rather, women seem to bring forward pregnancies to the time of the mass layoff. This is also supported by Figure A.11 where we plot the distribution of age at first observed birth.[21] Women give birth at a younger age on average in the treated group (30.5 years old) than in the control group for mass layoffs (32.1 years old), while the age at first observed births is more similar for closures (30.8 years for treated and 30.2 for control women). On the other hand, the net effect on abortions is significantly increased in the case of closures, by 4 per 1000 women yearly.

To further study which fertility margin is affected we split the sample to mothers with and without children at the time of the layoff event, and re-estimate equation 1.3 on these two subgroups. In the sample of women affected by closures, 26 percent are already mothers at the time of the event, while the same share is 22 percent for mass layoffs. The estimates for the treatment effect in year 0 are not statistically significantly different from the main results or from each other, but there are some differences in the point estimates. Specifically, for closures, pregnancies increase the same way (by 0.010), for women with and without children, but the large abortion response seems to be driven more by women without children (0.008 versus 0.005 for women who have children). In the case of mass layoffs, births increase by 0.009 in year 0 for women without children, and by a smaller amount, 0.006 for women who have children. Thus, first-time mothers seem to react a bit more, but we see similar but slightly smaller estimates for women who already have children. In the next section, we look at further heterogeneous effects.

[21] As we only observe pregnancies for a 9-year window, this is not necessarily the same as the age at first birth

Table 1.3: Three period DID regression results for the effect of closures and mass layoffs on fertility outcomes

Sample	Closure			Mass Layoff		
Outcome	Pregnancies	Births	Abortions	Pregnancies	Births	Abortions
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.002	-0.000	-0.002	-0.001	-0.000	-0.000
	(0.002)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)
Year 0	0.016***	0.018***	-0.002	0.018***	0.017***	0.001
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Year 1-3	0.020***	0.025***	-0.005***	0.022***	0.024***	-0.003***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Treated X Year 0	0.010**	0.003	0.007**	0.005	0.008***	-0.003
	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)
Treated X Year 1-3	0.002	-0.001	0.003	0.001	-0.001	0.002
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Exact matched set FE	YES	YES	YES	YES	YES	YES
Propensity score	YES	YES	YES	YES	YES	YES
Bootstrapped p-value of Treated X Year 0	0.027	0.378	0.016	0.181	0.002	0.151
Bootstrapped p-value of Treated X Year 1-3	0.489	0.751	0.093	0.739	0.594	0.244
R-squared	0.074	0.073	0.057	0.083	0.086	0.061
Pre-treatment mean in control group	0.015	0.003	0.012	0.01	0.001	0.009
Observations	136,647			164,047		
N treated	2496			4068		
N control	16860			19763		

Note: Standard errors clustered by exact match set in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates from regression equation 1.3. Births and abortions are measured at the estimated times of conception. Pregnancies are the sum of births and abortions.

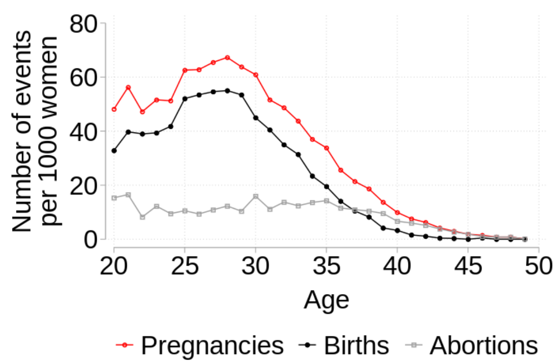
Hungarian DOI: 10.14754/CEU.2023.04

1.6.3 Heterogeneity analysis

In this section we provide additional evidence supporting our interpretation of the large fertility responses as precautionary pregnancies in anticipation of a firm closure or mass layoff event. In particular, we focus on women who are more flexible in timing their pregnancies or more willing to use abortions as a method of birth control.

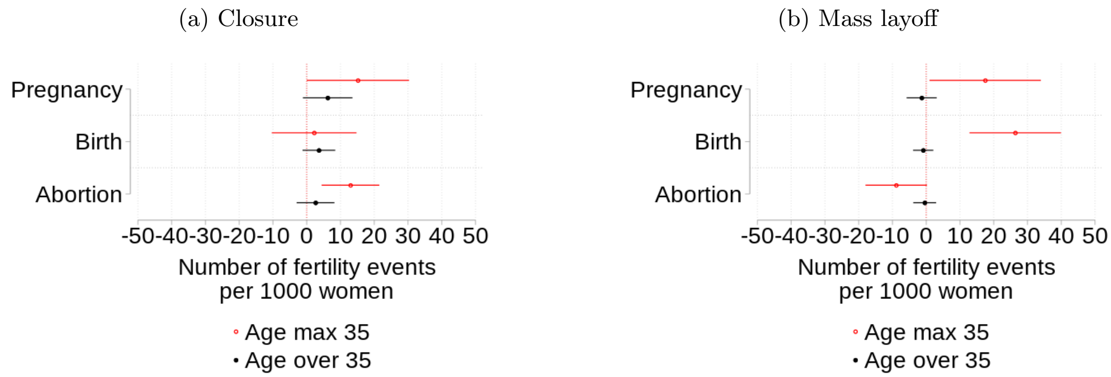
First, we check whether young women respond more in anticipation of the layoff events. We argue that women feeling threatened by job loss may respond by increasing pregnancies. This response is only possible if they can get pregnant relatively fast: after starting to suspect troubles at the firm, but before the actual shock happens. In addition, they have to be willing to have a child. Women approaching the end of their reproductive age span are more likely to have already achieved their desired fertility and even if they decide to get pregnant, they are less likely to succeed in doing so: while the chance of natural conception each month is 25 percent for 25-year-olds, it drops to 5 percent by the age of 40 (ASRM, 2012; Dunson et al., 2002; van Noord-Zaadstra et al., 1991). Figure 1.10 showing the number of fertility events by age in our control group confirms that pregnancy probabilities are at their maximum for women between 25 and 30 years of age (more than 60 pregnancies per 1000 women), and they start to drop fast after this age (to under 10 pregnancies after age 40).

Figure 1.10: Number of fertility events per 1000 women by age in the pooled control group



We split the sample at age 35, and estimate equation 1.3 separately for younger and older women. Figure 1.11 presents the anticipation effect (δ_1), for pregnancies, births, and abortions in case of closures and mass layoffs. The point estimates indicate that indeed women under age 35 drive the main results, while fertility effects are close to 0 for older women. Importantly, the magnitude of effects on pregnancies in the younger sub-sample is similarly large, and statistically significant ($p < 0.1$) in the case of closures (16 pregnancies per 1000 women) and mass layoffs (18 pregnancies per 1000 women). This is a piece of evidence that the pregnancy increase for mass layoffs is unlikely to arise randomly, even though it is statistically insignificant in our main specification. Supporting our main findings, while in case of closures, a larger part of the conceived pregnancies gets aborted, women affected by mass layoffs are more likely to give birth. The effect sizes are even larger for young women than for the whole sample (0.013 for abortions in case of closures ($p < 0.05$) and 0.027 for birth in the case of mass layoffs ($p < 0.01$)).

Figure 1.11: Anticipation effects by age, with 90 percent confidence intervals



Estimates for δ_1 in equation 1.3, with 90% confidence intervals.

While age is an important determinant of fertility it is not the only one. For example, Figure A.12 shows that white-collar women tend to give birth at an older age than blue-collar women. In the following, we split the sample into low- and high-pregnancy probability groups, to investigate whether high-pregnancy probability women drive our results. To obtain the groups, we estimate a logit regression of an indicator for pregnancy using the pooled sample of the control groups. The predictors are age, occupation, their interaction, tenure, an indicator of having a young child, place of living, and wage- and employment history. Based on the estimated coefficients, we predict probabilities for treated and control women and split the sample at the median pregnancy probability of the control group.[22]

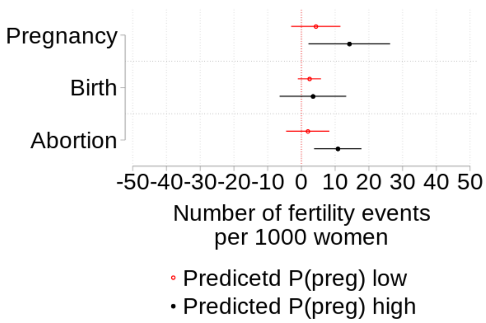
Figure 1.12 shows the estimates for the anticipation effect in these groups. This split produces very similar estimates to the split by age. The effects on all fertility variables are essentially zero for women with low predicted pregnancy probability. For women with high predicted pregnancy probability, the pregnancy effects are similarly large for mass layoffs and closures, but the effects on abortions and births markedly differ. This underlines that women who are more flexible in timing their pregnancies drive the anticipation effects and that the increase in precautionary pregnancies is similar before both types of shocks.

Next, we compare fertility responses by the woman's willingness to use abortions. We focus on women with high pregnancy probability and split them into groups with a low and high predicted probability of abortion. To define the groups we run a logit of an indicator of having an abortion in event year 0 in the sample of women who get pregnant in the pooled control group, using the same right-hand side variables as before. Based on these estimates we predict the probability of having an abortion conditional on getting pregnant for the whole high pregnancy probability group. We again split the sample at the median of the control group to define a group with high and a group with low conditional abortion probability. Figure 1.13 shows the coefficient estimate of the diff-in-diff model for these groups. For closure events, women with a high predicted probability of abortions are the ones who drive the increase in pregnancies and abortions. For mass layoffs, we do not observe a clear difference between the groups with different abortion probabilities, both

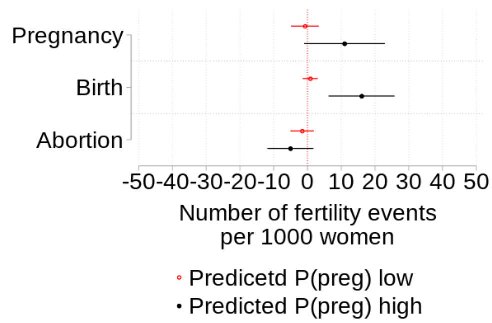
[22]The details of this analysis are available from the authors upon request.

Figure 1.12: Anticipation effects by predicted pregnancy probability, with 90 percent confidence intervals

(a) Closure



(b) Mass layoff

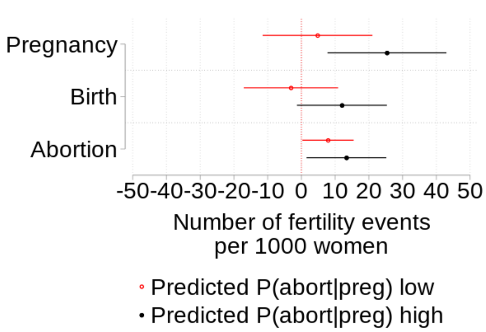


Estimates for δ_1 in equation 1.3, with 90% confidence intervals.

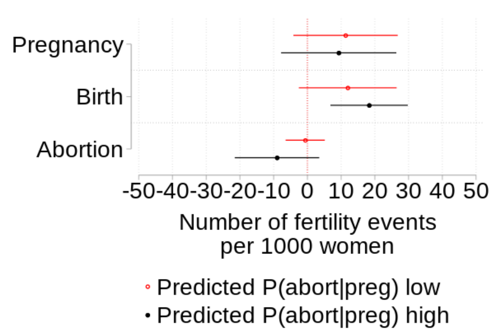
estimates of pregnancies are around 0.01 and insignificantly different from 0. This indicates that women who want to avoid abortions are less responsive in increasing pregnancies when the risk that the firm is closing - and thus the risk that the precautionary pregnancy strategy breaks down - is high. When the risk of firm closure is lower - in the case of mass layoffs - women less willing to take the risk of abortion also seem to respond to the threat of job loss.

Figure 1.13: Anticipation effects by predicted conditional abortion probability, with 90 percent confidence intervals

(a) Closure



(b) Mass layoff



Estimates for δ_1 in equation 1.3.

Our heterogeneity results should be taken with a grain of salt because even when we see large differences in point estimates between the groups, we cannot differentiate them in statistical terms. Nevertheless, the differences in the point estimates are consistent with our main explanation of the treatment effects in the year before the events: women strategically increasing pregnancies in the face of coming employment shocks.

1.7 Robustness checks

In this section, we look at alternative explanations and argue that our main findings are robust to a number of modifications in the empirical strategy.

Our interpretation of the differing fertility responses between mass layoffs and closures is that it can be attributed to institutional surroundings: the presence of employment protection and high maternity benefits if the firm survives the crisis, and the loss of these if the firm closes. An alternative explanation could be the different compositions of the two samples. The most important differences that are correlated with fertility decisions are that women in the closure sample are somewhat younger (mean age is 36, while it is 38 in the mass layoff sample), and a larger proportion of them has already at least one young child (26 percent vs 22 percent). To some extent, our heterogeneity analyses support that the differential effects between the two events are present if we take these variables under control. But to further analyze whether the different composition of the two samples is sufficient to explain the differences in the effects, we run regressions similar to the one specified in equation 1.3, using the pooled sample of women affected by either shock. A modification compared to equation 1.3 is that we do not include exact matched set fixed effects in these specifications, because then we would lose sufficient overlap between the mass layoff and the closure samples. As without exact match set dummies calendar time is not controlled for automatically, we include calendar year fixed effects in these regressions. The results in Table A.7 show that our estimates from the pooled samples are similar to our main results (positive pregnancy effect of 0.01 ($p < 0.05$) for closures and 0.005 ($p > 0.1$) for mass layoffs, positive birth effect of 0.008 ($p < 0.00$) for mass layoffs, and positive abortion effect of 0.006 ($p < 0.05$) for closures), supporting that it is not the different composition of the two samples that drive the differences in the fertility responses.

Next, we show that the main results are not sensitive to our choices in defining the sample. First, we exactly match on a maximum of 4 years of birth and abortion history in event years -2 to -5. The pre-treatment fertility is already similar in our main specification for most variables, but in the closures sample, there are significantly fewer abortions in the pre-treatment period .[23] This robustness check is important as by enforcing similarity in all fertility variables during the pre-treatment period, we make it more plausible that parallel trends are also satisfied post-treatment. Second, we employ two additional matching approaches without using any caliper and with a stricter caliper set at half the size used in the main specification. The choice of the caliper was subjective and aimed to minimize economically significant differences between the treatment and control groups while retaining a sufficiently large sample size. Thus, ensuring that the main results are not overly sensitive to this caliper choice is important. We re-estimate equation 1.3 on the samples using these three different sample selection approaches.

Figures 1.14, 1.15 and 1.16 summarise the regression estimates and reveal that our results are mainly robust to these modifications. None of the estimates differ from the results in the original regressions in statistical terms, and they are of a similar magnitude. However, for some of our estimates the statistical significance changes. In particular, when we match on pre-treatment fertility, the abortion effect in case of closures becomes somewhat smaller, and insignificant, and the effect on births becomes significantly positive when we use different calipers in the matching.

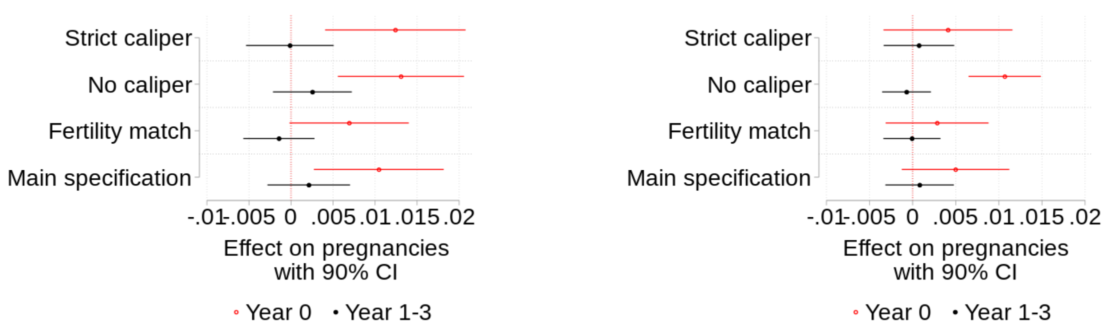
[23]For every woman we can only use the available pre-treatment years.

These results are also consistent with our proposed mechanism that women increase pregnancies in anticipation of the unknown shock, but some pregnancies are aborted when women learn that their current firm is about to stop existing. However, the share of pregnancies that get aborted might be smaller than what is suggested by our main specification. In addition, when we match on pre-treatment fertility, the estimate on abortions in case of mass layoffs becomes significantly negative. This can indicate that increased births in case of mass layoffs are driven more by the different resolutions in unplanned pregnancies, than what is suggested by our main results.

Regardless of these changes in significance, for all specifications, our estimates are markedly different in case of closures and mass layoffs and are consistently larger for responses in anticipation of the shocks compared to the response observed after the shock. These features support the presence of an anticipatory response in fertility, and a different response to different shocks.

Figure 1.14: Pregnancies: the effect of employment shocks - robustness checks

(a) Closure (b) Mass layoff

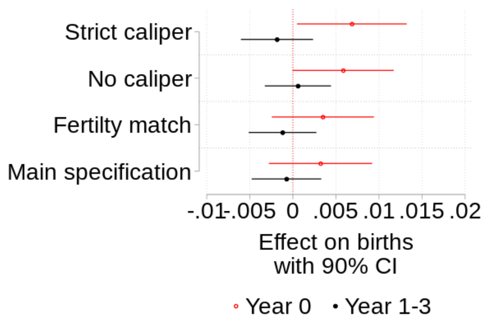


Estimates based on equation 1.3.

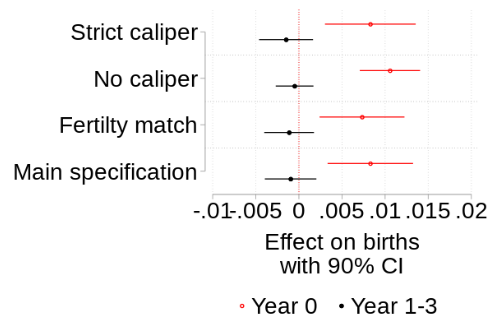
As we noted earlier, miscarriages are not included among the pregnancies, and this could lead to a measurement bias of the main results. In Section A.2 in the Appendix, we provide a calculation showing that this measurement error is too small to substantially influence our results.

Figure 1.15: Births: the effect of employment shocks - robustness checks

(a) Closure



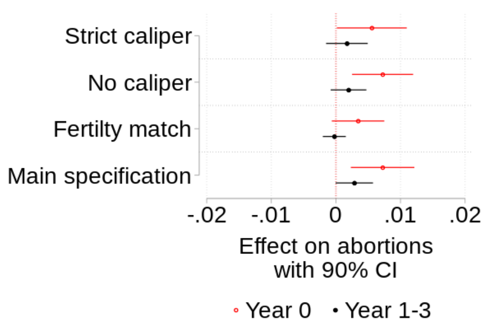
(b) Mass layoff



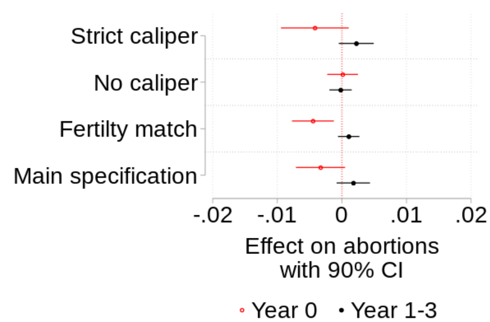
Estimates based on equation 1.3.

Figure 1.16: Abortions: the effect of employment shocks - robustness checks

(a) Closure



(b) Mass layoff



Estimates based on equation 1.3.

1.8 Conclusion

In this chapter, we analyze women's fertility responses to two different types of employment shocks, firm closures, and mass layoffs. We argue that these shocks may have different impacts because of institutions that provide dismissal protection and financial benefits during pregnancy and after childbirth. We find strong evidence of precautionary fertility responses: women increase pregnancies before coming employment shocks, and the increase is even larger for young women. The resolution of the pregnancies is different though. If the firm survives the crisis, women are covered by dismissal protection and high maternity benefits, and they keep their pregnancies and use them as insurance against layoff. We present suggestive evidence that women time their pregnancies to an earlier date, while overall fertility does not seem to increase as a result of the shock. If the firm closes though, dismissal protection and high maternity benefits become unavailable, and we find that the probability of abortion increases significantly in this case. The post-shock responses are smaller and insignificantly different from zero, but the effect sizes are consistent with previous findings in the literature: we estimate negative effects on births and positive effects on abortions for both types of events.

The novelty of our study is that we analyze fertility responses before the actual shock happens, and demonstrate the phenomenon of precautionary fertility behavior. Moreover, we present evidence that anticipatory effects are highly sensitive to institutional surroundings. In addition, while other studies already provided reliable causal micro evidence of the effect of economic shocks on the number of births, our research is the first that can look at the number of abortions as well. An advantage of our research is that we observe abortions and births at the individual level in administrative data and that the linked employment-employee nature of the data makes it possible to identify individuals affected by closures and mass layoffs. As these are likely to be conditionally exogenous shocks, we can make use of the large pool of women to match very similar controls to the treated group and identify causal effects under plausibly satisfied assumptions.

A limitation of our study is that even though we observe relatively large samples of women affected by closures or mass layoffs, as abortions and births are rare, we have to identify fertility effects from a small number of fertility events. As the effects are large in the main regressions, we can still reliably differentiate them from 0 despite our power problem, but when we split the sample further, the estimates become so imprecise, that even large effects are statistically indistinguishable from 0 and each other. Future research can shed more light on the importance of the precautionary response in different socio-economic groups of women.

Our results are relevant for the increasing share of women who take into account career and employment conditions when planning their fertility. In particular, the strategy of precautionary pregnancy can be potentially used by any woman who assigns some positive probability of future job loss in a context where employment protection is available. As we have argued, it is likely that workers can foresee the coming employment shocks and that a substantial fraction of young women are able to conceive in a few months.

A surprising aspect of our findings is that a decision that affects potentially the entire life of a person – having or not having a child at a given time – can be highly sensitive to experiencing a random shock in the labor market. This indicates that women striving to create a financially stable period around their childbirth and maternity are pushed to react fast in an increasingly volatile labor market. Policies that increase stability around childbirth and maternity can empower women to be less dependent on unexpected economic shocks during their careers. For example, in the context of Hungary, the precautionary response might be especially large due to the policy requiring that women must be employed at the time of childbirth to qualify for high maternity benefits. By detaching maternity and child benefits from current employment status, women would have larger freedom to time their pregnancies without the fear of the adverse consequences of losing employment at the time of childbirth.

Our results for increased abortions in the case of firm closures indicate that if employment protection and maternity benefits are unavailable, women fully suffer the consequences of employment shocks. For Hungary, as employment protection for the pregnant is largely universal, this affects only the narrow set of women who remain unprotected for rare reasons like firm closures, or illegal dismissal. But in countries with weaker protection, they might be relevant for many more. If women are not yet pregnant, they may postpone childbearing and decrease lifetime fertility (Currie and Schwandt, 2014). If they are pregnant, they may turn to abortion, or, if abortion is not possible, they suffer serious financial consequences as shown by Miller et al. (2023). In an environment where institutions protect pregnant women and mothers, women can time their pregnancies to younger ages, which can shrink the gap between desired and completed fertility (Beaujouan and Berghammer, 2019). Our findings support the view that dismissal protection and maternity leave policies are powerful tools in incentivizing women to keep pregnancies in times of economic shocks.

2 Chapter 2: The Gains from Family Foster Care: Evidence from Hungary

Joint work with Gábor Kertesi

2.1 Introduction

When a child is neglected, abused, or lacks a protective family for some other reason, she is placed into state care. Children who spend time in state care are a vulnerable group. Compared to the general population their average outcomes related to mental- and physical health, education, employment, homelessness, substance usage and contact with the criminal justice system are worse (McDonald et al., 1996, Bald, Doyle Jr., et al., 2022).

The two main forms of state care are residential care and family foster care. In residential care, groups of children live together in an institution and are cared for by a team of child protection workers. In foster care, professional parents raise the children in their own homes, sometimes together with their biological children. In the United States and other Anglophone countries, residential care is thought of as a last resort for children with severe mental or behavioral problems, unfit for foster care (del Valle, 2014). But in developing countries and in the majority of industrialized countries as well, institutional care is still common: worldwide 2.7 million children live in such settings (Petrowski et al., 2017).

There is ample evidence that psycho-social deprivation experienced by young children raised in institutions is harmful to brain development (Nelson, 2014). But, the majority of children living in state care spend their early years in their biological families and are only removed to live in state care later in childhood.[24] Evidence comparing the effects of being raised in an institution and in a foster family as an older child is scarce with mixed results on few outcomes (see the review Lee et al., 2011). Our research provides evidence on the effect of foster care placement of older children on a large set of new adult outcomes. Specifically, we ask how education-, labor market- and health outcomes are affected by the type of care children are raised in as teenagers.

To answer, we analyze data from Hungary, a high-income country where foster care and residential care co-exist, with 56 percent of children raised in the former and 44 percent in the latter.[25] In Hungary, the majority of children in both foster care and residential care do not exhibit severe behavioral problems, creating a favorable context to analyze the differential effects of these two care types. One of our data sources is the National Assessment of Basic Competencies, a standardized mathematics and reading test conducted annually on the universe of 6^{th} 8^{th} and 10^{th} grade students. The test is accompanied by a questionnaire with a rich set of variables. Most importantly, we observe the home type (own family, foster care, or residential care), and indicators of cognitive[26] and non-cognitive skills[27] of the children. We select children who live in the same home type in every observed year and link the data to an individual-level monthly administrative panel on employment, wages, doctor's visits, and purchases of prescription drugs.

[24] In the US, 2013, 61 percent of children were over 5 at first entryhttps://cascw.umn.edu/policy/new-data-on-adoption-and-foster-care-in-the-us/ (downloaded 2023.04.03). In the UK, 2021, 63 percent https://explore-education-statistics.service.gov.uk/find-statistics/children-looked-after-in-england-including-adoptions/2021 (downloaded 2023.04.03)

[25] At the beginning of our observation period, 2008.

[26] Standardized test scores in mathematics and reading, and school grades in four subjects.

[27] Official diagnoses on special educational needs, with a separate category for mild mental disability, physical disability, speech impairment, autism, psychological problems, and behavior or learning problems; and school grades in behavior and effort.

First, we document that the average outcomes of children raised in foster care are substantially better than the outcomes of children raised in residential care. By age 19, 40 percent of children raised in foster care finish secondary school, while the same number for children raised in residential care is only 21 percent. At age 19, 37 percent of youth raised in foster families spends at least 6 months without either working or studying, which is 16 percentage points lower than the figure in residential care-raised children (53 percent). Indicators of mental health are also better for teenagers raised in foster care: they are 5 percentage points less likely to have ever bought antidepressants, and 4 percentage points less likely to have ever bought tranquilizers. When examining teenage motherhood rates, 22 percent of girls raised in foster families become mothers by age 19, while the corresponding figure for girls raised in residential care is a troubling 40 percent. Additionally, abortions are less frequent among girls raised in foster families, with a rate of 15 percent, compared to a rate of 30 percent among girls in residential care.

However, these large raw differences in the means are likely to overestimate the positive effect of foster care on future outcomes. The reason for this is that foster parents can decide not to take in, or to let go of a child any time. Foster parents in Hungary are scarce, and foster parents capable and willing to care for children with disabilities and special needs, or very troubled children, are even more so. Thus, children living in residential care are on average more problematic to start with, with worse potential outcomes as adults.

In our main empirical strategy, we control for baseline differences, running OLS regressions of adult outcomes on home type and a large set of variables observed in 6$^{\text{th}}$ grade (around age 13): indicators of cognitive and non-cognitive skills, gender, year dummies, and county of residence.

The OLS estimates show that by age 19, young adults who were raised in family foster care are 8.3 percentage points more likely to finish secondary education (p=0.021) than comparable youth raised in residential care. They are 4.9 percentage points less likely to have ever bought tranquilizers (p=0.039), and 11.4 percentage points less likely to spend more than 6 months without either working or studying at age 19 (p=0.0043). For girls, teenage pregnancy, birth, and abortion are less likely, by 15.7, 11.5, and 12.1 percentage points respectively (p=0.012, 0.047, and 0.023). The estimate on antidepressant usage is not statistically significant but also points to a beneficial effect of foster care (-3.8 percentage points). To address the problem of testing multiple outcomes, we run the regression of a composite index of the outcomes, and we find a large and significant gain from foster care on this outcome as well (0.216 with p=0.0001). In addition, we calculate q values for false discovery rates, which are below 10 percent for all outcomes, and below 5 percent for all of our previously statistically significant outcomes.

These results still might be biased estimates of the true causal effects for two reasons. First, because we use control variables measured at 6$^{\text{th}}$ grade, which themselves can be outcomes of living in a given home type from an early age. We argue that in our case we have a special case of the bad control problem (proxy controls), which makes us underestimate the true beneficial effect of foster care. Second, the presence of important omitted variables can also bias the estimates. We argue that it is plausible that the most important omitted variables we can detect, Roma ethnicity and age of first placement, cannot explain the large differences between the two groups. While we cannot rule out the existence of other omitted variables, the large effect sizes combined with the factors biasing our estimates downward suggest that it is unlikely that only omitted variables explain the large differences between foster youth and residential care youth.

To provide further support for the causal interpretation of the results, we run IV regressions using spatial variation in the local capacity of the foster mothers as an instrument for being placed into foster care. The reduced form from this strategy shows that comparable children raised in state care have better outcomes as adults in counties with higher foster mother capacities. Specifically, young adults raised in state care in counties where foster mothers are more abundant are 7 percentage points more likely to finish secondary education, 4.2 percentage points less likely to use tranquilizers, and spend 0.78 fewer months as NEETs, on average. These are significant estimates at the 10 percent level, and while estimates for the other outcomes are insignificant, they also show better outcomes for children who live in counties with higher foster mother capacities. A placebo test on children living with their own parents confirms that omitted county features cannot explain these differences. The estimates from the structural equation also show large gains from foster care. However, the results from these regressions should be interpreted with caution as the exclusion restriction is likely not to be fully satisfied. Omitted county-level features of the child protection system, which do not affect the general population of children and thus are not reflected in the placebo, might confound the estimates for children in state care. In addition, the estimates are noisy in this specification, and no results are statistically significant when we correct for multiple testing. Still, the IV results pointing in the same direction as the OLS, lend additional support to the beneficial effect of foster placement.

We provide evidence that the quality of foster care matters as well. Children placed to foster mothers with at least high school education exhibit better outcomes than comparable children placed to lower educated foster mothers. This difference by the foster mother's education is only statistically significant for the outcome of births though: girls living with educated foster mothers are 15 percentage points less likely to become teenage mothers than girls living with foster mothers with low education.

We investigate if specific subgroups of children can gain more from foster care. We find that foster care affects boys more positively for the outcomes we observe for both genders but the differences are statistically insignificant. Children both with and without special needs and disabilities gain from foster care, and there is no statistically significantly different effect between these groups.

The results are robust to a number of modifications in the empirical strategy. Estimating regressions on the subsample where overlap in the control variables is satisfied, using different sample selection criteria, and changing the set of control variables do not affect the estimates substantially.

Our results contribute to the literature on the economics of foster care (see Bald, Doyle Jr., et al., 2022). This literature mainly focused on assessing the causal effect of removing a child from her own family, with mixed results. No economics paper we know of compared the outcomes of children raised in residential care and foster care. One possible reason behind this is that in the US and the UK, residential care is uncommon and reserved for troubled children. The fact that in Hungary, similarly to other European countries, many children without severe behavioral problems live in residential settings offers a possibility to compare children raised in foster care and residential care, who are similar otherwise. Another potential reason behind the scarcity of results is that disadvantaged groups, like homeless people, criminal offenders, and drug addicts – groups which are over-represented among former foster and residential care youth – are notoriously difficult to reach with surveys (Lambert, 1990). Having access to administrative data with information on past home types provides a unique opportunity to learn something about the population of adults who were raised in state care as children.

We also contribute to the literature on the outcomes of children in state care from other disciplines, psychology, social work, and sociology. There is convincing causal evidence from a recent randomized controlled trial, the Bucharest Early Intervention Project, that children below the age of 3 who are placed into foster families are better off in terms of a large set of psychological outcomes than children who stayed in institutional care (Nelson, 2014). However, children are typically removed from their birth families at older ages, and the needs of older children may differ substantially compared to babies. For example, the constant presence of a primary caregiver might be less important for children over 3. Evidence on older children, however, is mostly descriptive, and documents that children in residential care exhibit worse outcomes, but without controlling for baseline differences (Barth, 2002, McDonald et al., 1996). The small number of papers comparing children with similar characteristics in residential care and foster care, find mixed evidence for a limited number of outcomes (Lee et al., 2011). Our research contributes to this literature by providing new evidence on the effect of foster care placement of adolescents on a large set of adult outcomes.

Our results inform policy by showing that governments can potentially increase the well-being of children raised in state care by placing more children in foster families. To achieve this, one approach could be to increase the supply of foster parents by incentivizing more people to choose this occupation. Since foster care is typically less expensive than residential care (Eurochild, 2015), decreasing the number of children in residential care would make it possible to reallocate resources to foster care even within existing budgetary constraints.[28] Some candidate policies are increasing foster parent wages, or providing better information about the possibility of becoming a foster parent for the general population.[29] Another approach could be to decrease the demand for state care, by reducing the number of children needed to be removed from their birth families in the first place. If there were fewer children in the system, the currently available foster parents could raise a higher share of them. In Hungary, one way could be the better enforcement of the law prohibiting the removal of children on the sheer basis of adverse economic circumstances.[30] It is estimated that in Hungary one-third of the children raised in state care are displaced from their birth families for reasons related to poverty, even though this practice is illegal. In addition, generally reducing poverty would decrease the number of dysfunctional families, and thus the number of children needed to be placed in state care. Yet another way which can decrase the number of children removed from their birth families is highlighted by our result on decreased teenage births. Children of teenage mothers without family support are more at risk of ending up living in state care themselves (Wall-Wieler et al., 2018). Thus, providing quality family foster care for teenage girls in state care has the potential to break the cycle of involvement with child protection services.

The rest of this chapter is organized as follows: Section 2.2 reviews the literature, and Section 2.3 summarizes the Hungarian institutional background. Section 2.4 describes the data and the descriptive differences between individuals raised in foster care and residential care. Section 2.5 presents the empirical strategy and the results from the OLS and IV strategies. In Section 2.6 we analyze heterogeneous effects, in Section 2.7 we present robustness checks, and we conclude in Section 2.8.

2.2 Literature review

In this section we review the empirical literature analyzing the outcomes of children in state care on future outcomes, focusing on the effect of placement type.

[28]In Hungary, child protection experts estimate that the cost/child in residential care is 1.3 times higher than in foster care colorbluehttps://index.hu/belfold/2019/05/06/allami_gondozas_gyerek_kiemeles_tamogatas/(downloaded 2023.07.16.)

[29]Misconceptions and prejudice about this occupation are widespread in Hungary, see for example this article: https://abcug.hu/barataiktol-rokonaiktol-felnek-a-neveloszuloseget-fontolgato-csaladok/(downloaded 2023.06.16.)

[30]1997. évi XXXI. törvény a gyermekek védelméről és a gyámügyi igazgatásról, 7. § (1).

The recent summary of Bald, Doyle Jr., et al. (2022) about the economics of foster care highlighted that more research is needed on the effect of placement type on future outcomes. This strand of literature concentrated on estimating the causal effect of removing a child from her birth family, using the placement tendency of the investigators as an instrument. The results are mixed: Doyle Jr. (2007) finds a negative insignificant effect, Gross and Baron (2022) a positive effect, and Bald, Chyn, et al. (2022) a positive effect for girls, and no effect for boys. Recently, the emergence of research has shed light on the effects of different placement types, with an emphasis on the distinction between (non-kinship) foster care and kinship foster care. Lovett and Xue (2020) find a positive effect of kinship care on long-run outcomes including education and employment, while results on short-term outcomes (for example test scores and placement type) are mixed (Hayduk, 2017).

However, evidence about a different and common care type, residential care[31] as an alternative to foster care is more scarce. In the context of the United States, Lee et al. (2011) reviewed the literature comparing residential care to other care types. Many of the reviewed papers concentrate on specialized interventions to provide alternatives for congregate care for troubled, for example placing them in treatment foster care. In treatment foster care, foster parents receive specialized training to be able to care for children with serious mental health and behavioral problems. The reviewed eight studies, among which six used randomized assignment, find consistently better outcomes for children placed in treatment foster care: contact with the criminal justice system, the number of days spent in lockup, and the frequency of running away decreased compared to controls in congregate care. More closely related to our setting, three studies using observational data compared the effect of (general) foster care versus group care. All three find better outcomes for children placed in foster care. DeSena et al. (2005) evaluated the Safe Homes program, during which children between 3 and 12 years of age were placed in group care instead of foster care, as a first placement. They found that children placed in group care experienced more placements during the following year than matched children in foster care. Friedrich et al. (2005), use panel data with two time periods, and find that problematic sexualized behaviors tend to persist more for youth placed in residential care. Ryan et al. (2008) find that youth placed into residential care are more likely to be arrested in the 5 following the placement than matched youth placed in foster care. Results for the United Kingdom are also available. Using the 1970 British Cohort Study survey data, Dregan and Gulliford (2012) analyze the outcomes of adults raised in different types of state care. They find that children who spent time in residential care are more likely to be convicted for crimes as adults and to experience depression compared to the general population of children in residential care, however, the estimates are not statistically significantly different from each other.

All in all, findings from the United States and the United Kingdom generally point to a beneficial effect of foster care on some outcomes. However, in these countries, residential care is thought of as a last resort (del Valle, 2014), only reserved for children with severe mental or behavioral problems (for example, juvenile offenders), or children with disabilities. Thus, finding a suitable control group for institutionalized children is challenging. In addition, evidence from these countries on the effects of foster care can be generalized only for this special, problematic group of children.

Outside the United States and the United Kingdom, the highest quality casual evidence comes from the Bucharest Early Intervention Project, the only randomized controlled trial to date comparing foster care to institutional care. During the BEIP, from 136 previously institutionalized babies and toddlers in Romania in 2001, 68 were selected to live with foster parents, highly trained and

[31] In the literature, residential care is often interchangeably referred to as group care, congregate care, or institutional care.

paid by the project. For other children, care as usual continued: most of them stayed in residential care until the age of 5, and later in childhood some of them still lived in institutions, while others were reunited with their birth families, or were living with regular foster families of the Romanian child protection system (Wade et al., 2019). Children in the project participated in follow-up psychological assessments at 30, 42, and 54 months, 8 years, 12 years, and 16 years, and further follow-up studies are planned for the future as well. A large series of papers documents that the children who stayed in the institutions have worse outcomes: for example, delays and abnormalities in brain development, lower IQ, problems with attachment, and higher incidence of psychiatric disorders (Wade et al., 2019, Wade et al., 2023, Smyke et al., 2010, Nelson, 2014). The problems persisted to later ages as well. Research from quasi-experimental studies reviewed by Dozier et al. (2012) confirms that it is harmful to babies to live in institutional settings, and children placed in foster families achieve better outcomes.

We contribute to this literature in multiple ways. First, our analysis is in a country where both family foster care and residential care are common. This means that contrary to research in Anglophone countries, which can only analyze problematic children in this context, our research can provide evidence about a wider group, the general population of children living in state care. While children in this group are disproportionately affected by mental health and behavioral problems, and disabilities, and come from difficult backgrounds, the large majority of them are fit to live in a family-like environment. In our data, 81 percent of children in foster care, and 74 percent in residential care do not exhibit behavioral, psychological, or learning problems, and are not diagnosed with any special needs. On the one hand, it is possible that children without behavioral problems can gain even more from living in a family environment, than problematic kids, who probably would have bad outcomes no matter where they lived. On the other hand, it is also possible that children with less special needs are affected less negatively by institutional settings, as they probably need less personal attention and care to thrive. In addition, in contexts where less problematic children live together in groups, negative peer effects might prove to be less of a concern.

Second, our research sheds light on a rich set of adult outcomes, at an age (the end of age 19) when most of the children already left the child protection system, with the usage of administrative data. Data about adults with a history of state care can be typically obtained by surveys. However, disadvantaged groups of adults, like drug addicts, homeless people, or criminal offenders are hard to reach by surveys Lambert (1990). Maybe because of this, it is not clear yet from the literature how adult outcomes are affected by the type of home children are raised in.

Finally, our research contributes to the literature by concentrating on the effect of home type where children live as adolescents. Results about the effect of foster care as opposed to residential care are most abundant and convincing in the context of babies. In this group, it is well established that institutional settings are harmful. Unlike babies though, older children might not need the constant presence of a primary caregiver and might benefit more from the expertise of child protection professionals. The typical child in state care is removed from her biological family not as a baby, but at an older age [32]. Thus obtaining information about how older children are affected by different types of care is important for policies aiming to improve the outcomes of children raised in state care.

[32]In the US, 2013, 61 percent of children were over 5 at first entryhttps://cascw.umn.edu/policy/new-data-on-adoption-and-foster-care-in-the-us/ (downloaded 2023.04.03). In the UK, 2021, 63 percent https://explore-education-statistics.service.gov.uk/find-statistics/children-looked-after-in-england-including-adoptions/2021 (downloaded 2023.04.03). For Hungary, no official country-level statistic is available for the age of first placement. In a 2015 dataset covering the universe of children in state care from one Hungarian county, Nógrád, we see that 77 percent of children were over 5 at the first placement (Darvas, Farkas, Ágnes, et al., 2016)

2.3 Hungarian background

In this section, we describe the main features of state care in Hungary and the decision process for the placement of children into foster families or residential care.

2.3.1 Children in state care in Hungary

Children who lack a protective family environment, and do not have relatives or friends willing to take care of them, are placed into state care, temporarily or permanently. Permanent placement is rare (8 percent of all placements in 2011), and only possible if the parents do not have custody rights anymore, for example, because of a court decision or in case the parents die. The most frequent reason (70 percent of all cases) for temporal placements is emotional or physical neglect of the child. It is less frequent that the child is taken into state care because of behavioral problems (20 percent of the cases); physical abuse as a reason for removal from the biological family is even rarer (6 percent) (KSH, 2012). Legally, children cannot be removed from their families just because of financial difficulties in their family. However, in practice, it is estimated that every third child is taken into state care for reasons closely related to poverty.[33]

In principle, state care is a temporal solution until the child's own family is fit to take care of her again, or, in case of permanent placements, until the child is adopted. In practice, however, the biological parents can rarely solve the problems fast leading to the removal of the child, and children older than 10 are almost never adopted. Thus children typically spend many years (5.5 on average before they turn 18) in state care (SOS, 2015).

There is no official statistic about the age of the first placements, but 2009 data from Nógrád, a Hungarian county shows that 68 percent are over 3 years old, and 50 percent are over 6 when she first enters into state care (Darvas, Farkas, Kende, et al., 2016). Around 8-10 thousand children left the system annually between 2006 and 2020, with a slightly higher number of entrants. Thus the number of children raised in state care slowly increased in this period from 17 thousand to 21 thousand (or from around 1000 to 1200 children per 100000 children in Hungary). This is a relatively high figure in the European context, where the number of children in state care per 100000 children varies from less than 500 in Mediterranean countries to more than 1000 in post-socialist countries (see Figure 2.1).

When a child is taken into state care, she can be placed into residential care or family foster care. We first present the main features of the two types of care in Hungary, and then describe the process of making a decision about the placement.

2.3.2 Family foster care and residential care

Family foster care is a type of care in which children are placed to live in the private home of trained and paid professional temporary parents. In Hungary, foster parents can raise maximum 7 children in their home, including their biological children. To become a foster parent, they have to complete a 60-hour training and comply with other eligibility criteria (age between 24 and 65, passing a psychological test, and suitable living conditions). They work under local foster parent networks maintained by the state, churches or NGOs.

[33]See the ombudsman's report at the following link: https://www.ajbh.hu/-/eroforrasokat-a-megelozesre-es-az-alapellatasra-az-ombudsman-a-gyermekek-csaladbol-valo-elsodlegesen-anyagi-okbol-torteno-kiemelesek-gyakorlatarol (downloaded 03.07.2023)

The number of foster parents slowly increased from 5300 in 2006 to 5700 in 2020. Foster parents raise an increasing number of foster children in their homes: in 2006, 17 percent of them raised 4 or more foster children, in 2020 30 percent of them. Foster parent wages are low: the base salary is 30 percent of the minimum wage, with an increase of 20 percent of the minimum wage per every additional foster child. On top of the base salary, foster parents get allowances they have to spend on the needs of their foster children.

Considering the low wages, and the high-responsibility and emotionally challenging nature of the job, it is not surprising that there is a constant shortage of foster parents. According to experts at least a thousand foster parents were missing from the system in 2021[34], suggesting that the shortage was even more severe in our observation period when there were fewer foster parents. The scarcity is also apparent from the fact that even though all children under 12 should be placed to foster families according to a 2014 legislation, 10 percent of children under 12 still lived in residential care in 2018.[35] Almost all (96 percent) of foster parents are women. Table B.1 in the appendix compares foster mothers to biological mothers in our sample of mothers of 6$^{\text{th}}$ grade children and shows that foster mothers are generally older (mean age of 49 vs 39) and less educated than biological mothers (36 percent with at least secondary education vs 52 percent), and they typically live in larger households (6.8 vs 4.5).

In Hungary a growing number of children in state care are placed into foster care (11 thousand in 2004, 16 thousand in 2020). The ratio of children living in foster families among all children in state care steadily increased from less than 40 percent in 1996 to over 70 percent in 2020 (Figure B.1 in the appendix).

The 30 percent in residential care is a high figure compared to the UK and the US, where only around 10 percent of state-raised children live in such settings. But compared to other European countries the ratio is relatively low, as it is apparent in Figure 2.1). Still in 2019, in some countries (for example, in Germany and Spain) more children lived in residential care than in foster care.

In the following, we describe the features of residential care in Hungary. The main form of residential care is the children's group home. Group homes can accommodate a minimum of 12 and a maximum of 40 children, who live in small and typically mixed-age and mixed-gender groups of a maximum of 12 children. A team of childcare professionals takes care of them 24 hours a day, but none of them lives together with the children. According to the child protection legislation, the team should consist of highly trained educator(s) (teachers, psychologists or social workers, trained child protection assistants, child supervisors, developmental teachers, and psychologists).[36] However, wages are low with a heavy workload, and the Hungarian system struggles with a shortage and frequent fluctuation of childcare workers. For example, in 2012, one-third of psychologist jobs were vacant, and there was a shortage of child supervisors as well (KSH, 2012). The problem of labor shortage seems persistent, in 2019 14 percent of the jobs at children's group homes were vacant.[37] In addition, child supervisor positions are sometimes filled with employees without any training (390 in 2020).[38]

[34]See the following article in Népszava: https://nepszava.hu/3134409_ma-van-az-allami-gondoskodasban-elo-gyermekek-napja-figyeljunk-oda-rajuk (downloaded: 03.07.2023)

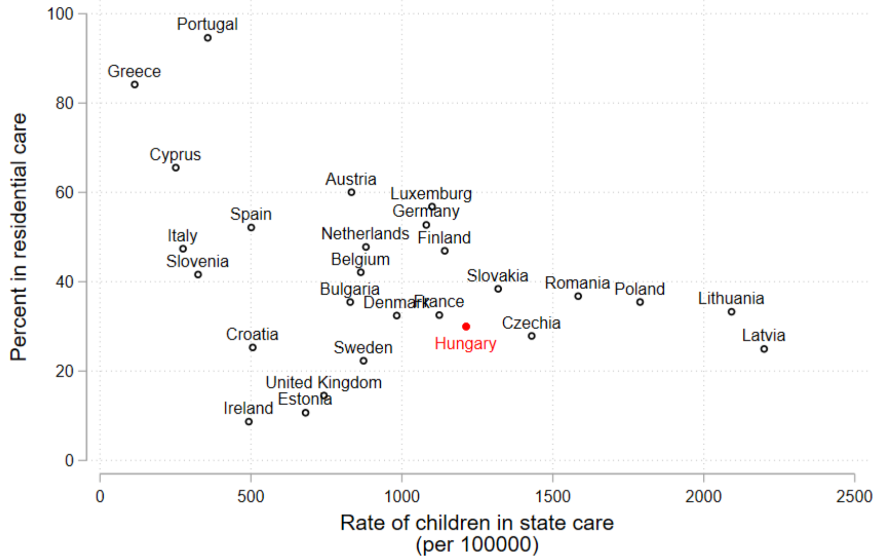
[35]Source: https://www.parlament.hu/documents/10181/4464848/Infojegyzet_2020_44_gyermekotthonok.pdf/963af491-78b4-7e46-55df-0c53df5fdeae?t=1590051507104

[36]15/1998. (IV. 30.) NM rendelet a személyes gondoskodást nyújtó gyermekjóléti, gyermekvédelmi intézmények, valamint személyek szakmai feladatairól és működésük feltételeiről, 2010-ben hatályos változata

[37]Source: Report of Hintalaovon ALapítvány: https://hintalovon.hu/wp-content/uploads/2021/06/Hintalovon_Alapitvany_Gyermekjogi_jelentes_2020.pdf downloaded (03.07.2023)

[38]Source: https://merce.hu/2020/10/02/szel-bernadett-brutalis-szakemberhiany-van-a-gyermekotthonokban/

Figure 2.1: Rate of children in state care, and percent living in residential care in Europe in 2019



Source: Eurochild DataCare Project https://eurochild.org/initiative/datacare/

To sum up, in the European context, an above-average share of children live in state care in Hungary. The number of foster placements steadily increased in the last decades in Hungary, putting the country among the countries with the highest foster care ratios in Europe. Still, many children live in residential care, and the fact that the majority of them do not exhibit severe behavioral problems suggests that most of them would be fit to live in a foster family as well. The conditions for both foster families and group homes are far from ideal. There is a shortage of workers in both types of homes. The professionals working in the system often feel overwhelmed and underpaid. The human capital of foster parents is lower than that of the mothers in the general population, while in group homes there is a shortage of trained professionals. In the next section, we describe how authorities decide about the placement of children into different types of homes.

2.3.3 Placement Decision

When a child enters into state care, or when her previous placement has to be changed for some reason, the local authorities make a decision about the placement of the child. According to the law, children should be primarily placed into foster families, and if it is not possible, in residential care. During the decision process, the county child protection authority (TEGYESZ by its Hungarian abbreviation) considers the age of the child, the placement of siblings, the location of the placement relative to the biological parents and previous school, and the special needs and disabilities of

downloaded 03.07.2023

the child. The younger the child, the more authorities prefer placement into foster care.[39] When multiple siblings need to be placed, the authorities attempt to keep them together. As children are often removed from large families with many siblings it is not always possible to achieve this. Placing the child close to her previous home is also important, as there are compulsory regular meetings with the biological parents, and placements outside 50 kilometers of the previous home are scarce (Vígh, 2015).

In practice, TEGYESZ announces to foster parent networks the need to place a child, typically with information about the age, gender, special needs and health condition of the child (Vígh, 2015). If multiple networks respond, the TEGYESZ decides considering the needs of the child listed above.

As a result of this decision process, and the scarcity of foster parents, similar children can end up being placed in different types of homes. Children who would be fit to live in foster families are placed in residential care because not enough foster families are available locally at the time they need to be placed.

2.4 Data and Sample

The main data source is an administrative individual-level monthly panel hosted by the Databank of the Centre for Economic and Regional Studies. The data set contains information on a 50 percent random sample (5.17 million people) of the Hungarian population[40]. The sample is aging: the original sample of people taken in 2003 is followed until 2017, and no new observations are recruited. Administrative records of individuals from the National Health Insurance Fund Administration, the Hungarian State Treasury, the National Tax and Customs Administration, the Ministry of Finance and the Educational Authority are linked based on social security numbers (Sebők, 2019).

Most importantly for our research, data from the 2008-2017 waves of the National Assessment of Basic Competencies (NABC) is also linked to the administrative data. The NABC is an annual standardized mathematics and reading test conducted on the whole population of 6^{th} 8^{th} and 10^{th} grade students. A background survey accompanies the tests with a rich set of questions about the student and her family environment. We use this data source to identify the home type: foster family or group home, and in some specifications own family as well. Using the NABC to identify home type means that we can observe the home type of the child 3 points in time: 6^{th} grade, 8^{th} grade, and 10^{th} grade.[41]

We use two questions from the background questionnaire to define the home type. In one question children choose their family type, from three options: own family, foster family, and children's group home. In another, they list people living with them. We define a child as one who lives in a foster family if she indicates living in a foster family, lives with a foster mother, and does not live with any biological adult family members (for example biological mother, father, grandmother, grandfather). We categorize children's home types into group homes if they indicate living in a group home and they do not live with any biological adult family members or foster parents. We exclude children with inconsistent or insufficient data about home type. Table 2.1

[39] Since 2014, legally, children under 12 must be placed into foster homes, but this legislation was not in place yet in our observation period for 6^{th} grade students.

[40] The administrative database used in this chapter is a property of the National Health Insurance Fund Administration, the Central Administration of National Pension Insurance, the National Tax and Customs Administration, the National Employment Service, and the Educational Authority of Hungary. The data was processed by the Databank of the Centre for Economic and Regional Studies.

[41] More observations are possible if the child is repeating grades, and fewer observations are possible if the child is not taking the test, or does not answer the background questionnaire in a given year

shows the details of the definition.

Table 2.1: Definition of home type

Household Members			
Lives with	Does not live with	Survey Home Type	Cleaned Home Type
Biological parent		own family	own parents
		foster family	own parents
		grouphome	missing (inconsistent)
		missing	own parents
Grandparent and/or other relative (e.g. uncle, aunt)	Biological parent, foster parent	own family	kinship care
		foster family	kinship care
		grouphome	missing (inconsistent)
		missing	kinship care
Foster mother	Biological parent, grandparent, other relative (e.g. uncle, aunt)	own family	missing (inconsistent)
		foster family	foster care
		grouphome	missing (inconsistent)
		missing	missing (insufficient)
	Biological parent, foster parent, grandparent other relative	own family	missing (insufficient)
		foster family	missing (insufficient)
		grouphome	grouphome
		missing	missing (insufficient)
All other combinations			missing (inconsistent)

Even though the NABC covers the full population of students in specific grades, we do not observe the home type for every student, because of non-response. It is hard to give precise estimates about what percentage of the foster children and residential care children we observe in the NABC, as there is no aggregate data available about the number of children in different types of homes by grade. The Hungarian Statistical Office reports data on the overall age distribution of children in state care in a given year, but no conditional distributions by home type. Along with this, the aggregate number of children in different types of state care is reported. In the first row of Table 2.2 we estimate the number of 6^{th} grade children in different care types, using the third of the 12-14 age group as the full aggregate number of children state care, and assuming that the same age distribution for foster and residential care. These are crude estimates, but still, they can help to get a sense of the non-response rate for the different home types. We estimate that we observe around 80 percent of 6^{th} children who live with their own parents, 65 percent of all children in foster care, and 39 percent of children in group homes. The high non-response rate can arise from the fact children have to take the questionnaire home and answer it with their caregivers. Thus the sample we observe is plausibly less problematic and has more involved and less overwhelmed caregivers than the whole sample of children in state care. This means that we probably observe a sample with more favorable outcomes than the full population of children in state care in Hungary, especially in the case of children in group homes.

From this pool of children, we include those in our sample whom we observe as young adults – until the end of age 19 – in the administrative data (which ends in 2017). As most children are 12-14 in grade 6 this means that most of our sample consists of children who are 6^{th} grade students between 2008 and 2010. As many children repeat grades in the sample of children in state care, we observe some children at age 19, who are 6^{th} grade students later than 2010. We include these children in our main analysis to maximize the number of observations but we also run robustness checks where we only include 2008-2010 6th graders.

A last step in our sample selection is that we keep only those children, who did not change home type. This criterion applies to 86 percent of children in state care within our sample. By doing so, we focus specifically on children who have been in long-term foster care or long-term residential care during their teenage years, excluding those who have only had short-term stays in state care. This can introduce some selection problems: for example, we exclude those who started to live in foster care, but for some reason, they continued to live in residential care. It is possible that these children are especially problematic, and their foster parents let go of them. To make sure that the exclusion of these children is not an important issue in our analysis, we run robustness checks on the sample where we only take into account the home type in 6th grade.

Our final sample consists of 890 children, 616 children in foster care, and 274 in residential care. In the next section, we show the descriptive differences between these two groups.

Table 2.2: Sample

		All children	Own parents	Foster care	Group homes
2008					
Estimated number of 6th graders based on KSH.	N	110875	108231	758	564
Observed in NABC	N	107 654	85296	474	202
	Ratio to full population	97.1	78.8	62.5	35.8
Observed in Admin3 until age 19: Sample used in robustness checks	N	50927	40568	222	87
	Ratio to NABC population	47.3	47.6	46.8	43.1
Does not change home type: Sample used in main analysis	N		40219	173	78
	Ratio to Admin3+NABC population		99.1	77.9	89.7
	Ratio to full population		37.2	22.8	13.8
2009					
Estimated number of 6th graders based on KSH.	N	104712	102081	785	530
Observed in NABC	N	100620	79295	488	211
	Ratio to full population	96.1	77.7	62.1	39.8
Observed in Admin3 until age 19: Sample used in robustness checks	N	47254	37723	213	81
	Ratio to NABC population	47.0	47.6	43.6	38.4
Does not change home type: Sample used in main analysis	N		37394	174	73
	Ratio to Admin3+NABC population		99.1	81.7	90.1
	Ratio to full population		36.6	22.2	13.8
2010					
Estimated number of 6th graders based on KSH.	N	98756	96179	773	516
Observed in NABC	N	96898	78935	562	204
	Ratio to full population	98.1	82.1	72.7	39.5
Observed in Admin3 until age 19: Sample used in robustness checks	N	39715	32354	230	77
	Ratio to NABC population	41.0	41.0	40.9	37.7
Does not change home type: Sample used in main analysis	N		32055	194	68
	Ratio to Admin3+NABC population		99.1	84.3	88.3
	Ratio to full population		33.3	25.1	13.2

Note: In our main analysis we include 75 additional foster children and 55 in residential care, who are 6th graders in 2011 or later, but are observed until age 19.

2.4.1 Descriptive differences

Table 2.3 shows the means of variables observed at 6^{th} grade, and the means of the outcomes at age 19 for children raised in foster care and residential care.

Table 2.3: Means of variables by care type

	Foster care	Residential care	Difference
Control variables			
Age in 6th grade	12.9	13.3	-0.45***
Boy	0.49	0.49	-0.004
Number of siblings	3.24	3.50	-0.26*
Mathematics testscore	1343	1324	18
Reading testscore	1342	1302	40**
Mathematics grade	2.65	2.41	0.24***
Grammar grade	2.96	2.72	0.24***
Literature grade	3.29	2.89	0.40***
Behaviour grade	3.73	3.44	0.30***
Effort grade	3.32	3.02	0.30***
No special need	0.807	0.741	0.07*
Mild mental disability	0.003	0.015	-0.011
Physical disability	0.003	0.004	-0.000
Speech impairment	0.002	0.007	-0.006
Psychological problems	0.117	0.161	-0.044+
Behaviour or learning problems	0.068	0.073	-0.005
Starts primary school late	0.60	0.58	0.02
Outcome variables			
Secondary education	0.40	0.21	0.193***
High school education	0.17	0.06	0.110***
Used antidepressants	0.11	0.16	-0.05*
Used tranquillizers	0.07	0.11	-0.03
Pregnancy	0.31	0.56	-0.26***
Birth	0.22	0.40	-0.18***
Abortion	0.15	0.30	-0.15***
NEET months at age 19	4.22	5.88	-1.67***
NEET at least 6 months	0.37	0.53	-0.15***
Index of all outcomes	0.11	-0.23	0.34***
N	616	274	

Robust standard errors, $+\ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001$.
In the main regressions, age, test scores, and grades are included as categorical dummies

It is clear from the table that children in foster care and children in residential care are already different on average by 6^{th} grade. For instance, children in residential care tend to be older, with an average age of 13.3 years. As the share the children in the two groups start school at the same age on average, this indicates that children in residential care are more likely to have repeated grades than children in foster care. In both groups, gender ratios are roughly the same (49 percent boys). Most outcomes measuring cognitive skills show that children in foster care have an advantage. Their grades in mathematics, grammar, and literature are better (by 0.24-0.4, or around a third of the standard deviation), and their reading test scores are higher (by 40 points, or 0.2 standard deviations). There is no statistically significant difference in mathematics test scores. Note that the value of specific grades can depend on the context. The highest grade of 5 can be harder to achieve in a more demanding school. However, test scores are standardized: every child completes the same test in a given calendar year. Children in foster care have better results in terms of non-cognitive measures as well. Their effort and behavior grade is higher (0.3, or one-third of the standard deviation). In addition, fewer of them are diagnosed with any special educational needs (SEN) (81 percent versus 74 percent). Students categorized as having special educational needs encompass various conditions, including physical disabilities, sensory disabilities, mild to moderate mental disabilities, speech impairments, autism, psychological problems (including dyslexia, dysgraphia, and hyperactivity), as well as students with learning and behavioral difficulties. The ratio categorized to these groups is higher in the case of all special needs for children in residential care, however, the difference is only statistically significant in one category: 12 percent of children in foster care exhibit psychological problems and 16 percent of students in residential care. Though we have limited information about early childhood, we have a variable for delayed primary school start, an early indicator of ability.[42] Children appear to have been more similar before the start of primary school: 60 percent in foster care and 58 percent in residential care started primary school late.

Thus generally, foster children have better outcomes already at grade 6. One reason could be that these variables are already outcomes of the different types of care. The fact children appear to have been similar at the start of primary school supports this mechanism. Another possible mechanism is that children with special education needs and psychological problems are more likely to be placed into residential care, as foster parents willing and capable to take care of these children are even more scarce than regular foster parents.

The differences between foster-raised and residential care youth appear to be even larger for adult outcomes. For instance, despite having similar standardized mathematics test scores in 6$^{\text{th}}$ grade, it is remarkable that children in foster care are twice as likely to complete secondary education by the age of 19 (40 percent versus 20 percent). Furthermore, a significantly higher proportion of foster care children, three times as many, are expected to graduate with an academic secondary degree (17 percent versus 6 percent), which makes it possible to continue their studies at a university. These outcomes are important because those who have any secondary education have better labor market prospects than those who do not. Graduating from high school (either the academic [43] or the vocational type [44]) is a precondition for starting university, and those who obtained a high school degree have much better prospects than those without (Hajdu et al., 2015).

We look at two outcomes related to mental health: an indicator of ever buying prescribed antidepressants and tranquilizers. In both groups, the purchase of these drugs happens more frequently than in the general population pointing to more psychological problems. The rates are even higher in residential care: 16 percent used antidepressants, versus 11 percent in foster

[42] According to the regulation, children turning 6 before June 1 in any calendar year are required to start primary school on September 1 of that particular calendar year. Children born in September-May are considered to start primary school late if they start school after they have turned 7. Children born between June and August are considered to start late if they enter primary school after they have turned 8.

[43] Gimnázium in Hungarian.

[44] Szakközépiskola in Hungarian.

care. For tranquilizers, the difference is not statistically significant, but the mean is again higher for youth raised in residential care (11 percent versus 7 percent)

The difference is also striking in terms of fertility outcomes, which we can identify from the ISCED codes recorded by the doctors during patient visits: 56 percent of girls in residential care get pregnant in their teens, as opposed to 31 percent in foster care. Births and abortions are both less frequent for foster-raised girls (22 percent and 15 percent) than for girls in residential care (40 percent and 30 percent), indicating that both wanted and unwanted pregnancies occur less for girls in foster care.

To capture adverse labor market situation we use two outcomes related to NEET (neither employed, education or training) status: the months spent as NEET, and an indicator of being NEET for at least 6 months during age 19. We prefer this outcome to the share of employed individuals, as successful youth might not work but still study at this age. Former foster youth spends 4.22 months as NEET on average, while residential care youth are NEET for 5.88 months, on average. Individuals raised in residential care are 15 percentage points more likely to spend at least 6 months as NEETS at age 19.

In the regressions, we use an index as an outcome as well, calculated as the mean of four standardized variables: finishing secondary education, an indicator of using antidepressants or tranquilizers, an indicator for pregnancy, and months spent as NEET, where a larger value shows a more favorable outcome (for example, not getting pregnant). Former foster youth has a higher value (0.11) on this index than former residential care youth (-0.23).

2.5 Empirical strategy and results

While we observe large descriptive differences in the outcomes of children placed into foster care and residential care, we cannot conclude that it is the different types of placements causing these differences. Older children, and children with special needs, or disabilities are less likely to be placed to foster care because of the scarcity of suitable foster parents. In addition, foster parents can decide to let go of a child and they are more likely to make that decision if the child is more problematic.

We use two approaches to deal with this selection problem. The first is an OLS strategy controlling for a rich set of variables measuring cognitive and non-cognitive skills in 6^{th} grade to capture and control for the different underlying abilities of the children. The second is an instrumental variable strategy, using local foster parent capacity as an IV, to filter out the effect of further potential confounders. In the following sections, we present these empirical strategies and the results we get using them.

2.5.1 OLS

2.5.1.1 Empirical strategy To identify the causal effect of placement type on future outcomes, ideally, we would run an experiment where children taken into state care are randomly assigned to foster care or residential care and compare the means of adult outcomes. The next best would be to observe all features of the children which influence the placement decision, at the time of the decision (location of the biological parents, age, special needs and the number of siblings need to be placed), and control for these variables. However, we first observe rich measures of ability of the children in 6^{th} grade, when they already spent some time in state care. Our strategy is to control for these measures of late ability and discuss how the late measurement biases our estimates. We use the following OLS model:

$$Y_i = \alpha + \beta foster_i + X'_{i,gr6}\gamma + u_i, \tag{2.1}$$

where Y_i can be a large set of different outcomes, related to education, mental health, fertility, and labor market situation (see Table 2.3). The main explanatory variable is $foster_i$, an indicator of living in long-term foster care as a teenager, with living in residential care as the reference group. $X'_{i,gr6}$ is a set of controls measured in 6th grade: gender, age, number of siblings, county indicators, calendar year indicators, measures of cognitive ability (standardized mathematics, and reading test scores, grades in mathematics, literature and grammar), and measures of non-cognitive ability (grades for behavior and effort in school, and the presence of the following special education needs: mild mental disability, autism, physical disability, speech impairment, psychological problems, and behavior or learning problems). We include grades and test scores as categorical dummies (6 categories for grades, 8 for test scores) with a missing value as a separate category.[45]

We use White robust standard errors to account for the heteroskedasticity in the models (necessarily present for most of our outcomes, which are binary variables). As we use multiple outcomes, the probability of Type I error is higher than the simple p-values. We tackle this problem by two methods: first, we estimate the regression of a composite index, which combines our outcomes, to check whether this variable is affected statistically significantly. In addition, we calculate Bonferroni corrected p-values and Benjamini-Krieger-Yekutieli sharpened two-stage q values for false discovery rates based on Benjamini et al. (2006).

The coefficient of interest is β, which would identify the effect of placement type if conditional on the variables in $X'_{i,gr6}$, care type was independent of potential outcomes. While there is some randomness in the placement decision, as described in Section 2.3, this assumption is not fully satisfied. First, because some variables in $X'_{i,gr6}$ are bad controls: cognitive and non-cognitive ability in 6th grade can be already outcomes of the earlier placement. In the appendix, we show, that under plausible assumptions, we can say that these are special versions of bad controls, proxy controls, biasing gains from foster care downwards. There is a straightforward intuition behind this result. By controlling for 6th grade variables we compare children, who are similar at this time. If we consider the findings from the literature that residential care affects young children negatively, comparing children who are similar in 6th grade means that the child in foster care started out with a *worse* early ability: thus comparing her to a child with better underlying ability in residential care underestimates the true gains from foster care.

A second, and more important threat to the assumption of independent placement conditional on observables, is the presence of omitted variables that can affect placement type and future outcomes as well. An important omission is the history of previous placements, with the age when the child is removed from the biological family. While we do not observe the age of first placement in our data, we can gain some insight into this from a separate dataset on children in state care from one Hungarian county (Nógrád), from the study of Darvas, Farkas, Ágnes, et al. (2016). In these cross-sectional data, we observe the whole population of children in state care in the county in 2011. In this dataset, the mean age of first placement for children in foster care is 4.2, and it is 9.1 for children in residential care. Thus children in foster care tend to be removed from their biological families substantially younger. The direction of the bias caused by this omitted variable depends ultimately on the question of whether children are better off in their biological families or in state care. If they are better off in state care, children removed earlier and placed into foster

[45] For grades in different subjects, 6-7 percent of our sample has missing values, and for test scores less than 1 percent, but every child has a non-missing value at least on one of these variables. By allowing for missing values on some of the variables, we can keep more observations.

care might have better outcomes not because of the foster placement, but because of the earlier removal. However, the literature shows that the effect of removal is ambiguous: some find that the marginal child is better off in the biological family (Doyle Jr., 2007), while others find the opposite (Bald, Chyn, et al., 2022). The studies looking into the correlation between placement age and future outcomes mostly find no significant relationship between age of placement, with some studies finding small negative or small positive correlations (see the review of McDonald et al., 1996). To further study the role of this omitted variable we rerun the regressions on a small sample who is taken into state care between 6^{th} and 8^{th} grade (156 children) to take the age of placement under control. The effects have the same size and similar magnitude as in our main specification (insignificant though). This result taken together with the mixed findings in the literature shows that it is unlikely that the unobserved difference of placement age explains the difference in the outcomes between children in foster care and residential care.

Another potentially important unobserved variable that could be correlated with placement type and future outcomes is ethnicity. Roma children are overrepresented among children in state care. In 2005, it was estimated that 32 percent of children in state care were members of the Roma ethnic minority, while the Roma ratio was 13 percent in the general population (ERRC, 2007). If Roma children were more overrepresented in residential care than in foster care, we could find worse outcomes for children in residential care because of discrimination against the Roma, rather than residential placement itself. In the above-mentioned dataset from Nógrád, we have information about ethnicity, so we can use it to get a better understanding of the distribution of Roma children in different types of state care. In Nógrád, Roma children were more overrepresented in foster care (80 percent), than in residential care (73 percent). Nógrád is a less developed county within Hungary, with one of the highest Roma ratios, thus it is not clear whether this pattern is similar in other parts of the country as well. Nevertheless, these data provide evidence suggesting that unobserved ethnicity is unlikely to lead to an overestimation of the beneficial effect of foster care, considering that Roma children are not expected to be substantially overrepresented in residential care.

Taken together the proxy control problem, and the most important omitted variables we can detect, we think it is likely that β measures a lower bound of the average treatment effect of foster care on future outcomes.

2.5.1.2 OLS results Table 2.7 shows the OLS regression results for different outcomes. We show the estimate for β from equation 2.1, and in a separate line, we show the uncontrolled difference between children in foster care and residential care for the given outcome. Because of the above-mentioned factors biasing β downwards, the true effect of foster care is likely to be in between these two numbers. First, we show education outcomes: finishing secondary education (any type) and graduating from high school (a subset of those who finish any type of secondary education) by the end of age 19. Model (1) shows that those who receive foster care are 19 percentage points more likely to finish secondary school, and 8.3 percentage points of the difference remain if we control for differences in 6^{th} grade. This is a statistically significant difference (p=0.021), and we consider it large: it is a 40 percent increase compared to the baseline share of 21 percent of secondary school graduates in residential care. To gain even more insight into the effect size, we also ran a version of equation 2.1 where we included children who live with their own parents as well, and the main explanatory variable is a dummy variable for living location (own family, foster care, residential care). We show the results of these regressions in Table B.2 in the appendix. In that specification, we measure that the difference between children raised in foster care and residential care (5.3 pp in that specification), is even larger than the difference between children raised in their birth families and children raised in foster care (3.8 percentage points). The estimated coefficient for finishing high school is also positive (1.7 percentage points), but insignificant.

In columns (3) and (4) of Table 2.7 we show the coefficient estimates on mental health outcomes. The estimated effect of foster care on using antidepressants is negative but insignificant. On tranquilizers, we find a negative 5 percentage points effect (p=0.039). We measure mental health outcomes from data from pharmacies, where we see purchases of prescription drugs. Generally, not buying prescription drugs can show the lack of a mental health problem, but also the presence of an untreated one, thus whether the negative coefficient indicates real differences in mental health is unclear. We think that the share of untreated children in the sample of children in state care should be small while they are in the child protection system: both children in foster care and in residential care regularly meet psychologists, thus it is likely that psychological problems are discovered and treated in this population. However, it is also possible that children with milder mental health problems are over-treated with tranquilizers in institutions to minimize aggressive behavior. In our data, we see that tranquilizer usage is at the maximum at around 16-17 and it drops at ages 18 and 19 when the children leave the child protection system, and the drop is larger in the case of children in residential care. This is consistent with over-treatment, but also with children experiencing more severe psychological problems as a result of living in an institution. Thus, it is not clear whether the negative effect on tranquilizer usage signals that children in foster care have better mental health or that they have the same mental health but are not overtreated with tranquilizers. Whichever channel explains the negative effect on tranquilizer usage, living in foster care seems beneficial to children.

Columns (5) to (7) show the effects on pregnancies births and abortions, estimated on the sample of girls. We consider these outcomes important as teenage motherhood has an adverse effect on teen mothers' future prospects (see for example Ashcraft et al., 2013), and their children are more likely to have adverse outcomes as well (see for example Aizer et al., 2022). In the context of children in state care, teenage motherhood is possibly even more difficult, as the girls lack a supportive family background. In a recent study, children of teenage mothers in state care were found to be more than seven times more likely to lose custody of their babies by the time children were two years old (Wall-Wieler et al., 2018)), creating a cycle of involvement with the child protection services. Even if the pregnancy is terminated, having an abortion as a teenager can be painful both mentally and physically. The estimates show that foster care decreases the probability of teenage pregnancies by 16 percentage points, and births and abortions by 12 percentage points (p=0.012, 0.047, and 0.023 respectively). This indicates that both wanted and unwanted pregnancies happen less frequently among children in foster care.

In columns (8) and (9) we estimate the effects on two outcomes related to NEET (not in education employment or training) status. Similarly to previous outcomes, we find large possible gains from foster care. Conditional on 6^{th} grade ability foster youth spend 1.1 fewer months as NEET at age 19 than residential youth (p=0.005), and they are 11 pp points less likely to spend at least 6 months as NEET at age 19 (p=0.004).

Lastly, in column (10) we re-estimate equation 2.1 on a standardized index of four outcomes (secondary education, mental health, pregnancy, NEET months), larger for better outcomes. The estimate is large (0.0216) and highly statistically significant for this composite outcome as well (p=0.0001), supporting that significant results on the other outcomes are not an artifact from multiple hypothesis testing. To further support this, we show Benjamini-Krieger-Yekutieli sharpened two-stage q values for false discovery rates in Table B.3 in the appendix, calculated with the code of

Anderson (2008). Using a 5 percent threshold for false discovery, all previously statistically significant outcomes remain significant. If we use the Bonferroni correction (i.e. simply multiplying the p values with the number of tests), most outcomes become non-significant at the 5 percent level, but some, including the index calculated from other outcomes, retain statistical significance even with this overly conservative (and less powerful) method.

Table 2.4: Regression estimates for the effect of foster care on all outcomes

Outcome	(1) Secondary School	(2) High School	(3) Antidper	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes
Foster care	0.083*	0.017	-0.038	-0.049*	-0.157*	-0.115*	-0.121*	-1.108**	-0.114**	0.216***
	(0.036)	(0.021)	(0.029)	(0.024)	(0.062)	(0.058)	(0.053)	(0.394)	(0.040)	(0.054)
Raw difference	0.19***	0.11***	-0.05*	-0.03	-0.26***	-0.18***	-0.15***	-1.67***	0.15***	0.34***
adj. R^2	0.156	0.300	0.033	0.007	0.100	0.097	0.043	0.138	0.123	0.176
N	890	890	890	890	457	457	457	890	890	890

The following controls (measured at 6$^{\text{th}}$ grade) are included in all models: gender; age; number of siblings; county indicators; calendar year indicators; standardized mathematics-, and reading testscores; grades in mathematics, literature, behavior and effort; indicators for the presence of mild mental disability, autism, physical disability, speech impairment, psychological problems, and behavior or learning problems. Robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Hungarian DOI: 10.14754/CEU.2023.04

2.5.2 IV

2.5.2.1 Empirical strategy To provide further support for the causal interpretation of the results, we turn to an IV specification. In this specification, we use variation in county-level foster parent capacity to instrument getting into foster care.

We define foster parent capacity as the number of foster mothers in the county per 100 children living in state care in 2015. The number of foster mothers is calculated from the administrative data where we observe the occupation code [46] and the living location of employees. We use 2015 because we observe foster parents as a separate occupation only since then.[47] The number of children in state care per county in 2015 is reported by the Hungarian Statistical Office. We split counties into two categories, at the median foster mother capacity (24 per 100 children in state care). This division allows us to use the Wald estimator instead of a continuous instrumental variable (IV), which we prefer, as it provides a more straightforward interpretation of the estimates.

The idea is that counties that are similar otherwise, can have low or high foster parent capacity. In Figure 2.2 we show that generally, in more developed counties, there are fewer foster parents relative to the number of children living in state care.[48] Still, there is variation in foster parent capacity even within similarly developed counties. For example, Vas and Fejér counties are similarly developed, but while in Vas there were only 10 foster parents per 100 children in state care, in Fejér, the same number is 26.

As there are compulsory meetings with the biological parents, children need to be placed close to their previous living location. Indeed, in the sample of children who live with their biological family in 6$^{\text{th}}$ grade and in state care in 8$^{\text{th}}$ grade, we observe that 95 percent of children stay in the same county. Thus, authorities in counties with few foster parents are less likely to find foster placements for children living there. In our IV approach, we compare children in state care who are similar but live in counties with different foster mother capacities. Specifically, we estimate the following model:

$$Y_i = \alpha + \beta foster_{i,gr6} + X'_{i,gr6}\gamma + C'_{s,gr6}\delta + u_i, \tag{2.2}$$

where $foster_{i,gr6}$ is instrumented by $fostermotherc - high_s$, an indicator of living in a county with high foster parent capacity. Y_i denotes the same outcomes as in equation 2.1. In $X'_{i,gr6}\gamma$ we include the same controls as in 2.1 except for county dummies. We cannot include county dummies because the instrument varies at the county level only. Instead, we use county-level controls, $C_{s,gr6}$ measured at grade 6 (employment, unemployment, mean wage, birth rate, and a composite development indicator). Including the county-level controls is necessary to get meaningful estimates, as foster mother capacity is correlated with county features.

We report unclustered, robust standard errors, which prove to be more conservative than the ones we get if we cluster the errors by county.

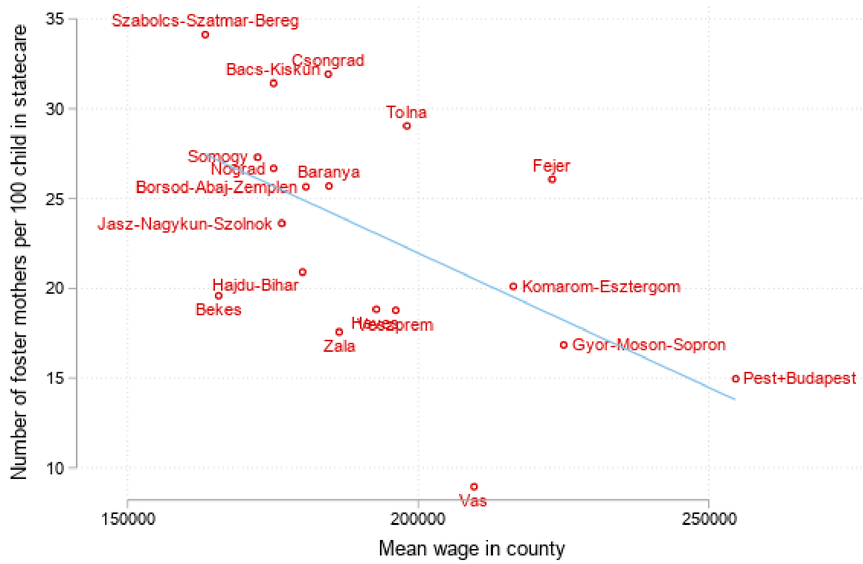
The coefficient β measures the local average treatment effect on compliers – the effect of foster care on those children who would be placed into residential care in a county with low foster parent

[46] Occupation code (FEOR) 3512

[47] Before 2015, being a foster parent was a special type of employment without social security insurance, and is unobserved in our data because of this.

[48] In the analysis, we pool Pest and Budapest, which are administratively separate entities (with separate child protection services), but geographically, the city of Budapest is located inside Pest county. For this reason, children removed from their biological families in Budapest can be placed to live in foster families living in Pest, while still complying with the rule of compulsory meetings with biological parents.

Figure 2.2: Number of foster mothers per 100 children in state care by mean wage in county



Note: Linear fit with light blue. Data from Pest and Budapest are pooled.

capacity, and into foster care with high foster parent capacity – if the usual IV assumptions hold. In the following, we discuss these assumptions.

Table 2.5 reports the first stage results and shows that the instrument is relevant – controlling for county features, children who live in counties with high foster parent capacity are 15 percentage points more likely to be placed into foster care (p=0.0001), controlling for individual and county characteristics. The same number using only the sample of girls (the relevant sample for some of our estimates) is 16 percentage points (p=0.0001). But, even though foster parent capacity is correlated with getting into foster care, the instrument is somewhat weak in the sense that the F values are close to the usual rule-of-thumb value of 10 in the whole sample (15.3) and a bit smaller than that in the sample of girls (9.7). The Stock and Yogo critical values for a maximum 10 and 15 percent size distortion are 16.38 and 8.96, and the F values falling between these two numbers show that the IV is moderately weak, but the bias the weakness causes is likely not to be large. To characterize the compliers we run the first stage on specific subgroups of children, and divide it by the overall first stage to get the relative likelihood that the complier is a member of the given subgroup. Table B.5 in the appendix shows the results and indicates that compliers are somewhat more likely to have special needs than the overall population of children in state care, and to have below median math scores and at least two siblings. The differences are not substantial in terms of gender.

The other key IV assumption is exogeneity, or (conditionally) independent treatment assignment, which is threatened if county-level foster parent capacity is correlated with other variables affecting the outcomes of children living there. For example, suppose that unmeasured school quality is lower in counties with many foster parents. In this case, we would underestimate the positive

Table 2.5: First stage regression: foster care placement on foster mother capacity

	Raised in in foster care	
	(1)	(2)
High foster mother capacity	0.150***	0.159**
	(0.038)	(0.051)
Sample	Full	Girls
CD Wald F	15.32	9.699
N	869	443

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

County and individual controls are included in all regressions

effect of foster care. To argue that omitted variables affecting the outcomes of the whole population do not play a role, we run a placebo regression on the outcomes: a reduced form of the IV on the sample of children who live with their own parents:

$$Y_i = \alpha + \beta fostermotherc - high_{s,gr6} + X'_{i,gr6}\gamma + C'_{s,gr6}\delta + u_i. \tag{2.3}$$

In these regressions, β should be 0, if foster mother capacity is uncorrelated with unobserved variables correlated with the outcomes. Indeed, in Table B.4 in the appendix we see that the β coefficients are small and insignificant in most cases for children living with their own parents. The few significant estimates are negligible in terms of size, and alternate in sign.

We show the reduced form – equation 2.3 run on the sample of children in state care – in Table 2.6. The coefficients on high foster care capacity on future outcomes are an order of magnitude larger than in the placebo, and all outcomes of children in state care are better on average in counties with high foster mother capacity. The estimates are significant at the 10 percent level for finishing secondary school (7 percentage points more finishes secondary school in high foster mother capacity counties, p=0.076), and for NEET months at age 19 (-0.78 months spent as NEET in counties with high foster mother capacity, p=0.054). The coefficient on the outcome index is also positive (0.136) and significant (p=0.016). We consider these estimates important in their own right. The reduced form shows that the outcomes of children in state care are better in places with high foster mother capacity, and the placebo confirms that it is not some county-level feature affecting the whole child population (like local labor market opportunities, or local school quality) which lies behind this difference. Thus keeping county features constant, a child protection system with more foster parents is better for children in state care.

Table 2.6: Regression of the outcomes on foster mother capacity on the sample of children in state care - reduced form

	(1) Secondary school	(2) Highschool	(3) Antidepr	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes
High fmother capacity	0.070^{+}	0.033	-0.033	-0.042^{+}	-0.011	-0.025	-0.001	-0.781^{+}	-0.060	0.136^{*}
	(0.039)	(0.026)	(0.029)	(0.025)	(0.058)	(0.054)	(0.049)	(0.411)	(0.041)	(0.056)
N	869	869	869	869	443	443	443	869	869	869
adj. R^2	0.148	0.289	0.016	-0.006	0.062	0.042	0.023	0.125	0.109	0.962

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hungarian DOI: 10.14754/CEU.2023.04

In order for the IV estimation to show the real causal effect of foster care, the exclusion restriction should also be satisfied: the correlation in the reduced form should be only caused by the higher share of children who live in foster care. And even though the placebo confirms that the instrument is conditionally uncorrelated with outcomes of the general population of children, a remaining issue with the IV is that there can be omitted county-level features of the child protection system which do not affect the outcomes of children who live with their own parents, but can confound the estimates for children who live in state care. It is not immediately clear how this problem biases our results. For example, one can imagine that the number of foster parents is negatively correlated with the quality of group homes because the local child protection system concentrates on recruiting foster parents rather than increasing group home quality. In this case, we would overestimate the beneficial effect of foster care. On the other hand, it is also possible that due to the smaller number of children in group homes, the shortage of childcare professionals in group homes is less of a problem in places with high foster parent capacity, which would make us underestimate the beneficial effects of foster care.

2.5.2.2 IV Results From the IV, we estimate that compliers are 47 percentage points more likely to finish secondary education (column 1) if they are raised in foster care. The estimate is large but noisy, only significant at the 10 percent level (p=0.083). On tranquilizer usage, we estimate 28 percentage point difference favoring foster care, and NEET months at age 19 are estimated to be smaller by 5.2 (p=0.065). The estimated effect on the index of the outcomes is 0.91 (p=0.022). For other outcomes, we also estimate a coefficient indicating a beneficial effect on foster care, but these estimates are not statistically significantly different from zero. For all estimates the Anderson Rubin Wald p values are reported as well, which are robust to the presence of weak instruments. For all of the statistically significant estimates mentioned above, the A-R p values are below 10 percent as well. However, the q values for false discovery rates are all above 20 percent (see Table B.6 in the appendix), indicating that there is a fair chance that we find some statistically significant estimates by chance.

To sum up, the reduced form from the IV strategy provides evidence that children in counties with higher foster mother capacity are better off in terms of education, mental health, and labor market outcomes. The final IV estimates show large gains from foster care. However, due to the small complier population, the noisy estimates, and the fact that the exclusion restriction is likely to be violated, we cannot pin down precise quantitative estimates for the average effect of foster care on children from this speciafication.

Table 2.7: Instrumental variable estimates for the effect of foster care on all outcomes

Outcome	(1) Secondary School	(2) High School	(3) Antidper	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes
Foster care	0.465^{+}	0.223	-0.221	-0.281^{+}	-0.069	-0.157	-0.006	-5.219^{+}	-0.403	0.910*
	(0.268)	(0.173)	(0.194)	(0.169)	(0.336)	(0.313)	(0.282)	(2.852)	(0.277)	(0.401)
Anderson Rubin Wald p	0.074	0.194	0.259	0.090	0.852	0.645	0.984	0.058	0.147	0.015
First stage CD Wald F	15.3	15.3	15.3	15.3	9.7	9.7	9.7	15.3	15.3	15.3
N	890	890	890	890	457	457	457	890	890	890

The following individual controls (measured at 6th grade) are included in all models: gender; age; number of siblings; calendar year indicators; standardized mathematics-, and reading testscores; grades in mathematics, literature, behavior and effort; indicators for the presence of mild mental disability, autism, physical disability, speech impairment, psychological problems, and behavior or learning problems. The following county-level controls are used: employment, unemployment, mean wage, birthrate, development. Robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Hungarian DOI: 10.14754/CEU.2023.04

Both the OLS and the IV are imperfect with untestable identifying assumptions boiling down to not omitting important individual-level variables in the case of the OLS, and not omitting important county-level features of the child protection system in the case of the IV. The OLS results are both large and statistically significant, the IV results are also large, but noisy. The two strategies pointing to the same direction support that it is not some omitted variable causing the large difference between youth raised in foster care and in residential care, but indeed children raised in foster care are better off because of the home type they were raised in as adolescents.

2.5.3 Possible mechanisms behind the main results

Delving into a comprehensive analysis of why being raised in foster care can offer greater benefits compared to residential care is beyond the scope of this chapter. Nonetheless, we briefly mention some of the possible mechanisms behind our large estimates.

Probably one of the most important channels that can explain the large gains from foster care is the larger opportunity for the child to form a trusting and caring relationship with a supportive adult figure in this type of care. These relationships can be important for children to successfully navigate in life, and having a loving adult figure is not only important for young children, but also for teenagers (Nagaoka et al., 2015) and even for young adults who already left the child protection system (Gypen et al., 2017). Even in well-functioning group homes with highly trained professionals, especially if the child-adult ratio is high, such relationships are less likely to form (UNICEF, 2022).

In addition, residential care can lead to worse outcomes through a number of other mechanisms. One is the possibility of negative peer effects. In the context of groups of deviant teenagers, the presence of negative peer influences is well established (Dodge et al., 2006). In residential care in Hungary, only a minority of children exhibit behavioral problems. Still, this ratio is higher than in foster homes, and also, the size of peer groups is larger, thus children are more likely to be affected by deviant peers. This can be an important mechanism behind their worse outcomes. Material circumstances in group homes are also worse than in foster homes. For example, the share of children having a stable internet connection, or owning a desk is lower for children in residential care in our data. This shows that the conditions to succeed in school are far from ideal in residential care.

In the following, we turn to estimating heterogeneous effects, to see what type of children can gain the most from foster care, and how the features of the foster home can affect the outcomes of the children.

2.6 Heterogeneous effects

2.6.1 Quality of foster care

In the following, we analyze the differences in the outcomes by home type, separately by the foster mother's education. Specifically, instead of a dummy indicating residential or foster care in equation 2.1, we include a dummy with three categories: living in residential care, living in foster care with a foster mother who has less than secondary education, and living in foster care with a foster mother who has at least secondary education. The education of the foster mother can be thought of as an indicator of foster care quality. Children with more educated foster mothers live in smaller households, have better material circumstances, and spend more time with their foster parents. [49] Table 2.8 shows the effects by the foster mother's education.

We present the results in Table 2.8. The first takeaway from the estimates is that the beneficial effect of foster care is present regardless of the foster mother's education: children growing up with low educated foster mothers perform better than children in residential care: they are significantly more likely to finish secondary school (8.2 percentage points), spend 1.12 fewer months as NEETS, less likely to use tranquilizers (-4.8 percentage points), and have fewer pregnancies (-13.5 percentage points) and abortions (-15.1 percentage points). The estimate on the index of the outcomes is also

[49] For 35 children we do not observe the education of the foster mother, we exclude them from this analysis

positive and statistically significant (0.210). The effect on the other outcomes is insignificant but also in the direction showing gains from foster care.

The other takeaway is that the beneficial effects seem even more positive for children living with foster mothers with at least secondary education. The point estimates for the beneficial effect of foster care are larger for most outcomes. But, as Table B.7 in the appendix shows, the difference by foster mother's education is only significant in the case of teenage births. Girls living with highly educated foster mothers are 21 percentage points less likely to become teenage mothers than girls in residential care, and 15 percentage points less likely than girls living with low-educated foster mothers.

Thus, the quality of foster care seems to matter, but even foster parents with low human capital seem to yield better outcomes for the children compared to residential care. Other than the features of foster care itself, it is likely that specific groups of children are affected differently by the two types of care. In the following, we analyze differential effects by gender and special education needs.

Table 2.8: Regression estimates on the effect of foster care by foster mother's education

	(1) Secondary school	(2) Highschool	(3) NEET months	(4) NEET 6m	(5) Antidepr	(6) Tranquil	(7) Pregnancy	(8) Birth	(9) Abortion	(10) Index of outcomes
Low Educ Foster M.	0.082*	0.013	-1.123**	-0.123**	-0.037	-0.048^{+}	-0.135*	-0.070	-0.151**	0.210***
	(0.039)	(0.023)	(0.424)	(0.043)	(0.031)	(0.025)	(0.067)	(0.063)	(0.058)	(0.058)
High Educ Foster M.	0.094*	0.021	-1.263*	-0.113*	-0.045	-0.059*	-0.234**	-0.208***	-0.122*	0.249***
	(0.048)	(0.030)	(0.490)	(0.049)	(0.033)	(0.029)	(0.072)	(0.063)	(0.060)	(0.068)
N	855	855	855	855	855	855	438	438	438	855
adj. R^2	0.158	0.298	0.127	0.113	0.027	-0.001	0.109	0.109	0.049	0.172

Robust standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hungarian DOI: 10.14754/CEU.2023.04

2.6.2 Heterogeneous effects for subgroups of children

In this subsection, we re-estimate equation 2.1 with the inclusion of interaction terms to see whether specific groups of children benefit more from foster care. First, we estimate heterogeneous effects by gender. Table 2.9 shows that the gains from foster care seem to be smaller for girls for all outcomes. While these differential effects are not significant for any outcome, they are large for finishing secondary school and mental health outcomes.

Table 2.9: Regression estimates on the effect of foster care by gender

	(1) Secondary school	(2) Highschool	(3) Antidepr	(4) Tranquil	(5) NEET months	(6) NEET 6m	(7) Index of outcomes
Foster care	0.126*	0.005	-0.065	-0.065+	-1.146*	-0.119*	0.238**
	(0.050)	(0.028)	(0.042)	(0.035)	(0.534)	(0.056)	(0.080)
Foster care X Girl	-0.084	0.022	0.053	0.032	0.075	0.010	-0.044
	(0.066)	(0.040)	(0.054)	(0.047)	(0.727)	(0.074)	(0.102)
N	890	890	890	890	890	890	890
adj. R^2	0.157	0.299	0.034	0.007	0.137	0.122	0.175

Note: The following controls (measured at 6^{th} grade) are included in all models: gender; age; number of siblings; county indicators; calendar year indicators; standardized mathematics-, and reading test scores; grades in mathematics, literature, behavior and effort; indicators for the presence of mild mental disability, autism, physical disability, speech impairment, psychological problems, and behavior or learning problems. Robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

In Table 2.10 we show heterogeneities by the presence of special education needs. The coefficient of the interaction term is alternating in sign, and close to 0 in most cases, except for outcomes related to fertility.

Table 2.10: The effect of foster care by special needs

	(1) Secondary school	(2) Highschool	(3) Antidepr	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes
Foster	0.089*	0.007	-0.038	-0.048+	-0.141*	-0.132*	-0.082	-1.206**	-0.124**	0.228***
	(0.040)	(0.025)	(0.031)	(0.027)	(0.068)	(0.064)	(0.059)	(0.432)	(0.044)	(0.060)
Foster X Special needs	-0.021	0.041	0.002	-0.004	-0.122	0.010	-0.172	0.446	0.044	-0.056
	(0.074)	(0.040)	(0.068)	(0.055)	(0.122)	(0.120)	(0.107)	(0.833)	(0.088)	(0.118)
N	890	890	890	890	457	457	457	890	890	890
adj. R^2	0.151	0.299	0.035	0.009	0.082	0.079	0.039	0.140	0.124	0.179

Note: The following controls (measured at 6th grade) are included in all models: gender; age; number of siblings; county indicators; calendar year indicators; standardized mathematics-, and reading test scores; grades in mathematics, literature, behavior and effort; indicators for the presence of mild mental disability, autism, physical disability, speech impairment, psychological problems, and behavior or learning problems. Robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Hungarian DOI: 10.14754/CEU.2023.04

These results suggest that regardless of gender and the presence of special needs, children can benefit from being raised in family foster care. The beneficial effect seems to be a bit larger for boys, but the results are not significant.

2.7 Robustness checks

In this section, we show that the main OLS results are robust to a number of modifications. In our first robustness check, we refine our sample to satisfy the overlap of the control variables. Specifically, we restrict the sample to those children, who based on their observed features in 6^{th} grade would have a positive probability to be placed either into residential care or foster care. To achieve this, we estimate the following probit model:

$$P(foster = 1|X'_{i,gr6}) = \Phi(X'_{i,gr6}\beta), \tag{2.4}$$

where $X'_{i,gr6}$ contains the same variables as in equation 2.1: gender, age, number of siblings, county indicators, calendar year indicators, measures of cognitive ability - standardized mathematics, and reading test scores, grades in mathematics, literature, and grammar -, and measures of non-cognitive ability: grades for behavior and effort in school, and the presence of the following special education needs: mild mental disability, autism, physical disability, speech impairment, psychological problems, and behavior or learning problems. The distribution of the predicted propensities by placement type are shown in Figure 2.3. Based on the figure, we drop observations with a propensity score over 0.9 and below 0.3. This way we only include children, who based on their observed individual-level characteristics and county where they live, could be placed into either type of care.

Figure 2.3: Overlap of the control variables

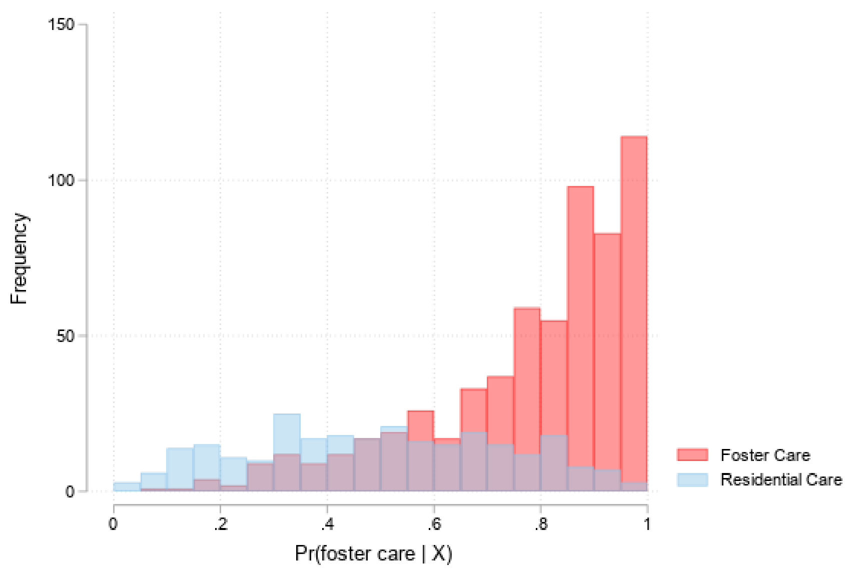


Table 2.11 shows the results of re-estimating equation 2.1 with a full set of controls on this restricted sample. The estimates are similar in magnitude to our main OLS results on the whole sample, and they point in the same direction. For some variables, the point estimates are even larger (though not statistically significant). For example, in this subsample children in foster care are 4.6 percentage points more likely to graduate from high school. This robustness check shows that the gains from foster care are present in the group of children, who, based on their observable features could be placed into either type of care.

In our second robustness check we use a larger sample than in our main results, and include every child of whom we observe home type in 6^{th} grade, and estimate the effect of care type in grade 6. The worry with the main results is that by excluding children who are in foster care in grade 6, and in group homes later we exclude problematic foster children, and thus overestimate the beneficial effect of foster care. Table 2.12 confirms that this effect is not driving our main estimates, as the results are similar in magnitude in this sample too.

In our final robustness check, we add an indicator of early ability, specifically delayed school start, as an additional control variable. This inclusion aims to provide evidence that our results are not driven by early differences in ability. The estimates are again qualitatively the same as our main results (see Table 2.13)

Table 2.11: Regression estimates on the effect of foster care on the subsample where overlap is satisfied

	(1) Secondary school	(2) Highschool	(3) Antidepr	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes
Foster care	0.100*	0.046*	-0.016	-0.038	-0.134+	-0.069	-0.106+	-1.004*	-0.101*	0.198***
	(0.041)	(0.022)	(0.032)	(0.027)	(0.072)	(0.067)	(0.059)	(0.433)	(0.044)	(0.060)
N	595	595	595	595	299	299	299	595	595	595
adj. R^2	0.100	0.275	0.064	0.026	0.124	0.104	0.043	0.160	0.144	0.157

Note: The following controls (measured at 6^{th} grade) are included in all models: gender; age; number of siblings; county indicators; calendar year indicators; standardized mathematics-, and reading test scores; grades in mathematics, literature, behavior and effort; indicators for the presence of mild mental disability, autism, physical disability, speech impairment, psychological problems, and behavior or learning problems. Robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table 2.12: Regression estimates on the effect of foster care on sample in foster care in 6th grade

	(1) Secondary school	(2) Highschool	(3) Antidepr	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes
Foster	0.080*	0.024	-0.019	-0.033	-0.188**	-0.144**	-0.123**	-1.106**	-0.108**	0.196***
	(0.032)	(0.019)	(0.025)	(0.021)	(0.057)	(0.053)	(0.047)	(0.357)	(0.037)	(0.050)
N	1061	1061	1061	1061	540	540	540	1061	1061	1061
adj. R^2	0.141	0.291	0.034	0.030	0.088	0.097	0.043	0.126	0.113	0.165

Note: The following controls (measured at 6^{th} grade) are included in all models: gender; age; number of siblings; county indicators; calendar year indicators; standardized mathematics-, and reading test scores; grades in mathematics, literature, behavior and effort; indicators for the presence of mild mental disability, autism, physical disability, speech impairment, psychological problems, and behavior or learning problems. Robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Hungarian DOI: 10.14754/CEU.2023.04

Table 2.13: Regression estimates for the effect of foster care on all outcomes with the inclusion of delayed school start

Outcome	(1) Secondary School	(2) High School	(3) Antidepr	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of out-comes
Foster care	0.067^{+}	0.012	-0.035	-0.053^{*}	-0.138^{*}	-0.102^{+}	-0.123^{*}	-0.949^{*}	-0.098^{*}	0.186^{**}
	(0.038)	(0.022)	(0.030)	(0.026)	(0.065)	(0.059)	(0.056)	(0.415)	(0.042)	(0.057)
adj. R^2	0.153	0.307	0.029	-0.002	0.104	0.114	0.053	0.141	0.126	0.179
N	840	840	840	840	431	431	431	840	840	840

Note: The following controls (measured at 6$^{\text{th}}$ grade) are included in all models: gender; age; number of siblings; county indicators; calendar year indicators; indicator for delayed school start, standardized mathematics-, and reading test scores; grades in mathematics, literature, behavior and effort; indicators for the presence of mild mental disability, autism, physical disability, speech impairment, psychological problems, and behavior or learning problems. Robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Hungarian DOI: 10.14754/CEU.2023.04

2.8 Conclusion

In this chapter, we compare the outcomes of young adults who spent time in foster care during their adolescence with those who were placed in residential care in Hungary. We make use of individual administrative panel data linking information of indicators of ability at around age 13 and a large set of adult outcomes. We present evidence that young adults raised in foster care experience significantly better outcomes compared to similar youth raised in residential care, conditional on indicators of cognitive ability (for example test scores) and non-cognitive ability (for example behavioral problems). They are substantially more likely to complete secondary education, less likely to use medication for mental health issues, less likely to become pregnant as teenagers and spend less time without either employment or enrollment in education.

A limitation in our analysis is that we first observe children in grade 6, and we do not have information on previous placement history, and ethnicity. We use an IV strategy to eliminate this problem, instrumenting foster placement by local foster mother capacity. The results point in the same direction as the OLS results but are not statistically significant for any of the outcomes when we correct for multiple testing.

Prior research has shown the beneficial effects of foster care compared to residential care, for children who were placed in state care before the age of 3 (Nelson, 2014). Existing research on older children suggests that foster care is likely to have a positive impact on children (see Lee et al., 2011). However, these findings are valid for the group of children with severe behavioral problems and pertain to only a few specific outcomes. Our study provides the first evidence supporting the positive effect of foster care on a wide range of adult outcomes for adolescents raised in such settings.

In English-speaking countries residential care is only reserved as a last resort for children whose placement into foster families is not possible (del Valle, 2014). But in most other countries, including Hungary, a large number of children who would be fit to live with families, live in residential care. Child welfare organizations have long called for the de-institutionalization of these children (see e.g. UNICEF, 2022). Our study shows that living in a family-like environment during adolescence can not only be thought of as a normative principle but also as a means to improve outcomes. Efforts to increase the number of foster parents to be able to place more children in foster care could be important policies in those countries, where residential care is still common (for example Germany, France, the Benelux countries, and the Mediterranean countries in Europe). In these countries, the effect of foster care might be different because of the local details of the child protection system - for example, the quality of both residential and foster care can be substantially different. Still, the fact that we find better outcomes for foster youth regardless of gender, special needs, and the quality of foster care, suggests that gains from living in a family-like environment might be quite universal.

Future research in different child protection settings can shed more light on what features of children and foster families make it more likely that children raised in state care grow up to be successful adults. To pin down precise causal effects, randomized controlled trials in the nature of the Bucharest Early Intervention Project could provide high-quality evidence about the effectiveness of different care types for different subgroups of children. Finally, it is important to acknowledge that while foster care has the potential to improve the well-being of children living in state care, even foster youth has much worse outcomes than the general population of children. More research is needed to gain a better understanding of how governments can improve the overall quality of state care in order to raise successful individuals when their birth families fail them.

3 Chapter 3: The Effect of Physical Education Time on Students' Body Composition

3.1 Introduction

Childhood obesity is a growing problem worldwide. The prevalence of obesity among children aged 10-19 in European countries increased from around 2 percent in the 1980s to 7 percent in 2016. This is a large concern, as obesity is linked to numerous health problems (Bray, 2004) as well as adverse economic outcomes (Cawley, 2015). It is also well-established that only around one in five obese children can outgrow obesity as an adult (Inchley et al., 2017). Thus, finding ways to reduce obesity among children is an increasingly important policy goal.

One natural policy solution to reduce childhood obesity is to increase the physical activity of children through the public school system. Multiple policies attempted to do just this in the US and Europe, but the literature analyzing the effect of increased physical education (PE) time on the obesity of children finds that the effect of these policies is statistically indistinguishable from zero (see for example Cawley et al., 2007, Knaus et al., 2020 Taber et al., 2013 and Sabia et al., 2017), or at best positive for some subgroups of students (Cawley et al., 2013 find an effect for 5$^{\text{th}}$ grader boys).

However, the studies establishing the ineffectiveness of PE time share a common measurement problem, as they almost exclusively rely on body mass index (BMI) as an outcome. The calculation of BMI includes body weight and height and does not take into account the composition of body mass. This may cause a systematic underestimation of the effectiveness of policies aiming to increase physical activity, as moving more can reduce body weight by burning fat, but it can also increase it, by building muscles. Thus, even when children lose fat mass as a result of increased physical activity, it is possible that average BMI increases or stays unchanged if they also build muscle mass.

A superior health indicator is body fat percentage (BF%) (see Burkhauser and Cawley, 2008), which directly shows excess fat accumulation. However, while BMI data is easily collected through surveys by asking for the weight and height of respondents, collecting data on BF% requires specialized equipment. Only a handful of small sample studies (see the reviews of Harris et al., 2009, Lavelle et al., 2012 and Mura et al., 2015) could use BF% as a health outcome to evaluate interventions increasing physical activity. The results in these studies are mixed, with zero, negative, and positive estimated effects as well, and the small sample sizes and specialized nature of the interventions make the external validity of these results limited. Hence, the impact of PE time on students' BF% remains uncertain.

In my analysis, I pose the question of how PE time affects students' body composition. I can use a unique database (called NETFIT) on children's fitness, containing information on BF% and BMI as well. The data is collected from the entire population of Hungarian students in grades 5 to 12 between the end of the 2013/2014 and 2018/2019 school years. The measurement of BF%, height, and weight is done by PE teachers with standardized equipment provided free of charge to every school. The data are published aggregated to age and gender groups for each school. I use these aggregated data with the following outcomes: mean body fat percentage, mean BMI, and two measures for the share of obese students in the given age-gender-school group (one based on BF% and one on BMI thresholds).

To uncover the effect of PE time, I make use of a 2012 Hungarian policy that introduced daily PE classes in every school. Prior to the policy, students typically had 2 or 3 PE classes per week, thus PE time increased substantially, by 72-92 hours per year, depending on grade. The identification of the effect is possible as daily PE was introduced in a phasing-in system for every 1^{st}, 5^{th} and 9^{th} grader. This created a large variation in the number of PE classes between three cohort pairs that are plausibly similar in other aspects. Specifically, the cohorts starting first grade in 2012, 2008 and 2004[50] had 4 more years of daily PE classes than the cohorts of 2011, 2007 and 2003 respectively[51]. This translates to large differences in PE hours, for example, 2012/2013 9^{th} graders spent around 270 more hours with PE lessons by the time they finished high school than 2012/2013 10^{th} graders.

In my empirical strategy, I estimate OLS models of the body composition outcomes on past PE time controlling for gender, age, and school and year fixed effects. The staggered introduction of the policy means that for different age groups, PE time jumps in different calendar years, making it possible to include year fixed effects in the regressions. Accounting for year fixed effects in establishing counterfactual outcomes are important, as obesity is expected to grow in the absence of the policy. The identification assumption is that without the increase in PE time, students' body composition would be the same after controlling for age, gender, time, and school. A threat to this assumption is that other curriculum changes were introduced in the same phasing-in system as the daily PE classes policy. I argue that these changes do not plausibly have a large effect on body composition. In addition, I run a placebo test using data on students from a town where daily PE classes were already in place, to show that these other curriculum changes did not increase BF%, BMI, or the prevalence of obesity.

By using past PE time as an explanatory variable, my analysis assumes that PE time not only impacts the current body composition of students but also has a lasting effect. This assumption can be interpreted as the notion that time spent with PE accumulates as health capital and does not depreciate over time. In a robustness check, I relax this assumption and allow for the depreciation of health capital, but the results from these regressions are also indistinguishable from zero.

The regression results show that increased PE time did not have a statistically significant effect on any of the outcomes. The estimated effect of 100 additional PE hours on mean body fat percentage is 0.017. This is statistically insignificant (p=0.274) and close to 0. Compared to the mean body fat percentage of 21.93 in the sample, it is only a 0.07 percent increase and a larger than 0.05 percent decrease as a response to 100 additional PE hours can be ruled out with 95% confidence. The effect on average BMI is also negligible and statistically zero (p=0.488).

The estimated effect on the prevalence of obesity is a precise zero, too, regardless of whether it is measured based on a BF% threshold or a BMI threshold. The estimated effect on obesity defined by BF% is 0.004 percentage points (p=0.939), and a larger than 0.09 percentage point decrease in obesity can be ruled out by 95% confidence.

While increased PE time seems ineffective in reducing obesity for the general population of students, it is possible that there are some specific groups who benefit more from it. I check heterogeneous effects by age, gender, and socio-economic status of the students, and the number of gym rooms per one hundred students in the school, but I find no substantial effect on body composition for any of the subgroups.

These results contribute to the literature on the effect of PE time on obesity by using a new outcome, BF%, in the context of large-scale policy. I show that PE time is not only ineffective in reducing BMI but also in reducing BF%. This suggests that the lack of significant findings in previous literature is not only due to the limitations of BMI as a measure of obesity but also because increased physical education time alone does not actually have an impact on child obesity. To stop or reverse the increase in childhood obesity, governments need to work on more complex

[50] 1^{st}, 5^{th} and 9^{th} graders in 2012
[51] 2^{nd}, 6^{th} and 10^{th} graders in 2012

interventions.

The rest of this chapter is organized as follows. In Section 3.2 I review the literature, and in Section 3.3 I describe the Hungarian context along with the details of the 2012 daily PE classes policy. Section 3.4 presents the data; and some descriptive results about the body composition outcomes and their relationship. The empirical strategy and the results are presented in Sections 3.5 and 3.6. Robustness checks are described in Section 3.7 and I conclude in Section 3.8.

3.2 Literature Review

3.2.1 Why policies targeting increased physical activity can be important

It is well established in the literature that obesity has negative consequences on physical health, including increased risk of type 2 diabetes, hypertension, sleep apnea, cancer, and cardiovascular diseases, leading to reduced life expectancy of obese people (Bray, 2004). Moreover, obesity has been found to have adverse effects on emotional well-being and self-reported quality of life(Inchley et al., 2017), as well as various economic outcomes including wages, employment, and medical care costs (Cawley, 2015). Thus, it is definitely a large concern for societies that obesity has been growing steadily in the last decades. Along with the general increase in obesity, child obesity has been on the rise as well, especially in developed countries (Abarca-Gómez et al., 2017). Keeping in mind the negative consequences of obesity and the fact that 4 in 5 obese children stay obese as an adult, it is not surprising that the WHO calls childhood obesity one of the most serious public health challenges of the 21st century (Inchley et al., 2017).

It is less clear though what are the exact causes of the increasing rates of obesity. The literature acknowledges many possible factors, with a particular emphasis on reduced physical activity, food marketing practices, and technology (Keith et al., 2006). According to Keith et al. (2006) the excessive focus on these two primary causes – physical activity and food –, often referred to as the "Big Two," may divert attention from other key contributors to obesity [52], but they acknowledge the importance of these two factors as well.

It is not clear how big of a role the reduction in physical activity played in the increasing obesity in recent decades. Some studies suggest that reduction in physical activity played little role in people becoming more obese in developed countries: e.g. Cutler et al. (2003) argue that increased calorie intake caused by the technological shift in food production is the driver of the rise in obesity, and decreased energy expenditure is plausibly not an important factor. However, studies looking at specific activities like walking (Bassett et al., 2010) or employment-related activities (Church et al., 2011) find that the reduction in these activities can explain a substantial amount in rising obesity. The review of Dollman et al. (2005) presents some evidence that not only adults' but also children's lifestyles became more sedentary in the last decades. Though survey evidence shows no decrease in vigorous physical activity of children, and no consistent pattern in participation in organized sport and PE in schools, physical activity related to transportation (like walking or cycling) decreased substantially, while time spent with sedentary activities (watching TV and using the computer) increased. Still, it is generally hard to uncover time trends in physical activity as

[52] These include for example increased sleep deprivation, increased exposure to endocrine disruptors, reduced exposure to extreme cold and heat, decrease in smoking, increase in the usage of pharmaceuticals increasing weight, increasing age of mothers when the child is born, the higher reproduction rate of people genetically more prone to obesity. Other possible economic causes of obesity collected by Cawley (2015) are the decreasing relative price of energy-dense foods, increased income (which is relevant in terms of obesity only for low-income individuals), and peer and neighborhood effects (on the weight of women)

consistent measurements of time spent physically active are not available for long time periods. Therefore, the extent to which the increasing obesity rates can be attributed to the decline in physical activity remains unclear.

But, even if reduced physical activity is not an important cause behind the increase in child obesity in recent decades, policies aiming to increase energy expenditure among children might prove to be successful in reversing the trend. Hill et al. (2012) argue that it is much easier for people to avoid obesity, with a high physical activity and high calorie intake lifestyle than with a sedentary lifestyle combined with low calorie intake. Furthermore, they propose that increasing energy expenditure might be a necessary condition for policies targeting decreased calorie intake to work.

Thus, policies aiming at increasing energy expenditure are possibly promising attempts to decrease obesity in the population. In addition, targeting such policies toward children has the potential to reduce the probability of becoming obese early. As children rarely outgrow obesity, and preventing obesity is easier than losing weight after somebody is already obese (Hill et al., 2012), early interventions can prove to be especially beneficial.

As children spend much of their time in school, increasing the time of physical education classes to make children move more is a promising candidate policy to reduce obesity. However, research analyzing the effect of increased PE time on child obesity usually fails to detect a substantial and statistically significant effect. In the following, I summarize the findings of this literature.

3.2.2 What do we know about the effectiveness of policies increasing PE time

A number of papers analyzed the effect of in-school physical education requirements in the US on body weight. One set of papers focuses on high schools, and none of them finds a significant effect on the prevalence of obesity. The studies of Cawley et al. (2007) and J. Kim (2012) rely on cross-state variation in PE time requirements and find no significant effect on obesity and body weight for high-school students. Sabia et al. (2017) states that the above-mentioned studies might face an endogeneity problem, and they tackle it by using in-state variation in laws regulating PE time. Still, they find 0 effect of increased PE time as well. Packham and Street (2019) also identifies the effect of a large-scale policy targeting low-income schools in Texas. They find no effect of increased PE time on the average BMI of students, but they detect a small decrease in the prevalence of obesity among students in schools around the eligibility cutoff of the policy. These studies (along with the paper of Perna et al. (2012)) find that the state laws about PE have indeed resulted in a substantial increase in the amount of time dedicated to in-school physical activity. However, it is inconclusive whether these policies increased overall (inside and outside school) physical activity: Cawley et al. (2007) finds an overall increase in physical activity only for girls while Sabia et al. (2017) only for boys.

For 5^{th} grader boys in the US, Cawley et al. (2013) find a reduction in BMI as a response to increased PE time, but no such effect for girls, while Taber et al. 2013 finds no evidence that PE laws reduce student weight gain.

Outside the US, Knaus et al. (2020) analyzes data on German students and uses cross-state variation in PE requirements to instrument the actual number of PE classes a student has for identification of the effect of PE time on a number of outcomes. While they find effects for many outcomes (e.g., grades, motor skills, and non-cognitive skills), they find no effect on health-related factors such as BMI and obesity. They report that time spent with physical activity increases as a response to increased PE time for girls but not for boys. Knaus et al. (2020) checks the heterogeneity of the effect by household income of students and they find no relevant difference in the effects on health.

Beyond the literature above analyzing the effect of increased PE time in regular schools as a result of widespread country or state policies, a large body of literature evaluates the effect of smaller local (school- or town-based) interventions. Eighteen of such interventions are included in the meta-analysis of Harris et al. (2009) and none of these studies find an effect on BMI. Three of these studies measure body fat percentage as well, with one of them finding a 0, one a positive, and one a negative effect. The review of Lavelle et al. (2012) finds a small average positive effect on BMI of school-based interventions (a mean of 0.13 points reduction in BMI is estimated as a result of increased school-based physical activity). Mura et al. (2015) reviews 47 papers analyzing the effect of physical activity interventions, among which only 3 found a positive effect on BMI, while more found an effect on body composition and fitness outcomes. The large majority of these 47 interventions were multi-component, focusing not only on increasing physical activity but also on increasing healthy dietary habits, so their effect cannot be attributed to increased physical activity on its own.

The bottom line of this literature is that it seems like that increasing PE time does not have an effect distinguishable from zero on the obesity of students in general. When studies do find an effect, they detect it for only a subgroup of children: Cawley et al. (2013) find an effect for 5^{th} grade boys, Packham and Street (2019) measure a LATE for students attending low-income schools around the cutoff of the policy. The vast majority of studies on small-scale interventions do not detect an effect on the BMI of students. The small number of small-scale interventions that have achieved small effects on reducing child obesity may have limited external validity. These interventions often analyze small, non-representative samples of students and schools, and are likely implemented under more controlled conditions with highly trained teachers than typical large-scale policies. In addition, the small number of successful programs differ substantially in terms of implementation and content, so it is not clear which element of them makes them more effective (Mura et al., 2015).

There are many possible reasons that can explain the general ineffectiveness of policies aiming to increase PE time. For example, Knaus et al. 2020 propose that the variation in PE class time they use (45 minutes weekly) might not be enough to induce measurable effects on obesity. Furthermore, even if students spend substantially more time with physical activity in school, it is possible that they simultaneously decrease out-of-school physical activity. Indeed, studies examining this have found evidence suggesting that for a certain group of students[53], in-school physical activity and out-of-school physical activity may act as substitutes. And, even if substitution is not an important issue, and students do spend more time with physical activity when they have more PE lessons, they might simply eat more as well, leaving their net calorie consumption and thus BMI unchanged.

Another, possibly less important channel through which the effect of increased PE time can be small is proposed by Packham and Street (2019) who find that increased PE class time increased disciplinary problems among children, which can be an indication of more bullying. Children bullied during PE classes might become more stressed, less healthy, and more obese, which can offset (some of) the positive effect of increased physical activity time on health.

Other than these explanations, it can also be the case that increased PE time *does* have an effect on the body composition and obesity of the students, but the above-mentioned studies are unable to detect it because of a measurement problem. The literature establishing the zero effect

[53]Cawley et al. (2013) and Sabia et al. (2017) find these are the girls, while the results of Cawley et al. (2007) indicate these are the boys, while Knaus et al. (2020) finds no evidence for such substitution

of these policies almost exclusively relies on BMI, usually calculated from self-reported weight and height (with the exception of 3 small school-based interventions with inconclusive outcomes).[54] One small potential problem with this is that it is well documented that people under-report their weight and that heavier people under-report it more (see for example Ng, 2019). As the main outcome is measured noisily, and especially for obese individuals, it might be more difficult to detect the effect of PE policies statistically significantly even if such an effect exists. Another, and possibly a more important measurement problem (acknowledged by more of the authors of the above-mentioned papers as well) is that interventions aiming to increase physical activity can reduce fat mass and increase muscle mass at the same time. Thus, these policies can increase the BMI of some individuals, or leave their BMI unchanged while making their body composition healthier. Because of this measurement problem, the literature analyzing the effect of interventions increasing PE time might systematically underestimate the real potential of these policies in changing the body composition and the prevalence of obesity among students.

I contribute to this literature by using body fat percentage along with BMI in assessing the effect of a country-wide intervention increasing PE time for Hungarian school children. As body weight and body fat percentage are measured by PE teachers with standard equipment, the noise coming from self-reported weight is not an issue in my analysis. In addition, the increase in PE time during the 2012 Hungarian policy analyzed here was quite large – 2 classes (90 minutes) per week for kids below 5$^{\text{th}}$ grade and 3 classes (135 minutes) from 5$^{\text{th}}$ to 12$^{\text{th}}$ graders – making it more likely to detect an effect of increased PE time if it exists.

3.3 Hungarian context and policy details

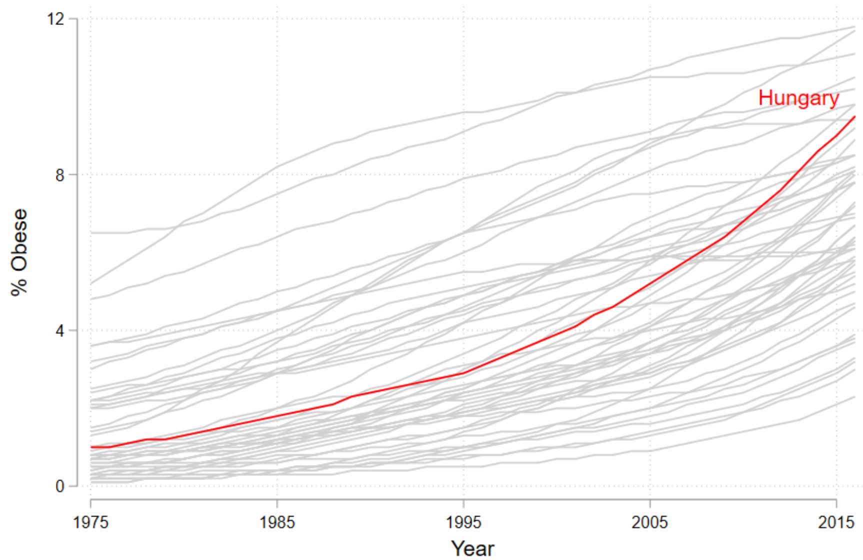
In this section, I present general information about child obesity and PE classes in Hungary, to provide context for the results. Then, I describe the implementation details of the 2012 policy increasing PE time.

3.3.1 Childhood obesity and PE classes in Hungary in the European context

Obesity among children has been steadily increasing in Hungary since the seventies. According to WHO data, 27 percent of 10-19-year-old Hungarian children were overweight[55] in 2016, and 9 percent of children were obese.[56] Similarly to other post-socialist countries, the prevalence of childhood obesity in Hungary increased more rapidly since the 1990s than in Western Europe, making Hungary one of the most obese countries in Europe by 2016 (see Figure 3.1).

[54] Packham and Street (2019) also have data on body fat percentage, but they do not present results based on that measure.
[55] BMI at least 1 standard deviation above the median.
[56] BMI at least 2 standard deviations above the median.

Figure 3.1: Obesity among 10-19-year-old children in Hungary and in other European countries



Source: World Health Organization, Global Health Observatory Indicators (https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-obesity-among-children-and-adolescents-bmi-2-standard-deviations-above-the-median-(crude-estimate)). Children are categorized as obese if they are at least 2 standard deviations above the age-specific median BMI

Multiple policies were implemented in the 2010s in Hungary reacting to this problem, with the majority of them focusing on achieving healthier eating habits among children.[57] In addition to these policies, a 2012 nation-wide reform targeted the physical activity of children, introducing daily PE classes in public schools.

In Hungary, a PE class lasts for 45 minutes with activities such as athletics, gymnastics, and games. In primary schools below 5th grade, generalist teachers are allowed to hold PE lessons while from 5th grade, only teachers specialized in PE can teach. Before the 2012 policy, elementary school children up to grade 4 had at least 3 physical education classes per week, while older students typically had 2 or 3 classes.[58] The amount of time allocated to PE (both in absolute terms and relative to the overall instructional time) prior to 2012, along with the curriculum and the required

[57] The most important of these policies – according to the National Institute of Pharmacy and Nutrition in Hungary – are the following (link, downloaded: 11.12.2020): a decree providing free fruit for some schools (50/2012. (V. 25.) VM rendelet); a decree forbidding selling unhealthy food inside schools (20/2012. (VIII.31.) EMMI rendelet), a decree regulating a maximum amount of trans-fats in foods (71/2013. (XI. 20.) EMMI rendelet); a decree making school mass catering healthier by regulating the quality of foods; and the amount of nutrients provided in 10-day windows (37/2014. (IV.30.) EMMI rendelet)

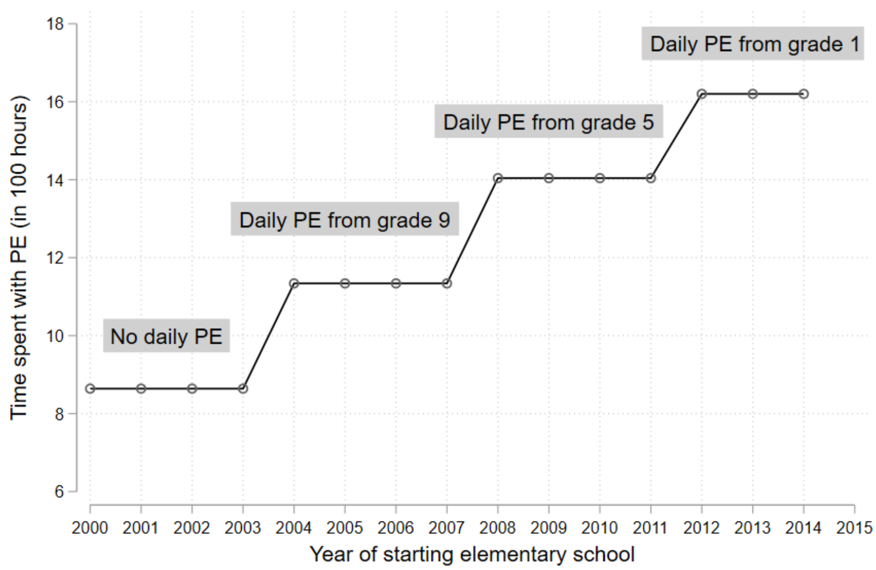
[58] In some cases, the number of PE classes was already higher: in Hódmezővásárhely, a Hungarian town, daily PE classes were introduced already in 2005. In addition, a few specialized schools in Budapest already had daily PE classes before 2012.

teacher training, can be considered fairly standard in the European context (see Buhovac, 2018). This makes the Hungarian experience of PE class time increase relevant for policymakers in other European countries considering implementing similar policies.

3.3.2 Details and implementation of the 2012 daily PE classes policy

The law introducing daily PE classes in September 2012 was passed in December 2011, leaving schools about 6 months to prepare for the change. The extra PE classes were added on top of existing classes, thus the overall time children have to spend at school increased as well. To make the transition smooth, daily PE classes were introduced only in 1st, 5th, and 9th grades in a phasing-in system in the 2012/2013 school year. Figure 3.2 shows the number of hours different cohorts will have spent with PE classes if they stayed in school for 12 years. The cohorts which started elementary school in or before 2003, were not affected by the daily PE classes policy and spent 864 hours with PE classes by the time they left secondary school. The cohorts starting elementary school in 2012 or later have daily PE classes in each of their school years, meaning they will have spent 1620 hours with PE when they leave secondary school. The cohorts in-between were partially affected by the policy: those students who started 1st grade between 2008 and 2011 started to have daily PE in 5th grade, and 2004-2007 1st graders started to have daily PE in 9th grade.

Figure 3.2: Total time spent with PE upon leaving secondary school after 12 years for different cohorts



The typical Hungarian student has indeed attended daily PE classes since the policy was implemented. However, under special conditions, students are not required to be present at every PE class. If they can verify out-of-school physical activity in a sports organization, or in-school elective physical activity, parents can ask for an exemption from 2 of the 5 PE classes weekly.

Survey evidence suggests that only a small minority of students use these exemptions: a sample of around a thousand students from grades 5th to 8th (Fintor, 2017) shows that 91 percent of all students attended all of the five weekly PE classes. These students were from the Northern Great Plain region, which is the 2nd poorest of the 7 regions of Hungary. While these results may not generalize to children living in richer regions or to older children, they are indicative of the fact that the majority of children probably do not use the exemptions to skip the extra PE classes.

Upon introduction, the policy was criticized for the lack of sufficient suitable gym rooms for such a large number of PE classes, especially in city schools where space is scarcer. In these schools, PE classes have to be held on the corridors and stairs of the schools, or outside (in a yard or on the street), which can be sub-optimal for effective workout. Indeed, in the 2015/2016 school year – the first year when every student had to have daily PE classes – the number of gym rooms per one hundred students was below 0.5 in 70 percent of the schools and was 0 in 17 percent of the schools.[59] In the heterogeneity analysis, I estimate the effect of PE time for schools with and without enough gym rooms separately to see if the lack of gym rooms is an important determinant of the effectiveness of the policy.

One of the main costs of the policy is the additional wages that need to be paid to accommodate the increased number of PE classes. To estimate this cost, I use the number of students in 2014/2015[60] (395344 in grades 1-4 and 679488 in 5-12 grades), and an average class size of 30, which translates into weekly around 33000 PE classes in grades 1-4, and extra 58000 classes in grades 5-12. A full-time teacher can teach 26 hours a week, meaning that as a result of the policy, the wage of around 3900 extra PE teachers (around 4 percent of the whole body of teachers) has to be financed by the central budget yearly. As the number of PE teachers did not increase according to official statistics[61] in this period, in practice, schools probably solved the issue of extra PE classes not with new hires, but with overtime work, or reorganizing the classes among current staff members. Still, the overall wage costs should be similar. Another possible cost could be the infrastructural developments needed to accommodate the new PE classes. However, based on official statistics, the number of gym rooms did not increase in this period, suggesting that there were typically no investments in extra infrastructure in schools as a response to the policy. Considering only the increased wage costs, I estimate that the annual cost of this policy is 9.6 - 16.7 billion HUF annually from the 2015/2016 school year.[62] Compared to other interventions (for example, increasing taxes on unhealthy food), this measure can be considered relatively expensive.

To sum up, the daily PE classes policy roughly doubled PE time in Hungary, increasing it from an average level to the highest in Europe. While infrastructural problems posed challenges for holding daily PE classes in some schools, survey evidence shows that the vast majority of children do attend PE classes daily as a result of the policy. Two years after the introduction of the policy, the Hungarian School Sports Federation started to collect comprehensive data on students' fitness, opening the opportunity to analyze the effect of increased PE time.

[59] With classes around the size of 33, and 6 classes a day, a school can organize gym classes in gym rooms during the day for every kid if it has one gym room per 200 students.

[60] The number of students between 2014 and 2019 roughly stagnated

[61] These calculations use the KIR-STAT database of the Hungarian Educational Authority. The calculations and the conclusions within the document are the intellectual product of the author.

[62] The exact cost depends on the qualification, and experience of teachers, and the exact class sizes of PE classes.

3.4 Data and descriptive statistics

In my analysis, I use data from the Hungarian National Fitness Assessment Program (abbreviated NETFIT). NETFIT is funded by the European Union and the Hungarian Government and is developed by the Hungarian School Sport Federation. Within the framework of the program, all Hungarian schools are obliged by a law [63] to provide fitness data on 5th to 12th grader students every year since the 2014/2015 school year.

The data, collected at the individual level, includes measures of body composition, aerobic fitness, musculoskeletal fitness, and flexibility along with the age and gender of the students. Physical education teachers are responsible for collecting the data and uploading it to the system of NETFIT. Standard equipment for the measurements is provided free of charge for every school. Before the 2014/2015 school year, almost 8000 PE teachers received training to use the equipment, and information on the usage of the equipment is constantly available on the website of the Hungarian School Sport Federation.[64] PE teachers have to collect fitness data on students between every January and May. I do not have information on the exact time of data collection, but more than 80 percent of data is uploaded in May indicating that PE teachers plausibly make the measurements by the end of the school year.

The Hungarian School Sport Federation makes the data publicly available at the school level. I scraped the data in September 2020 from their website.[65]

Fitness outcomes are published separately for every school, aggregated by gender and age group. Thus, the level of observation in the database I use is a group of students with the same gender and age in each school (and I refer to it as a school-gender-age group from now on). Table 3.1 shows the number of these school-gender-age groups by school year, with the number of students whose measurements were taken in a given year.[66] The data covers the vast majority of children: in the observation period, depending on the exact year, data is collected from 84 to 89 percent of all Hungarian students from 5th to 12th grades.

Table 3.1: Number of students and number of school-age-gender groups by School year

School year	Number of 5-12 grade students*	Number of students in data	Percent Observed	Number of school-gender-age groups in data
2014/2015	679488	577732	85	38660
2015/2016	661515	586729	89	39120
2016/2017	679223	595417	88	39456
2017/2018	684881	576754	84	38426
2018/2019	679754	587733	86	38040

*Source: Hungarian Statistical Office

In the following, I describe the outcome and explanatory variables used in the analysis in more detail.

[63] 2011.évi CXC. törvény 80. § available here: https://net.jogtar.hu/jogszabaly?docid=a1100190.tv (downloaded: 1/2/2020)

[64] Around 20000 PE teachers are registered now in the system of NETFIT

[65] https://www.netfit.eu/public/pb_riport_iskola.php (downloaded: 11.12.2020)

[66] I use here the number of students who got their body fat percentage measured, the number of students for whom BMI is measured is marginally higher.

3.4.1 Outcome variables

The outcome variables are based on two measures to define body type, body mass index (BMI) and body fat percentage (BF%). BMI is total body mass divided by the square of the body height, and BF% is total fat mass divided by the total mass. Height, weight, and BF% of students are measured by the PE teachers. To measure BF%, schools were provided with specialized equipment (OMRON BF511 body composition monitors) free of charge.

In the data, children are categorized as normal, overweight, or obese based on BF% and BMI as well. For the categorization based on BF%, thresholds proposed by Laurson et al. (2011) are used, for the BMI categorization the thresholds are by Cole and Lobstein (2012). These thresholds vary by age and gender but do not vary by calendar year.

I use the following four measures as outcomes in my analysis: share of obese students in the given school-gender-age group, based on BF%; share of obese students based on BMI; mean BF% and mean BMI. Obesity is by definition the accumulation of excess body fat to an extent which can impair health.[67] Because of this, I consider the prevalence of obesity as the most important health outcome in my analysis, which is most accurately measured by BF%. Still, changes in mean BF% are also of interest, beyond the prevalence of obesity. There is some evidence that having excess body fat in the healthy weight range is associated with greater health risks (see the Literature review of Wickelgren, (1998)), and that among obese children the severity of obesity matters for health (Skinner et al., 2015).

In the following I present descriptive statistics of the obesity outcomes defined by BMI and BF%. I also show corresponding figures about the mean outcomes in the appendix (Figures C.1 - C.3).

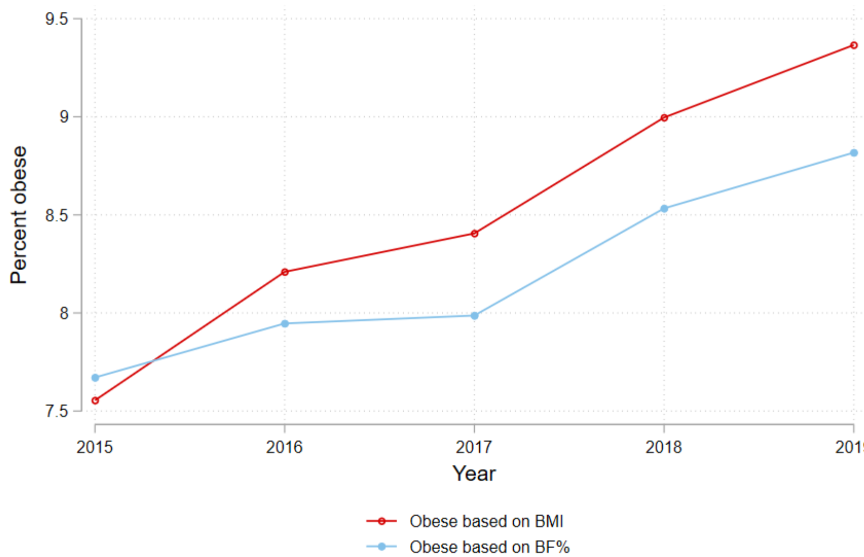
Figure 3.3 shows how obesity defined by the two measures changed between 2015 and 2019 in the data. The 2016 estimate of child obesity by the WHO (9%), defined by the share of children being at least two standard deviations above the median BMI, and obesity defined by absolute BMI and BF% thresholds are similar (8.0 by the BF%, and 8.2 by the BMI threshold). Both BF% and BMI measures show that obesity increased, but the increase was faster for the BMI-based obesity measure. This could happen, for example, if children became more muscular in the observation period, and more of them were mistakenly categorized as obese based on their BMI.

To describe the relationship between the two measures of obesity, I show the scatterplot of them separately for boys and girls. Note that the level of observation is the school-gender-age group. Figure 3.4 reveals that the two measures are correlated (the correlation coefficient is 0.61) but are not the same. If the two measures of obesity were largely the same, the observations would lie close to the 45-degree line (grey on the plot). However, there is substantial variation around the lines, showing that obesity rates defined by BMI and BF% can be substantially different in groups. Moreover, in groups of boys, obesity defined by BF% tends to be lower than the obesity rate by BMI. For example, in groups of boys where 25 percent is categorized as obese based on their BMI, the expected obesity rate by the BF% measure (predicted by a quadratic polynomial) is only 17 percent. This suggests that some boys with large muscle mass are mistakenly categorized as obese. For girls, focusing on the part of the distribution where 90 percent of the observations fall – groups with a maximum of 25 percent obesity rates – obesity rates defined by BMI and BF% are on average close to each other, but again with a considerable variation.

The age-gender specific obesity thresholds used in the NETFIT data reflect that children are

[67] See the definition of WHO here https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight (downloaded 01/02/2021)

Figure 3.3: Time trend in obesity defined by the two obesity measures



expected to have different levels of normal body mass and body fat percentage based on their age and sex. Still, Figure 3.5 reveals that regardless of this, the means of the two obesity measures are substantially different by age and gender. Compared to obesity defined by BMI, the BF% based obesity measures show a lower prevalence of obesity for boys, and higher for girls. In addition, for girls, while obesity decreases with age when the BMI measure is used, there is a sharp increase by the BF% measure. Specifically, while at age 11 around 9 percent of girls are categorized as obese by both measures, at age 18, only 6 percent is obese by the BMI measure and 16 percent by the BF% measure.

The descriptive comparisons underline the importance to use BF% as an outcome, as for boys, the BMI-based measure seems to overestimate obesity, and for girls over 14, it seems to underestimate it.[68]

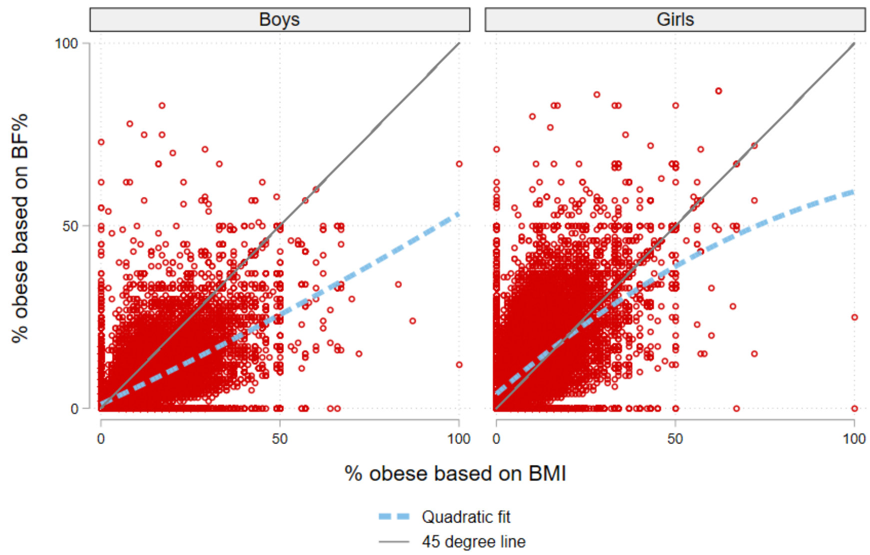
3.4.2 Explanatory variable

The explanatory variable I use is the total time spent on PE classes by a student at the end of a given school year in 100 hours.

Note that in the time period, the data covers (from the end of the 2014/2015 school year until the end of the 2018/2019 school year) the yearly increase in PE time is constant. This means that I cannot use variation within cohorts to uncover the effect of PE classes on body composition.

[68] This can have consequences which are outside the scope of this paper. For example, based on BMI measures, some studies conclude that obesity among boys tends to be higher than among girls. The graphs suggest that this finding can be an artifact resulting from the overestimation of real obesity among boys, and the underestimation of it for girls.

Figure 3.4: Obesity defined by BMI and BF% among groups of boys and girls



Note: groups of at least 5 students, from years 2015 to 2019 in all age groups are used to construct the scatterplots

The variation between cohorts comes from the fact that in the past they have spent very different amounts of time in PE classes.

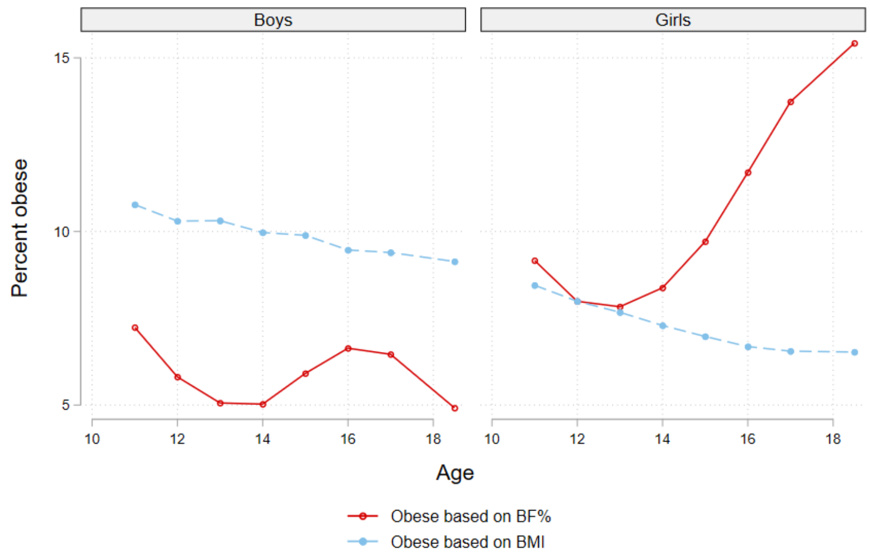
A difficulty in the analysis is that the data are aggregated by age, and the policy created variation in PE time not by age, but by grade. If the data were aggregated by grade, I would be able to calculate past PE time almost exactly for every grade and school year.[69]. In the age groups I observe, a mix of students from different grades is represented. For example, the group of 12-year-old students consists of 6th graders and 5th graders with different amounts of accumulated PE time. For each school, I make an estimate for lower grade students in each age group based on the annual National Assessment of Basic Competencies data [70] The NABC covers 6th, 8th, and 10th grade students, and provides information about their age. I estimate the average share of lower grade students in a given age group for every school based on this data.

Table 3.2 shows the estimated average time of PE classes students have ever had in each year and age-group in 100 hours. In Table C.1 in the appendix I show the actual hours by grade, which are used to calculate the hours based on age considering the share of students in different grades in the given school. Consider for example students who are 13 years old at the end of the 2017/2018 school year. On average, 37 percent of them are 6th graders in this school year, and they started school in 2012/2013 when daily PE classes were introduced. As there are 180 schooldays yearly,

[69] A small error would be introduced because of students who repeat grades, but this affects only 2-3 percent of students yearly in Hungary

[70] These calculations use the 2008-2017 waves of the National Assessment of Basic Competencies (OKM) database of the Hungarian Educational Authority. The calculations and the conclusions within the document are the intellectual product of the author.

Figure 3.5: Prevalence of obesity by age and gender based on BF% and BMI obesity measures



Note: groups of at least 5 students, from years 2015 to 2019 in all age groups are used to construct the scatterplots

this means that a 6^{th} grader has spent 810 hours (180*45 minutes for 6 years) at PE classes by the end of the 2017/2018 school year. The other 63 percent of 13-year-old students are 7^{th} graders in 2017/2018 who started school in 2011/2012, so from 1^{st} to 4^{th} grade they have had 3 PE classes weekly, and since 5^{th} grade, they have had daily PE classes. So, they have had $108*4*45+180*3*45$ minutes, or 729 hours of PE classes during their lives. Taken these two numbers together 13-year-old students by the end of the 2017/2018 school year have had $0.37*810+0.63*729 = 759$ hours of PE on average, which is the number reported in 100 hours in Table 3.2. I calculated this similarly for every age-group in every year, except for small modifications in the calculations for the 11-year-old and the 18^{+}-year-old groups, detailed in the table note.

As Table 3.2 shows the 2012/2013 daily PE classes policy introduced large jumps between cohorts in the number of PE classes they have ever had at a given age. Specifically, there are big differences between 2011/2012 1^{st}, 5^{th} and 9^{th} graders 2012/2013 and 1^{st}, 5^{th} and 9^{th} graders in the number of PE classes they have ever had in a given age, which translates into smaller jumps within different age groups. This means that there is still variation left in the number of PE classes somebody has ever had even after controlling for time and the age of students.

Table 3.2: Average time spent at PE classes by different age groups by the end of the given school year (in 100 hours)

Age	2014/2015	2015/2016	School year 2016/2017	2017/2018	2018/2019
11	4.59	4.59	6.75	6.75	6.75
12	5.44	5.44	6.24	7.60	7.60
13	6.79	6.79	6.79	7.59	8.95
14	6.44	8.13	8.13	8.13	8.94
15	6.79	7.79	9.49	9.49	9.49
16	8.14	8.14	9.14	10.84	10.84
17	9.49	9.49	9.49	10.49	12.19
18+	8.95	11.03	11.03	11.03	11.65

Every 11-year-old is considered a 5th grader as 4th graders are not included in the data.
18-year-old and older students are grouped into one category (18+) by NETFIT, so
all 12 grade students and a share of 11th grade students are categorized here.

3.5 Empirical strategy

I use the following model to uncover the effect of extra PE time on the body composition of students:

$$Y_{gt} = \alpha + \beta PEhours100_{gt} + \delta_2 girl_{gt} + \sum_{k=12}^{18} \delta_k age_{kgt} + \sum_{k=12}^{18} \pi_k (girl_{gt} \times age_{kgt}) + \\ + \gamma_t + \mu_s + u_{gt} \quad (3.1)$$

where Y_{gt} is the mean of BF% (or BMI), or proportion of obese students based on BF% (or BMI) in school-gender-age group g at year t. $PEhours100_{gt}$ is the overall time a school-gender-age group has ever spent with PE classes on average at time t in 100 hours (see Table 3.2 for the mean value of this variable by age and year). The coefficient of interest is β which shows how additional 100 hours spent on PE classes affects the outcome. Note that $PEhours100_{gt}$ measures past accumulated PE time, which means that the empirical strategy assumes that PE time affects body composition permanently. We can think about this strategy as estimating a health production function where the output is BF% and it is affected by health capital measured by time spent with PE accumulated in the past. Equation 3.1 assumes that current PE time affects body fat percentage the same way as past PE time, or in other words, that health capital does not depreciate.[71] This may be a strong assumption and in a robustness check, I relax it, and allow the effect of past PE classes to diminish over time.

Year fixed effects (γ_t), gender ($girl$ and age dummies (age_{kgt} is equal to one if age is k) are also included on the right hand side.In addition, I control for the interaction of gender and age to

[71]So, for example, the cohort who started school in 2012/2013 and had 4 more years of daily PE classes than the cohort who started 2011/2012. If PE time had a positive effect on their body composition, I expect the later cohort to be less obese (conditional on the time trend) as 5^{th} graders, as 6^{th} graders, as 7^{th} graders, etc (even though in grade 5^{th} and 6^{th} etc both of these cohorts have daily PE classes).

account for the fact that girls tend to become more obese during their teenage years while boys tend to become less obese. In some specifications, school-fixed effects (μ_s) are included as well.

The regression is weighted by the number of students in each school-gender-age group cell.I cluster standard errors to account for the possible serial correlation in the error term. Note that children are measured multiple times in the data. For example, the group of 12-year-old girls in school "A" in 2015, and the group of 13-year-old girls in school "A" in 2016 is composed of (almost) the same students measured at different times. Thus the error term is correlated by school-gender-cohort groups. In addition to that, the error term may be correlated by school as well. In the main specification, I cluster standard errors by school and show in the appendix that the results are not affected if the standard errors are clustered by school-gender-cohort groups.

The identification assumption is that after controlling for the above-mentioned variables the assignment of the number of PE classes is as good as random. Or in other words, if we take care of an aggregate time effect, in the absence of the PE classes policy students with the same age and gender in the same schools would be the same in terms of body composition. The phasing-in introduction of this policy means that there are multiple cohorts where the time spent on PE classes jumps, and in the absence of the policy we would not expect that something else affects the body composition of these specific cohorts, but not the others. So, for example, the fact that multiple policies possibly affecting the body composition of children were introduced in Hungary after 2010 (see Section 2.1) should not threaten the identification, as these policies were introduced to every student regardless of their age at the time, thus their potential effect is picked up by the year fixed effects.

One threat to this identification is that some other policies were introduced in 2012/2013 in the same phasing-in system, which could possibly affect the body composition of the students. Specifically, there have been some curriculum changes in other subjects (including large changes in the Literature curriculum), daily art classes were introduced, and schools were required to offer elective classes up until 4 p.m. every day. While these changes should not directly influence the body weight of students, indirectly they can affect it. For example, if the overall workload of students increased as a result of the curriculum changes, it could lead increased level of stress for students or could leave less time for students to engage in extracurricular sports activities. [72]

To check if other phased-in curriculum changes were important in terms of body composition, I estimate equation 3.1 on students in a Hungarian town, Hódmezővásárhely, where daily PE classes were already in place when the country-wide policy was introduced, but the other curriculum changes affected them as well. In these placebo regressions, I expect β to be 0 unless the other curriculum changes are important confounders. The estimates in these placebo regressions are negative, but statistically not different from 0, supporting that the other policies introduced in the same timing did not increase obesity (see Tables C.2 and C.3 in the Appendix).

In addition, the literature linking obesity to stress and (psychological) workload indicates that the magnitude of the bias introduced by the other policies is probably small. For example, increased time spent on homework is not related to increased obesity according to the review of Sturm (2004). In the case of adults, the review of Overgaard et al. (2004) shows that there is no association between psychological workload and obesity. Considering these results in the literature and the fact that there is no evidence that the curriculum changes substantially increased the workload or stress of students in the Hungarian case, it is a plausible assumption that the effect of these changes on students' body composition is much smaller in magnitude than the possible effect of directly

[72]There is no data to confirm or reject if the policy really increased workload, and opinions in the media differ about this.

increasing time spent with physical activity.

In the empirical strategy I measure how variation in past PE time affects the current body composition of students. We can think about this strategy as estimating a health production function where the output is BF% and it is affected by health capital measured by time spent with PE accumulated in the past. Equation 3.1 assumes that current PE time affects body fat percentage the same way as past PE time, or in other words, that health capital does not depreciate.[73] This may be a strong assumption and in a robustness check, I relax it, and allow the effect of past PE classes to diminish over time.

3.6 Results

Before turning to estimate equation 3.1, I show descriptive graphs to provide a better intuition for the empirical strategy and to provide a first glimpse into the effects. Figure 3.6 shows how obesity changed in different age groups between 2015 and 2019. In all graphs, grey dashed lines indicate the time when past PE time jumped in the given age group. If the policy had a large effect we could expect a decrease or a smaller-than-usual increase in obesity after the PE time jumps. For example, in the graph of 11-year-old students in the upper left panel, it can be observed that between 2016 and 2017 – when past PE time jumped from 459 hours to 675 hours – the obesity rate of 11-year-old students decreased from 8.2 percent to 7.8 percent. However, just based on data from 11-year-old students we could not be sure that it is really increased PE time that lead to the decrease in obesity, as anything else could have happened in that year that decreased obesity. By looking at the graphs of other age groups, we indeed see that obesity decreased in 2017 in other age groups as well, where PE time did not increase (for 13-year-olds and 17-year-olds). This suggests that in 2017 something else other than the PE time policy also affected children's body composition. Model 3.1 filters out the factors affecting all students in a specific year, by including year fixed effects.

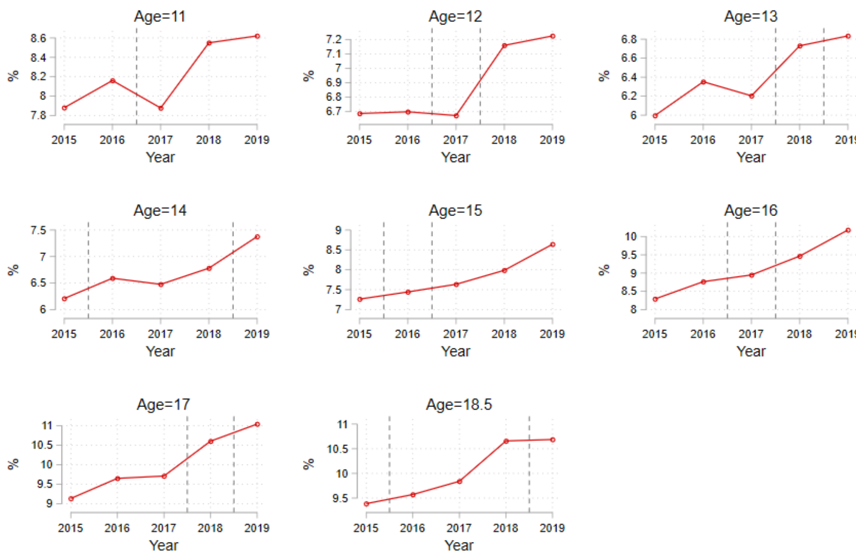
Only by looking at the panels of Figure 3.6 it is suggestive that the policy did not substantially decrease obesity. There are typically visible year-to-year increases in obesity, even in years when past PE time increased. In the appendix, Figures C.4-C.6 plot the other outcomes in the same manner, and similarly to obesity defined by BF%, we see no consistent pattern of decrease after PE time jumps, neither in obesity defined by BMI, nor mean BF%, nor mean BMI. The regression estimates in the following tables make this observation more precise.

Table 3.3 shows the regression results on outcomes based on BF%. In Model (1) the outcome variable is the mean BF% in school-gender-age groups, and gender, age indicators, their interaction, and year fixed effects are included as controls. Supporting the visual results in Figure 3.6, the point estimate on past PE time is very close to zero and statistically indistinguishable from it (0.015, p=0.340). In Model (2) I also include school fixed effects on the right-hand side, but this does not change the statistical significance and magnitude of the estimate (0.017, p=0.274). To get a better sense of the magnitude of this estimate: compared to the mean BF% of 21.93 in the sample, this is a very small, 0.07 percent increase in mean body fat percentage and a larger than 0.05 percent reduction in mean body fat percentage as a response to 100 additional PE hours can be ruled out with 95% confidence.

Similarly to the results on mean BF%, the estimates on the prevalence of obesity are also

[73] So, for example, the cohort who started school in 2012/2013 and had 4 more years of daily PE classes than the cohort who started 2011/2012. If PE time had a positive effect on their body composition, I expect the later cohort to be less obese (conditional on the time trend) as 5th graders, as 6th graders, as 7th graders, etc (even though in grade 5th and 6th etc both of these cohorts have daily PE classes).

Figure 3.6: Prevalence of obesity by age and year based on the BF% obesity measure



Dashed grey lines indicate a jump in past PE hours

statistically insignificant, and close to 0 (0.014 (p=0.771) in Model (3) without school fixed effects, and 0.004 (p=0.939) in Model (4) with school fixed effects). Based on this estimate, it can be ruled out with 95% confidence that 100 additional PE hours in the past decreases obesity by more than 0.09 percentage points compared to the baseline obesity of 8.19 percent. To put this estimate more in context, consider that obesity typically increases by 0.5 percentage points every year, thus even in the best-case scenario, 100 extra hours of PE per year could offset less than 20 percent of the general expected yearly increase in obesity.

Table 3.4 shows the results of the same regressions but now the outcomes are mean BMI and percentage of obese students based on BMI thresholds. The results in Model (1) and Model (2) show that one hundred hours of extra PE increased average BMI by 0.005, or 0.006 if school fixed effects are included. These are again statistically insignificant (p=0.570 and 0.488) and close to zero estimates. The effect of increased PE time on obesity based on a BMI cutoff is also very close to 0 and insignificant (-0.033 (p=0.481) or -0.035 (p=0.451) depending on the exact model).

Table 3.3: Regressions of PE time on mean BF% and obesity

	Mean BF%		Percent obese[1]	
	(1)	(2)	(3)	(4)
PEhours100	0.015	0.017	0.014	0.004
	(0.016)	(0.015)	(0.046)	(0.046)
School FE	No	Yes	No	Yes
adj. R-sq	0.692	0.774	0.142	0.313
Mean of outcome	21.93		8.19	
N	179476		179534	

Year FE, age dummies, gender, and age and gender interactions are included in all models. Standard errors (in parentheses) are clustered by school. $+p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001$

[1] Categorization based on BF% thresholds

Table 3.4: Regressions of PE time on mean BMI and obesity

	Mean BMI		Obese based on BMI	
	(1)	(2)	(3)	(4)
PEhours100	0.005	0.006	-0.033	-0.035
	(0.009)	(0.009)	(0.047)	(0.046)
School FE	No	Yes	No	Yes
adj. R-sq	0.386	0.539	0.036	0.216
Mean of outcome	21.48		8.51	
N	179177		179197	

Year FE, age dummies, gender, and age and gender interactions are included in all models. Standard errors (in parentheses) are clustered by school. $+p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001$

[1] Categorization based on BMI thresholds

Taking the coefficient estimates of 100 hours (β), it can be calculated how the whole daily PE classes reform changed body composition outcomes. As the policy increased hours spent at PE classes by 67.5 hours yearly for students above 5$^{\text{th}}$ grade and by 54 hours for students below it, a child who is affected by the policy throughout her studies and goes to school for 12 years, is expected to spend 756 more hours with PE than a child who is fully unaffected by the policy. Thus the overall estimated change in body composition outcomes as a response to the policy is 7.56 times the β. The estimates and their confidence intervals, based on the specifications containing school fixed effects as well, are the following for specific outcomes: 0.129 (-0.094;0.351) for mean BF%, 0.030 (-0.651;0.712) for obesity based on BF%, 0.045 (-0.608;0.699) for mean BMI, -0.265 (-0.598;0.069) for obesity based on BMI.

Overall, the results show that increased PE time could not reduce obesity. Similarly to most previous findings, BMI did not decrease as a result of increased PE time. BF%, and obesity defined by BF% also stayed unchanged, indicating that it is not only weight that is unaffected by PE time, but also body composition in general. But, even though the policy was ineffective in reducing obesity in the general population of children, it is possible that there are some specific socio-economic groups where obesity decreased because of the policy.

3.6.1 Heterogeneous effects on the prevalence of obesity

In this section, I analyze whether the policy was more successful in specific groups of children. Having information about heterogeneous effects can help to target future policies at those groups who would benefit the most of them and can shed light on the mechanisms of why the policy works or does not work. I present the results in Table 3.5 for one outcome, obesity based on BF%, and show the results for the other three outcomes in the appendix.

In column (1) of Table 3.5, I repeat the estimate for the total population (a statistically insignificant increase in obesity by 0.004 percentage points as a response to 100 extra PE classes) for reference. In the following columns, I estimate equation 3.1, with school fixed effects on specific subsamples. This means that in these specifications I allow the association between the explanatory variables (e.g. the time trend) to be different in the subsample than in the total population.

Column (2) shows the estimates for the effect of PE on obesity for children under 15 years of age. This is a potentially interesting group because there is some evidence that PE time has a larger effect on elementary school kids: e.g. Cawley et al. (2013) finds a positive effect for 5^{th} grade boys while they find no effect for high school students in a different paper (Cawley et al., 2007). However, in the case of the Hungarian policy, the effect of PE time on the prevalence of obesity is not significant either, and is in the wrong direction (0.051). For older children the estimate is negative, but again insignificant, and close to zero (-0.057).

Model (4) and (5) show the estimates separately for boys and girls. Previous studies suggest that it is possible that the impact of PE time differs by gender, and the only significant results in the literature so far are for boys. In my context, the effect is indeed negative for boys (-0.053) and positive for girls (0.035), but these effects are small, and statistically indistinguishable from zero, or from each other.

In columns (6) and (8) I check the effects for students in schools where the ratio of disadvantaged kids is high. To calculate the share of disadvantaged children I use the 2015 wave of the National Assessment of Basic Competencies data of the Hungarian Education Office[74]. Obesity disproportionately affects disadvantaged children in rich countries. Figure C.7 in the appendix shows that this is also the case in Hungary: the prevalence of obesity increases by the ratio of poor students in schools, and decreases by the ratio of kids with college-educated parents. The inequality in obesity can enhance existing inequalities, so it is important to check whether PE classes can decrease obesity for disadvantaged children despite their ineffectiveness in the general population. In addition, it is plausible to expect that disadvantaged students might benefit more from this policy. Studies have shown that children with worse social economic status gain more from certain school-based interventions (see Knaus et al., 2020). Further, poorer or less educated parents might not be able to pay for extracurricular sports activities and might be less efficient in enforcing or teaching healthy eating habits and the importance of physical activity to their children. Still, the results show that students in schools where a minimum of 50 percent of children are poor[75], or where a maximum of 15 percent of children have at least one college-educated parent have not benefited more from

[74]

[75] The child indicates in the questionnaire that they live in very bad material circumstances

increased PE time either: the effect is close to 0 and insignificant for these subgroups as well (0.177 and 0.148). The same estimates for kids in more advantageous schools by poverty (column (7)) or education of parents (column (9)) are also 0 (0.018, -0.023). This is in line with Knaus et al. (2020) which does not find differential health effects of PE time by social status of children either.

The last split I make in the sample is by the number of gym rooms in schools. Specifically, I compare children in schools where at least one gym room was available for 200 children (making it possible to hold every PE class in a gym room in the school), to schools where the gym room-student ratio was worse than this in 2015. A popular criticism frequently appearing in the Hungarian media[76] against daily gym classes is that schools do not have sufficient number of rooms suitable for PE lessons to provide meaningful activity daily for every student. However, column (10) in Table 3.5 shows that this is not sufficient to explain the lack of effect on obesity, as the effect in schools where there were enough rooms is still insignificantly different from 0 (-0.032), and indistinguishable from the effect in schools with fewer gym rooms (-0.036).

These results indicate that the policy was not only ineffective in reducing obesity in general but also for specific subgroups. Tables C.8 - C.10 in the appendix for the other outcomes also show that increased PE time has no statistically significant effect on the body composition outcomes of special groups of students.

[76]See for example https://www.origo.hu/gazdasag/20141006-megbukhat-a-mindennapos-testneveles.html and https://444.hu/2020/07/29/nyolc-eve-vezette-be-a-kormany-a-mindennapos-testnevelest-de-az-iskolakban-meg-mindig-nincs-eleg-tornaterem(downloaded: 01/02/2020)

Table 3.5: Heterogeneous effects on obesity by BF%

Subsample	(1) Total	(2) Age Under 15	(3) Age Min 15	(4) Boy	(5) Girl	(6) Min 50% poor	(7) Less than 50% Poor	(8) Max 15 % college edu-cated	(9) More than 15% coll. educ.	(10) Min 1 gym room per 200	(11) Less than 1 gym room per 200
PEHours100	0.004	0.051	-0.057	-0.053	0.035	0.177	0.018	0.148	-0.023	-0.032	0.036
	(0.046)	(0.074)	(0.057)	(0.049)	(0.069)	(0.212)	(0.053)	(0.103)	(0.056)	(0.147)	(0.055)
N	179534	87378	92156	91103	88431	19565	130800	74035	86380	45356	104516
adj. R-sq	0.313	0.169	0.384	0.182	0.345	0.157	0.281	0.192	0.310	0.133	0.311

Standard errors are clustered by school, *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

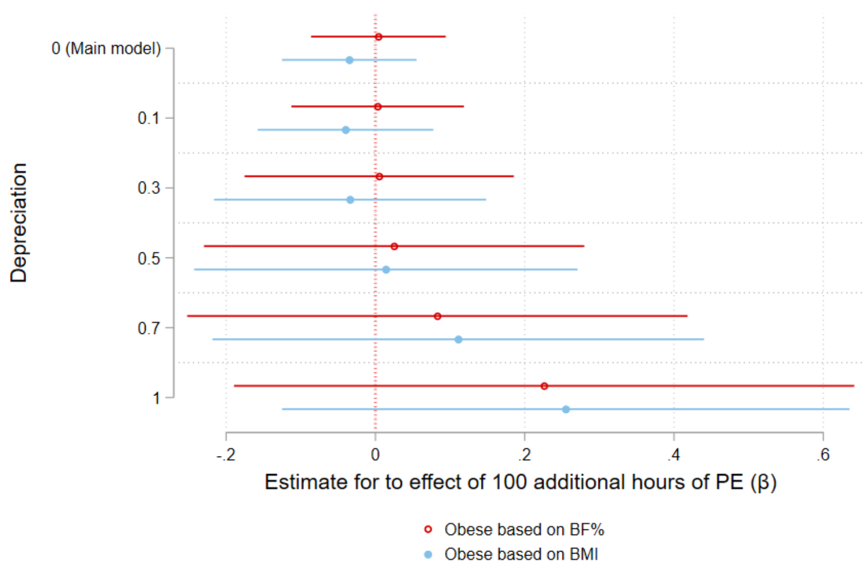
Hungarian DOI: 10.14754/CEU.2023.04

3.7 Robustness checks

As a first robustness check, I show in the appendix (Tables C.4-C.5) that the results are virtually the same if I cluster the standard errors at the school-gender-cohort level instead of the school level. In addition, the results are not sensitive to weighting, the unweighted results are again very similar to the main results (Tables C.6-C.7).

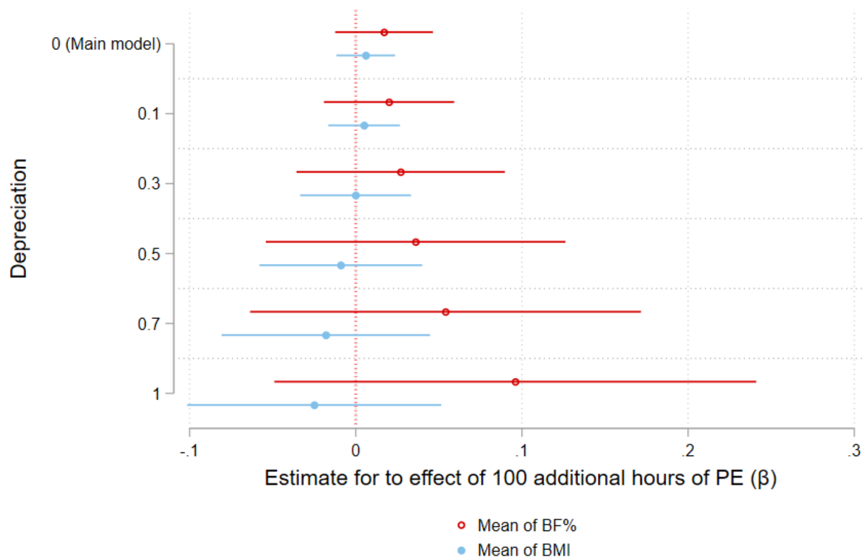
The last robustness check relaxes the assumption that past PE time keeps on affecting body composition in later years as well. As detailed in Section 3.5, the model described by equation 3.1 assumes that PE time accumulates as health capital, and keeps on affecting body composition over time. There are pieces of evidence supporting this view, for example, it is well established that obesity is persistent (Daouli et al., 2014), suggesting that avoiding obesity in a specific year can have a positive effect in the following years as well. Thus the notion that present gains or losses in fat mass keep on affecting body composition years later makes sense. However, it might also be the case that current investments in health matter more for body composition than past investments. In the following, I estimate versions of model 3.1 where I account for the possibility that health capital depreciates over time. Specifically, I multiply past PE hours by $(1 - \textit{depreciation rate})^t$ where t shows the number of years since the student had the given number of PE classes. A deprecation rate of 0 corresponds to the main model, and a depreciation rate of 1 would mean that it is only current physical activity which matters in current body composition.

Figure 3.7: Regression estimates for the effect of PE time on obesity with different depreciation rates



As from 2015 everybody already had daily PE classes, most variation in accumulated PE time difference comes from the past. This means that the larger the depreciation rate, the smaller the variation in past PE time. This is reflected in Figures 3.7 and 3.8, where it is visible that with larger

Figure 3.8: Regression estimates for the effect of PE time on obesity with different depreciation rates



depreciation rates standard errors are getting larger as well. The figures also show that regardless of the depreciation rate, the point estimates for β in equation 3.1 are still statistically insignificant and close to zero for the prevalence of obesity (Figure 3.7) and for mean BMI and BF% as well (Figure 3.8).

These results support that the close-to-zero estimates in Tables 3.3 and 3.4 are not coming by because of the choice of the specific depreciation rate of 1. Further, they suggest that PE time is not only ineffective in the long run, but in the short run as well.

3.8 Conclusion

This is the first study that uses data on body fat percentage to evaluate the effect of increasing physical education time for elementary and secondary school students, in the context of a large-scale policy. The inclusion of this new outcome is important, as BMI is a flawed measure of obesity (Burkhauser and Cawley, 2008). In the context of physical activity, the usage of BMI as an outcome can hide decreases in obesity, as increased physical activity can lead to a healthier body composition while leaving weight, and thus BMI unchanged.

In the analysis, I show that even though BMI and BF% are correlated, they do measure different things, and obesity rates defined based on thresholds of these two measures can be substantially different, especially in some subgroups (e.g. girls over 16). In line with previous findings on the relationship between PE classes and student health (e.g. Cawley et al., 2007 S.-Y. Kim and So, 2012 and Sabia et al., 2017), my results also show that increasing PE time was unable to reduce BMI or the prevalence of obesity measured by BMI among children. In addition to those results, I also show that PE time is not only ineffective in reducing the weight of students but also more generally, in improving the body composition of students: the effects of PE time on body fat percentage and obesity measured by BF% thresholds are also zero.

While some previous studies showed that specific subgroups of students can benefit more from increased PE time (e.g. Packham and Street, 2019), heterogeneity analysis in the case of the 2012 Hungarian policy revealed that prevalence of obesity among boys, elementary school students, students in schools where the proportion of disadvantaged students is high, and among students in well-equipped schools has not decreased either.

A difficulty in the analysis is that the data are aggregated to school-gender-age groups. With individual-level data and information on grades, more precise estimates could be calculated, and a richer heterogeneity analysis could be conducted. For example, with quantile regressions, one could check whether children at specific parts of the BF% distribution benefit more or less from increased PE time.

The mechanism behind the general finding that PE time in itself does not decrease obesity is still unclear. My results on BF% suggest that the measurement problem introduced by the usage of BMI as an outcome cannot explain the lack of significant results in the literature. Moreover, the analysis suggests that in the Hungarian case, the ineffectiveness cannot be attributed to the lack of suitable infrastructure either. Increased PE time in itself is clearly not enough to tackle the alarming levels of increasing childhood obesity. Future research with more detailed data can explore the elements that could make similar large universal PE time policies more effective in reducing the prevalence of obesity among children.

References

Abadie, A., & Spiess, J. (2020). Robust Post-Matching Inference [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2020.1840383]. *Journal of the American Statistical Association*, *0*(0), 1–13. https://doi.org/10.1080/01621459.2020.1840383

Abarca-Gómez, L., Abdeen, Z. A., Hamid, Z. A., Abu-Rmeileh, N. M., Acosta-Cazares, B., Acuin, C., Adams, R. J., Aekplakorn, W., Afsana, K., Aguilar-Salinas, C. A., Agyemang, C., Ahmadvand, A., Ahrens, W., Ajlouni, K., Akhtaeva, N., Al-Hazzaa, H. M., Al-Othman, A. R., Al-Raddadi, R., Al Buhairan, F., ... Ezzati, M. (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: A pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *The Lancet*, *390*(10113), 2627–2642. https://doi.org/10.1016/S0140-6736(17)32129-3

Adsera, A. (2005). Vanishing children: From high unemployment to low fertility in developed countries. *American Economic Review Papers and Proceedings*, *95*, 189–193.

Aizer, A., Devereux, P., & Salvanes, K. (2022). Grandparents, moms, or dads? why children of teen mothers do worse in life. *Journal of Human Resources*, *57*(6), 2012–2047.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, *103*(484), 1481–1495.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. https://ideas.repec.org/b/pup/pbooks/8769.html

Ashcraft, A., Fernández-Val, I., & Lang, K. (2013). The consequences of teenage childbearing: Consistent estimates when abortion makes miscarriage non-random. *The Economic Journal*, *123*(571), 875–905.

ASRM. (2012). Age and Fertility - A Guide for Patients. https://www.asrm.org/globalassets/rf/news-and-publications/bookletsfact-sheets/english-fact-sheets-and-info-booklets/Age_and_Fertility.pdf

Bald, A., Chyn, E., Hastings, J., & Machelett, M. (2022). The Causal Impact of Removing Children from Abusive and Neglectful Homes [Publisher: The University of Chicago Press]. *Journal of Political Economy*, *130*(7), 1919–1962. https://doi.org/10.1086/719856

Bald, A., Doyle Jr., J. J., Gross, M., & Jacob, B. A. (2022). Economics of Foster Care. *Journal of Economic Perspectives*, *36*(2), 223–246. https://doi.org/10.1257/jep.36.2.223

Bárdits, A., Adamecz-Völgyi, A., Bisztray, M., Weber, A., & Szabo, A. (2023a). Precautionary fertility: Conceptions, births, and abortions around employment shocks. *IZA Discussion Paper*.

Bárdits, A., Adamecz-Völgyi, A., Bisztray, M., Weber, A., & Szabo, A. (2023b). Precautionary fertility: Conceptions, births, and abortions around employment shocks. *KRTK KTI Working Paper Series*.

Barth, R. P. (2002). Institutions vs. foster homes. *The empirical base for a century of action.*

Bassett, D. R., Wyatt, H. R., Thompson, H., Peters, J. C., & Hill, J. O. (2010). Pedometer-Measured Physical Activity and Health Behaviors in United States Adults. *Medicine and science in sports and exercise*, *42*(10), 1819–1825. https://doi.org/10.1249/MSS.0b013e3181dc2e54

Bearak, J. M., Popinchalk, A., Beavin, C., Ganatra, B., Moller, A.-B., Tunçalp, Ö., & Alkema, L. (2022). Country-specific estimates of unintended pregnancy and abortion incidence: A global comparative analysis of levels in 2015–2019. *BMJ Global Health*, *7*(3), e007151.

Beaujouan, E., & Berghammer, C. (2019). The gap between lifetime fertility intentions and completed fertility in europe and the united states: A cohort approach. *Population Research and Policy Review*, *38*, 507–535.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), 491–507.

Bertheau, A., Acabbi, E. M., Lombardi, S., Barcelo, C., Gulyas, A., & Saggio, R. (2022). The unequal cost of job loss across countries. https://doi.org/10.3386/W29727

Blank, R. M., George, C. C., & London, R. A. (1996). State abortion rates the impact of policies, providers, politics, demographics, and economic environment. *Journal of Health Economics*, *15*(5), 513–553. https://doi.org/10.1016/S0167-6296(96)00494-8

Bray, G. A. (2004). Medical Consequences of Obesity [Publisher: Oxford Academic]. *The Journal of Clinical Endocrinology & Metabolism*, *89*(6), 2583–2589. https://doi.org/10.1210/jc.2004-0535

Buhovac, M. (2018). Physical Education and Sport at School in Europe. Retrieved December 9, 2020, from https://eacea.ec.europa.eu/national-policies/eurydice/content/physical-education-and-sport-school-europe_en

Burkhauser, R. V., & Cawley, J. (2008). Beyond bmi: The value of more accurate measures of fatness and obesity in social science research. *Journal of health economics*, *27*(2), 519–529.

Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics*, *43*, 244–268. https://doi.org/10.1016/j.jhealeco.2015.03.001

Cawley, J., Frisvold, D., & Meyerhoefer, C. (2013). The impact of physical education on obesity among elementary school children. *Journal of Health Economics*, *32*(4), 743–755. https://doi.org/10.1016/j.jhealeco.2013.04.006

Cawley, J., Meyerhoefer, C., & Newhouse, D. (2007). The impact of state physical education requirements on youth physical activity and overweight. *Health Economics*, *16*(12), 1287–1301. https://doi.org/10.1002/hec.1218

Church, T. S., Thomas, D. M., Tudor-Locke, C., Katzmarzyk, P. T., Earnest, C. P., Rodarte, R. Q., Martin, C. K., Blair, S. N., & Bouchard, C. (2011). Trends over 5 Decades in U.S. Occupation-Related Physical Activity and Their Associations with Obesity [Publisher: Public Library of Science]. *PLOS ONE*, *6*(5), e19657. https://doi.org/10.1371/journal.pone.0019657

Cole, T. J., & Lobstein, T. (2012). Extended international (iotf) body mass index cut-offs for thinness, overweight and obesity. *Pediatric obesity*, *7*(4), 284–294.

Currie, J., & Schwandt, H. (2014). Short- and long-term effects of unemployment on fertility [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *111*(41), 14734–14739. https://doi.org/10.1073/pnas.1408975111

Cutler, D. M., Glaeser, E. L., & Shapiro, J. M. (2003). Why Have Americans Become More Obese? *Journal of Economic Perspectives*, *17*(3), 93–118. https://doi.org/10.1257/089533003769204371

Daouli, J., Davillas, A., Demoussis, M., & Giannakopoulos, N. (2014). Obesity persistence and duration dependence: Evidence from a cohort of us adults (1985–2010). *Economics & Human Biology*, *12*, 30–44.

Darvas, Á., Farkas, Z., Ágnes, K., & Vígh, K. (2016). Roma gyerekek a szakellátásban. esély, 4: 52–82.

Darvas, Á., Farkas, Z., Kende, Á., & Vígh. (2016). Roma gyerekek a szakellátásban. *Esély*, *4*. https://www.esely.org/kiadvanyok/2016_4/2016-4_3-1_Darvas-Farkas-Kende-Vigh_Roma_gyerekek.pdf

De Paola, M., Nisticò, R., & Scoppa, V. (2021). *Fertility decisions and employment protection: The unintended consequences of the italian jobs act* (CSEF Working Papers). Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy. https://EconPapers.repec.org/RePEc:sef:csefwp:596

Dehejia, R., & Lleras-Muney, A. (2004). Booms, Busts, and Babies' Health [Publisher: Oxford Academic]. *The Quarterly Journal of Economics*, *119*(3), 1091–1130. https://doi.org/10.1162/0033553041502216

Del Bono, E., Weber, A., & Winter-Ebmer, R. (2012). Clash of Career and Family: Fertility Decisions After Job Displacement [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1542-4774.2012.01074.x]. *Journal of the European Economic Association*, *10*(4), 659–683. https://doi.org/10.1111/j.1542-4774.2012.01074.x

del Valle, J. F. (2014). Children in State Care. In A. Ben-Arieh, F. Casas, I. Frønes, & J. E. Korbin (Eds.), *Handbook of Child Well-Being* (pp. 2945–2963). Springer Netherlands. https://doi.org/10.1007/978-90-481-9063-8_119

DeSena, A. D., Murphy, R. A., Douglas-Palumberi, H., Blau, G., Kelly, B., Horwitz, S. M., & Kaufman, J. (2005). Safe homes: Is it worth the cost?: An evaluation of a group home permanency planning program for children who first enter out-of-home care. *Child Abuse & Neglect*, *29*(6), 627–643.

Dodge, K. A., Dishion, T. J., & Lansford, J. E. (2006). Deviant peer influences in intervention and public policy for youth. *Social Policy Report*, *20*(1), 1–20.

Doepke, M., Hannusch, A., Kindermann, F., & Tertilt, M. (2022). The economics of fertility: A new era. https://doi.org/10.3386/W29948

Dollman, J., Norton, K., & Norton, L. (2005). Evidence for secular trends in children's physical activity behaviour [Publisher: British Association of Sport and Excercise Medicine Section: Review]. *British Journal of Sports Medicine*, *39*(12), 892–897. https://doi.org/10.1136/bjsm.2004.016675

Doyle Jr., J. J. (2007). Child Protection and Child Outcomes: Measuring the Effects of Foster Care. *American Economic Review*, *97*(5), 1583–1610. https://doi.org/10.1257/aer.97.5.1583

Dozier, M., Zeanah, C. H., Wallin, A. R., & Shauffer, C. (2012). Institutional care for young children: Review of literature and policy implications. *Social issues and policy review*, *6*(1), 1–25.

Dregan, A., & Gulliford, M. C. (2012). Foster care, residential care and public care placement patterns are associated with adult life trajectories: Population-based cohort study. *Social psychiatry and psychiatric epidemiology*, *47*, 1517–1526.

Dunson, D. B., Colombo, B., & Baird, D. D. (2002). Changes with age in the level and duration of fertility in the menstrual cycle. *Human Reproduction (Oxford, England)*, *17*(5), 1399–1403. https://doi.org/10.1093/humrep/17.5.1399

Eliason, M., & Storrie, D. (2006). Lasting or Latent Scars? Swedish Evidence on the Long-Term Effects of Job Displacement [Publisher: University of Chicago Press]. *Journal of Labor Economics*, *24*(4), 831–856. Retrieved December 14, 2021, from https://econpapers.repec.org/article/ucpjlabec/v_3a24_3ay_3a2006_3ai_3a4_3ap_3a831-856.htm

ERRC. (2007). Fenntartott érdektelenség: Roma gyermekek a magyar gyermekvédelmi rendszerben. Retrieved July 3, 2023, from http://www.errc.org/uploads/upload_en/file/02/90/m00000290.pdf

Eurochild. (2015). *Towards a stronger economic evidence base to support child protection reform: From institutions to family based care and community level services.* (tech. rep.). https://www.ohchr.org/sites/default/files/Documents/Issues/Children/TowardsInvestment/EurochildHopeandHomesforChildrenandSOSChildrensVillages.pdf

Fintor, G. (2017). Implementáció és tanulói attitűdök : A mindennapos testnevelés funkciója egy hátrányos helyzetű régió iskoláiban. *doctoral dissertation at Debreceni Egyetem*.

Fitzenberger, B., & Seidlitz, A. (2023). Changing fertility and heterogeneous motherhood effects: Revisiting the effects of a parental benefits reform. https://www.asrm.org/globalassets/rf/news-and-publications/bookletsfact-sheets/english-fact-sheets-and-info-booklets/Age_and_Fertility.pdf

Friedrich, W. N., Baker, A. J., Parker, R., Schneiderman, M., Gries, L., & Archer, M. (2005). Youth with problematic sexualized behaviors in the child welfare system: A one-year longitudinal study. *Sexual Abuse: A Journal of Research and Treatment*, *17*, 391–406.

González, L., & Trommlerová, S. K. (2021). Cash Transfers and Fertility: How the Introduction and Cancellation of a Child Benefit Affected Births and Abortions [Publisher: University of Wisconsin Press]. *Journal of Human Resources*, 0220. https://doi.org/10.3368/jhr.59.1.0220-10725R2

Gross, M., & Baron, E. J. (2022). Temporary Stays and Persistent Gains: The Causal Effects of Foster Care. *American Economic Journal: Applied Economics*, *14*(2), 170–199. https://doi.org/10.1257/app.20200204

Gypen, L., Vanderfaeillie, J., De Maeyer, S., Belenger, L., & Van Holen, F. (2017). Outcomes of children who grew up in foster care: Systematic-review. *Children and Youth Services Review*, *76*, 74–83.

Hajdu, T., Hermann, Z., Horn, D., Kertesi, G., Kézdi, G., Köllő, J., & Varga, J. (2015). *Az érettségi védelmében* (tech. rep.). Budapest Working Papers on the Labour Market.

Halla, M., Schmieder, J., & Weber, A. (2020). Job displacement, family dynamics, and spousal labor supply. *American Economic Journal: Applied Economics*, *12*, 253–87. https://doi.org/10.1257/APP.20180671

Harris, K. C., Kuramoto, L. K., Schulzer, M., & Retallack, J. E. (2009). Effect of school-based physical activity interventions on body mass index in children: A meta-analysis [Publisher: CMAJ Section: Research]. *CMAJ*, *180*(7), 719–726. https://doi.org/10.1503/cmaj.080966

Hayduk, I. (2017). The Effect of Kinship Placement Laws on Foster Children's Well-Being [Publisher: De Gruyter]. *The B.E. Journal of Economic Analysis & Policy*, *17*(1). https://doi.org/10.1515/bejeap-2016-0196

Hendren, N. (2017). Knowledge of Future Job Loss and Implications for Unemployment Insurance. *American Economic Review*, *107*(7), 1778–1823. http://www.aeaweb.org/articles?id=10.1257/aer.20151655

Herbst, C. M. (2011). The earned income tax credit and abortion. *Social Science Research*, *40*, 1638–1651. https://doi.org/10.1016/J.SSRESEARCH.2011.05.003

Hill, J. O., Wyatt, H. R., & Peters, J. C. (2012). Energy Balance and Obesity. *Circulation*, *126*(1), 126–132. https://doi.org/10.1161/CIRCULATIONAHA.111.087213

Hotz, V. J., Klerman, J. A., & Willis, R. J. (1997). The economics of fertility in developed countries. In M. R. Rosenzweig & O. Stark (Eds.), *Handbook of population and familiy economics*. North Holland.

Huttunen, K., & Kellokumpu, J. (2016). The Effect of Job Displacement on Couples' Fertility Decisions [Publisher: The University of Chicago Press]. *Journal of Labor Economics*, *34*(2), 403–442. https://doi.org/10.1086/683645

Ichino, A., Schwerdt, G., Winter-Ebmer, R., & Zweimüller, J. (2017). Too old to work, too young to retire? *The Journal of the Economics of Ageing*, *9*, 14–29. https://doi.org/10.1016/j.jeoa.2016.07.001

Illing, H., Schmieder, J., & Trenkle, S. (2021). The gender gap in earnings losses after job displacement. www.iza.org

ILO. (2010). Maternity at work a review of national legislation findings from the ilo database of conditions of work and employment laws second edition. www.ilo.org/publns

ILO. (2022). ILO Travail - Conditions of Work and Employment Programme. https://www.ilo.org/dyn/travail/travmain.sectionChoice2?p_lang=en&p_structure=3&p_sc_id=3050&p_countries=ALL#tab5

Inchley, J., Currie, D., Jewell, J., Breda, J., Barnekow, V., Bucksch, J., Elgar, F., Hamrik, Z., Keane, E., Kelly, C., & Rasmussen, M. (2017). *Adolescent obesity and related behaviours: Trends and inequalities in the WHO European Region 2002-2014: Observations from the Health Behaviour in School-aged Children (HBSC) WHO collaborative cross-national study*.

Index. (2012). Beszüntették a tablettás abortuszt / Abortion pill ceased in Hungary. https://index.hu/belfold/2012/09/13/beszuntettek_a_tablettas_abortuszt/

Jacobson, L. S., LaLonde, R. J., & Sullivan, D. G. (1993a). Earnings Losses of Displaced Workers. *American Economic Review*, *83*(4), 685–709. http://www.jstor.org/stable/2117574

Jacobson, L. S., LaLonde, R. J., & Sullivan, D. G. (1993b). Earnings Losses of Displaced Workers [Publisher: American Economic Association]. *The American Economic Review*, *83*(4), 685–709. Retrieved January 4, 2023, from https://www.jstor.org/stable/2117574

Keith, S. W., Redden, D. T., Katzmarzyk, P. T., Boggiano, M. M., Hanlon, E. C., Benca, R. M., Ruden, D., Pietrobelli, A., Barger, J. L., Fontaine, K. R., Wang, C., Aronne, L. J., Wright, S. M., Baskin, M., Dhurandhar, N. V., Lijoi, M. C., Grilo, C. M., DeLuca, M., Westfall, A. O., & Allison, D. B. (2006). Putative contributors to the secular increase in obesity: Exploring the roads less traveled [Number: 11 Publisher: Nature Publishing Group]. *International Journal of Obesity*, *30*(11), 1585–1594. https://doi.org/10.1038/sj.ijo.0803326

Kim, J. (2012). Are Physical Education-Related State Policies and Schools' Physical Education Requirement Related to Children's Physical Activity and Obesity? *Journal of School Health*, *82*(6), 268–276. https://doi.org/https://doi.org/10.1111/j.1746-1561.2012.00697.x

Kim, S.-Y., & So, W.-Y. (2012). The Relationship Between School Performance and the Number of Physical Education Classes Attended by Korean Adolescent Students. *Journal of Sports Science & Medicine*, *11*(2), 226–230. Retrieved February 11, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3737878/

Knaus, M. C., Lechner, M., & Reimers, A. K. (2020). For better or worse? – The effects of physical education on child development. *Labour Economics*, *67*, 101904. https://doi.org/10.1016/j.labeco.2020.101904

KSH. (2012). Állami gondoskodástól a mai gyermekvédelemig. https://www.ksh.hu/docs/hun/xftp/idoszaki/pdf/allamigondoskodas.pdf

KSH. (2022a). Ezer megfelelő korú nőre jutó élveszületések / Live births per 1000 females of age 15-49. https://www.ksh.hu/stadat_files/nep/hu/nep0008.html

KSH. (2022b). Ezer megfelelő korú nőre jutó terhességmegszakítás / Abortions per 1000 females of age 15-49. https://www.ksh.hu/stadat_files/nep/hu/nep0014.html

Kuhn, P. J. (2002). Summary and Synthesis (P. J. Kuhn, Ed.). *Upjohn Press Book Chapters*. https://doi.org/10.17848/9781417505333.ch1

Lalive, R., & Zweimüller, J. (2009). How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments [Publisher: Oxford University Press]. *The Quarterly Journal of Economics*, *124*(3), 1363–1402. Retrieved June 15, 2022, from https://www.jstor.org/stable/40506259

Lambert, E. Y. (1990). *The collection and interpretation of data from hidden populations* (Vol. 98). US Department of Health; Human Services, Public Health Service, Alcohol ...

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011). Development of youth percent body fat standards using receiver operating characteristic curves. *American Journal of Preventive Medicine*, *41*(4 Suppl 2), S93–99. https://doi.org/10.1016/j.amepre.2011.07.003

Lavelle, H. V., Mackay, D. F., & Pell, J. P. (2012). Systematic review and meta-analysis of school-based interventions to reduce body mass index [Publisher: Oxford Academic]. *Journal of Public Health*, *34*(3), 360–369. https://doi.org/10.1093/pubmed/fdr116

Lee, B. R., Bright, C. L., Svoboda, D. V., Fakunmoju, S., & Barth, R. P. (2011). Outcomes of group care for youth: A review of comparative studies. *Research on Social Work Practice*, *21*(2), 177–189. https://doi.org/10.1177/1049731510386243

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. https://ideas.repec.org/c/boc/bocode/s432001.html

Lovett, N., & Xue, Y. (2020). Family first or the kindness of strangers? Foster care placements and adult outcomes. *Labour Economics*, *65*, 101840. https://doi.org/10.1016/j.labeco.2020.101840

McDonald, T. P., et al. (1996). *Assessing the long-term effects of foster care: A research synthesis.* ERIC.

Meekes, J., & Hassink, W. (2020). Fired and pregnant: Gender differences in job flexibility outcomes after job loss. *Life Course Centre Working Paper*, (2020-06).

Miller, S., Wherry, L. R., & Foster, D. G. (2023). The Economic Consequences of Being Denied an Abortion. *American Economic Journal: Economic Policy*, *15*(1), 394–437. https://doi.org/10.1257/pol.20210159

Mueller, A. I., & Spinnewijn, J. (2022). Expectations data, labor market and job search. https://personal.lse.ac.uk/spinnewi/HB_expectations.pdf

Mura, G., Rocha, N. B., Helmich, I., Budde, H., Machado, S., Wegner, M., Nardi, A. E., Arias-Carrión, O., Vellante, M., Baum, A., Guicciardi, M., Patten, S. B., & Carta, M. G. (2015). Physical Activity Interventions in Schools for Improving Lifestyle in European Countries. *Clinical Practice and Epidemiology in Mental Health : CP & EMH*, *11*(Suppl 1 M5), 77–101. https://doi.org/10.2174/1745017901511010077

Nagaoka, J., Farrington, C. A., Ehrlich, S. B., & Heath, R. D. (2015). *Foundations for young adult success: A developmental framework. concept paper for research and practice.* ERIC.

Nelson, C. A. (2014). *Romania's abandoned children*. Harvard University Press.

Ng, C. D. (2019). Biases in self-reported height and weight measurements and their effects on modeling health outcomes. *SSM - Population Health*, *7*. https://doi.org/10.1016/j.ssmph.2019.100405

OECD. (2022). *OECD Family Database* (Database). https://www.oecd.org/els/family/database.htm

Overgaard, D., Gyntelberg, F., & Heitmann, B. (2004). Psychological workload and body weight: Is there an association? A review of the literature. *Occupational medicine (Oxford, England)*, *54*, 35–41. https://doi.org/10.1093/occmed/kqg135

Packham, A., & Street, B. (2019). The effects of physical education on student fitness, achievement, and behavior. *Economics of Education Review*, *72*, 1–18. https://doi.org/10.1016/j.econedurev.2019.04.003

Perna, F. M., Oh, A., Chriqui, J. F., Mâsse, L. C., Atienza, A. A., Nebeling, L., Agurs-Collins, T., Moser, R. P., & Dodd, K. W. (2012). The Association of State Law to Physical Education Time Allocation in US Public Schools [Publisher: American Public Health Association]. *American Journal of Public Health*, *102*(8), 1594–1599. https://doi.org/10.2105/AJPH.2011.300587

Petrowski, N., Cappa, C., & Gross, P. (2017). Estimating the number of children in formal alternative care: Challenges and results. *Child Abuse & Neglect*, *70*, 388–398. https://doi.org/10.1016/j.chiabu.2016.11.026

Ryan, J. P., Marshall, J. M., Herz, D., & Hernandez, P. M. (2008). Juvenile delinquency in child welfare: Investigating group home effects. *Children and Youth Services Review*, *30*(9), 1088–1099.

Sabia, J. J., Nguyen, T. T., & Rosenberg, O. (2017). High School Physical Education Requirements and Youth Body Weight: New Evidence from the YRBS. *Health Economics*, *26*(10), 1291–1306. https://doi.org/10.1002/hec.3399

Schwerdt, G. (2011). Labor turnover before plant closure: "Leaving the sinking ship" vs. "Captain throwing ballast overboard". *Labour Economics*, *18*(1), 93–101. https://doi.org/10.1016/j.labeco.2010.08.003

Sebők, A. (2019). The panel of linked administrative data of cers databank. *Budapest Working Papers On The Labour Market*, (2). https://www.mtakti.hu/wp-content/uploads/2019/12/BWP1902.pdf

Singh, S., Remez, L., Sedgh, G., Kwok, L., & Onda, T. (2018). Abortion Worldwide 2017: Uneven Progress and Unequal Access. https://doi.org/10.1363/2018.29199

Skinner, A. C., Perrin, E. M., Moss, L. A., & Skelton, J. A. (2015). Cardiometabolic Risks and Severity of Obesity in Children and Young Adults. *New England Journal of Medicine*, *373*(14), 1307–1317. https://doi.org/10.1056/NEJMoa1502821

Smyke, A. T., Zeanah, C. H., Fox, N. A., Nelson, C. A., & Guthrie, D. (2010). Placement in foster care enhances quality of attachment among young institutionalized children. *Child Development*, *81*(1), 212–223. https://doi.org/https://doi.org/10.1111/j.1467-8624.2009.01390.x

SOS. (2015). 15 anomália a gyermekvédelemben. Retrieved July 3, 2023, from https://www.sos.hu/gyereksorsok/15-anomalia-a-gyermekvedelemben/

Stevens, A. H. (1997). Persistent Effects of Job Displacement: The Importance of Multiple Job Losses [Publisher: [University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago]]. *Journal of Labor Economics*, *15*(1), 165–188. Retrieved January 4, 2023, from https://www.jstor.org/stable/2535319

Sturm, R. (2004). Childhood Obesity — What We Can Learn From Existing Data on Societal Trends, Part 1. *Preventing Chronic Disease*, *2*(1). Retrieved March 23, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1323315/

Taber, D. R., Chriqui, J. F., Perna, F. M., Powell, L. M., Slater, S. J., & Chaloupka, F. J. (2013). Association between state physical education (PE) requirements and PE participation, physical activity, and body mass index change. *Preventive Medicine*, *57*(5), 629–633. https://doi.org/10.1016/j.ypmed.2013.08.018

UNICEF, E. (2022). *Better data for better child protection systems in europe: Mapping how data on children in alternative care are collected, analysed and published across 28 european countries - technical report of the datacare project* (tech. rep.). UNICEF, Eurochild.

van Noord-Zaadstra, B. M., Looman, C. W., Alsbach, H., Habbema, J. D., te Velde, E. R., & Karbaat, J. (1991). Delaying childbearing: Effect of age on fecundity and outcome of pregnancy. *BMJ (Clinical research ed.)*, *302*(6789), 1361–1365. https://doi.org/10.1136/bmj.302.6789.1361

Vígh, K. (2015). Törvényi változások a gyakorlatban–gyermekvédelmi szakellátás a fővárosban 2014 után.

Wade, M., McLaughlin, K. A., Buzzell, G. A., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2023). Family-based care buffers the stress sensitizing effect of early deprivation on executive functioning difficulties in adolescence. *Child Development*, *94*(1), e43–e56. https://doi.org/https://doi.org/10.1111/cdev.13863

Wade, M., Zeanah, C. H., Fox, N. A., Tibu, F., Ciolan, L. E., & Nelson, C. A. (2019). Stress sensitization among severely neglected children and protection by social enrichment [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *10*(1), 5771. https://doi.org/10.1038/s41467-019-13622-3

Wall-Wieler, E., Brownell, M., Singal, D., Nickel, N., & Roos, L. L. (2018). The cycle of child protection services involvement: A cohort study of adolescent mothers. *Pediatrics*, *141*(6).

Wickelgren, I. (1998). Obesity: How Big a Problem? *Science*, *280*(5368), 1364–1367. https://doi.org/10.1126/science.280.5368.1364

A Appendix for Chapter 1

A.1 Supplementary tables and figures

Table A.1: Number of births and abortions in official statistics and in our data

Year	Official number of abortions (KSH, 2022b)	Official number of births (KSH, 2022a)	Expected number of abortions in 50% admin data	Expected number of live births in 50% admin data	Observed number of abortions in 50% admin data	Observed number of live births in 50% admin data	Observed abortions (%)	Observed live births (%)
2009	43181	94707	21590.5	47353.5	20921	43464	97	92
2010	40449	88758	20224.5	44379	19406	41148	96	93
2011	38443	86632	19221.5	43316	18387	39388	96	91
2012	36118	88783	18059	44391.5	17592	40088	97	90
2013	34891	87189	17445.5	43594.5	17066	38928	98	89
2014	32663	90010	16331.5	45005	15709	39814	96	88
2015	31176	90190	15588	45095	14947	39649	96	88
2016	30439	91563	15219.5	45781.5	14453	39519	95	86
2017	28496	90077	14248	45038.5	13522	38615	95	86

Note: The number of births is corrected by twin births.

Hungarian DOI: 10.14754/CEU.2023.04

Hungarian DOI: 10.14754/CEU.2023.04

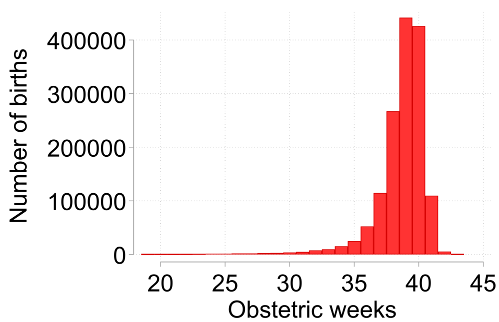
Table A.2: Child benefit rules

State child benefit	Availability at child age	Eligibility	Monthly sum	Monthly average in 2009[d]
Baby-care allowance[a]	0 to 0.5	employed at giving birth; worked at least 360 days in the past two years	70% of the previous wage	HUF 110,411 (USD 368)
Childcare benefit[b]	0.5 to 2	employed at giving birth; worked at least 360 days in the past two years	70% of the previous wage, maximum HUF 100,000 (about USD 334)	HUF 91,050 (USD 303)
Baby-care allowance[a]	0 to 0.5	on job search subsidy at giving birth; worked at least 360 days in the past two years	70% of the minimum wage	HUF 50,050 (USD 166)
Childcare allowance[c]	0 to 3	worked less than 360 days in the past two years	The amount of minimum pension	HUF 28,500 (USD 95)

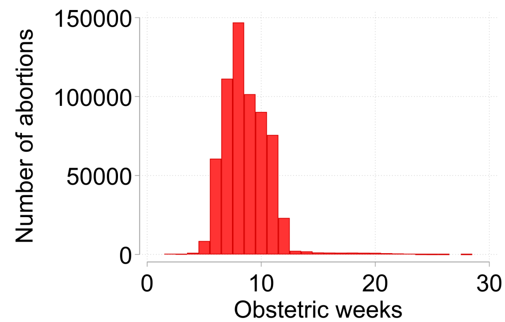
(a) Csecsemőgondozási díj (CSED), Terhességi-gyermekágyi segély (TGYAS) before 2015
Gyermekgondozási díj (GYED)
(c) Gyermekgondozást segítő ellátás (GYES)
(d) Based on data of the Hungarian Central Statistical Office

Figure A.1: Obstetric weeks of births and abortions

(a) Births



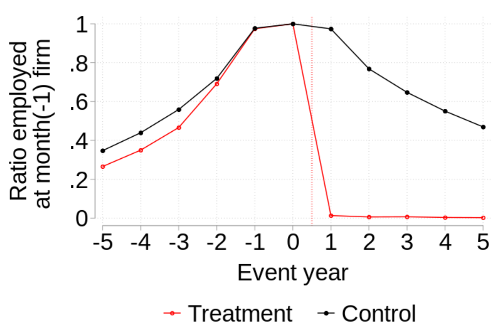
(b) Abortions



Source: Live birth and terminated birth databases of the Hungarian Statistical Office

Figure A.2: Percent working at the same firm as in event year 0 in the treated and the control groups

(a) Closure



(b) Mass layoff

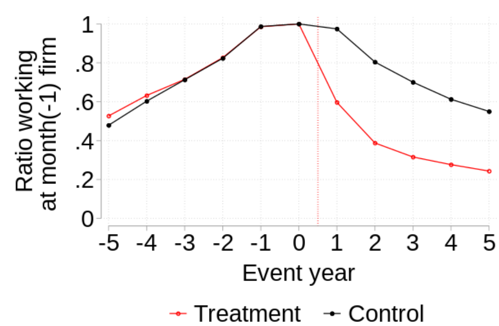
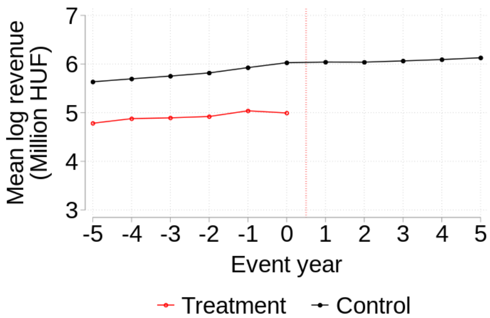
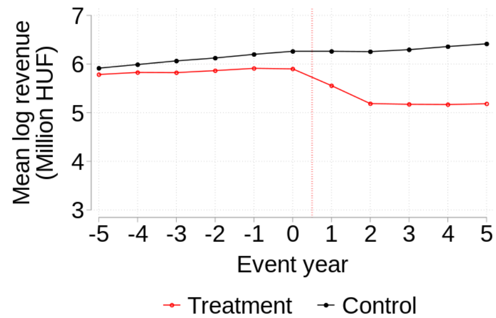


Figure A.3: Firm revenues around the layoff event

(a) Closure



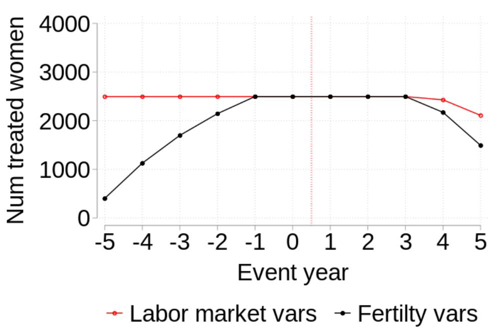
(b) Mass layoff



Firm revenues are available at a yearly frequency. Event year 0 is the calendar year of the layoff event. For control firms, the date of the pseudo-event is set to the year when the most control women are matched

Figure A.4: Number of observations in the treated groups by event year

(a) Closure



(b) Mass layoff

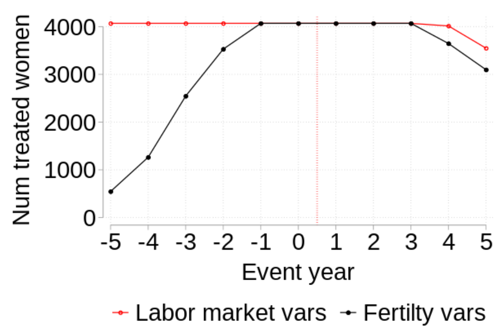
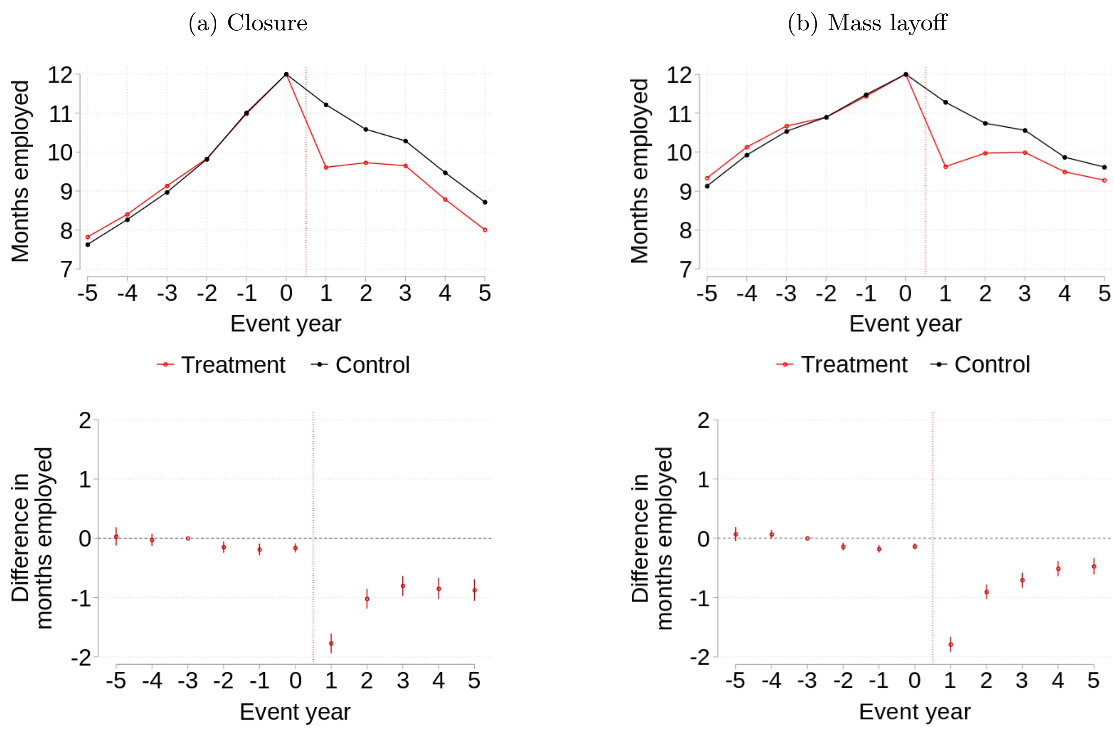


Figure A.5: Months spent employed in the treatment and control group before and after the shocks: raw means and regression estimates

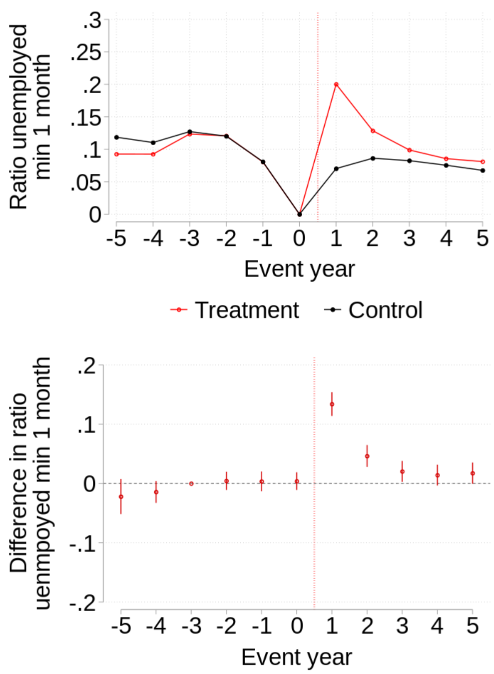


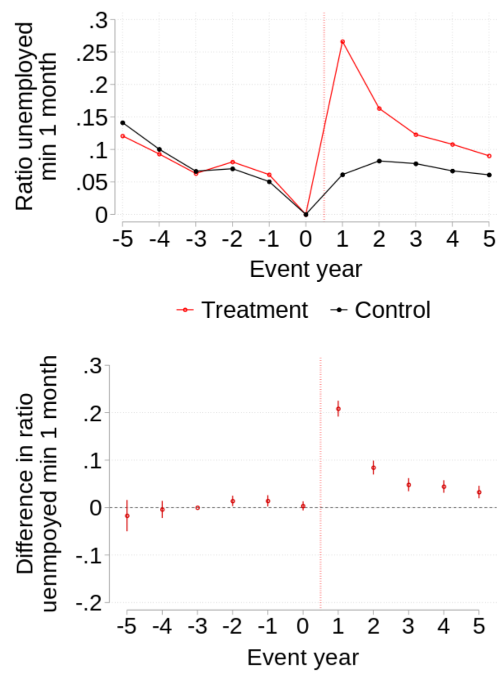
The last month of event year 0 is the time of matching. Number of observations by event years is shown on Figure A.4

Figure A.6: Unemployment in the treatment and control group before and after the shocks: raw means and regression estimates

(a) Closure

(b) Mass layoff

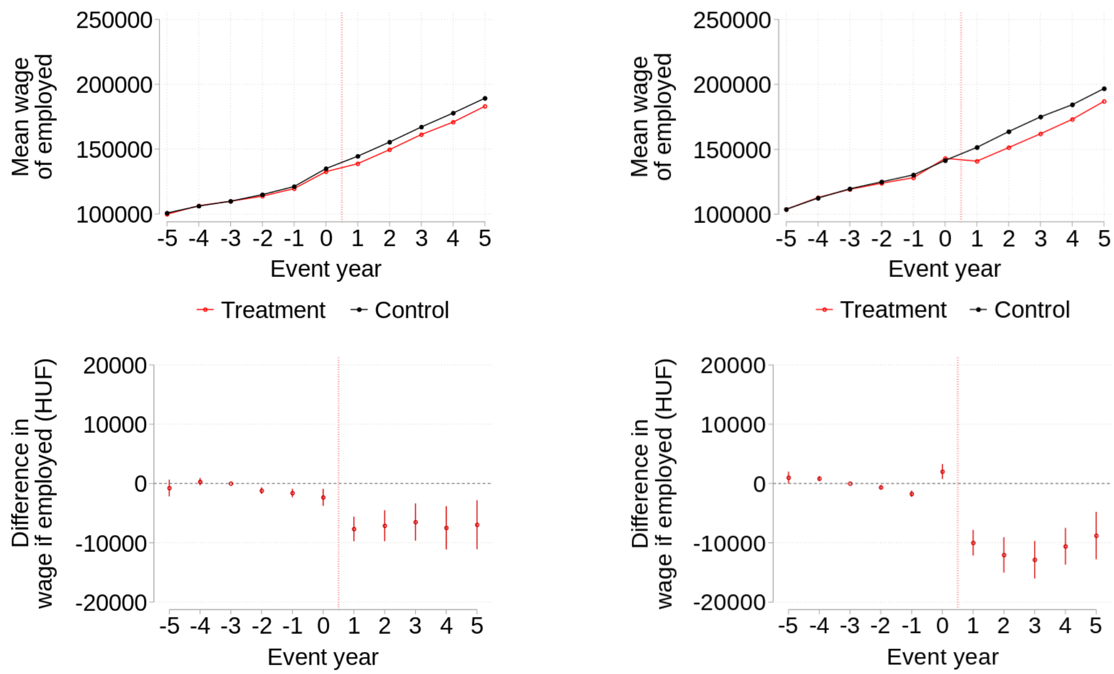




The last month of event year 0 is the time of matching. Number of observations by event years is shown in Figure A.4A.4

Figure A.7: Wages of working women in the treatment and control group before and after the shocks: raw means and regression estimates

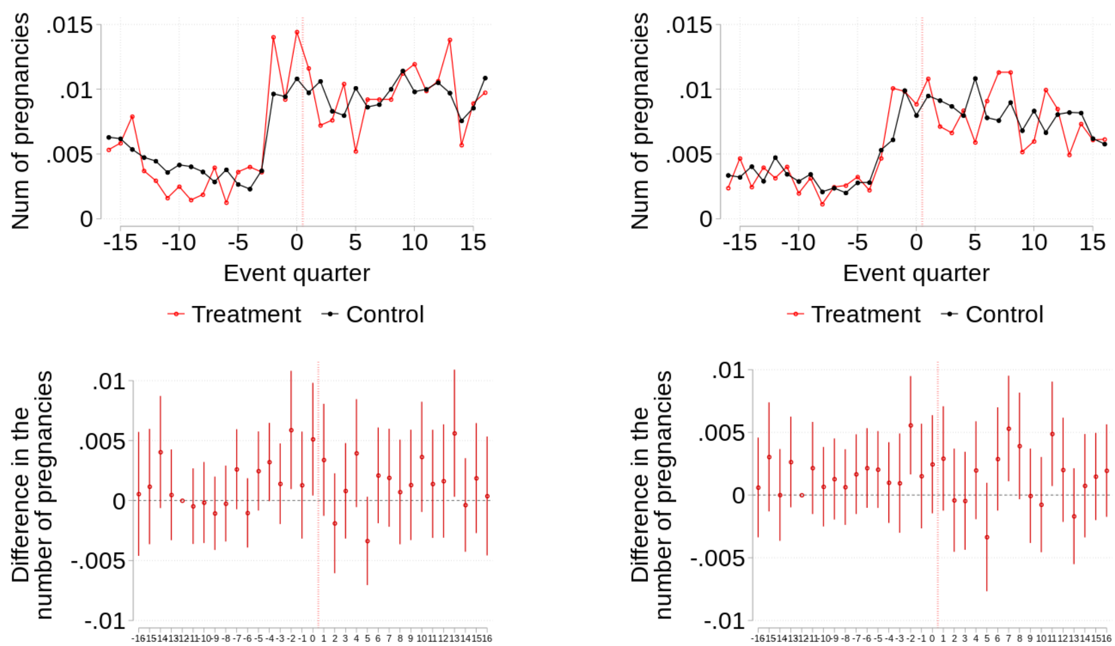
(a) Closure (b) Mass layoff



The last month of event year 0 is the time of matching. Number of observations by event years is shown on Figure A.4

Figure A.8: Pregnancies: raw means in the treatment and the control group and regression estimates by event quarter

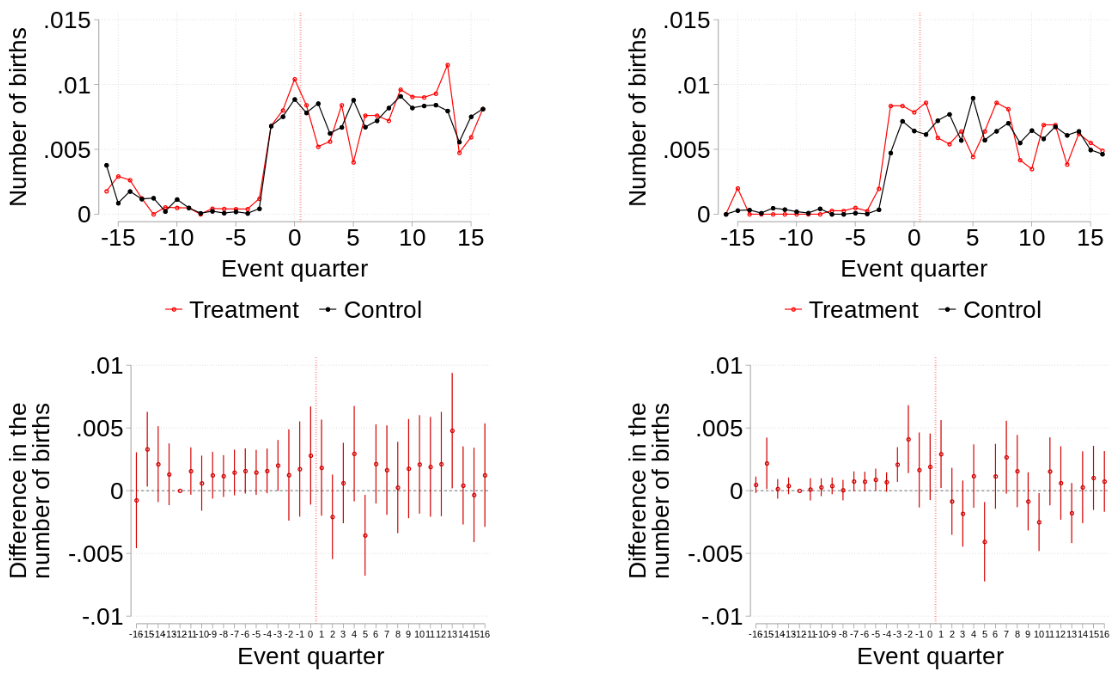
(a) Closure (b) Mass layoff



The red vertical line indicates the time of the layoff event. The last month of event quarter 0 is the time of matching. The number of observations by event years is shown in Figure A.4. The pregnancies, births, and abortions are counted in the quarter of conception.

Figure A.9: Births: raw means in the treatment and the control group and regression estimates by event quarter

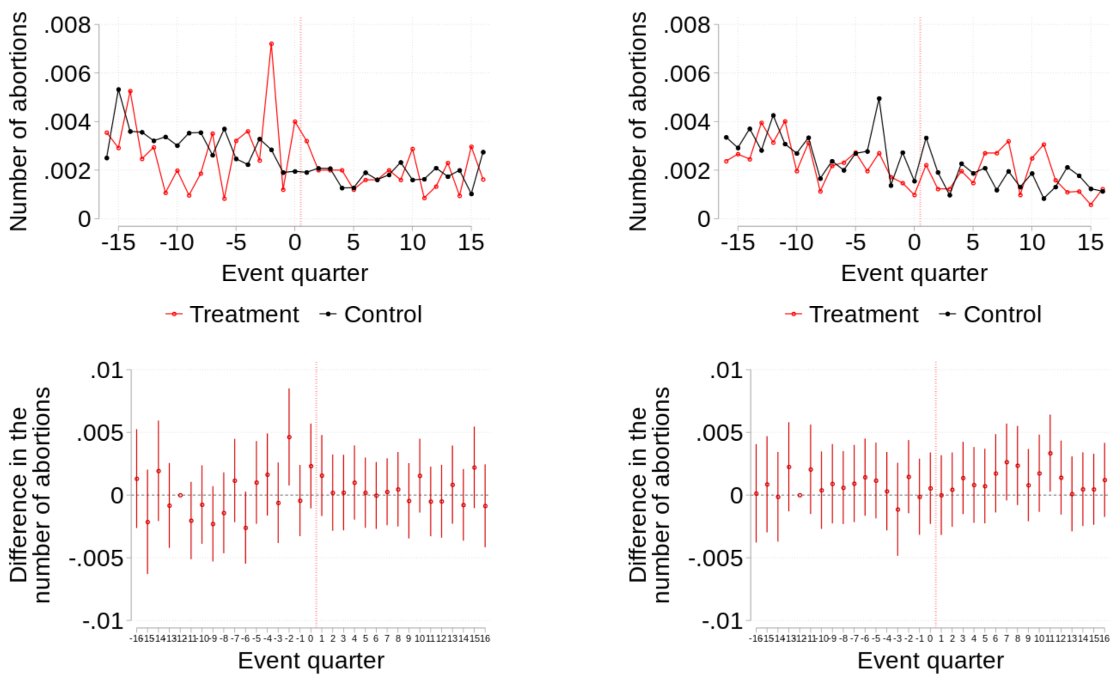
(a) Closure (b) Mass layoff



The red vertical line indicates the time of the layoff event. The last month of event quarter 0 is the time of matching. The number of observations by event years is shown in Figure A.4. The pregnancies, births, and abortions are counted in the quarter of conception.

Figure A.10: Abortions: raw means in the treatment and the control group and regression estimates by event quarter

(a) Closure (b) Mass layoff



The red vertical line indicates the time of the layoff event. The last month of event quarter 0 is the time of matching. The number of observations by event years is shown in Figure A.4. The pregnancies, births, and abortions are counted in the quarter of conception.

Table A.3: Three-period DID estimates for the effects of employment shocks on labor market outcomes

Sample	Closure		Mass layoff	
Outcome	Works	Wage	Works	Wage
	(1)	(2)	(3)	(4)
Treated	0.014***	0.062*	0.011***	0.034
	(0.005)	(0.037)	(0.003)	(0.027)
Year0	0.359***	4.639***	0.224***	3.592***
	(0.008)	(0.090)	(0.005)	(0.062)
Post1	0.149***	5.359***	0.038***	4.445***
	(0.007)	(0.110)	(0.005)	(0.101)
Treated X Year0	-0.005	-0.176**	-0.011***	0.182**
	(0.004)	(0.083)	(0.003)	(0.074)
Treated X Post1	-0.148***	-1.540***	-0.153***	-2.134***
	(0.009)	(0.163)	(0.007)	(0.145)
Exact matched set FE	YES	YES	YES	YES
Propensity score	YES	YES	YES	YES
Bootstrapped p-value of Treated X Year0	0.255	0.028	0.002	0.016
Bootstrapped p-value of Treated X Post1	0.000	0.000	0.000	0.000
R-squared	0.290	0.673	0.250	0.708
Pre-treatment mean in control group	0.672	8.728	0.851	10.569
Observations	174,204		214,479	
N treated	2496		4068	
N control	16860		19763	

Note: Standard errors clustered by exact match set in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Monthly wage is measured in 10000 HUF.

Hungarian DOI: 10.14754/CEU.2023.04

Table A.4: DID regression results for the net effect of closures and mass layoffs on the number of births, abortions, and pregnancies

	Closure			Mass Layoff		
	Births	Abortions	Pregnancies	Births	Abortions	Pregnancies
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.000	-0.002	-0.002	-0.000	-0.000	-0.001
	(0.001)	(0.002)	(0.002)	(0.000)	(0.001)	(0.001)
Year 0-3	0.023***	-0.004***	0.019***	0.023***	-0.002**	0.021***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Treated X Year 0-3	0.000	0.004**	0.004	0.001	0.000	0.002
	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)
R-squared	0.073	0.057	0.074	0.086	0.061	0.083
Exact matched set FE	YES	YES	YES	YES	YES	YES
Propensity score	YES	YES	YES	YES	YES	YES
Bootstrapped p-value of Treated x After	0.912	0.02	0.116	0.391	0.731	0.393
Pre-treatment mean in control group	0.003	0.012	0.015	0.001	0.009	0.01
Observations		136,647			164,047	
N treated		2496			4068	
N control		16860			19763	

Note: Standard errors clustered by exact match set in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.5: Three period DID regression results for the effect of closures and mass layoffs on fertility outcomes for women without previous births

	Closure			Mass layoff		
	(1)	(2)	(3)	(4)	(5)	(6)
	Pregnancies	Births	Abortions	Pregnancies	Births	Abortions
Treated	0.001	0.001	-0.000	0.001	0.001*	-0.000
	(0.002)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)
Year 0	0.023***	0.024***	-0.001	0.021***	0.019***	0.002
	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)
Year 1-3	0.032***	0.035***	-0.003***	0.026***	0.029***	-0.003***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Treated X Year 0	0.010*	0.002	0.008**	0.004	0.009**	-0.004*
	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)
Treated X Year 1-3	-0.004	-0.004	0.000	-0.001	-0.003*	0.002
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)
Observations	99,332	99,332	99,332	126,505	126,505	126,505
R-squared	0.087	0.084	0.077	0.093	0.096	0.067
Exact matched set FE	YES	YES	YES	YES	YES	YES
Propensity score	YES	YES	YES	YES	YES	YES
p-value of bootstraped t x year 0	.056	.58	.02	.305	.012	.069
p-value of bootstraped t x year 1-3	.214	.172	.914	.674	.106	.133
Pre-treatment mean in control group	.008	0	.008	.007	0	.007
N treated	1838	1838	1838	3184	3184	3184
N control	12401	12401	12401	15387	15387	15387

Note: Standard errors clustered by exact match set in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates from regression Eq. 1.3 on the subsample of women who have no children (defined by the absence of any past child tarnsfer). Births and abortions are measured at the estimated times of conception. Pregnancies are the sum of births and abortions.

Table A.6: Three period DID regression results for the effect of closures and mass layoffs on fertility outcomes for women who have a child

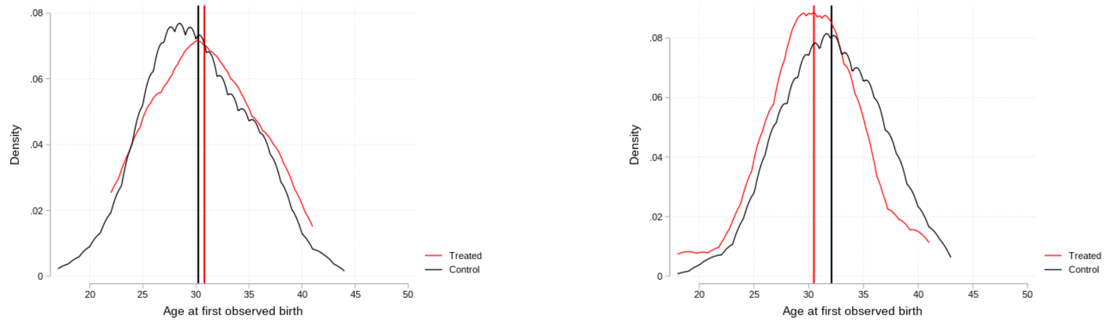
	Closure			Mass layoff		
	(1)	(2)	(3)	(4)	(5)	(6)
	Pregnancies	Births	Abortions	Pregnancies	Births	Abortions
Treated	-0.009*	-0.002	-0.007*	-0.005	-0.002	-0.003
	(0.005)	(0.003)	(0.004)	(0.004)	(0.002)	(0.004)
Year 0	-0.002	0.003	-0.005	0.008	0.012***	-0.004
	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)	(0.003)
Year 1-3	-0.009***	0.001	-0.010***	0.007*	0.011***	-0.004
	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)	(0.003)
Treated X Year 0	0.010	0.005	0.005	0.006	0.006	0.000
	(0.010)	(0.008)	(0.007)	(0.008)	(0.006)	(0.006)
Treated X Year 1-3	0.017***	0.007	0.010**	0.007	0.007	0.000
	(0.006)	(0.005)	(0.004)	(0.006)	(0.004)	(0.005)
Observations	37,315	37,315	37,315	37,542	37,542	37,542
R-squared	0.107	0.097	0.086	0.121	0.119	0.106
Exact matched set FE	YES	YES	YES	YES	YES	YES
Propensity score	YES	YES	YES	YES	YES	YES
p-value of bootstraped t x year0	.321	.467	.479	.411	.327	.979
p-value of bootstraped t x post1	.005	.137	.022	.272	.084	.972
Pre-treatment mean in control group	.034	.012	.022	.018	.002	.016
N treated	658	658	658	884	884	884
N control	4459	4459	4459	4376	4376	4376

Note: Standard errors clustered by exact match set in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates from regression Eq. 1.3 on the subsample of women who have no children (defined by the absence of any past child transfer). Births and abortions are measured at the estimated times of conception. Pregnancies are the sum of births and abortions.

Hungarian DOI: 10.14754/CEU.2023.04

Figure A.11: Age at first observed birth

(a) Closure (b) Mass layoff



Note: Kernel densities for the distribution age at first observed birth in the period of 2009-2017 for women who give birth in this period. Means are shown with vertical lines.

Figure A.12: Births and abortions by age for women in white collar and blue collar occupations

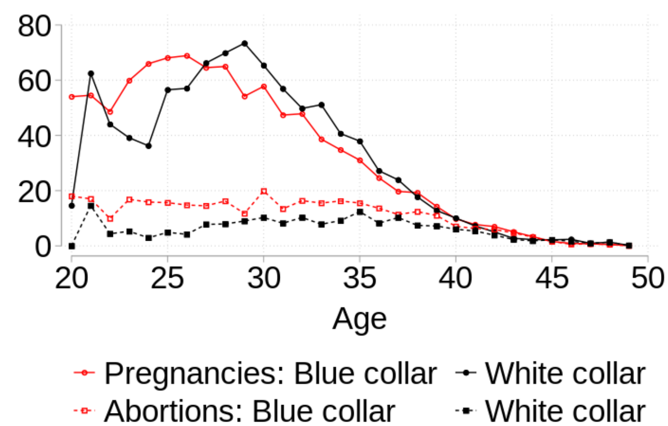


Table A.7: Three-period DID regression results in the pooled sample

	(1) Pregnancies	(2) Births	(3) Abortions
Closure	-0.004**	-0.002**	-0.002
	(0.002)	(0.001)	(0.001)
Mass Layoff	-0.002	-0.001*	-0.001
	(0.001)	(0.000)	(0.001)
Year 0	0.020***	0.021***	-0.001
	(0.002)	(0.001)	(0.001)
Year 1-3	0.028***	0.030***	-0.003**
	(0.002)	(0.002)	(0.001)
Closure X Year 0	0.010**	0.004	0.006**
	(0.004)	(0.003)	(0.003)
Closure X Year 1-3	0.002	0.000	0.002
	(0.003)	(0.002)	(0.002)
Mass Layoff X Year 0	0.005	0.008***	-0.002
	(0.003)	(0.003)	(0.002)
Mass Layoff X Year 1-3	0.001	-0.001	0.002
	(0.002)	(0.002)	(0.001)
R-squared	0.006	0.010	0.001
Bootstrapped p-value if Closure X Year 0	0.025	0.237	0.034
Bootstrapped p-value if Closure X Year 1-3	0.441	0.966	0.22
Bootstrapped p-value if Mass layoff X Year 0	0.118	0.007	0.207
Bootstrapped p-value if Mass Layoff X Year 1-3	0.68	0.415	0.097
Exact matched set FE	NO	NO	NO
Propensity score	YES	YES	YES
Calendar year FE	YES	YES	YES
Observations		300,694	

Note: Standard errors clustered by exact match set in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.2 Measurement error due to unobserved miscarriages

In the data, we do not have accurate information on miscarriages so our measure of pregnancies defined as the number of births plus the number of abortions is measured with error. Here we assess the potential bias of our results due to this measurement error. The main concern is that an increase in abortion will mechanically increase observed pregnancies if some of the aborted pregnancies would have been miscarriages.

Let the true number of pregnancies be P and assume it is not changed by a job displacement. We call $\tilde{P}$ the number of observed pregnancies, that is births plus abortions. The share of miscarriages among all pregnancies is m and a is the share of abortions. If there are no abortions $\tilde{P} = (1-m)P$. In case there are abortions

$$\tilde{P} = (a + (1-m)(1-a))P = (1 + am - m)P$$

This assumes all abortions happen before a miscarriage and only pregnancies that are not aborted are at risk of miscarriage.

We assume that the only difference between control and displaced women is the rate of abortions $a_0 \neq a_1$ and everything else is the same for both groups. In this case, we get

$$\begin{aligned} \Delta A &= P(a_1 - a_0) \\ \Delta \tilde{P} &= Pm(a_1 - a_0) \\ \frac{\Delta \tilde{P}}{\Delta A} &= m \end{aligned}$$

If $m = 0.1$, meaning that 10% of all pregnancies result in a miscarriage, an increase in the number of abortions by 10 would result in a mechanical increase in the number of observed pregnancies of 1. This calculation indicates that the implied mechanical increase of abortions from unobserved miscarriages is too small to explain the estimated effect of job displacement on observed pregnancies.

A.3 Theoretical model derivations

I. General patterns

The timing of information and decisions within a period:

1. Women start as employed (E) or unemployed (U), depending on getting hired or fired at the end of the previous period
2. Women learn if the firm is in trouble and form their expectations on the probability of a within-period layoff: q_p for pregnant and q_n for non-pregnant
3. Women decide about pregnancy probability (p_0 or p_1 with $p_0 < p_1$)
4. Women get pregnant with probability p_0 or p_1 and learn about their pregnancy status
5. Women update their expectations about the probability of a within-period layoff (updated q_p and q_n)
6. Women decide on abortion if pregnant
7. Flow payoffs are realized: w for the employed, z for the unemployed (with $w > z$), and additional $B(\theta)$ if getting and saying pregnant or $-C(\theta)$ if getting pregnant but aborting (B is the discounted net value of having a child, including non-monetary costs and benefits, $C \geq 0$ is abortion cost, including non-monetary costs as well, both $B(\theta)$ and $C(\theta)$ increase in heterogeneity parameter θ)
8. For the employed within-period layoffs are realized with actual probability q_p^a or q_n^a if the firm is in trouble, these are zero in normal times
9. Women get hired with probability h_p (if stayed pregnant) or h_n (if not pregnant or aborted) if started as unemployed or were laid off within the period, and get laid off with probability f_p (if stayed pregnant) or f_n (if not regnant or aborted) if started as employed and were not laid off within the period, assuming $h_p < h_n$, $f_p \leq f_n$ and $h_n + f_n \leq 1$

The value function for an employed woman with heterogeneity parameter θ in scenario s (where $E_s(\theta) = E(\theta)$ is the baseline scenario) with discount rate r is

$$rE_s(\theta) = w + (1-p)V_n + p\max\{B(\theta) + V_p, -C(\theta) + V_n\} + E(\theta) - E_s(\theta) \tag{A.1}$$

with

$$V_n = (q_n(1-h_n) + (1-q_n)f_n)(U(\theta) - E(\theta)) \tag{A.2}$$

$$V_p = (q_p(1-h_p) + (1-q_p)f_p)(U(\theta) - E(\theta)) \tag{A.3}$$

The value function for an unemployed woman with heterogeneity parameter θ and discount rate r is

 Hungarian DOI: 10.14754/CEU.2023.04

$$rU(\theta) = z + (1-p)\,Y_n + p\max\{B(\theta) + Y_p, -C(\theta) + Y_n\} \tag{A.4}$$

with

$$Y_n = h_n\,(E(\theta) - U(\theta)) \tag{A.5}$$

$$Y_p = h_p\,(E(\theta) - U(\theta)) \tag{A.6}$$

Proposition 1. *A woman who chooses abortion when she starts the period as being employed will also choose abortion when she starts the period as being unemployed.*

Proof. Women decide to abort when employed if

$$B(\theta) + C(\theta) < ((1 - h_n - f_n)(q_n - q_p) + (1 - q_p)(f_n - f_p) + q_p(h_p - h_n))(U(\theta) - E(\theta)) \tag{A.7}$$

Women decide to abort when unemployed if

$$B(\theta) + C(\theta) < (h_p - h_n)(U(\theta) - E(\theta)) \tag{A.8}$$

If θ satisfies inequality A.7, it also satisfies inequality A.8, as with $E(\theta) > U(\theta)$ the following inequality holds:

$$((1 - h_n - f_n)(q_n - q_p) + (1 - q_p)(f_n - f_p) + q_p(h_p - h_n))(U(\theta) - E(\theta)) < (h_p - h_n)(U(\theta) - E(\theta))$$

We will show that $E(\theta) > U(\theta)$ holds $\forall\theta$. □

Proposition 2. *Intended pregnancies are never aborted if q_n and q_p do not change.*

Proof. If a woman chooses to abort when becoming pregnant, then she is necessarily better off when she does not become pregnant, as $C(\theta) > 0$. Consequently, a woman who would want to abort upon becoming pregnant has no reason to increase the probability of becoming pregnant. □

Proposition 3. *Some of the unintended pregnancies will not be aborted even if q_n and q_p do not change.*

Proof. An employed woman is better off not getting pregnant, but she is also better off keeping the child upon becoming pregnant if two inequalities hold at the same time:

$$(q_n(1 - h_n) + (1 - q_n)f_n)(U(\theta) - E(\theta)) > B(\theta) + (q_p(1 - h_p) + (1 - q_p)f_p)(U(\theta) - E(\theta)) \tag{A.9}$$

$$B(\theta) + (q_p(1 - h_p) + (1 - q_p)f_p)(U(\theta) - E(\theta)) > -C(\theta) + (q_n(1 - h_n) + (1 - q_n)f_n)(U(\theta) - E(\theta)) \tag{A.10}$$

$\exists\theta$ satisfying both inequalities, as $C(\theta) > 0, \forall\theta$. Given inequality A.9, a woman with such θ does not increase her pregnancy probability, and her pregnancy will be unintended. Given inequality A.10 the woman will still be better off keeping the child upon becoming pregnant due to the high abortion cost. The same is true for an unemployed woman with θ for which the following holds:

$$h_n(E(\theta) - U(\theta)) > B(\theta) + h_p(E(\theta) - U(\theta)) \tag{A.11}$$

$$B(\theta) + h_p(E(\theta) - U(\theta)) > -C(\theta) + h_n(E(\theta) - U(\theta)) \tag{A.12}$$

□

Proposition 4. *A woman who starts the period as being employed and chooses not to increase her pregnancy probability would make the same decision if she started the period as being unemployed.*

Proof. An employed woman chooses not to increase her pregnancy probability if inequality A.9 holds. An unemployed woman makes the same decision if inequality A.11 hods. Given our assumptions on the parameters, if a θ satisfies inequality A.9, with $U(\theta) < E(\theta)$ it also satisfies inequality A.11, as

$$(1 - h_p - f_p)(1 - q_p) - (1 - h_n - f_n)(1 - q_n) > 0$$

□

II. Baseline scenario

In the baseline scenario with $q_n = q_p = 0$, women with $\theta < \underline{\theta}$ will always abort, with $\underline{\theta} < \theta < \overline{\theta}$ only abort if being unemployed and with $\overline{\theta} < \theta$ will never abort. Similarly, women with $\theta < \hat{\theta}$ will never increase their pregnancy probability, with $\hat{\theta} < \theta < \ddot{\theta}$ only increase their pregnancy probability if being employed and with $\ddot{\theta} < \theta$ will always increase their pregnancy probability. With all parameter values satisfying the initial assumptions, we have $\underline{\theta} < \hat{\theta} < \ddot{\theta}$ and $\underline{\theta} < \overline{\theta} < \ddot{\theta}$.

III. General scenario

A specific scenario only affects employed women, but not the unemployed.

Proposition 5. *An employed woman with θ will keep the child in any scenario if $\theta > \overline{\theta}$, i.e. if she would keep it even when being unemployed.*

Proof. $\theta > \overline{\theta}$ always satisfies the condition for keeping the baby in any scenario with $0 \leq q_p \leq q_n \leq 1$:

$$B(\theta) + C(\theta) > ((1 - h_n - f_n)(q_n - q_p) + (1 - q_p)(f_n - f_p) + q_p(h_p - h_n))(U(\theta) - E(\theta)) \quad \text{(A.13)}$$

The inequality holds for $\theta = \overline{\theta}$ and due to monotonicity, $B(\theta) + C(\theta) > B(\overline{\theta}) + C(\overline{\theta})$ if $\theta > \overline{\theta}$.

□

Proposition 6. *An employed woman with θ will increase her pregnancy probability in any scenario if $\theta > \ddot{\theta}$, i.e. if she would increase her pregnancy probability even when being unemployed.*

Proof. $\theta > \ddot{\theta}$ always satisfies the condition for increasing pregnancy probability in any scenario with $0 \leq q_p \leq q_n \leq 1$:

$$B(\theta) > (q_n(1 - h_n) + (1 - q_n)f_n - (q_p(1 - h_p) + (1 - q_p)f_p))(U(\theta) - E(\theta)) \quad \text{(A.14)}$$

The inequality holds for $\theta = \ddot{\theta}$ and due to monotonicity, $B(\theta) > B(\ddot{\theta})$ if $\theta > \ddot{\theta}$.

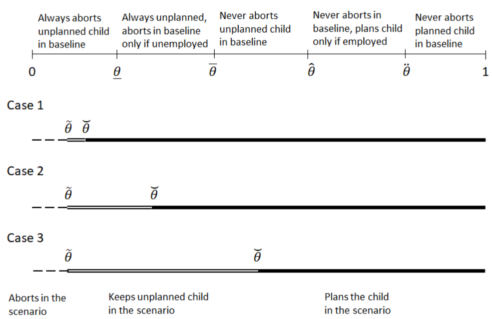
□

As for the baseline scenario, we can define cutoff values for the abortion and pregnancy probability increase decisions: women with $\theta < \widetilde{\theta}_s$ will always abort. Similarly, women with $\theta < \breve{\theta}_s$ will never increase their pregnancy probability.

Figures A.13 and A.14 summarize the potential order of the different abortion and planned pregnancy cutoffs in the various scenarios. The potential order is similar for the closure and mass layoff with no dismissal protection scenarios but it is different for the mass layoff with a full dismissal protection scenario. In the case of a mass layoff with partial dismissal protection, any of the presented cutoff orderings are possible.

Figure A.13: Cutoffs for abortion and planned pregnancies in the mass layoff with full or partial dismissal protection scenarios

(a) Closure



(b) Mass layoff

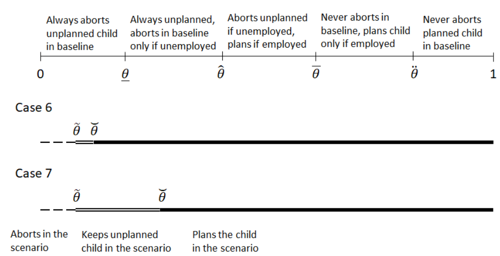
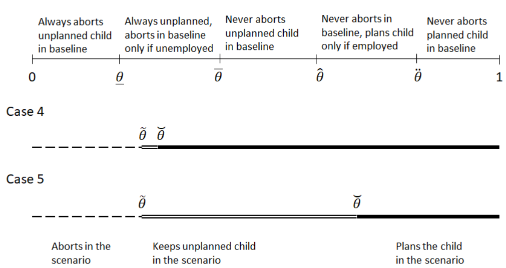
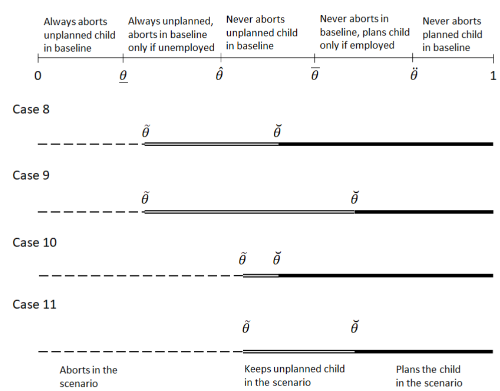


Figure A.14: Cutoffs for abortion and planned pregnancies in the closure and mass layoff with no or partial dismissal protection scenarios

(a) Closure



(b) Mass layoff



IV. Number of abortions and births

The number of abortions per N women is given by $p_0\underline{\theta}N$ for the baseline scenario and $p_0\widetilde{\theta}_sN$ for a specific scenario s. We can show that $\widetilde{\theta}_s > \underline{\theta}$[77] in a closure scenario, while it is the other way around in a mass layoff scenario with full protection of the pregnant. Consequently, if women are fully rational, and their expectations about within-period layoff probabilities and available pregnancy protection are close to the actual probabilities, we expect to have more abortions in the closure scenario and fewer in the mass layoff scenario with full protection compared to the baseline case.

The number of births per N women is given by $p_0(\hat{\theta} - \underline{\theta}) + p_1(1 - \hat{\theta})$ for the baseline scenario and $p_0(\breve{\theta} - \widetilde{\theta}) + p_1(1 - \breve{\theta})$ for a specific scenario s. We can show that the number of births is lower in a closure scenario than in the baseline and it is higher in a mass layoff scenario with full protection of the pregnant. At the same time, the expectation of women about within-period layoff probabilities might be less precise by the time when they make the decision about increasing their pregnancy probability compared to the time of the abortion decision. This difference in expectation precision might increase the number of actual births in a closure scenario.

[77]We can provide the calculations upon request.

B Appendix for Chapter 2

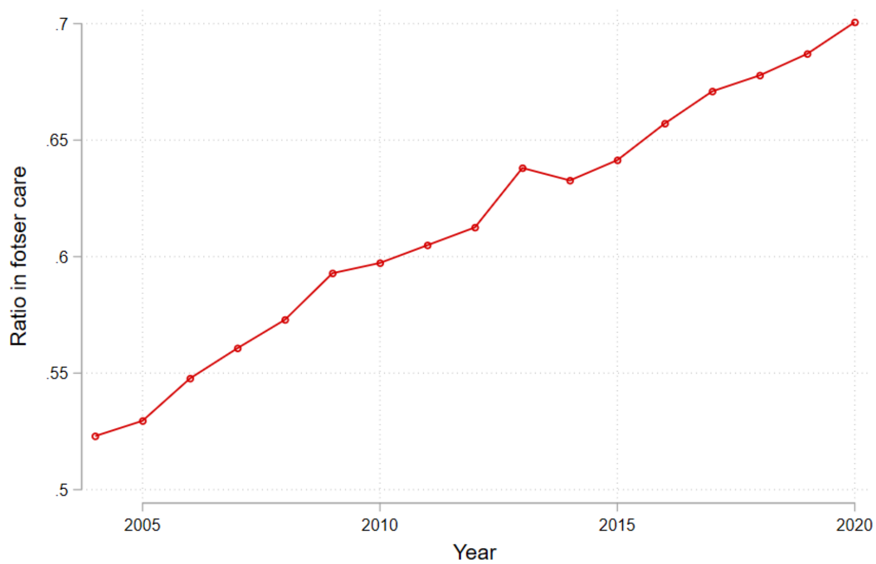
B.1 Supplementary tables and figures

Table B.1: Characteristics of foster mothers and biological mothers

	Education of (foster) mother				Age of (foster) mother	Household size
	Max 8 grades	Vocational	High School	College		
Foster family	25%	39%	22%	14%	49	6.8
Own parents	20%	28%	31%	21%	39	4.5

Note: Results are from the NABC data, using all children living with foster mothers or biological mothers between 2008 and 2014

Figure B.1: Percent in foster care among children in state care



Source: KSH https://www.ksh.hu/stadat_files/szo/hu/szo0017.html

Table B.2: Regression estimates for all outcomes on a sample including children in biological families

	(1) Secondary school	(2) Highschool	(3) Antidepr	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes*
Foster care	-0.038*	-0.101***	0.081***	0.046***	0.023	-0.001	0.023	0.390*	0.052**	-0.203***
	(0.019)	(0.013)	(0.012)	(0.010)	(0.025)	(0.023)	(0.020)	(0.186)	(0.019)	(0.036)
Residential Care	-0.091***	-0.085***	0.130***	0.076***	0.205***	0.114**	0.149***	1.358***	0.138***	-0.482***
	(0.025)	(0.017)	(0.022)	(0.019)	(0.042)	(0.041)	(0.039)	(0.285)	(0.029)	(0.057)
N	114934	114934	114934	114934	57233	57233	57233	114934	114934	114934
adj. R^2	0.267	0.413	0.012	0.005	0.186	0.193	0.060	0.112	0.101	0.226

Standard errors in parentheses + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimates of the equation 2.1 with the full set of controls, where for $foster$ a dummy with three categories is used
*Index calculated as the mean of the following four outcomes standardized over the full population: finishing secondary school, ever using antidepressants or tranquilizers by age 19 (-), ever getting pregnant by age 19(-), NEET months at age 19.

Table B.3: Original and corrected p-values for the main OLS regressions

Outcome	p-value	Bonferroni correction	q value
Secondary finished	0.0214	0.214	0.028
High school finished	0.4226	1	0.093
Antidepressants	0.1876	1	0.062
Tranquilizers	0.0392	0.392	0.035
Pregnancy	0.0122	0.122	0.022
Birth	0.0467	0.467	0.037
Abortion	0.0229	0.229	0.028
Neet months	0.005	0.05	0.016
Neet 6 month	0.0043	0.043	0.016
Index of all outcomes	0.0001	0.001	0.002

Hungarian DOI: 10.14754/CEU.2023.04

Table B.4: Regression of the outcomes on foster mother capacity on the sample of children living with their own parents - placebo

	(1) Secondary school	(2) Highschool	(3) Antidepr	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes
High fmother capacity	0.003	0.004	-0.003*	0.001	-0.006^{+}	-0.002	-0.004^{+}	0.040	0.007*	0.001
	(0.003)	(0.003)	(0.001)	(0.001)	(0.003)	(0.002)	(0.003)	(0.027)	(0.003)	(0.004)
Constant	0.911***	0.767***	-0.021	0.002	0.070	-0.499***	0.400	8.305***	0.641***	-0.313***
	(0.053)	(0.050)	(0.021)	(0.024)	(0.348)	(0.058)	(0.351)	(0.502)	(0.052)	(0.080)
N	113180	113180	113180	113180	56333	56333	56333	113180	113180	113180
adj. R^2	0.262	0.408	0.006	0.003	0.180	0.190	0.057	0.107	0.097	0.219

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hungarian DOI: 10.14754/CEU.2023.04

Table B.5: The relative likelihood a complier is a member of the given subgroup

Subgroup	First stage for the given group / Overall first stage
Boy	0.97
Girl	1.03
Any special need	1.36
No special need	0.88
Above median math score	0.99
Below media math score	1.04
Number of siblings max 1	0.84
Number of siblings over 1	1.05

Table B.6: Original and corrected p values for the IV regressions

Outcome	p value	Bonferroni correction	q value
Secondary finished	0.0787	0.787	0.31
Highschool finished	0.1957	1	0.356
Antidepressants	0.3125	1	0.406
Tranquilizers	0.1154	1	0.351
Pregnancy	0.8332	1	0.589
Birth	0.6125	1	0.485
Abortion	0.9813	1	0.647
Neet months	0.0655	0.655	0.31
Neet 6 month	0.1445	1	0.352
Index of all outcomes	0.0217	0.217	0.278

Table B.7: Regression estimates on the effect of foster care by foster mother's education on the sample in foster care

	(1) Secondary school	(2) Highschool	(3) Antidepr	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes
High educated foster mother	0.017	0.007	0.003	-0.006	-0.087	-0.149**	0.034	-0.053	0.021	0.025
	(0.045)	(0.032)	(0.027)	(0.024)	(0.065)	(0.054)	(0.053)	(0.434)	(0.043)	(0.059)
Constant	0.162	0.088	-0.118	0.080	-0.229	-0.349	0.079	3.655^{+}	0.383^{+}	0.403
	(0.209)	(0.117)	(0.129)	(0.083)	(0.310)	(0.288)	(0.242)	(2.056)	(0.206)	(0.258)
N	581	581	581	581	298	298	298	581	581	581
adj. R^2	0.121	0.303	0.043	-0.002	0.078	0.049	0.006	0.103	0.094	0.157

Robust standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Reference group: Children living with foster mothers with less than high school education

Hungarian DOI: 10.14754/CEU.2023.04

B.2 Bias resulting from the late measurement of ability

In the empirical strategy we argue that the fact that we measure our outcomes late, when children already spent some time in residential or foster care, makes us underestimate the true gains from foster care. In the following we spell out the argument in detail following the derivation of Angrist and Pischke, (2009 p50).

Say that we could measure cognitive and psychological ability early, before the placement decision. For simplicity, instead of a vector of variables, we use one variable a_i in this derivation - we think of this as a composite measure of early cognitive and non-cognitive ability, where a larger value of a_i shows better ability. In this case we would estimate the following model:

$$Y_i = \alpha + \rho foster_i + \gamma a_i + \varepsilon_i. \tag{B.1}$$

If a_i perfectly captures ability related to future outcomes, ρ shows the true effect of foster care. Again for simplicity we think of Y_i now as a variable where a larger value means a better outcome,e.g. an indicator of finishing secondary school, or an indicator of *not* having a teenage pregnancy, etc. We suppose γ is positive, i.e children with better cognitive and non-cognitive early ability are more likely to have better outcomes. However, instead of a_i, we only observe ability later, denoted by $a_{i,gr6}$, when a child already spent some time in state care. We suppose that this late ability depends on early ability and placement the following way:

$$a_{i,gr6} = \pi_0 + \pi_1 foster_i + \pi_2 a_i. \tag{B.2}$$

It is plausible that π_2 is positive: children with better early ability have better ability later. The literature results on the beneficial nature of early foster care suggest that pi_1 is also positive. But, to gain better insight about these signs we also try to estimate equation B.2. Even though we do not have a rich set of indicators for early ability, we have an imperfect measure of it: starting primary school in time [78]. Children who start later than the age of 7.5 were not mature enough to start the school at age 6.5, thus starting late can be an indicator of early problems. We regress ability in grade 6[79] on living in foster care in 6$^{\text{th}}$ grade and starting primary school in time. In this regression, the coefficient on foster care and starting school in time are positive and significant indeed (see Table B.8), supporting that π_1 and π_2 are positive as well.

Rearranging equation B.2 to a_i, and plugging it in to equation B.1 we get the following equation:

$$Y_i = (\alpha - \gamma\frac{\pi_0}{\pi_2}) + (\rho - \gamma\frac{\pi_1}{\pi_2}) foster_i + \frac{\gamma}{\pi_2} a_{gr6,i} + \varepsilon_i. \tag{B.3}$$

This is a simplified version of equation 2.1, which we can actually estimate. We run this version of the main equation and we see that children with better ability in grade 6 have better outcomes (see Table B.9). This means that the sign of $\frac{\gamma}{\pi_2}$ is positive. As we argued π_1, π_2 are plausibly positive as well. This means that the coefficient of $foster_i$ is underestimated in our main equation compared to rho. Thus, because of the proxy control problem we are likely to underestimate true beneficial effect of foster care on future outcomes.

[78] Starting in time is defined as starting primary school before age 7.5.

[79] calculated as the mean of standardized test scores, standardized grades and (negative) special needs, so larger values show better ability

Table B.8: Regression of ability on 6th grade on placement into foster care and early abilty

	(1)
	Ability index in 6th grade*
Foster care	0.182***
	(0.045)
Starts primary school in time	0.087*
	(0.041)
Cons	-0.172***
	(0.044)
N	840
adj. R^2	0.022

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Calculated as the standardized mean of testscores, grade, and different indicators of (negative) special needs

Table B.9: Regression of the outcomes on placed into foster care and an index of ability

	(1) Secondary school	(2) Highschool	(3) Antidepr	(4) Tranquil	(5) Pregnancy	(6) Birth	(7) Abortion	(8) NEET months	(9) NEET 6m	(10) Index of outcomes
Foster	0.142***	0.051*	-0.045	-0.041+	-0.236***	-0.183***	-0.137**	-1.460***	-0.133***	0.282***
	(0.035)	(0.022)	(0.027)	(0.024)	(0.058)	(0.054)	(0.048)	(0.380)	(0.038)	(0.052)
Ability index in 6th grade	0.156***	0.233***	-0.077***	-0.029+	-0.153***	-0.096**	-0.093**	-1.998***	-0.189***	0.310***
	(0.025)	(0.020)	(0.019)	(0.015)	(0.038)	(0.036)	(0.033)	(0.264)	(0.026)	(0.037)
N	890	890	890	890	457	457	457	890	890	890
adj. R^2	0.092	0.202	0.049	0.021	0.081	0.076	0.037	0.111	0.106	0.146

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hungarian DOI: 10.14754/CEU.2023.04

C Appendix for Chapter 3

Figure C.1: Time trend in mean BF% and mean BMI

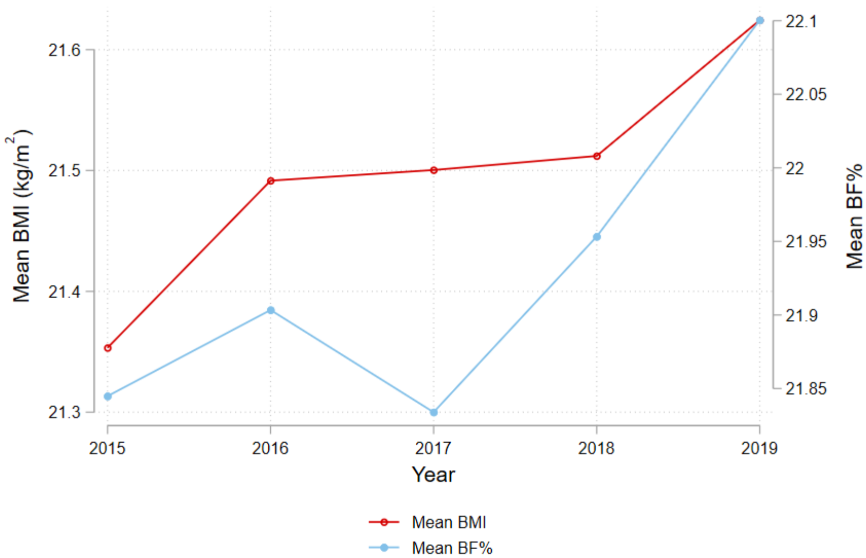


Table C.1: Past PE Time by Grade by the end of a School year

	School year				
	2014/2015	2015/2016	2016/2017	2017/2018	2018/2019
Grade					
5	459	459	675	675	675
6	594	594	594	810	810
7	729	729	729	729	945
8	594	864	864	864	864
9	729	729	999	999	999
10	864	864	864	1134	1134
11	999	999	999	999	1269
12	864	1134	1134	1134	1134

Figure C.2: Mean BMI and BF% among groups of boys and girls

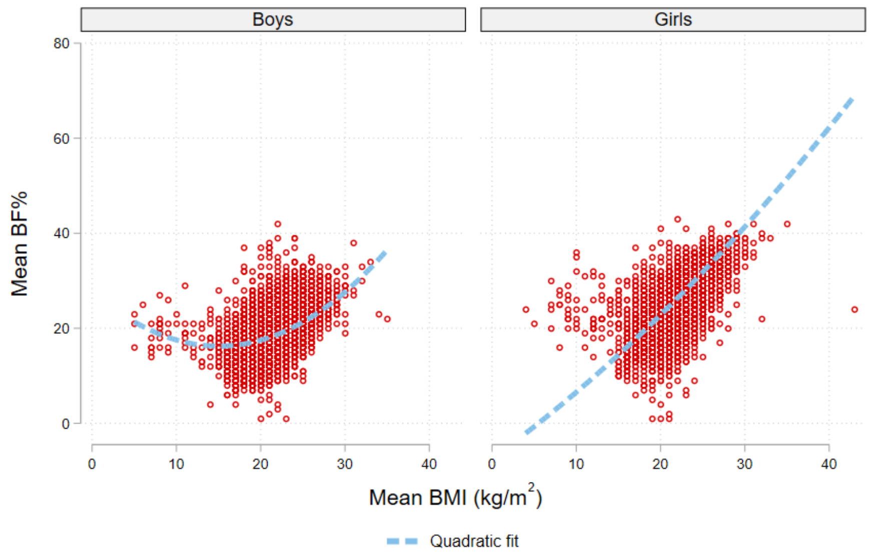


Figure C.3: Mean BF% and BMI by age and gender

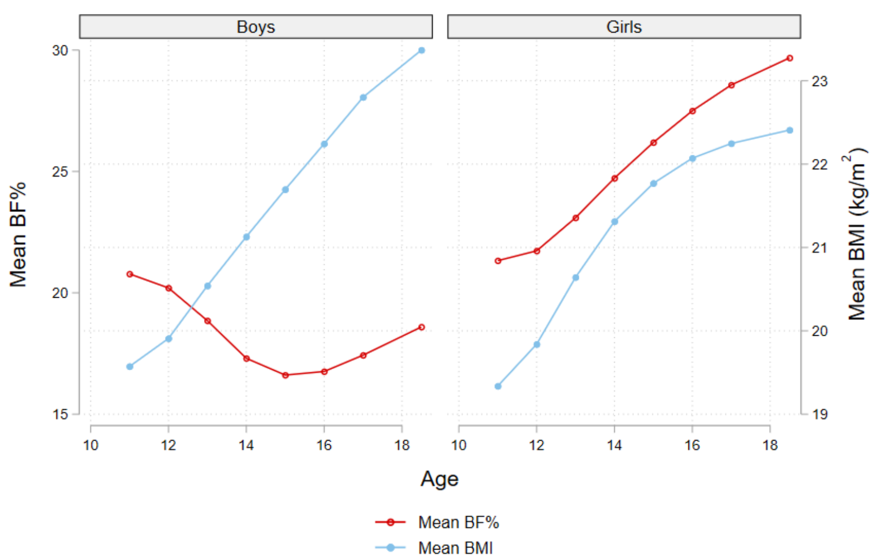


Table C.2: Placebo Regressions of PE time on mean BF% and obesity using schools in Hódmezvásárhely

	Mean BF%		Percent obese[1]	
	(1)	(2)	(3)	(4)
PEhours100	-0.175	-0.106	-0.718	-0.614
	(0.327)	(0.322)	(0.865)	(0.903)
School FE	No	Yes	No	Yes
adj. R-sq	0.828	0.851	0.202	0.277
Mean of outcome	21.67		7.05	
N	596		596	

Year FE, age dummies, gender, and age and gender interactions are included in all models. Standard errors (in parentheses) are clustered by school. $*p < 0.05, **p < 0.01, ***p < 0.001$

[1] Categorization based on BF% thresholds

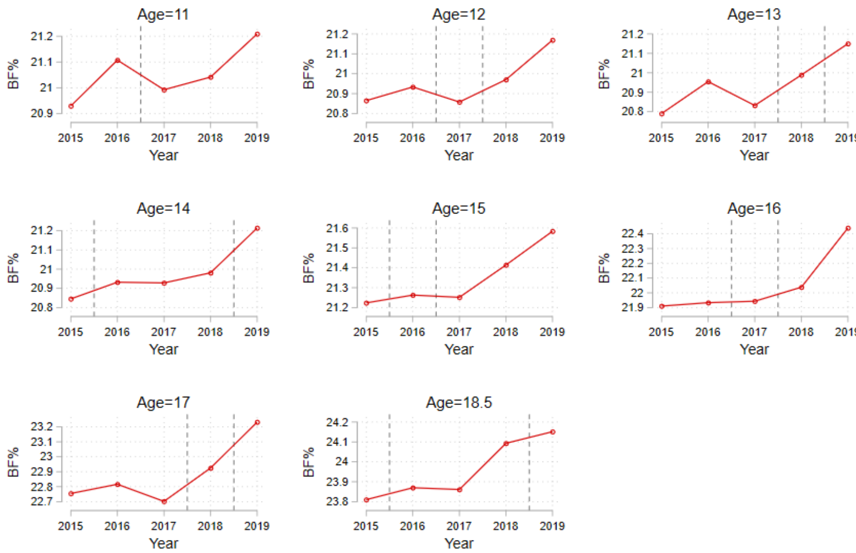
Table C.3: Placebo Regressions of PE time on mean BMI and obesity using schools in Hódmezvásárhely

	Mean BMI		Obese based on BMI	
	(1)	(2)	(3)	(4)
PEhours100	-0.173	-0.165	-0.933	-0.933
	(0.120)	(0.123)	(0.583)	(0.596)
School FE	No	Yes	No	Yes
adj. R-sq	0.501	0.585	0.010	0.142
Mean of outcome	21.32		7.68	
N	596		596	

Year FE, age dummies, gender, and age and gender interactions are included in all models. Standard errors (in parentheses) are clustered by school. $*p < 0.05, **p < 0.01, ***p < 0.001$

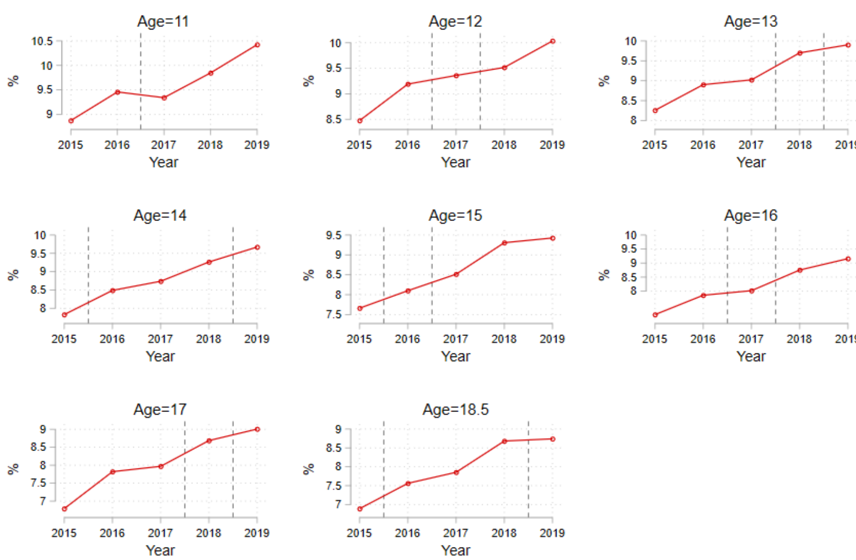
[1] Categorization based on BMI thresholds

Figure C.4: Mean BF% by age and calendar year



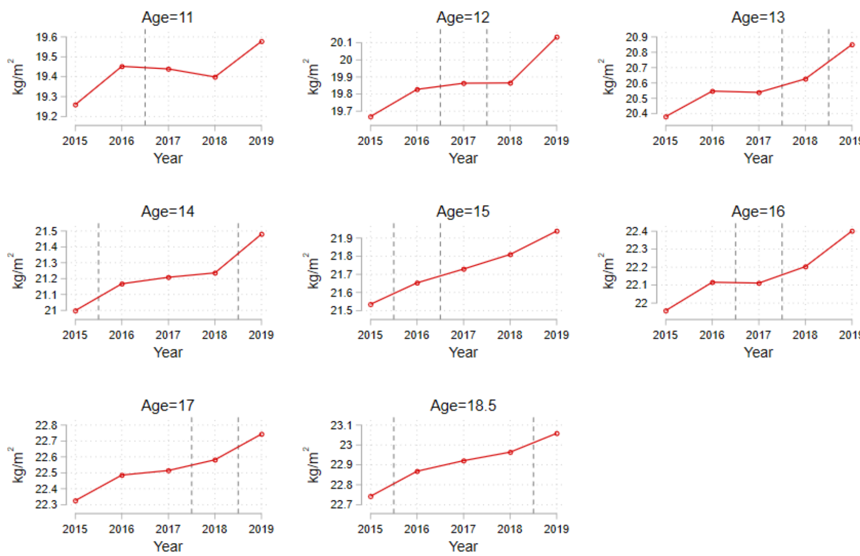
Dashed grey lines indicate jump in past PE hours

Figure C.5: Prevalence of obesity by age and year based on the BMI obesity measure



Dashed grey lines indicate jump in past PE hours

Figure C.6: Mean BMI by age and calendar year



Dashed grey lines indicate jump in past PE hours

Table C.4: Regressions of PE time on mean BF% and obesity with school-gender-cohort clustered standard errors

	Mean BF%		Percent obese[1]	
	(1)	(2)	(3)	(4)
PEhours100	0.015	0.017	0.014	0.004
	(0.025)	(0.018)	(0.065)	(0.046)
School FE	No	Yes	No	Yes
adj. R-sq	0.692	0.774	0.142	0.313
Mean of outcome	21.93		8.19	
N	179476		179534	

Year FE, age dummies, gender, and age and gender interactions are included in all models. Standard errors (in parentheses) are clustered by school-gender-cohort groups. $+p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001$ [1] Categorization based on BF% thresholds

Table C.5: Regressions of PE time on mean BMI and obesity with school-gender-cohort clustered standard errors

	Mean BMI		Percent obese[1]	
	(1)	(2)	(3)	(4)
PEhours100	0.005	0.006	-0.033	-0.035
	(0.012)	(0.009)	(0.062)	(0.045)
School FE	No	Yes	No	Yes
adj. R-sq	0.386	0.539	0.036	0.216
Mean of outcome	21.48		8.51	
N	179177		179197	

Year FE, age dummies, gender, and age and gender interactions are included in all models. Standard errors (in parentheses) are clustered by school-gender-cohort groups. $+p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001$ [1] Categorization based on BMI thresholds

Table C.6: Regressions of PE time on mean BF% and obesity with no weighting

	Mean BF%		Percent obese[1]	
	(1)	(2)	(3)	(4)
PEhours100	0.017	0.019	0.047	0.035
	(0.025)	(0.025)	(0.082)	(0.082)
School FE	No	Yes	No	Yes
adj. R-sq	0.434	0.509	0.040	0.108
Mean of outcome	21.93		8.19	
N	179476		179534	

Year FE, age dummies, gender, and age and gender interactions are included in all models. Standard errors (in parentheses) are clustered by school. $+p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001$

[1] Categorization based on BF% thresholds

Figure C.7: Prevalence of obesity in schools by proportion of disadvantaged children

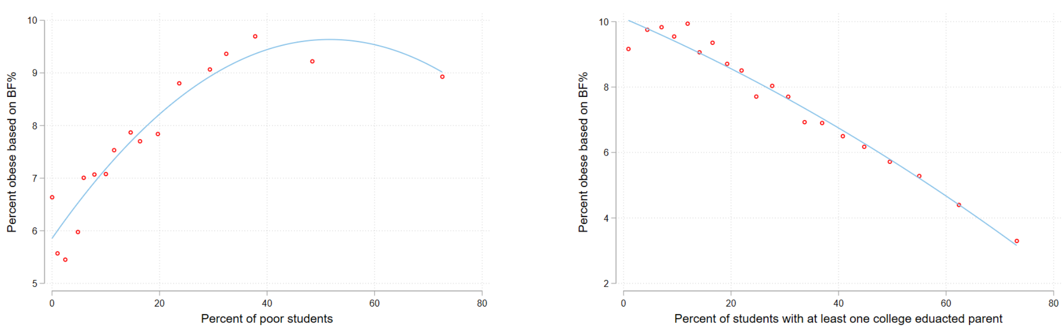


Table C.7: Regressions of PE time on mean BMI and obesity with no weighting

	Mean BMI		Percent obese[1]	
	(1)	(2)	(3)	(4)
PEhours100	0.019	0.018	0.064	0.050
	(0.014)	(0.014)	(0.087)	(0.088)
School FE	No	Yes	No	Yes
adj. R-sq	0.166	0.257	0.011	0.079
Mean of outcome	21.48		8.51	
N	179177		179197	

Year FE, age dummies, gender, and age and gender interactions are included in all models. Standard errors (in parentheses) are clustered by school. $+p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001$

[1] Categorization based on BMI thresholds

Table C.8: Heterogeneous effects on mean BF%

Subsample	(1) Total	(2) Age Under 15	(3) Age Min 15	(4) Boy	(5) Girl	(6) Min 50% poor	(7) Less than 50% Poor	(8) Max 15 % college edu-cated	(9) More than 15% coll. educ.	(10) Min 1 gym room per 200	(11) Less than 1 gym room per 200
PEHours100	0.017	0.006	0.015	0.005	0.024	0.091	0.008	0.052	-0.002	-0.017	0.018
	(0.015)	(0.028)	(0.022)	(0.021)	(0.022)	(0.079)	(0.020)	(0.038)	(0.022)	(0.051)	(0.022)
N	179476	87356	92120	91072	88404	19555	130761	74010	86351	45344	104479
adj. R-sq	0.774	0.524	0.839	0.422	0.684	0.576	0.773	0.630	0.812	0.547	0.792

Standard errors are clustered by school, *** p<0.001, ** p<0.01, * p<0.5

Hungarian DOI: 10.14754/CEU.2023.04

Table C.9: Heterogeneous effects on obesity by BMI

Subsample	(1) Total	(2) Age Under 15	(3) Age Min 15	(4) Boy	(5) Girl	(6) Min 50% poor	(7) Less than 50% Poor	(8) Max 15 % college educ	(9) More than 15% coll. educ.	(10) Min 1 gym room per 200	(11) Less than 1 gym room per 200
PEHours100	-0.035	-0.114	-0.006	-0.010	-0.054	0.242	-0.061	0.124	-0.096	-0.195	-0.014
	(0.046)	(0.086)	(0.049)	(0.064)	(0.056)	(0.217)	(0.053)	(0.108)	(0.055)	(0.161)	(0.054)
N	179197	87296	91901	90930	88267	19535	130547	73871	86233	45243	104347
adj. R-sq	0.216	0.166	0.259	0.224	0.213	0.116	0.213	0.113	0.229	0.105	0.235

Standard errors are clustered by school, *** p<0.001, ** p<0.01, * p<0.5

Table C.10: Heterogeneous effects on mean BMI

Subsample	(1) Total	(2) Age Under 15	(3) Age Min 15	(4) Boy	(5) Girl	(6) Min 50% poor	(7) Less than 50% Poor	(8) Max 15 % college edu-cated	(9) More than 15% coll. educ.	(10) Min 1 gym room per 200	(11) Less than 1 gym room per 200
PEHours100	0.006	0.003	0.014	0.011	0.004	0.030	-0.002	0.015	-0.004	-0.027	0.003
	(0.009)	(0.014)	(0.010)	(0.012)	(0.011)	(0.037)	(0.010)	(0.018)	(0.011)	(0.027)	(0.010)
N	179177	87295	91882	90917	88260	19530	130533	73861	86224	45237	104334
adj. R-sq	0.539	0.333	0.482	0.578	0.519	0.364	0.506	0.386	0.554	0.292	0.544

Standard errors are clustered by school, *** p<0.001, ** p<0.01, * p<0.5

Hungarian DOI: 10.14754/CEU.2023.04