NORTHWESTERN UNIVERSITY

Access to Birth-to-5 Child and Family Policies and Their Impact Beyond the Average

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Human Development and Social Policy

By

Timea Henriett Viragh

EVANSTON, ILLINOIS

September 2023

Abstract

Policies supporting families with young children provide an important context for human development. The two primary public policies available to caregivers around and after childbirth are parental leave and early care and education (ECE). A substantial body of research evaluates the effects of parental leave and ECE policies on various outcomes, but most existing studies focus on average treatment effects in the population. We know much less about how policy effects vary by family demographics or by child care program characteristics. The three studies in this dissertation aim to increase our understanding about the circumstances under which child and family policies are more or less effective in supporting families with young children.

Study 1 documents maternal wage dynamics around childbirth and their heterogeneity by education using German administrative data. Women with low education experience a smaller drop and a faster recovery of their wages than women with higher levels of education. These differences are largely explained by the fact that women with low education have their first child at a younger age. To investigate whether public policies can influence these wage dynamics, I exploit a paid family leave policy change in 2007 to set up a differences-in-regression-discontinuities design. The policy change shortened the duration of monthly benefit receipt, increased the amount of monthly transfers, and encouraged secondary caregivers to spend at least two months on leave. I find suggestive evidence that the policy did not have an effect on wage loss after childbirth for women with high education. Women with low education earn a larger share of their pre-birth wage in the second year after childbirth under the new policy regime. This is most likely because they return to work faster. My results suggest that the policy change only influenced labor market behavior of mothers with economic constraints.

Study 2 provides novel insights into ECE access by using a new indicator: distance traveled for child care. It documents trends in how far families travel to access care in the United States by neighborhood income using geographic mobility data for 2019 for n = 106,916 child care programs. The findings indicate that distance traveled follows an inverted U-shaped pattern. Families who live in the lowest and highest income neighborhoods tend to travel less for child care than families in the middle of the income distribution. In a case study examining the state of Illinois, the pattern is consistent with the physical availability of child care. Findings are discussed in terms of how distance traveled as an indicator can help define the child care market and its implications for early childhood policy.

Study 3 focuses on the effect of ECE on parental outcomes. In this paper, coauthored with Professor Terri Sabol, we use family stress theory to argue that the financial and logistical burden of paying for and managing child care arrangements can act as repeated stressors in families' lives, and may negatively influence parental mental well-being. We examine whether the provision of free and high-quality early care alleviates these stressors and leads to better parental health and well-being. We use data from the Head Start Impact Study, a nationally representative randomized controlled trial from the early 2000s, to answer the research question. Next to documenting the average effect of offering Head Start on parental well-being, we also explore heterogeneity in these effects by program characteristics, including family-centered services and supports in a program.

The three studies in this dissertation inform our understanding of how birth-to-5 child and family policies influence lives of families with young children within two different national contexts, Germany (with high public funding) and the United States (with lower public funding). Moving beyond documenting average effects, they showcase how policy impacts vary by family demographics and child care center characteristics. These insights contribute to policy design and could help researchers, practitioners, and policy makers to work together to build policies that better support families with young children.

Acknowledgment

Dissertation writing is a lonely experience, especially during a global pandemic and after moving to a new city. I am lucky to have had many people in my life who kept me on track and helped me finish this milestone.

First, I would like to thank my committee, Ofer Malamud, Terri Sabol, and Hannes Schwandt for providing useful insights and engaging in discussions about how to make the papers better. Thank you for helping me become a better scholar and for showing me good teaching practices. I would also like to thank Katie Gonzalez and Christina Padilla for sharing their Stata code and insights on the FIRC method – you made my third chapter much easier to write.

Next to faculty guidance, the most important contributors to one’s graduate school experience are fellow graduate students. I am grateful for the friendship of Claire Mackevicius, Olivia Healy, Amanda Cook, Nikki Guarino, Sheridan Fuller, Angel Bohannon, and Julia Honoroff. Thank you for numerous discussions about the meaning of applied policy research, about appropriate methods and data sources to use, and about countless other things like where to eat good food and how to enjoy life as a graduate student. I am also grateful for the DEEP lab community – thank you for listening to countless versions of my projects before I settled on the final, and hopefully best, ones.

I would like to thank two other professors, who contributed greatly to my embark on the PhD journey. Thank you, Gabor Kezdi, for teaching me my first causal inference class and making me excited about applied research. Thank you, Bernie Black, for giving me my first full-time job as a research assistant after I moved to the US. You’ve been a very generous mentor and showed me invaluable data documentation and project management skills that I profited from ever since.

It's customary to thank one’s close family last, however, this dissertation would not be finished without the support of my husband, Bence Bardoczy, and without the motivation that my son, Berci, provided. Thank you, Bence, for always being willing to discuss issues, questions, trying to trouble-shoot data problems, even though the topics we study and methods we use are quite different. Thank you for being there as an advisor, a colleague, and a partner during the long years of the pandemic lockdown. Thank you, Berci, for being a relatively good sleeper, so I could work on these papers for a few uninterrupted hours even when you were tiny. Thank you for keeping me on track and focused, so I could finish my tasks fast and then spend time with you. I love you both.

Table of Contents

List of Tables

List of Figures

Introduction

Policies supporting families with young children provide an important context for human development (Teti et al., 2017). The two primary public policies available to caregivers around and after childbirth are parental leave and early care and education. Parental leave provides time off of work to ensure that caregivers are able to spend time with their children in the early weeks and months, but at the same time stay attached to the labor market so their transition back to employment is possible (Rossin-Slater, 2018). Early care and education (ECE) enables caregivers to work outside the home, and provides a conducive environment for child development (Blau & Currie, 2006; E. U. Cascio, 2021).

Around childbirth, countries across the world support parents to varying degrees. Member countries of the Organisation for Economic Co-operation and Development (OECD) spend on average 2.1% of their Gross Domestic Product (GDP) on family benefits, but there are large differences across countries (OECD, 2022). While the United States spent only 0.61% in 2018 on public financial support for families and children, Germany spent 2.33% (OECD, 2022). A substantial body of research evaluates the effects of parental leave and ECE policies on various outcomes. Findings from these studies provide information on how successful these policies are in supporting parental employment and wages, parent and child health, and child development (see reviews by Morrissey, 2017a; Olivetti & Petrongolo, 2017; Rossin-Slater, 2018; Rossin-Slater & Uniat, 2019). However, most existing studies focus on average treatment effects in the population. We know much less about how policy effects vary by family demographics or by child care program characteristics. This dissertation aims to increase our understanding about the circumstances under which child and family policies are more or less effective in supporting families with young children. It is important to document these differences to help refine existing policies and to target public funds to build a policy environment that serves families better.

Study 1 in this dissertation focuses on maternal wage dynamics around childbirth and the degree to which parental leave policies can support female income after women have children. I study this question in the German context. Germany is a large Western-European country with generous parental leave policies. The large majority of mothers take time off of work to care for their children in the early years, and they may or may not return to the labor market full-time (Collins, 2019). The policy environment over the past two decades have been encouraging women to participate in the labor force after having children and they introduced measures that aim to increase gender equality in the labor market and in caretaking (Huebener et al., 2016). The changing policy context along with available administrative data make Germany an ideal context for my analyses. I use data on labor market histories to document maternal wages before and after childbirth. I extend prior work by disaggregating average patterns by education. I show that women with lower levels of education experience a smaller drop and a faster recovery in their wages following childbirth than women with higher levels of education. These differences are largely explained by the fact that women with lower levels of education have children at a younger age. To investigate whether public policies can influence these wage dynamics, I exploit a paid family leave policy change in 2007 to set up a differences-in-regression-discontinuities design. The policy change shortened the duration of monthly benefit receipt, increased the amount of monthly transfers, and encouraged secondary caregivers to spend at least two months on leave. I find suggestive evidence that the policy did not have an effect on wage loss after childbirth for women with high education. Women with low education earn a larger share of their pre-birth wage in the second year after childbirth under the new policy regime. This is most likely because they returned to work faster. My results suggest that the policy change only influenced labor market behavior of mothers with economic constraints.

Returning to the labor market following childbirth is only possible if caregivers have access to non-parental care for their children. Next to enabling caregiver work outside the home, early care and education (ECE) is often framed as an important tool to help children, especially children from disadvantaged families, to develop their academic skills (Phillips et al., 2017). ECE can also influence children's development of socio-emotional skills like executive function, self-regulation, and relationship with others (Sabol et al., 2021; Weiland & Yoshikawa, 2013). It may also contribute to the well-being of the whole family, not just to that of children (Teti et al., 2017). Furthermore, ECE is an important part of the social safety net (Bitler et al., 2020; Small, 2006a). Studies 2 and 3 of this dissertation focus on the early care and education landscape of the United States, a country that does not have a universal system of ECE. This stands in contrast to many European settings where publicly funded ECE is often universally available. In the U.S. context, effective investment is critical and my work seeks to inform the broader line of research on how best to structure an underfunded system and better understand the significance of child care in the lives of families with young children.

Over the past few decades federal, state, and local investments into early care and education have focused on increasing access to child care (Friedman-Krauss et al., 2021; Friese et al., 2017). Access to child care is a multi-dimensional concept including the physical availability, cost, and quality of child care programs, and how well they serve families' needs (Friese et al., 2017). Common indicators to measure ECE access are physical availability, i.e. the number of programs in a given geographic area (Cochi Ficano, 2006; Gordon & Chase-Lansdale, 2001; Malik et al., 2018), and enrollment in ECE (Bainbridge et al., 2005; Magnuson & Waldfogel, 2016). These indicators capture certain aspects of ECE access, however, they cannot provide a comprehensive picture of the child care market families face across the income spectrum. Study 2 introduces a novel indicator of access: distance traveled for child care. By using fine-grained geographic mobility data, I document trends in how far families travel to access care in the state of Illinois. I also examine differences by income across the whole income distribution. This new indicator provides a data-driven method to assess the size of the child care market and allows additional insights into current patterns of care. Findings from this study contribute to our understanding about where the highest need for child care is and where to target future funds.

The majority of the research on the effects of early care and education on human development focuses on children. We know much less about how ECE influences parental development and well-being. In study 3 of my dissertation, which is a coauthored paper, we use family stress theory (Conger & Elder, 1994; Elder & Caspi, 1988) to argue that the financial and logistical burden of paying for and managing child care arrangements can act as repeated stressors in families' lives, and may negatively influence parental mental well-being. We examine whether the provision of free and high-quality early care alleviates these stressors and leads to better parental health and well-being. We study Head Start, the only federally funded preschool program, which has been available for low-income families since 1965 (Vinovskis, 2008). Head Start promotes school readiness and family well-being by providing educational, as well as nutritional, health, and social services. In 2019, Head Start programs around the U.S. served about 1 million children and their families (Administration for Children and Families, 2019). We use data from the Head Start Impact Study, a nationally representative randomized controlled trial from the early 2000s, to answer the research question. Next to documenting the average effect of offering Head Start on parental well-being, we also explore heterogeneity in these effects by program characteristics, including family-centered services and supports in a program. Investigating treatment impact heterogeneity has become an important tool in education effectiveness research to understand “under what circumstances” and “for whom” interventions such as Head Start work best (Reardon & Stuart, 2017).

The three studies in this dissertation inform our understanding of how birth-to-5 child and family policies influence lives of families with young children within two different national contexts, Germany (with high public funding) and the United States (with lower public funding). Moving beyond documenting average effects, they showcase how policy impacts vary by family demographics and child care center characteristics. These insights contribute to policy design and could help researchers, practitioners, and policy makers to work together to build policies that better support families with young children.

1) Chapter 1 – The effects of a paid parental leave policy change on maternal labor market outcomes: Evidence from Germany

Introduction

Despite convergence over the past several decades, differences in labor market participation and wages between men and women still exist (Olivetti & Petrongolo, 2016). Among the traditional explanations are the fact that women spend more time on non-paid household activities and child care, they prefer more flexible work arrangements, and work in different occupations than men (Cortes & Pan, 2020). More recently, this literature has been focusing on the role of children in explaining the remaining gender gaps. After the arrival of children, mothers, compared to men or to childless women, experience a sizeable drop in earnings, as well as a decrease in hours worked and employment. This is often called the “child penalty” and has been well established across a wide range of countries (Aguilar-Gomez et al., 2019; Andresen & Nix, 2021; Bertrand et al., 2010; Kleven, Landais, & Søgaard, 2019; Kleven, Landais, Posch, et al., 2019). An important question is whether and to what extent public policies can mitigate the effects of having children on maternal labor market outcomes.

Parental leave policies can support parental employment after childbirth or adoption. They provide time off of work to care for children and at the same time, they allow caregivers to stay employed and to return to their pre-child jobs once their leave ends. Many countries pay benefits to caregivers on leave to protect their income. While parental leave policies often allow either parent to take time off of work, mothers are more likely to take leave (Bana et al., 2018; Bünning, 2015).

There is a large literature on the effects of parental leave policies on female labor market attachment (recent summaries are provided by Olivetti & Petrongolo, 2017; Rossin-Slater, 2018). In general, papers find a relatively small short-term effect on maternal employment (0-2 years after childbirth), and no change in labor market behavior on average in the long-term (starting 2-3 years after having children and observing behavior up to 10 years). Both extensions of job protected leave and increased benefit payments encouraged women to return to the labor market later, but the effects were small in magnitude in Austria (Lalive et al., 2014; Lalive & Zweimüller, 2009), in Germany (Schönberg & Ludsteck, 2014), and in Sweden (Ginja et al., 2020). Moreover, parental leave policy changes did not reduce the "child penalty" in Austria (Kleven et al., 2020) or in Norway (Andresen & Nix, 2022).

Two potential explanations for the small and null effects of paid leave policy changes on maternal labor market outcomes are (1) strong preferences about time allocation between market work and home production; and (2) the presence of heterogeneous effects. Most papers do not analyze policy effects on sub-groups of mothers, and while they find no impact on average, there might be differences in labor market responses based on budgetary or institutional constraints families face. For example, higher-income mothers may face lower pressure to return to the labor market to protect household income, or lower-income mothers could have lower access to child care which prevents them from working outside the home. There is some evidence that mothers with different education and pre-birth income levels respond to policies differently. In a cross-country analysis of a sample of Organisation for Economic Co-operation and Development (OECD) members, Olivetti and Petrongolo (2017) find that correlations between the number of weeks of job protected leave and female employment rate are only statistically significantly different from 0 for women with the lowest level of education. However, other papers find similar responses in labor market behavior to paid leave extensions across income groups (Ginja et al., 2020; Lalive & Zweimüller, 2009).

The current paper expands our understanding about the heterogeneous effects of childbirth and paid leave policy changes on maternal labor market outcomes. I investigate maternal wage dynamics around childbirth and analyze the impact of a 2007 policy change in Germany on maternal wages by level of education. Multiple aspects of the policy supported female labor force participation following childbirth, but it influenced families differently based on pre-birth income (Huebener et al., 2016). The new policy regime reduced the length of transfer receipt from 24 to 12 months. It also increased the transfer amount. The prior regime provided a means-tested 300-euro monthly transfer to eligible families, while the new regime introduced a 2/3 replacement of pre-birth earnings and stopped means-testing. This meant that lower-income families were made worse off and higher income families were made better-off in terms of income during leave (Huebener et al., 2019). These differences may have led to heterogeneous effects by pre-birth income. Since family income and education are highly correlated, I use maternal education at birth as a proxy for income.

I first document maternal wage dynamics around childbirth using administrative data from Germany. I replicate trends from prior papers that used survey data for the overall population (Kleven, Landais, Posch, et al., 2019), and provide new findings on sub-groups by educational attainment. My dataset covers a 2% random sample of individuals who paid social security contributions, which is about 2 million people, and includes their complete labor market histories since 1975. I identify mothers with a first birth between 2003-2007 and use an event-study design to describe their wages 5 years before to 10 years after childbirth. Mothers experience a 53% drop in their raw earnings in the first year after childbirth compared to the year right before childbirth, and they only recover about half of this loss by 10 years after childbirth. However, once I account for age and year effects, this recovery disappears, and maternal wages remain at around 60% of their pre-birth levels throughout the observed 10-year post-birth period. Mothers with no vocational training or university diploma (mothers with low education) have a faster recovery of earnings. They start to earn as much as they did before childbirth by 8 years after childbirth. Mothers with at least a university degree (mothers with high education) and mothers who have vocational training (mothers with mid-level education) have similar wage dynamics to that of the overall sample. Post-birth dynamics including age and year effects are similar for all four groups (all mothers, and mothers with all three levels of education).. This indicates that the increasing wage post-childbirth is mainly due to age and economic growth. The faster recovery for mothers with low levels of education is mostly explained by the fact that they are younger when they have children.

In the second half of the paper, I analyze whether a 2007 change in the paid parental leave policy of Germany had an effect on maternal wage dynamics around childbirth. The new policy came into effect on January 1, 2007. I employ a differences-in-regression-discontinuities framework to analyze the effect of the policy on maternal wages. I compare mothers who gave birth in the last quarter (October-December) of 2006, who were subject to the old policy regime, to mothers who gave birth in the first quarter of 2007 (January-March), who were eligible for the new benefits. Because of the seasonality of births around January (Buckles & Hungerman, 2013; Currie & Schwandt, 2013), I use the same two quarters in prior years (2003, 2004, and 2005) to account for seasonal differences in demographic characteristics of mothers who gave birth at the end of the year versus those in the beginning of the year. My main outcome of interest is the change in earnings loss mothers experience following childbirth. I find suggestive evidence that maternal wages fall by less under the new policy regime than they did under the old policy regime, but the point estimates are imprecise. This downward trend is true up to 15 months after the arrival of children. The results are driven by mothers with low education. I find no effect of the policy on wage loss of mothers who have a high level of education.

This paper contributes to two strands of the literature. First, it adds to our understanding about maternal wage dynamics around childbirth (often called the “child penalty”). Several papers across a range of countries document a large and immediate decrease in female wages, hours worked, and participation following childbirth, with no convergence between men and women even several years after the birth of children (Aguilar-Gomez et al., 2019; Andresen & Nix, 2021; Kleven, Landais, Posch, et al., 2019; Sandler & Szembrot, 2019). Kleven et al. (2019) estimate a long-run child penalty of 61% for women in Germany using the Socio-Economic Panel survey and births between 1985-2003. I bring more descriptive evidence on maternal wage dynamics in Germany by using an administrative data source for births between 2003-2007. This dataset provides a larger sample size and is less susceptible to misreporting of income. I also document trends for women with different levels of education. It is important to understand heterogeneities in wage dynamics as they can help understand differences in policy impacts.

Second, my paper extends analyses of parental leave policy effects on maternal labor market participation by focusing on women with different levels of education. In general, parental leave shorter than 1 year encourages women to stay employed and is beneficial for female labor force participation. Longer leave tends to reduce long-term employment of women (Rossin-Slater, 2018). Extension of leave duration and increase in the number of months parents receive transfers for while on leave lengthens the amount of time mothers stay home in Austria, but only to a small extent and in the short-run (Lalive et al., 2014; Lalive & Zweimüller, 2009). Women have similar behaviors regardless of whether their incomes are below or above the median before childbirth. Furthermore, only 20-25% of women return to work when their leave ends in the Austrian sample (Lalive & Zweimüller, 2009). Findings are qualitatively similar in Germany. Schonberg & Ludsteck (2014) estimate a 1-month extension of time at home after childbirth when benefit receipt increased by 4 months, and a 3-month extension with a 16-month increase in benefit receipt. These papers suggest that it is difficult to change maternal labor market behavior by paid leave policies, at least on the aggregate level. Mothers may have relatively strong preferences about market work and home production that cannot be altered by policy changes, or social norms could be highly internalized around caregiving. Lack of child care to enable parental work outside the home can also be an explanation for the small effects of paid leave on employment.

An active strand of the “child penalty” literature investigates whether public policies like parental leave or the provision of child care can influence parental wage changes around childbirth. Parental leave policy changes have been found to have no effect on maternal earnings in Austria (Kleven et al., 2020). In Norway, a family leave policy change that encouraged both mothers and fathers to take time off of work to care for their children had no effect on parental earnings, either (Andresen & Nix, 2022). In this Norwegian context, the provision of public child care led to a reduction in the child penalty for mothers in the short-run (Andresen & Nix, 2022). In Germany, mothers who lived in counties with low public child care provision experienced a larger drop in their earnings following childbirth than mothers in counties with a high provision of child care (Chhaochharia et al., 2020). The current paper contributes to this literature on policy effects on the “child penalty” the following ways. First, I generate the outcome of wage loss after childbirth at a monthly frequency. Most prior papers use annual data, which can mask important insights into wage dynamics. Monthly wage dynamics may be especially important given that prior papers found changes in return to the labor market at the monthly level. I also analyze whether the policy change had heterogeneous effects on wage loss by education. Women with different levels of education tend to experience different career and family life trajectories, as well as different constraints, so paid leave policies could have different impacts on their wage dynamics. It is important to understand these differences to refine existing policies to better serve families with young children.

Institutional background

In Germany, maternity leave is available for mothers, which starts 6 weeks before the due date of the child and lasts for 8 weeks after childbirth (Bergemann & Riphahn, 2022). The leave is job protected and mothers are only allowed to work in the weeks prior to childbirth with explicit written consent. They are not allowed to work in the 8 weeks after childbirth (Huebener et al., 2019). Employed mothers receive their full pay during their leave.

Next to the maternity leave, job protected parental leave has also been available in Germany for either parent. The 2007 reform called the *Elterngeld* changed the benefit structure associated with the existing parental leave policy (Bergemann & Riphahn, 2022). The policy change aimed (1) to increase family income in the first year of a child’s life; (2) to encourage mothers to participate on the labor market; and (3) to increase gender equality by encouraging fathers to take part in family life (Huebener et al., 2016, 2019). The new system went into effect on January 1, 2007. All parents whose children were born on or after this date were eligible for the new benefits. The government coalition decided on the reform in May 2006 and parliament agreed in September 2006 (Kluve & Tamm, 2013), which means that parents whose children are born around January 1, 2007 did not know about this reform when they decided to have children.

This policy changed several aspects of the prior system (Huebener et al., 2016). First, it changed the amount of paid time parents could take off work to care for their child. Before 2007, one parent could stay home for 24 months and receive some transfers. The leave was almost always taken by the mother, only 3.5 percent of fathers took any leave in 2006 (Bünning, 2015). The reform cut this time in half to 12 months. Second, under the prior scheme, there were no specific incentives in place for the second caregiver to stay home. The reform allowed 2 extra months, bringing the total to 14, if both parents took at least 2 months. These 2 months are referred to as the "partner months". Parents could allocate the 14 months however they saw fit if one parent took at least 2.[1] Third, the reform also changed financial incentives. Under the prior policy, the monthly transfer was 300 euros for the duration of the leave, which was means-tested. In 2006, 77% of families received the payments for 6 months and about 50% from month 7 on (Ehlert, 2008). The new policy incorporated earnings replacement, with parents receiving about 2/3 of their net monthly income for the duration of their leave. The minimum payment remained 300 euros and a cap of 1,800 euros was imposed. Those who were not employed prior to childbirth also received the 300-euro minimum transfers. 300 euros is about 11% of the average net household income pre-birth for the years 2005-2008 (Huebener et al., 2019).

By design, the policy had heterogeneous effects on family income based on pre-child earnings. Families who had higher pre-child earnings received more transfers from the state than families with lower earnings. Families who were eligible for the 300-euro transfer under the prior policy regime saw no change in their transfers in the first 12 months after the birth of their child, but they lost the monthly 300 euros for the second 12 months. There were some families who were eligible for the 300-euro transfer prior to the reform but had higher monthly earnings, so the

[1] Single parents are allowed to take all 14 months.

transfers they received in the first 12 months were higher. They also lost the 300-euro transfers for the second 12 months. Families who were ineligible for transfers prior to the policy change became eligible for at most 1,800 euros per months for the first 12 months after childbirth. These differences in transfers by pre-child family income can lead to heterogeneous effects on wage dynamics. I cannot calculate pre-child household income, so I use maternal pre-birth education as a proxy for the income category. I analyze the effects of the policy on mothers with no vocational training or university degree (mother with low education) and on mothers with at least a university diploma (mothers with high education).

Conceptual background on the effects of the Elterngeld on maternal wages

Labor force participation for mothers with young children has been relatively low in Germany. In 2006, 61.2% of women between the ages of 15-64 with at least one child (aged 0-14) were employed, compared to the average of 65.3% of the Organization of Economic Cooperation and Development (OECD, 2020). The paid parental leave reform changed the benefit structure with the specific aim to influence female labor force participation. The policy could achieve this goal via several pathways. First, it encouraged women to work prior to having children, so that the transfers they received while on leave were higher. Second, under the new policy regime more women, specifically those who had higher incomes pre-child, received transfers from the state in the first 12 months after childbirth. This could have led some women to take longer leaves given that their incomes were guaranteed at 2/3 of their pre-child earnings up to 12 months. Kluve and Tamm (2013) found a 6 percent reduction in the share of mothers who worked during the first year after giving birth induced by the reform. This pathway could lead to a larger loss in wages for mothers in the cohorts eligible for the reform in the first year after childbirth compared to previous cohorts, because while previous cohorts may have returned to their pre-child earnings before 12 months, the eligible cohorts would stay on leave and hence have lower wages for a longer time period. This pathway is expected to be more pronounced for women with higher pre-child earnings.

The third pathway of the policy influencing female labor force participation is through encouraging mothers to return to the labor force earlier. During the prior policy regime, eligible mothers received transfers up to 24 months, which under the new regime was reduced to 12 months. This implies that mothers have to go back to work earlier to retain their monthly income. This pathway is expected to apply especially to women with lower pre-child earnings as they would have received transfers during both systems. Bergemann & Riphahn (2022) found that those mothers who received transfers under both systems returned to the labor force 10 months earlier at the median after the reform. Mothers who were not eligible for transfers before 2007 returned 8 months earlier. Maternal earnings in cohorts eligible for the reform should be higher in the months when they work instead of staying at home than earnings of mothers in non-reform eligible cohorts. This would mean that the policy increased wages for at least some months between 12-24 after childbirth. Given the shorter disruption of employment, mothers post-reform might advance in their careers faster than the pre-reform cohort, which could lead to higher wages, as well. Thus, the policy could have reduced wage loss after children starting in the second year after childbirth. This reduction is likely more pronounced for women with lower pre-child earnings.

Post-child earnings of women and gender gaps in earnings is also influenced by paternal leave taking and fathers' participation in both family life and in the labor force. While parental leave for fathers was technically available prior to the reform, only 3.5% of fathers took any leave in 2006 (Bünning, 2015). The reform encouraged fathers to take time off work on two accounts. First, the 2/3 earnings replacement provided a higher income for many fathers during the months they were on leave. Second, if each parent took at least 2 months of leave, the total allowance increased from 12 to 14 months. In line with this incentive, the share of fathers who took leave increased to 34% by 2014 (Huebener et al., 2016). Fathers whose partners had a higher education were twice as likely to take time off as fathers with partners who had a lower education (Huebener et al., 2019). These changes in paternal behavior could have a direct effect on their wages, which is expected to decrease for the duration of their leave (as the replacement rate is 2/3 with cap). This by design makes wage loss after children larger for reform-eligible fathers. This could have an indirect effect on maternal wages if in turn women are encouraged to work during those months when the father is on parental leave to increase monthly family income. However, 55% of fathers took leave simultaneously with their partner (Bünning, 2015), and anecdotal evidence suggest that families used this time to take a long vacation together.

Paternal leave taking can influence maternal wages through an indirect channel if the change in paternal behavior around child care is persistent and leads to changes in norms around family life participation. Policies that incentivize fathers to stay home for some time following the birth of their children were found to increase time spent with children and on household duties in several countries (Haas & Hwang, 2008; Kotsadam & Finseraas, 2011; Nepomnyaschy & Waldfogel, 2007; Tanaka & Waldfogel, 2007). Bünning (2015) and Tamm (2019) both analyze the *Elterngeld* and find that fathers reduced their working hours in the short-term, and increased the time they spend with child care, which is persistent even after their return to work. If social norms around female labor force participation after having children change and women are more likely to continue working after having children, maternal wage loss after childbirth can be reduced in the longer-run. However, these changes in social norms are relatively slow, and even if fathers spend more time with their children when they are not at work, they may not change their behavior in the labor force.

While I am not able to measure gender gaps in earnings in the Sample of Integrated Labour Market Biographies (SIAB) (Berge et al., 2021) the policy change could have an effect on gender gaps in wage loss following children. The definition of the child penalty provided by Kleven, Landais, & Søgaard (2019) is "the percentage by which women fall behind men due to children" (p.182). Since this number includes both the changes in maternal and paternal wages, both a smaller penalty for women and a larger penalty for men would result in a smaller overall gender gap. As outlined above, the policy by design could increase paternal child penalty in the short-run (for those fathers who took leave), and maternal child penalty is expected to increase in the short-run and decrease in the long-run. Depending on the magnitudes of the increases in paternal and maternal child penalties in the short-run, the gender gap could either decrease, increase, or stay the same. In the longer run, the gender gap could decrease if maternal child penalty decreases or if social norms change in a way encourage fathers to substitute work time to more family time.

To summarize, I expect to see an increase in wage loss (a larger child penalty) for mothers in the first year following childbirth, especially those who had higher earnings pre-child. After the first 12 months, I hypothesize a decrease in wage loss (a smaller child penalty), more pronounced for women with lower earnings pre-child. I do not expect to see a change in wage loss in the long-term.

The *Elterngeld* specifically aimed to increase equality on the labor market between men and women and to encourage higher female labor force participation. If families faced no constraints when they made decisions about allocating time for market work, childcare, housework, and leisure, the outcome of gender equality in the labor force would be a desirable outcome without any qualifiers. However, given time and monetary constraints, as well as personal preferences around spending time with children and market work, even if the policy led to smaller child penalties for women and a more equal labor market, it might have reduced family welfare.

The effect of Elterngeld on household income and child development

The literature on the effects of the *Elterngeld* has focused on two other outcomes: household income and child development. Concerning household income, the policy influenced some, mostly low-earner, households negatively. Those families who would have been eligible for 24 months of transfers in the old policy regime but were only eligible for 12 months under the new regime could have experienced a decrease in their household income in the second year. Indeed, while household income in the first year remained largely unchanged for previously eligible families, it decreased by about a 1000 euros in the second year after childbirth (Huebener et al., 2019). Families who were ineligible during the prior reform (in general, higher-earner households) benefited from the reform. Huebener et al. (2019) estimate that household income for these families increased by about 7500 euros during the first year after childbirth. Changes in household income in the first few years after childbirth could indirectly influence decisions parents make about when and how much to participate in the labor market.

Child development can be influenced on three accounts: (1) shorter overall time with parents; (2) longer time spent with fathers; and (3) changes in household income. Changes in household income are relatively small for most households especially in the long-term (Huebener et al., 2019). In a model of children’s cognitive development Del Boca et al. (2014) show that time spent with both parents matters more for cognitive development than financial investments. These points make it more likely that overall time with parents and time with fathers are the more important channels for child development. The effect of these changes would depend on alternative care arrangements and the quality of both parental time and non-parental child care. Without directly measuring the quantity and quality of the types of care children receive, it is difficult to know how the potential changes in the amount of maternal, paternal, and non-parental care are going to influence child development.

Two papers analyze the effect of *Elterngeld* on child development. Huber (2019) finds that infant socio-emotional development was negatively affected by the reform. This effect is driven by families who became worse off following the reform. When looking at an index of physical health mostly relying on doctor's visits and whether children have medical problems, they find no changes in outcomes following the reform. Huebener et al. (2019) use data from school entry examinations to assess the effect of the policy on different child development indicators. These consist of four indicator variables of whether the child lags behind in language development (whether they can use prepositions, build plural words, or repeat pseudo-words); whether they lag behind in gross motor development (whether they can stand and jump on one feet); whether they have socio-emotional problems (whether they receive medical or psychological treatment); and whether they are ready for school overall assessed by a pediatrician. They find very small effects of the reform on these indicators. They also look at whether maternal employment, paternal leave taking, or the availability of child care in the county of residence mediates the effects, but they find no significant impacts. They conclude that the reform had no effect on children.

Dataset and Measures

I use the Sample of Integrated Labour Market Biographies (SIAB) from the Integrated Employment Biographies of the Institute for Employment Research (IAB) (Berge et al., 2021). The SIAB is produced by the Research Data Centre (FDZ) of the Federal Employment Agency at the IAB. The SIAB is an administrative dataset that includes labor market histories of individuals who paid social security contributions, which is about 80% of the working population. The dataset does not include civil servants and self-employed workers. A 2% sample of this population is available for research use, which covers 1,940,692 individuals and their employment biographies (Dorner et al., 2010). These biographies are recorded continuously, so indicators can be created up to daily frequency. The dataset includes time periods (spells) of employment, as well as periods when people receive unemployment or other benefits in accordance with the Social Code Book III and Social Code Book II. Periods of registered job search are also included. The data are available for Western Germany since 1975 and for Eastern Germany since 1992. I use labor market histories for the years 1998-2019.

One limitation of the SIAB dataset is that it does not contain specific information on childbirth for individuals. Women start maternity leave 6 weeks prior to their due date, which is recorded in the dataset. It is possible to impute children's date of birth using this information for women. It is not possible to infer when men become fathers using the SIAB only. Fathers would also deregister from employment when they go on parental leave, however, as there is no universal leave taking at a specific time relative to children's birth date, birth dates of children cannot be inferred for men.

I use the strategy and code provided by Müller, Filser & Frodermann (2022) to generate an imputed birth date for women who go on maternity leave. They use the "grund" variable to identify when someone deregisters from employment due to "wage compensation from a statutory health insurance", or code 151. The same code is used when someone goes on long-term sick leave, so they only impute childbirth for women who are at most 40 years old when they start their leave. The imputed birth date of children is the date when women start maternity leave plus 42 days (6 weeks). I use mothers' wages 2 months before giving birth as the baseline wage in my analyses below to account for the fact that they are in general already on leave in the month before giving birth. There are several cleaning steps, too. If someone stays on leave for fewer than 98 days (maternity protection lasts 6 weeks prior to giving birth and 8 weeks after birth), women are considered to be on sick-leave and not on maternity leave. The number of days between 2 potential childbirths is also checked to make sure that sufficient time passes between two maternity leave periods. This method can identify about 50% of all births (Müller et al., 2022), because (1) the dataset does not include civil servants and self-employed workers; (2) the method cannot differentiate between live- and stillbirths; (3) it cannot identify twin births; and (4) it cannot account for multiple births if the mother is not employed between subsequent children.

Employment histories are reported continuously as spells in the SIAB (Dorner et al., 2010). It means that one observation (for employment spells) is a period between two specific dates when a person is employed in the same position.[2] Their wages are reported as a daily wage during that period. It is also indicated whether they work part-time or full-time. I generate monthly wages using the following procedure. I first adjust wages for inflation using 2015 as the base year for the consumer price index (CPI), following Dauth & Eppelsheimer (2020). Then, for each day in each month between January 1998 and December 2019 I check whether the employment spell contains that specific day. If it does, I assign the daily wage to that day. Then, I sum the daily wages by month. This creates a monthly panel for the period January 1998 to December 2019. If an individual has multiple jobs, wages from all jobs are added up.

The SIAB contains a relatively limited set of demographic variables. I use birth year to measure people's age, and the information on country of birth to generate an indicator variable for

[2] Benefit Recipient Histories while one is on unemployment or other benefits and Jobseeker Histories when people are registered to be on job search are also reported in the dataset (Dorner et al., 2010).

whether one was born in Germany or in a different country. I also generate three variables for the highest level of education relying on the imputed education variable from the FDZ (following Dauth & Eppelsheimer (2020)): (1) no vocational training or university diploma (or "low education"); (2) completed vocational training (or "mid-level education"); (3) degree from a university (or "high education")[3]. I use these education categories to proxy for pre-birth household income. Monthly earnings of women in the high education category are 2,872 euros on average, monthly earnings of women in the middle education category are 1,797 euros on average, and monthly wages of women with low levels of education is 1,214 euros on average. In robustness checks, I also generate "high-income" and "low-income" categories using pre-birth earnings. Women who earn below the median are in the "low-income" category and women who earn above the median are in the "high-income" category.

My sample includes women who had a first birth between 2003-2007. The policy came into effect on January 1, 2007 and I use a differences-in-regressions-discontinuities design to measure its impact. As I outline in section 5.1 below, for this identification strategy I need information not just on women who gave birth around the cut-off date, but also on women who gave birth in the same period in prior years. I chose to include only first births because the change from no children to one child is likely different than the changes families experience with subsequent children. I further restrict my sample to include women who are between 18-40 years old when they have their first child. The imputation strategy for children's birth dates limits maternal age at 40, and I use 5 years of employment and wage history before childbirth. Women would be too young to work if I included mothers below 18. My analytic sample includes n =

[3] The imputed variable considers data errors and reporting differences across years. The sample size is similar if I use the raw education data provided in the dataset.

22,909 women with a first childbirth in 2003-2007 and age 18-40 at the time of first childbirth. For the regression discontinuity design, I only use births in the first (January-March) and last quarters (October-December) of the years.[4] This sub-sample with only Q1 and Q4 births includes 8,811 mothers.

Table 1 includes summary statistics for the sample. Column 1 includes women with a first birth in 2003-2007, column 2 includes women with only Q1 and Q4 births, and column 3 further restricts the sample to include only women with Q1 and Q4 births who also have baseline wage data 2 months before giving birth. Women are 29 years old on average when they have their first child, and about 9% of them were not born in Germany. 13.4% of them have a university degree and another 13.4% of them have low levels of education (no vocational training or university diploma). Mothers earn on average about 2,500 euros before giving birth. There are no statistically significant differences between the samples in column 1 and column 2. Column 3 (the sample with a baseline wage in t = -2) is slightly older (by 0.4 years) and better educated on average (1 percentage point fewer people have low education). These differences in sample means are statistically significant at 5%.

Outcome variable: wages

The main outcome variable of interest is monthly wage. When women are on maternity leave, they are still employed but their wages are set to 0 as they do not earn their monthly income from employment. If they do not return to work once their maternity leave ends, they disappear from the dataset. In general, when someone is not employed or is not actively looking for a job,

[4] In 2003 I only use the last quarter and in 2007 I only use the first quarter, so this sub-sample includes births in October-December 2003, January-March 2004, October-December 2004, January-March 2005, October-December 2005, January-March 2006, October-December 2006, and January-March 2007.

they are absent from the SIAB. If they are employed in a civil servant position or they are self-employed, there are also no observations. For mothers a typical employment history looks the following: (1) employment with non-zero wages; (2) maternity leave with 0 wages; (3) no observations for a period (when they stay out of the labor force to spend time with their children); (4) employment with non-zero wages (if they return to a non-civil-servant or non-self-employed position). For the purposes of my analyses, there is important information missing in periods when mothers are not observed in the dataset. Hence, I generate a balanced panel where I impute all months without observation in the dataset between January 1998 and December 2019 with 0 wages. This method will set some women's wages as 0 who are in effect earning wages as civil servants or are self-employed. I run the analyses both with and without the imputed zeros. Wages are top-coded as there is a ceiling for social security contributions. If one earns more than the contribution ceiling, their wage is reported as the ceiling. About 2.7% of the observed wages are top-coded in my sample of mothers.

In graphs and regressions describing wage dynamics of mothers around childbirth I use an annual frequency. I index the year of childbirth as $t = 0$. I generate the mean of monthly wages over the year for 5 years before and 10 years after childbirth as the outcome variable in these analyses.

In analyses investigating whether the paid leave policy change influenced maternal wages, I use monthly wages. I generate a *wage loss after childbirth* variable to capture by how much maternal wages change after childbirth compared to pre-child wages. I use $t = 0$ to index childbirth. To define wage loss in period $t + 1$, I use the following formula: $\frac{y_{base} - y_{t+1}}{y_{base}}$, where y_{t+1} is an individual's salary 1 month after childbirth, and y_{base} is the individual's salary 2 months before childbirth (as mothers are in general on leave the month before birth and their wages are 0 by design). For example, if someone earned 1800 euros 2 months before childbirth, and 300 euros 1 month after childbirth, the child penalty in period $t + 1$ is $\frac{1800-300}{1800} = 0.83$. This number means that the individual earns 83% less after childbirth than what they earned before giving birth. If 10 months after childbirth they earn 1200 euros, the child penalty in period $t + 10$ is $\frac{1800-1200}{1800} = 0.33$, meaning that they earn 33% less 10 months after childbirth than what they earned before childbirth. I generate this monthly *wage loss* variable for 60 months (5 years) after childbirth for each mother individually.

Notes on interpretation: I analyze whether the paid leave policy change influenced wage loss. If the reform decreased wage loss, the group affected by the reform should retain a higher percentage of their pre-child wages, making their wage loss outcome smaller. Staying with the previous example, let's say an individual in before-reform era earned 1800 euros before giving birth and 300 after giving birth, earning 83% less after childbirth. In a counterfactual scenario, the individual in the reform era earned more than 300, say 500. Wage loss for the counterfactual scenario would be $\frac{1800-500}{1800} = 0.72$. This means that wage loss in the reform era is smaller. The difference in wage loss is 0.72-0.83= -0.11. We would interpret this as the policy had an effect of -11 percentage points. Hence, if the policy decreased wage loss, effect sizes are going to be negative. From the point of view of the policy's goals a lower wage loss is a "positive outcome". Thus, a negative coefficient is interpreted as a "positive outcome". A positive coefficient would mean that the policy increased wage loss, so it is a "negative outcome".

How to interpret the coefficients if someone earns more after giving birth than before? Let's say they earn 1800 right before giving birth and 2500 after giving birth. Then their earnings difference would be $\frac{1800-2500}{1800} = -0.38$. This means a 38% increase in salary. If in the counterfactual scenario of the reform era they earn 2000 post-child (instead of 2500), their "wage loss" would be $\frac{1800-2000}{1800} = -0.11$ (still an increase, but of a smaller 11 percentage points). The difference between the two counterfactual outcomes is -11-(-38) = 27 percentage points, which is bigger than zero, so we can interpret it the same way as above: the policy had a negative effect on earnings (because the individual is earning 2000 as opposed to 2500 post-child). If in a different counterfactual scenario the wage went up to 3000, the "wage loss" variable would show $\frac{1800-3000}{1800} = -0.66$, or a 66% increase in wages from pre- to post-child. Here, the effect of the policy would be -66-(-38) = -28 (instead of earning 2500 the individual earns 3000 post-child compared to the 1800 base). This difference is smaller than 0, and the interpretation is the same as in the other cases: the policy had a "positive effect" on wages.

Descriptive statistics

Panel B of Table 1 reports descriptive statistics for a set of outcome variables. Wages 3 months after childbirth are between 700-900 euros per month across the samples, but only a small percentage of mothers have non-zero wages this close to childbirth. If we impute all missing observations as 0, wages drop to between 200-300 euros. About 30% of mothers are reported to work part-time 3 months after birth, but again, a large share of the sample has missing information. A year after giving birth, maternal wages are around 1,200 euros per month, which is about half of pre-child wages. Around 55% of all mothers in the sample have missing wage data 13 months after childbirth. 50% of those who work are employed part-time. Their wages increase somewhat to about 1,400 euros per month by month 24 (2 years after childbirth). 45% of mothers have no wage information 2 years after childbirth. Close to 60% of mothers work part-time 2 years after childbirth, but this information is only available for about 55% of the sample.

Documenting maternal wage dynamics around childbirth

In this section, I document maternal wage dynamics around childbirth in Germany for mothers whose first children were born in 2003-2007. I show that mothers experience a large and persistent drop in monthly wages right after childbirth, both overall and by education level. Monthly wages do not reach pre-birth levels for mothers in my analytic sample whom I follow for 10 years after childbirth, except for the group with the lowest level of education. However, once I account for the effect of age and economic growth in wages, the differences between the groups disappear.

I use event study regressions to document wage dynamics around childbirth. The event is the birth of the child. I set the year when the child is born to $t = 0$. Wages in each year are indexed to the event of childbirth. I report wages annually for 5 periods before and 10 periods after childbirth. To describe maternal mean wages before and after children I run the following regression:

$$(1)\ Y_{ist} = \sum_{j \neq -1} \alpha_j \, \mathbf{I}[j = t] + \varepsilon_{ist}$$

where Y_{ist} is the average monthly wage in year s for individual i in event time t. $\mathbf{I}[j = t]$ depicts a full set of event time dummies. I use the year before childbirth as the baseline wage, hence omit $t = -1$ from the regression. The $\widehat{\alpha_j}$ coefficients are the mean wages for each event time compared to the year before childbirth. I run the same regression for sub-samples of mothers with low levels of education, mid-level education, and mothers with high levels of education. Standard errors are clustered by individual.

Women earn on average 2,296 euros per month in the year before childbirth. Their wages decrease to 1,088 euros per month on average in the year after childbirth, which is a (1,088-

2,296)/2,296*100 = 53% drop. They continue to earn less than before childbirth during the 10-year period following childbirth. Even 10 years after the arrival of children mothers earn on average 23% less than before children. Women with high levels of education (at least a university diploma) earn 3,498 euros per month on average in the year before childbirth, which drops by 43% to 1,980 in the year after childbirth. Mothers in this group earn 16% less 10 years after childbirth than they did pre-child. Women with low levels of education (no vocational training or university diploma) experience a similarly large decrease in earnings after childbirth, a 46% drop (from 1,244 euros to 669 euros). However, their raw earnings recover faster than wages of women in the two other groups. They reach their pre-child earnings 8 years after childbirth and have higher monthly wages in years 9 and 10 after childbirth than they did pre-child. Women with mid-level education (completed vocational training) earn on average 2,358 euros per months before childbirth, which drops by 28% to 1,687 after childbirth. They earn 28% less 10 years after childbirth than what they were earning before having children. Figure 1 Panel A shows maternal wages 5 years before to 10 years after childbirth as percentages of the baseline wage right before childbirth for all four groups.

The means reported in the previous paragraph do not account for the fact that many mothers exit the labor force when they have children at least for some time. For example, 13 months after childbirth only 46% of women have any wage data reported in the dataset. To correct for this selection, I produce similar estimates using a balanced panel where all missing wages are imputed as zero. This provides a lower bound for the estimates of maternal wages around childbirth. The imputed mean wage in the period before childbirth for the full analytic sample is 2,144 euros per month. Imputed maternal wages drop by 77.5% to 483 euros. Even 10 years after childbirth mothers earn 38% less than they did before childbirth once I account for the missing wages. Women with low levels of education experience a similarly large drop as the overall sample (75%), but their wages recover faster. They do not reach pre-child levels of imputed wages, though. 10 years after childbirth women in this group earn 20% less than before children. Imputed wages of women with high levels of education decrease by 67.5% right after childbirth. They earn 33.5% less 10 years after childbirth. Imputed wages of women with mid-level education drop by 51%, and they earn 41% less 10 years after childbirth than they did before having children. Figure 1 Panel B shows imputed wage dynamics for all four groups of women.

These differences in maternal wage dynamics by educations groups around childbirth can be explained by several factors. For example, age can drive the results. Women have children at different ages, and women with lower levels of education have children earlier on average. In my sample, women with low education are 25 years old on average when they have their first child, and women with at least a university degree have their first child on average at age 32. Wages also tend to be higher when one is older, which contributes to the different levels of wages mothers have before childbirth. Women at different ages are also at different stages of their careers, which likely contributes to the differences in wage dynamics. To control for these age-related life-cycle effects in wages, I run event-study regressions with including not only the event time dummies, but also indicator variables for each observed age in the sample. The model I estimate is the following:

$$(2)\ Y_{ist} = \sum_{j \neq -1} \alpha_j \, \mathbf{I}[j = t] + \sum_k \beta_k \mathbf{I}[k = age_{is}] + \varepsilon_{ist}$$

For each group of women, I estimate equation (2) separately, and I omit the indicator variable for the mean age (29 for the full sample, 25 for the low-education group, 29 for the mid-level education group, and 32 for the high-education group). I also omit event time -1. The coefficient on my omitted variable (the constant in the regression) is an estimate of the average wage in $t = -1$ net of age effects. This can be interpreted as estimating the $\hat{\alpha}_j$ coefficients as if everyone in my sample had the mean age.

Figure 2 shows wage dynamics around childbirth adjusting for age effects. Panel A uses observations with non-missing wage data. After controlling for the effect of age in wages, mothers in all three groups experience very similar dynamics. Their wages drop by about 50% in the year after childbirth, and then stay relatively flat. Women with low education experience an increase in their wages, but even 10 years after childbirth they earn 32% less than before childbirth. Dynamics are qualitatively similar if I use imputed wages. The drop in wages after childbirth is very large, between 70-80% across the four groups. Women with low levels of education again recover some of their lost wages, but they still earn only about half of what they did pre-child.

A second factor I consider is year effects. Maternal wages are likely influenced by economic growth and events such as the Great Recession of 2007-2009, with heterogeneous effects on groups by levels of education. To control for such events, I include a set of year dummies in my event-study regressions next to the event time and the age dummies, making the full specification of my model the following:

$$(3)\ Y_{ist} = \sum_{j \neq 1} \alpha_j\, \mathbf{I}[j = t] + \sum_k \beta_k \mathbf{I}[k = age_{is}] + \sum_y \gamma_y \mathbf{I}[y = s] + \varepsilon_{ist}$$

I omit the mean of the year variable, 2008, to make the $\hat{\alpha}_j$ coefficients comparable to estimates from the other equations. I calculate the percentage change in wages by event time using the coefficient on the omitted variable as the baseline.

Figure 3 shows the $\hat{\alpha}_j$ coefficients as a percentage of the baseline wage. Controlling for macroeconomic factors makes the estimated wage-drop larger. The recovery that women experience over time disappears, which can be interpreted that it was mostly driven by economic growth (real wage growth). Women with low levels of education experience the lowest drop in their earnings, they earn about 40% of their pre-child wage throughout the observed period. Women in the full sample and those with higher levels of education experience a 55% drop that increases to 60-65% during the 10 years after first childbirth.

The effect of the policy change on maternal wages

In this section of the paper, I investigate whether the paid family leave policy change influenced the wage drop mothers experience around childbirth. As outlined in section 2.1, I expect to see an increase in wage loss in the first year following childbirth (the drop in wages is going to be larger under the new policy regime than what it was under the old regime). I expect this increase in wage loss to influence especially those who had higher earnings pre-child. After the first 12 months, I hypothesize a decrease in wage loss: the drop from pre-child to post-child wages in the 2nd year after childbirth to be smaller in the new regime than what they were in the old regime. I expect this effect to be more pronounced for women with lower earnings pre-child. I do not expect to see a change a change in wage loss in the long-term.

Identification strategy

This section answers the research question of whether the paid family leave policy change had an effect on labor market outcomes. The ideal experiment would randomly assign future potential parents to either the old (control group) or to the new policy regime (treatment group) and would look at the differences in average outcomes between these two groups. Since this ideal experiment is not feasible, I rely on the date when the policy became effective to allocate people to treatment and control groups. Parents whose children were born on or after January 1st, 2007 were eligible for the new paid parental leave benefits (treatment group), and people whose children were born on or before December 31st, 2006 stayed under the old regime (control group). This set-up would make a regression discontinuity (RD) design possible.

The identifying assumption for the RD design is that the potential outcomes are continuous with respect to the assignment variable, or in other words, there are no discrete changes in the outcome variables except due to the treatment. In this specific case, this means that in the absence of the policy change we would not expect to see any jumps or discontinuous changes in labor market outcomes of parents whose children are born in December compared to parents whose children are born in January. This also means that all observed and unobserved characteristics are expected to be continuous (or "vary smoothly") around the cut-off (Lee & Lemieux, 2010).

In Table 2, panel A, I compare mothers who gave birth in the year 2006 to mothers who gave birth in the year 2007. Their observable characteristics are mostly similar. Both are around 29 years old when they have their first child, about 9% of them were born not in Germany, 14.5% have at least a bachelors degree, 13% have no vocational training or university education. They earn around 2200 euros a month before giving birth. The pre-birth wages including the imputed zeros of mothers who gave birth in 2006 are 1951 euros, versus 2013 for the mothers who gave birth in 2007, which is the only statistically significant difference between the two groups.

To analyze the effect of the policy using a 12-month bandwidth around the cut-off date, I specify the following relationship between the variables of interest:

$$(4)\ y_i = \beta_1 T_i + f(m) + f(m)T + \gamma \boldsymbol{X}_i + \varepsilon_i,$$

where y_i refers to the outcomes of individual i (wage loss in period 1, 2, etc., each of which is a separate outcome in a separate equation). T is an indicator variable =1 for the year 2007, $f(m)$

refers to the function of the running variable m. More specifically, it indexes the month of birth relative to the cut-off of January 1st, so December is -1, January is 1, February is 2, etc. I allow this function to be different after the cut-off in T. $\boldsymbol{X}$ contains a set of demographic control variables (age at first birth, being foreign-born, indicators for low and high education, baseline wage 2 months prior to giving birth, part-time employment 2 months prior to giving birth), and ε is the error-term. I estimate this regression using Ordinary Least Squares.

Prior research has documented a seasonality in births in the United States, which would make the identifying assumption invalid. Women who give birth in January tend to be less educated, younger, and are less likely to be married (Buckles & Hungerman, 2013; Currie & Schwandt, 2013). These findings could bias my results the following way. In general, people with fewer years of education tend to earn lower wages. If more people with lower wages give birth in January, the average wages in January are going to be lower. Attributing all of the differences in wages between December and January to the policy could overestimate the effect of the policy.[5]

To overcome this limitation, I use a differences-in-regression-discontinuities design to estimate the effect of the policy change on maternal labor market outcomes, which is a popular method to study the effects of family leave policy changes on parental labor market and health outcomes (Lalive et al., 2014; Lalive & Zweimüller, 2009; Persson & Rossin-Slater, 2021; Schönberg & Ludsteck, 2014). For the RD set-up, I use October-December births in 2006 just before the policy as the control group, and January-March births in 2007 as the treatment group. I

[5] Misreporting of birth month could be another concern. Torun & Tumen (2017) show that in Turkey 20% more births are reported for January than for other months driven by families with lower-socioeconomic status due to geography and institutional reasons. This fact is relevant for my project because 3% of the nationally representative German Socio-Economic Panel survey sample report a Turkish nationality, which is the largest ethnic group after German (86%). However, the strategy I use to infer birth dates relies on administrative data on when mothers deregister from employment to go on maternity leave, so misreporting of birth dates is not of concern for this project.

call these births together the “reform sample”. I also use the same time periods in the three preceding years to control for seasonality, which constitute the “non-reform” sample. I use births from October 2003-March 2004, October 2004-March 2005, October 2005-March 2006 as the “non-reform sample”, with October-December births as the “control group” and January-March births as the “treatment group”, and October 2006-March 2007 births as the “reform sample”, with the same treatment-control set-up. I use a robust data-driven estimation to select the optimal bandwidth around the cut-off date (Calonico et al., 2014, 2017). The optimal number of days is 84 around the January 1st introduction date. Since this is very close to using three months on each side, I keep the three-month bandwidth in the below analyses.

I specify the following equation to describe the relationship of interest:

$$(5)\ y_i = \beta_1 Q_{1i} + \beta_2 R_i + \beta_3 Q_{1i} R_i + f(m) + f(m) Q_1 + \gamma \boldsymbol{X}_i + \sigma_p + \varepsilon_i,$$

where y_i refers to the outcomes of individual i (wage loss in period 1, 2, etc., each of which is a separate outcome in a separate equation). $Q_1$is an indicator variable =1 for the months of January-March, R is an indicator variable =1 for the reform sample (people whose children was born between October 2006 and March 2007), $f(m)$ refers to the function of the running variable m. More specifically, it indexes the month of birth relative to the cut-off of January 1st, so December is -1, January is 1, October is -2, etc. I allow this function to be different after the cut-off in Q_1. $\boldsymbol{X}$ contains a set of demographic control variables (age at first birth, being foreign-born, indicators for low and high education, baseline wage 2 months prior to giving birth, part-time employment 2 months prior to giving birth), σ_p is period fixed-effects, and ε is the error-term. I estimate this regression using Ordinary Least Squares. β_3 is the coefficient of interest, $\hat{\beta}_3$ is the estimate of the effect of the policy change on the reform sample if the identifying assumptions hold.

I build up my analyses to the full specification in (4) step-by-step. First, I estimate the RD models separately for the reform and non-reform groups by adding each term sequentially. I start with only the Q_1 dummy, then add control variables, then add the trends:

(6) $y_i = \beta_1 Q_{1i} + f(m) + f(m)Q_1 + \gamma \boldsymbol{X}_i + \varepsilon_i$,

then I pool the samples and add the "differences" (the reform dummy and the interaction term between reform and quarter 1) with the period fixed-effects.

Table 2 shows differences in means of baseline variables by reform/non-reform group status. There are no statistically significant differences between observable characteristics of women who gave birth in October-December 2006 (reform sample, control group, n = 1054) and women who gave birth in January-March 2007 (reform sample, treatment group, n = 1026). Women who gave birth in January-March of 2004, 2005, or 2006 (non-reform sample, treatment group, n = 3289) are 0.6 years older on average than women who gave birth in October-December 2003, 2004, 2005 (non-reform sample, control group, n = 3442). Women in the non-reform treatment group are 2 percentage points more likely to have at least a university degree and 3 percentage points less likely to have low education.

Manipulation of birth dates of children is a concern in this institutional set-up. Some women may have tried to give birth in the new year so that they become eligible for the new benefits. This would be possible by delaying labor inductions for example. As long as individuals are not able to precisely manipulate the date of the birth of their children, there would be a source of randomized variation in the treatment status very close to the threshold, which would make the RD strategy theoretically valid (Lee & Lemieux, 2010). It is not possible to precisely control the start and the length of all births, so if the sample had enough births one minute before midnight and one minute after midnight, the two groups of mothers would be expected to be very similar to each other. However, researchers often extend the time period around the cut-off to include parents of children who are born in the few weeks or months on the two sides of the threshold to increase sample size. As one increases the time period considered around the policy introduction date, it becomes less likely that the treatment and control groups stay comparable to each other.

Looking at a 7-day period around the January 1, 2007 threshold, papers have shown that about 1000 births, which account for about 8% of births in the general population, were shifted from the last week of December to the first week of January (Neugart & Ohlsson, 2013; Tamm, 2013). This shift was driven by working women. On average about 40% of births are a result of C-section or a labor induction on a working day, and 25% on a weekend or public holiday. Jürges (2017) shows that 80% of the "missing" births in December and 90% of the excess births in January can be explained by delayed elective C-sections and labor inductions. This means that the majority of the births are still expected to fall randomly on either side of the threshold. I could over- or underestimate the effect of the policy if people who achieved the delayed C-sections and labor inductions have observable characteristics that correlate with potential outcomes. However, I use the date when mothers go on maternity leave 6 weeks prior to their due date to infer the birth date of children. As long as mothers do not adjust the first day of their maternity leave systematically, I will infer non-manipulated dates of birth. This will allow me to estimate intent-to-treat effects.

Looking at the number of births by week compared to the January 1, 2007 cut-off would provide some evidence that the manipulation of birth is not of concern in my set-up. Figure 4 graphs the number of births by week for one year before and one year after the cut-off date. While there appears to be more weeks with a higher number of births after the reform, there is no discrete jump at the cut-off date. The local linear smoothing shows somewhat fewer births in the first week of January (which is the opposite of the expected direction if women delay their births), but the 95% confidence intervals are highly overlapping.

As a last threat to identification, the policy could have induced some families to have (more) children. The government coalition decided on the reform in May 2006, and parliament agreed in September 2006 (Kluve & Tamm, 2013), which means that parents whose children were born around the January 1, 2007 date did not know about the policy change when they decided to have children. This implies that changes in fertility close to the cut-off date do not threaten the identifying assumptions.

Results: The effect of the policy change on maternal wage loss

I first run the Regression Discontinuity analyses of maternal wage loss (the difference in earnings between pre- and post-child) on giving birth in the first quarter of the year. I define a separate outcome for each month after childbirth. I show results for 3 outcomes: wage loss in t=3 (3 months after childbirth), wage loss in t=13 (13 months, or a little over a year), and wage loss in t=24 (2 years after childbirth). I also use imputed wages for these time periods (where I include wage observations of mothers who are not in the dataset in a specific time period as having 0 wages for that period).

Figure 5 shows the change in wage loss in t = 13 (panels A and B) and wage loss in t =24 (panels C and D) around the cut-off date with local linear smoothing and 95% confidence intervals. Panels A and C show the reform sample (when the policy was introduced) and Panels B and D show the same period in prior years (with no change in policy). Wage loss 13 months after childbirth is lower after the cut-off date in the reform sample, while there is no discontinuous change in the non-reform years. Wage loss 24 months after childbirth looks similar in both reform and non-reform years.[6]

Table 3 shows regression coefficients for the RD models. Panel A reports coefficients with a 12-month bandwidth, Panel B reports coefficients for the Reform sample with a 3-month bandwidth, and Panel C reports coefficients for the Non-reform sample with a 3-month bandwidth. There is no statistically significant difference between wage loss 3, 13, or 24 months after childbirth for women who gave birth in 2006 versus 2007, or between women who gave birth in Q4 and women who gave birth in Q1. This is true both for the reform and non-reform sample. There are significant differences in imputed wage loss 13 and 24 months after childbirth. Women who gave birth in 2007 experience a larger wage drop 3 months after childbirth than women who gave birth in 2006. Women who gave birth in 2007 also experience a smaller wage drop 13 and 24 months after childbirth than women who gave birth in 2006. Women in the reform sample who gave birth in January experience a 3 percentage point lower wage drop between their pre-child wages and wages 13 months after birth than women who gave birth in December based on the model with baseline control variables and time trends. The drop is statistically significantly different in the other specifications, too (without controls and/or time trends), but the coefficient changes a lot between the specifications. The coefficients on the 24-month-post-birth wage loss variables also indicate a decrease in wage loss as a result of the policy. However, the point estimates are very large: 10 and 26 percentage points, respectively, in models with controls and with time trends. A change in imputed wage loss but not in the raw wage loss indicates that fewer women have zero wages among those who give birth in Q1 compared to those who give birth in

[6] I cannot produce similar graphs for wage loss 3 months after childbirth because the sample size for non-missing wage information by week is too small.

Q4. The fact that there are statistically significant differences in wage outcomes between women who gave birth in Q4 and Q1 in the non-reform sample, too, makes it less credible that the changes in the reform sample are due to the policy itself. Hence, I turn to the full specification of the model where I check whether the differences in the regression discontinuity estimates are statistically significantly different from zero.

Table 5 reports regression results from the full model specification with a 3-month bandwidth from equation (5) for the same set of outcomes: wage loss 3, 13, and 24 months after childbirth, and imputed wage loss 3, 13, and 24 months after childbirth. Panel A uses the full sample, panel B the sample of mothers who have no vocational training or university degree, and panel C uses the sample of mothers with at least a university degree at the time of childbirth.

I also generate wage loss outcomes for each month up to 60 months (5 years) after childbirth and graph the coefficients on the "treated" dummy for the 12-month bandwidth, and on the interaction term for the 3-month bandwidth with the 95% confidence intervals. Figures 6, 7, and 8 show these estimated coefficients for the full sample, for the low education sample, and for the high education sample, respectively. The estimated coefficients on the figures for periods 3, 13, and 24 are the same as the estimated coefficients in Table 5.

For estimates using a 12-month bandwidth, the wage loss is larger in the initial periods under the new policy regime, but then becomes smaller under the new regime after month 4. This is true both for the raw wages and for the imputed wages in the full sample. For women with low education, the initially higher wage loss than smaller wage loss under the new regime is apparent for the raw wages, but not for the imputed wages. The confidence intervals for women with high levels of education are very wide and include zero for all periods.

In estimates using a 3-month bandwidth, for all three samples, confidence intervals are quite wide for all periods and with very few exceptions contain 0. This means that we cannot reject the null hypotheses that the policy had no effect on wage loss for any period. For the full sample, the point estimates are negative for periods 5-20 for raw wage loss and 9-25 for imputed wage loss, which would be interpreted as wage loss decreasing as an effect of the policy (the desirable outcome from the policy's standpoint). However, these are not statistically significantly different from zero.

The only group where the point estimates show a consistent pattern with my hypotheses is women with low levels of education. The sample size is relatively small, so the estimates are noisy, but they imply that wage loss decreased in each period in the first two years after childbirth. The imputed wage loss starts to decrease after 8 periods. The policy stopped transfers to this group after the first 12 months, encouraging mothers to return to work after the transfers stop to retain monthly earnings. If mothers in the old regime stayed home for the second year, but under the new regime went back to work, we would expect their wage loss 12-24 months after childbirth to be smaller in the new regime. The data support this story. The point estimates on imputed wages (Figure 8, Panel B) suggest that the gap between maternal wages pre-child and 15 months post-child decreased by about 50 percentage points (the average gap is about 70%). The 95% confidence intervals are very wide, but they do not contain 0 for 15-18 months after childbirth.

Considering the sample of women with high levels of education, the coefficients in the first 3 years after childbirth are all very close to zero, with relatively tight confidence intervals (the coefficients are not precisely estimated as the sample size is relatively small). This could be interpreted as the policy had no effect on wage loss for women with high levels of education. No significant change is indicated for the imputed wage loss variable either, which implies that they also did not change their responses on the extensive margin. These findings mean that wage dynamics of women in this group did not change as a result of the policy. They experience a similar drop in earnings compared to their pre-child levels for all periods under the new regime than what they experienced under the old regime. The policy change encouraged this group of women to work before having children (so they have higher transfers during their leave), and to return to work after the transfers stop, however, the data does not support a change in their behavior. I do not test whether levels of wages change, only if the difference changes. It is possible that someone earned 1000 in the old regime in t-2 and 600 in t+3, which is a 40% drop. In the counterfactual scenario, in the new regime, they could have earned 1100 in t-2 and 660 in t+3, which is the same 40% drop. Their earnings are higher both pre- and post-child, but the gap is still 40%, so my analyses would show no effect of the policy. In terms of wage loss after children, their outcome is unchanged.

Conclusion

Women experience a large and persistent drop in their wages after childbirth (Aguilar-Gomez et al., 2019; Andresen & Nix, 2021; Kleven, Landais, Posch, et al., 2019; Sandler & Szembrot, 2019). I document a 53% decrease in monthly wage one year after childbirth compared to their earnings in the year right before childbirth for German mothers who had a first birth between 2003-2007. Their wages are still 23% less 10 years after childbirth than what they were before having children. This increase of 30 percentage points between year 1 and year 10 after childbirth is explained by age effects and economic growth. Women with high and low education levels have similar dynamics, although women with low education experience a smaller drop in their earnings in percentage terms than women with high education.

The decision of how much market work, housework, and child care to provide after children arrive depends on many factors including personal preferences and social norms, as well as institutional, financial, and time constraints. Family and child policies, like paid family leave, can directly or indirectly influence these factors. The German paid family leave policy change in 2007 reduced the number of months parents received transfers for but increased the amount they received. Low-income households received transfers under both regimes, for two years under the prior regime, and for one year after the new regime. High-income households received no transfers under the old regime, and 2/3 of their income in the new regime (Huebener et al., 2019). I use education levels as a proxy to sort women into these two groups.

One specific aim of the policy was to encourage mothers to return to work earlier, which could also lead to a reduction in gender inequality in the labor market. If mothers indeed returned to work earlier under the new regime, their earnings should recover faster after childbirth. I find suggestive evidence that the paid family leave policy change of 2007 reduced maternal wage loss after childbirth in the second year after children were born. This change is driven by mothers with low levels of education. Wage loss of highly educated mothers is not influenced by the policy change. Even though prior research shows that women in both groups returned to work several months earlier under the new policy (Bergemann & Riphahn, 2022), I do not find evidence of this in my dataset and sample. Highly educated mothers who gave birth in 2006 and 2007 have a similarly long period of about 2 years of no market work after the arrival of their children. Mothers with low education return to market work about 3 months earlier in 2007 than in 2006. This suggests that policy incentives influence mothers with lower household income more. They are less likely to be able to stay home with their children in the absence of the transfers. High-income mothers seem to have more flexibility in whether they return to work or stay home; their decision is not so strongly influenced by the paid leave policy.

Paid family leave policy changes with incentives for secondary caregivers to take time off of work are promising tools to increase gender equality. Discussions about how to achieve gender equality often focus on enabling women to have similar labor market outcomes to men or on reducing the impact of children on maternal labor market experiences. However, gender equality can be achieved not just by a change in women's behavior. Policies such as the *Elterngeld* have the potential to change paternal choices around market work, child care, and housework, which could lead to more equitable outcomes between genders both in the short- and in the long-term. In this project, I am only able to study women and their labor market response to the policy change. Future research should focus on the effects of family policies on household decision-making to understand the channels that can lead to gender equality.

Chapter 1 – Tables

Table 1.1: Descriptive Statistics

Panel A: Baseline variables

	(1)	(2)	(3)
sample size	**22,909**	**8,811**	**7,665**
Age	29.1 (4.9)	29.1 (4.95)	29.4 (4.8)
Born abroad	9.2%	9.1%	8.5%
High education	13.6%	13.4%	13.7%
Low education	13.4%	13.4%	12.2%
Wage in t = -2	2,247.3 (1262.1)	2,249.50 (1262.95)	2,249.50 (1262.95)
Imputed wage t = -2	1,973.0 (1392.7)	1,956.9 (1400.1)	2,249.50 (1262.95)
Works part-time t = -2	19.8%	19.9%	19.9%

Panel B: Outcome variables

	(1)	(2)	(3)
sample size	**22,909**	**8,811**	**7,665**
Wage in t = 3	920.0 (1359.7)	723.5 (1201.9)	728.1 (1207.9)
Imputed wage t = 3	291.2 (876.5)	230.7 (757.7)	263.5 (806.4)
Wage in t = 13	1,240.80 (1308.5)	1,264.70 (1306.1)	1,296.50 (1319.1)
Imputed wage t = 13	578.10 (1086.7)	563.80 (1075.0)	636.50 (1128.9)
Wage in t = 24	1,415.20 (1258.1)	1,411.50 (1254.9)	1,451.90 (1266.2)
Imputed wage t = 24	753.50 (1158.2)	745.20 (1152.3)	828.90 (1196.5)

Notes: Sample includes mothers who gave birth between 01-01-2003 and 12-31-2007 and are between ages 18-40 at the time of first childbirth. Column (1) includes the full sample. Column (2) includes only women who gave birth in 2003 October-2004 March, 2004 October-2005 March, 2005 October-2006 March. Column (3) incudes the same women as column (2) who also have baseline wage data in t = -2. Table reports means (standard deviations) of demographic characteristics (age, country of birth, education) at the time of childbirth. Baseline labor market indicators are reported for the period 2 months before childbirth, and outcome variables are reported for periods 3, 13, and 24 months after childbirth. Data come from the German Sample of Integrated Labour Market Biographies (SIAB) (Berge et al., 2021). “Low education” refers to the group of mothers who have no vocational training or a university diploma. “High education” refers to the group of mothers with at least a university diploma. When mothers are not working for pay, when they work in a civil servant job or are self-employed, they have no observations in the dataset. To generate imputed wages, all missing wages are set to 0.

Table 1.2: Differences in means of baseline variables (balance checks)

Panel A: 2006 and 2007 births

	2007 birth "treated"	2006 birth "control"	diff (2006-2007)	p-value
sample size	**4442**	**4297**	**8739**	
Age at first birth	29.21	29.11	-0.10	0.31
Foreign born	8.58%	9.00%	0.41	0.51
High education	14.58%	14.53%	-0.05	0.95
Low education	13.44%	13.05%	-0.39	0.61
Wage in t = -2	2193.88	2170.80	-23.08	0.41
Imputed wage t = -2	2013.11	1951.04	-62.07	0.03
Works part-time t = -2	20.01%	21.28%	1.20	0.19

Notes: Table reports means of demographic characteristics (age, country of birth, education) at the time of childbirth for women who gave birth in 2007 (treated group) and women who gave birth in 2006 (control group). Data come from the German Sample of Integrated Labour Market Biographies (SIAB) (Berge et al., 2021). “Low education” refers to the group of mothers who have no vocational training or a university diploma. “High education” refers to the group of mothers with at least a university diploma. When mothers are not working for pay, when they work in a civil servant job or are self-employed, they have no observations in the dataset. To generate imputed wages, all missing wages are set to 0.

Table 1.3: Differences in means of baseline variables (balance checks, quarters)

Panel A: Reform sample (birth in 2006 October-2007 March)

	Q1 birth "treated"	Q4 birth "control"	diff (Q1-Q4)	p-value
sample size	**1026**	**1054**	**2080**	
Age at first birth	29.17	28.91	0.26	0.21
Foreign born	8.04%	7.87%	0.17	0.9
High education	13.98%	12.37%	0.02	0.33
Low education	14.27%	13.54%	0.01	0.67
Wage in t = -2	2160.66	2154.67	5.99	0.92
Imputed wage t = -2	1960.6	1958.42	2.18	0.97
Works part-time t = -2	19.27%	22.13%	-0.03	0.13

Panel B: Non-reform sample (birth in 2003 October-2004 March, 2004 October-2005 March, 2005 October-2006 March)

	Q1 birth "treated"	Q4 birth "control"	diff (Q1-Q4)	p-value
sample size	**3289**	**3442**	**6731**	
Age at first birth	29.30	28.78	0.59	<0.001
Foreign born	9.20%	9.66%	-0.005	0.53
High education	14.58%	12.64%	0.02	0.04
Low education	11.45%	14.53%	-0.03	<0.001
Wage in t = -2	2290.45	2269.31	21.14	0.52
Imputed wage t = -2	1956.18	1956.14	0.04	0.99
Works part-time t = -2	18.63%	16.27%	-0.02	0.08

Notes: Table reports means of demographic characteristics (age, country of birth, education) at the time of childbirth for women who gave birth in January-March (Q1, or treated group) and women who gave birth in October-December (Q4, or control group). Data come from the German Sample of Integrated Labour Market Biographies (SIAB) (Berge et al., 2021). “Low education” refers to the group of mothers who have no vocational training or a university diploma. “High education” refers to the group of mothers with at least a university diploma. When mothers are not working for pay, when they work in a civil servant job or are self-employed, they have no observations in the dataset. To generate imputed wages, all missing wages are set to 0.

Table 1.4: Regression results for the Regression Discontinuity

Panel A: 12-month bandwidth

	Raw wage			Imputed wage		
	t=3	t=13	t=24	t=3	t=13	t=24
2007 birth	0.055 (0.033)	-0.073 (0.054)	-0.063 (0.048)	0.042*** (0.010)	-0.041 (0.028)	-0.052* (0.030)
N obs.	2230	3945	4764	7935	7935	7935
2007 birth + controls	0.069** (0.032)	-0.057 (0.054)	-0.065 (0.050)	0.044*** (0.011)	-0.037 (0.030)	-0.053* (0.032)
N obs.	2187	3758	4496	7394	7394	7394
2007 birth + controls + trends	0.049 (0.062)	-0.155 (0.101)	-0.166* (0.094)	-0.004 (0.023)	-0.159*** (0.060)	-0.187*** (0.066)
N obs.	2187	3758	4496	7394	7394	7394

Panel B: 3-month bandwidth, reform sample

	Raw wage			Imputed wage		
	t=3	t=13	t=24	t=3	t=13	t=24
Q1 birth	-0.0085 (0.048)	-0.07 (0.041)	-0.019 (0.048)	0.018 (0.016)	-0.054** (0.025)	-0.053 (0.033)
N obs.	531	947	1132	1889	1889	1889
Q1 birth + controls	0.011 (0.045)	-0.068 (0.043)	-0.081 (0.049)	-0.011 (0.018)	-0.090*** (0.030)	-0.102*** (0.037)
N obs.	518	843	978	1634	1634	1634
Q1 birth + controls + trends	-0.141 (0.127)	-0.092 (0.136)	-0.095 (0.158)	-0.187*** (0.058)	-0.03*** (0.098)	-0.257** (0.123)
N obs.	518	843	978	1634	1634	1634

Panel C: 3-month bandwidth, non-reform sample

	Raw wage			Imputed wage		
	t=3	t=13	t=24	t=3	t=13	t=24
Q1 birth	-0.016 (0.02)	-0.015 (0.046)	-0.056 (0.062)	-0.017** (0.009)	0.002 (0.024)	-0.063* (0.037)
N obs.	2242	2811	3238	5770	5770	5770
Q1 birth + controls	-0.015 (0.02)	-0.039 (0.046)	-0.061 (0.066)	-0.046*** (0.008)	-0.046* (0.025)	-0.099** (0.043)
N obs.	2218	2668	3003	5217	5217	5217
Q1 birth + controls + trends	0.002 (0.054)	0.143 (0.127)	0.176 (0.138)	-0.112*** (0.029)	-0.081 (0.066)	-0.036 (0.085)
N obs.	2218	2668	3003	5217	5217	5217

Notes: Each coefficient is from a separate regression. The sample includes mothers who gave birth between 2003-2007, and are between ages 18-40 at the time of first childbirth. The Reform sample includes women who gave birth in October 2006-March 2007. The Non-reform sample includes women who gave birth in 2003 October-2004 March, 2004 October-2005 March, 2005 October-2006 March. Control variables are age at first birth, indicator variables for level of education, indicator variable for being born not in Germany, baseline wage in t=-2, and indicator for part-time employment in t=-2. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.001

Table 1.5: Regression results for differences-in-regression-discontinuity design

Full sample

	Raw wage			Imputed wage		
	t=3	t=13	t=24	t=3	t=13	t=24
Coeff on RxQ1	0.027 (0.052)	-0.038 (0.066)	-0.025 (0.083)	0.033 (0.21)	-0.048 (0.040)	-0.005 (0.057)
N. obs	2736	3511	3981	6851	6851	6851

Women with low education

	Raw wage			Imputed wage		
	t=3	t=13	t=24	t=3	t=13	t=24
Coeff on RxQ1	-0.106 (0.153)	-0.235 (0.210)	-0.012 (0.286)	0.071 (0.070)	-0.240 (0.137)	-0.199 (0.183)
N. obs	343	350	399	837	837	837

Women with high education

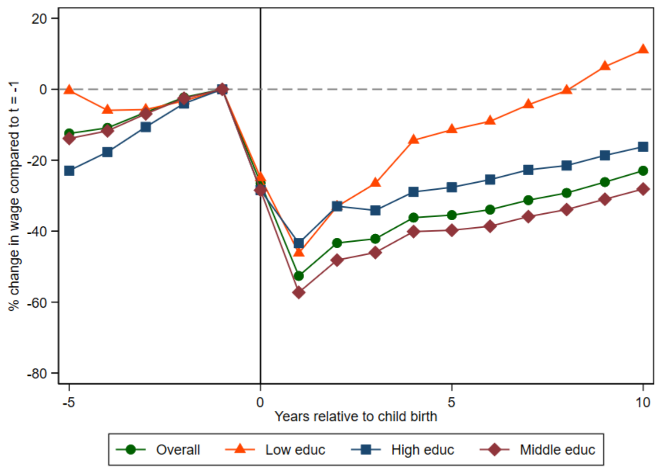
	Raw wage			Imputed wage		
	t=3	t=13	t=24	t=3	t=13	t=24
Coeff on RxQ1	0.078 (0.125)	0.015 (0.098)	-0.037 (0.127)	0.034 (0.057)	0.006 (0.081)	-0.004 (0.100)
N. obs	418	560	587	936	936	936

Notes: Each coefficient is from a separate regression. Regression equation is specified in equation (4) of the main text. The sample includes mothers who gave birth between 2003-2007, and are between ages 18-40 at the time of first childbirth. “Low education” refers to the group of mothers who have no vocational training or a university diploma. “High education” refers to the group of mothers with at least a university diploma. When mothers are not working for pay, when they work in a civil servant job or are self-employed, they have no observations in the dataset. To generate imputed wages, all missing wages are set to 0. Robust standard errors are in parentheses. *p<0.1, **p<0.05, ***p<0.001

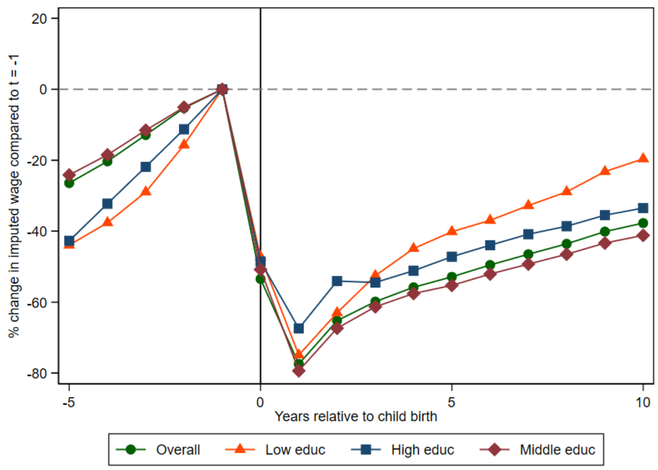
Chapter 1 – Figures

Figure 1.1: Maternal wage dynamics around childbirth

Panel A: Raw wages



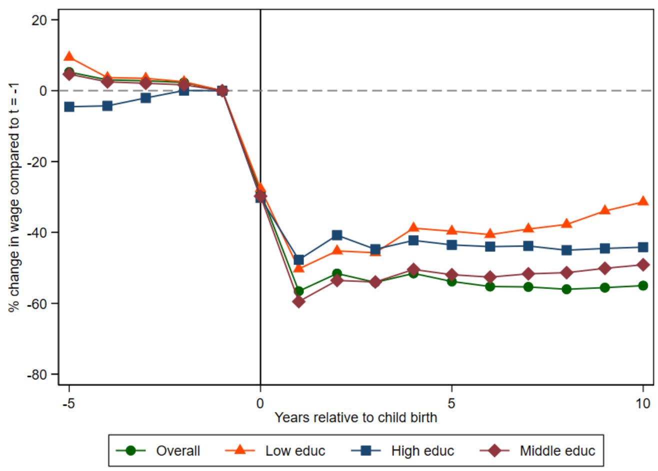
Panel B: Imputed wages



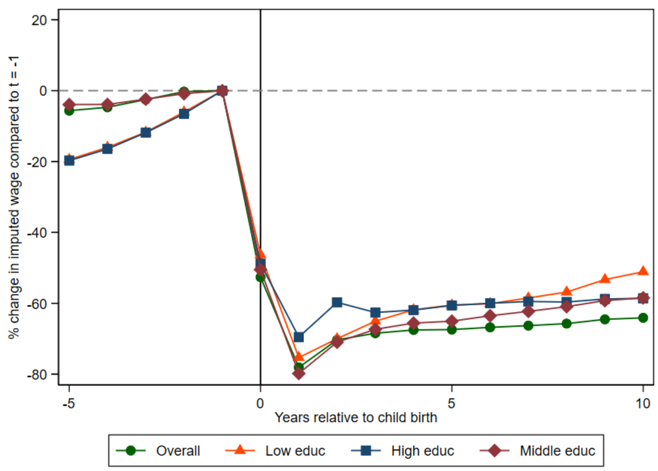
Notes: Figures show mean monthly wages annually for 5 years before and 10 years after childbirth. The analytic sample includes mothers with a first childbirth between 2003-2007 who were between the ages of 18-40 at childbirth. Wage data is only considered for women who are at least 16. Mean wage in the year right before childbirth ($t = -1$) is set as the baseline and all other periods are calculated as a percentage of this baseline wage. Wage data come from the German Sample of Integrated Labour Market Biographies (SIAB) (Berge et al., 2021). "Low education" refers to the group of mothers who have no vocational training or a university diploma. "Middle education" refers to the group of mothers who have a completed vocational training, but no university diploma. "High education" refers to the group of mothers with at least a university diploma. When mothers are not working for pay, when they work in a civil servant job or are self-employed, they have no observations in the dataset. Panel B shows wage dynamics with imputed wages, where all missing wages are set to 0.

Figure 1.2: Maternal wage dynamics around childbirth without age effects

Panel A: Raw wages



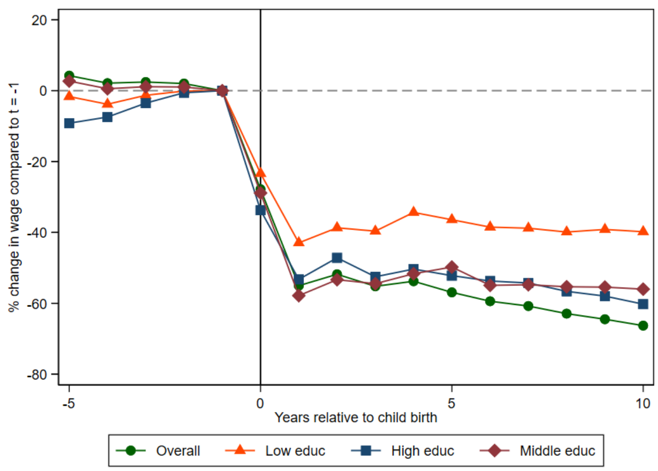
Panel B: Imputed wages



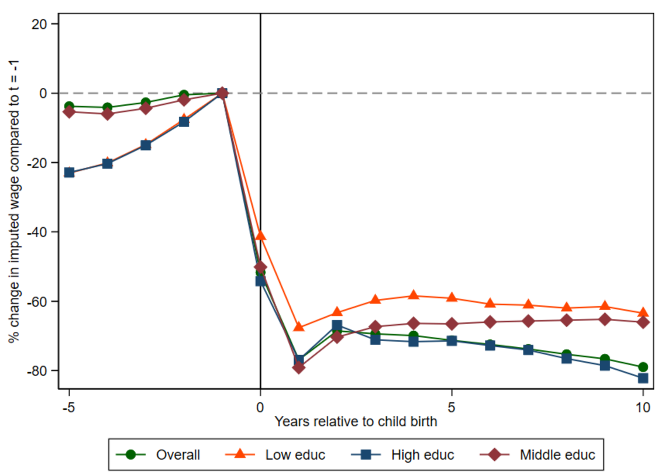
Notes: Figures show coefficients on the event time dummies from regressions of wages on event time dummies and age dummies. The graph can be interpreted as showing mean monthly wages annually for 5 years before and 10 years after childbirth keeping age constant. The analytic sample includes mothers with a first childbirth between 2003-2007 who were between the ages of 18-40 at childbirth. Wage data is only considered for women who are at least 16. Mean wage in the year right before childbirth ($t = -1$) and mean age for each group (29 for the full sample, 25 for the low-education group, 32 for the high-education group) is set as the baseline. All other periods are calculated as a percentage of this baseline wage. Wage data come from the German Sample of Integrated Labour Market Biographies (SIAB) (Berge et al., 2021). "Low education" refers to the group of mothers who have no vocational training or a university diploma. "Middle education" refers to the group of mothers who have a completed vocational training, but no university diploma. "High education" refers to the group of mothers with at least a university diploma. When mothers are not working for pay, when they work in a civil servant job or are self-employed, they have no observations in the dataset. Panel B shows wage dynamics with imputed wages, where all missing wages are set to 0.

Figure 1.3: Maternal wage dynamics around childbirth without age and year effects

Panel A: Raw wages

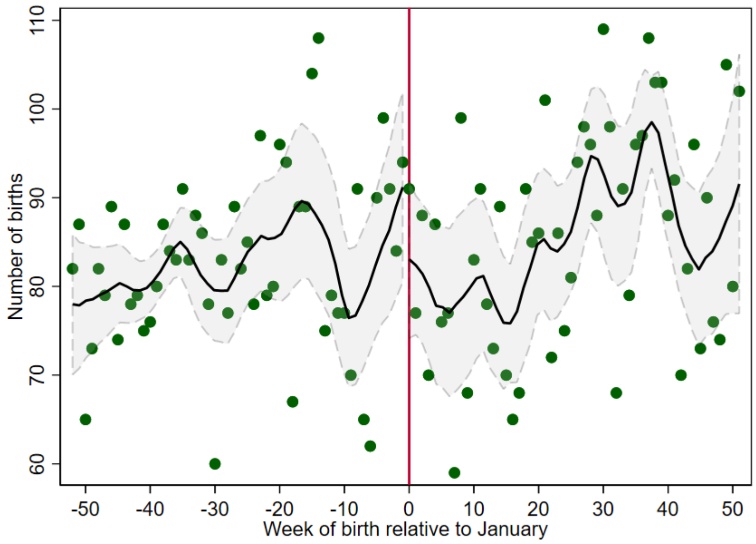


Panel B: Imputed wages



Notes: Figures show coefficients on the event time dummies from regressions of wages on event time dummies, age dummies, and year dummies. The graph can be interpreted as showing mean monthly wages annually for 5 years before and 10 years after childbirth keeping age and economic conditions constant. The analytic sample includes mothers with a first childbirth between 2003-2007 who were between the ages of 18-40 at childbirth. Wage data is only considered for women who are at least 16. Mean wage in the year right before childbirth ($t = -1$) , mean age for each group (29 for the full sample, 25 for the low-education group, 32 for the high-education group), and the mean year (2008) is set as the baseline. All other periods are calculated as a percentage of this baseline wage. Wage data come from the German Sample of Integrated Labour Market Biographies (SIAB) (Berge et al., 2021). "Low education" refers to the group of mothers who have no vocational training or a university diploma. "Middle education" refers to the group of mothers who have a completed vocational training, but no university diploma. "High education" refers to the group of mothers with at least a university diploma. When mothers are not working for pay, when they work in a civil servant job or are self-employed, they have no observations in the dataset. Panel B shows wage dynamics with imputed wages, where all missing wages are set to 0.

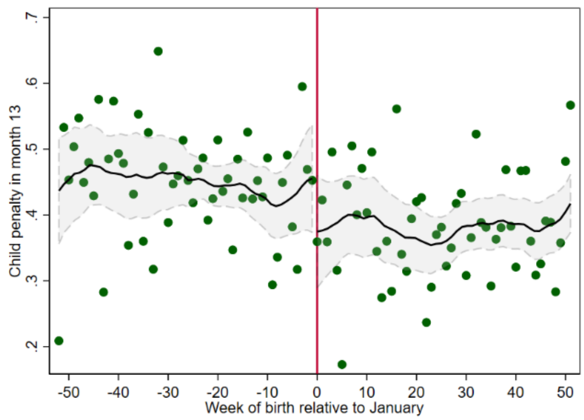
Figure 1.4: Smoothness of birth around policy introduction



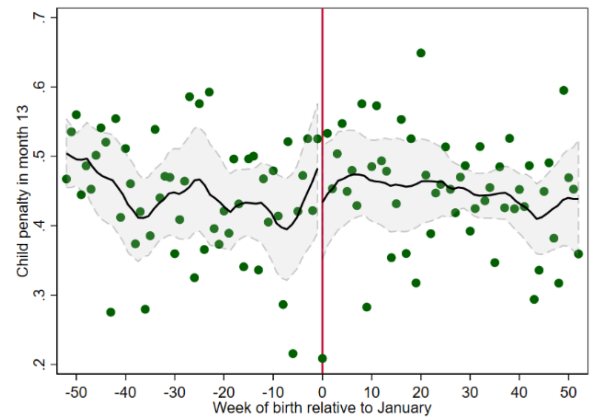
Notes: Graph shows the number of births by week relative to January 1, 2007, which is the date of the policy introduction analyzed in the paper. Data come from the German Sample of Integrated Labour Market Biographies (SIAB) (Berge et al., 2021). Birth date of children is inferred based on the date when mothers go on maternity leave, which is registered in the SIAB. Maternity leave starts 6 weeks prior to one's due date. Local mean-smoothing with 95% confidence intervals is also displayed on the graphs.

Figure 1.5: Wage loss after childbirth by week of childbirth

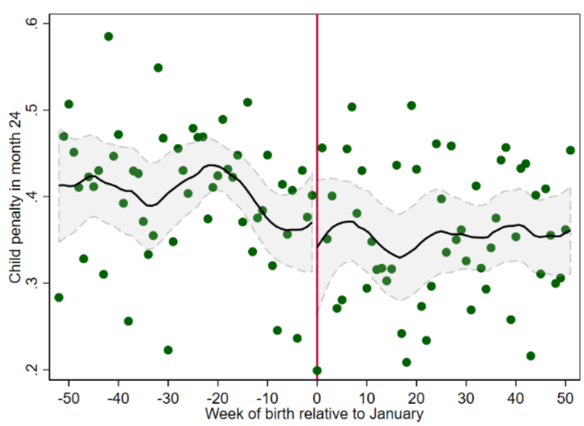
Panel A: t = 13, year 2007 (reform)



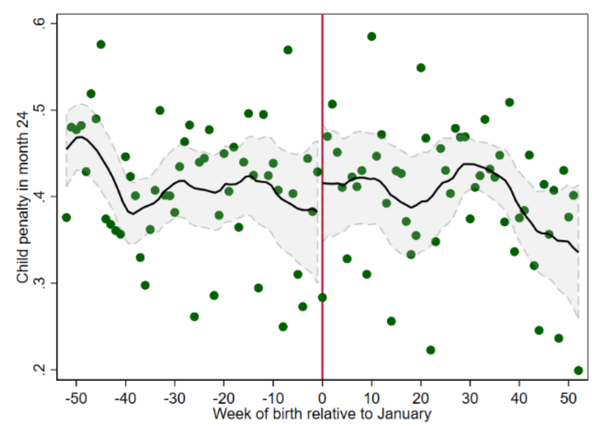
Panel B: t = 13, year 2004-2006 (non-reform)



Panel C: t = 24, year 2007 (reform)



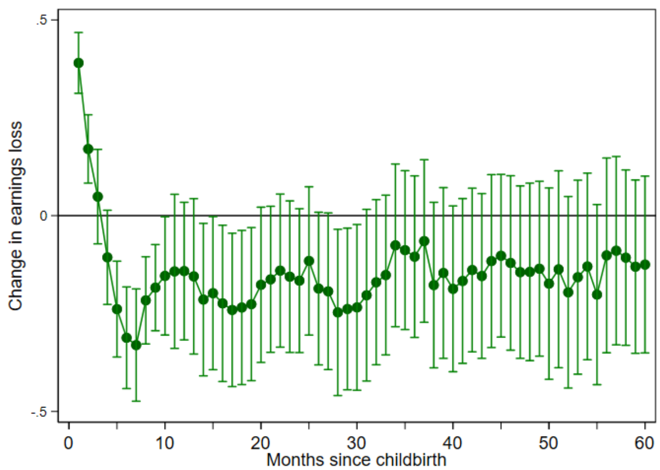
Panel D: t = 24, year 2004-2006 (non-reform)



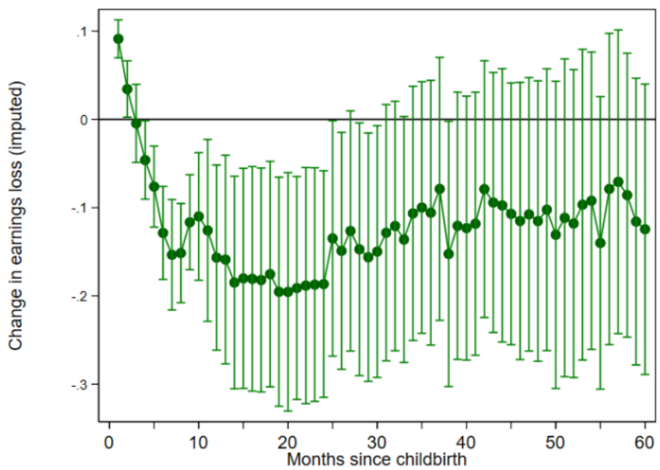
Notes: Figures show mean wage loss after childbirth (the “child penalty”) compared to wages 2 months before childbirth, by week of birth. Panels A and B show the wage loss variable 13 months after childbirth, and panels C and D show wage loss 24 months after childbirth. Panels A and C use 2007, the reform year, and panels B and D use 2004-2006, the non-reform years pooled. The vertical red line indicates January 1st. Local mean-smoothing with 95% confidence intervals is also displayed on the graphs.

Figure 1.6: The effect of the paid leave policy change on maternal wage loss

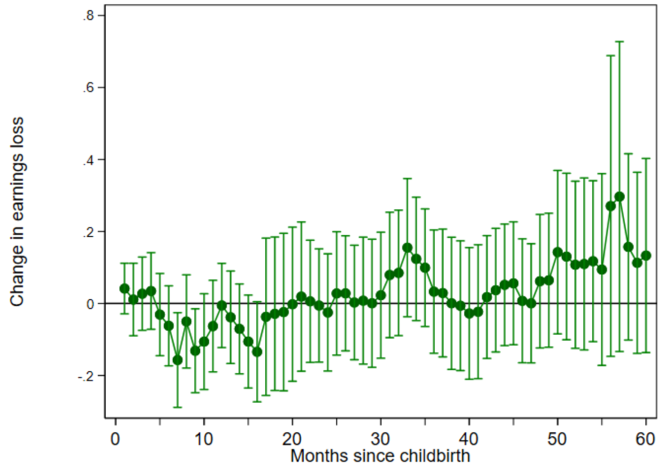
Panel A: Raw wages, 12-month bandwidth



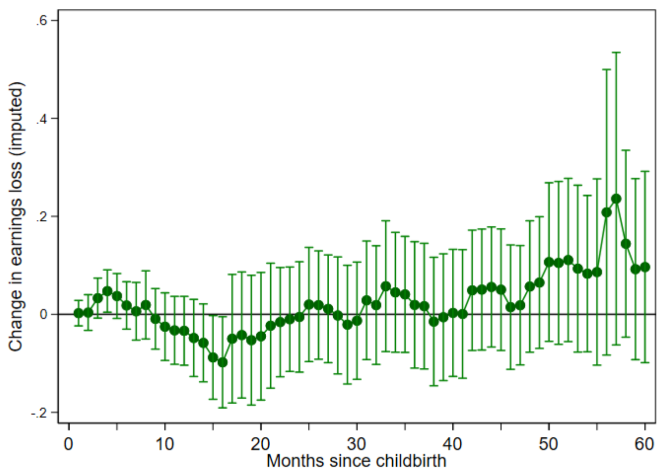
Panel B: Imputed wages, 12-month bandwidth



Panel C: Raw wages, 3-month bandwidth



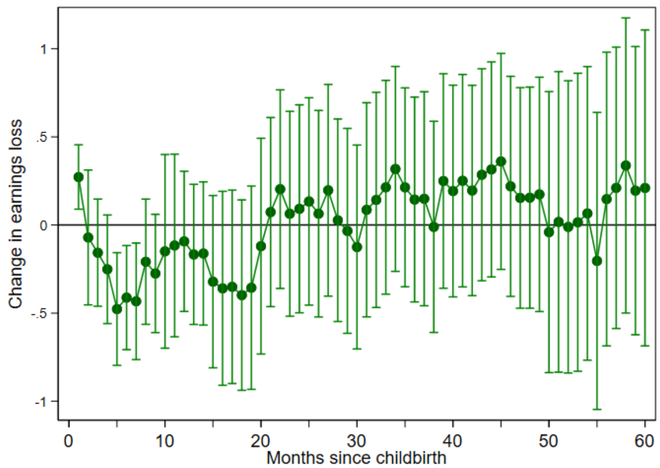
Panel D: Imputed wages, 3-month bandwidth



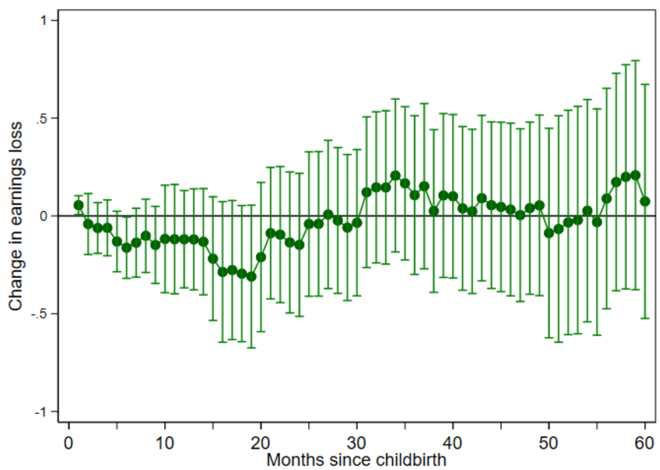
Notes: Figures show estimated coefficients from regressions in the fully specified form of equations (1) (Panel A and B) and equation (5) (Panel C and D). Point estimates and 95% confidence intervals come from separate regressions where the outcome is wage loss t months after childbirth (t = 1,2,..,60) compared to pre-birth wages. The coefficients should be interpreted as percentage point differences in wage loss as a result of the policy. Coefficients below 0 mean that wage loss became smaller, so the drop in wages between pre-child and post-child periods is smaller after the policy change. Panels A and B use sample of mthers who gave birth in 2006 and 2007. Panels C and D use the full sample of mothers who gave birth in 2003Q4, 2004Q1, 2004Q4, 2005Q1, 2005Q4, 2006Q1, 2006Q4, and 2007Q1. Panels A and C show raw wages as they appear in the dataset. Panels B and D use imputed wages where all missing observations are set to 0.

Figure 1.7: The effect of the paid leave policy change on maternal wage loss: mothers with low education

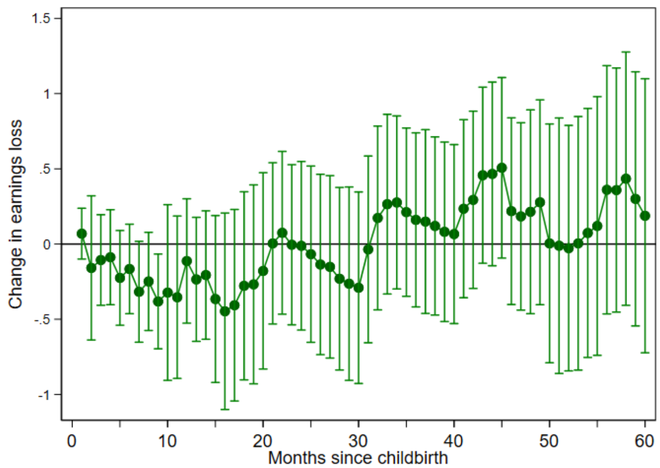
Panel A: Raw wages, 12-month bandwidth



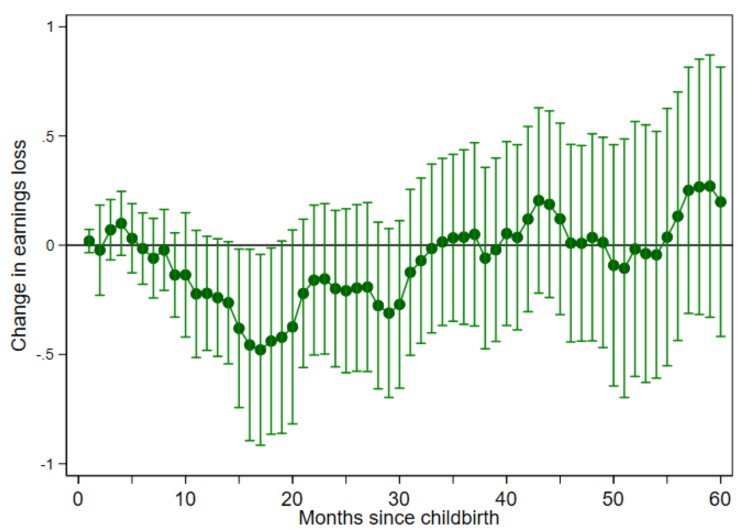
Panel B: Imputed wages, 12-month bandwidth



Panel C: Raw wages, 3-month bandwidth



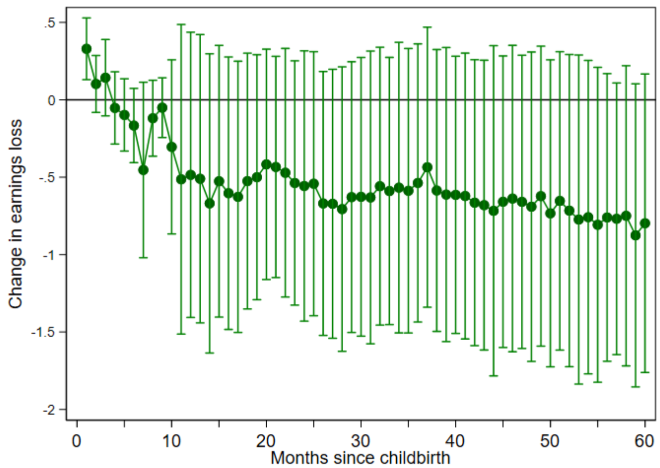
Panel D: Imputed wages, 3-month bandwidth



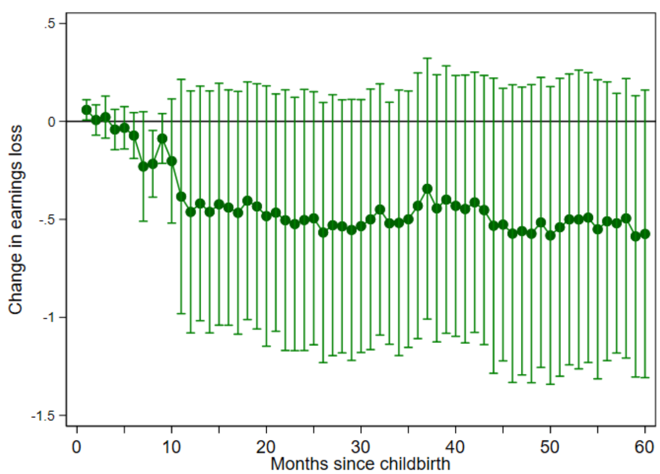
Notes: Figures show estimated coefficients from regressions in the fully specified form of equations (1) (Panel A and B) and equation (5) (Panel C and D). Point estimates and 95% confidence intervals come from separate regressions where the outcome is wage loss t months after childbirth (t = 1,2,..,60) compared to pre-birth wages. The coefficients should be interpreted as percentage point differences in wage loss as a result of the policy. Coefficients below 0 mean that wage loss became smaller, so the drop in wages between pre-child and post-child periods is smaller after the policy change. The graphs show estimates for mothers who did not have any vocational training or university degree at time of childbirth. Panels A and B use sample of mthers who gave birth in 2006 and 2007. Panels C and D use the full sample of mothers who gave birth in 2003Q4, 2004Q1, 2004Q4, 2005Q1, 2005Q4, 2006Q1, 2006Q4, and 2007Q1. Panels A and C show raw wages as they appear in the dataset. Panels B and D use imputed wages where all missing observations are set to 0.

Figure 1.8: The effect of the paid leave policy change on maternal wage loss: mothers with high education

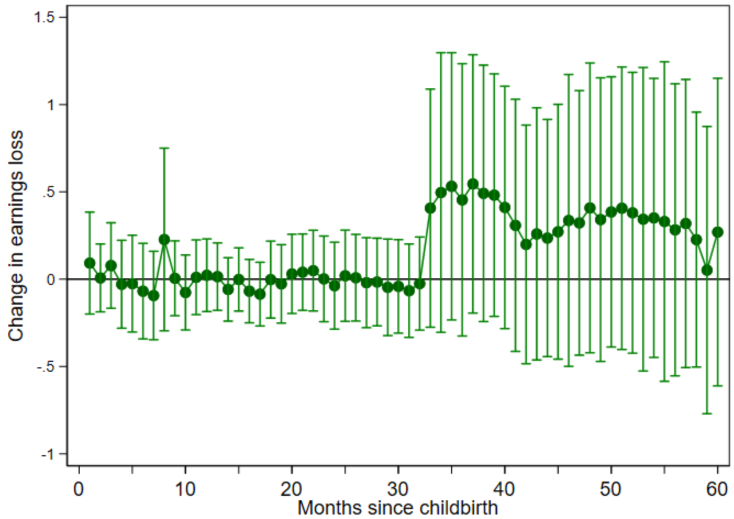
Panel A: Raw wages, 12-month bandwidth



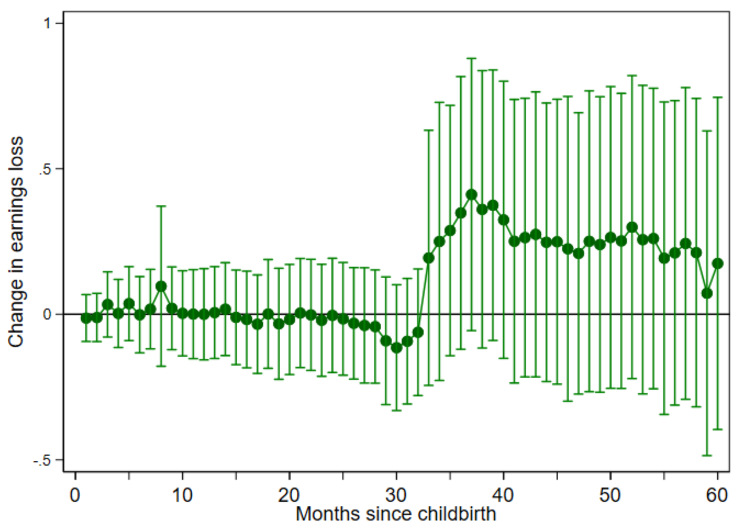
Panel B: Imputed wages, 12-month bandwidth



Panel C: Raw wages, 3-month bandwidth



Panel D: Imputed wages, 3-month bandwidth



Notes: Figures show estimated coefficients from regressions in the fully specified form of equations (1) (Panel A and B) and equation (5) (Panel C and D). Point estimates and 95% confidence intervals come from separate regressions where the outcome is wage loss t months after childbirth (t = 1,2,..,60) compared to pre-birth wages. The coefficients should be interpreted as percentage point differences in wage loss as a result of the policy. Coefficients below 0 mean that wage loss became smaller, so the drop in wages between pre-child and post-child periods is smaller after the policy change. The graphs show estimates for mothers who had at least a university degree at time of childbirth. Panels A and B use sample of mthers who gave birth in 2006 and 2007. Panels C and D use the full sample of mothers who gave birth in 2003Q4, 2004Q1, 2004Q4, 2005Q1, 2005Q4, 2006Q1, 2006Q4, and 2007Q1. Panels A and C show raw wages as they appear in the dataset. Panels B and D use imputed wages where all missing observations are set to 0.

2) Chapter 2 -- Access to child care: Novel insights using distance traveled for care as an indicator

Introduction

High-quality early childhood education (ECE) programs have the potential to provide a conducive environment for developing children's skills and consequently to promote success at each developmental stage (Brooks-Gunn et al., 2016; Chaudry et al., 2017; Phillips et al., 2017). Public money on child care programs can improve academic achievement (Brooks-Gunn et al., 2016; Duncan & Magnuson, 2013; Heckman, 2006), but also contribute to children's development of socio-emotional skills like executive function, self-regulation, and relationship with others (Sabol et al., 2021; Weiland & Yoshikawa, 2013). The United States spends less as a percentage of GDP on early care for 3-to-5-year-olds than most other member countries of the Organization of Economic Cooperation and Development (OECD) and also has lower-than-OECD-average enrollment rates in ECE (OECD, 2019). Furthermore, inequalities in ECE access within the U.S. have been widely documented (e.g. Chaudry et al., 2017; Magnuson & Waldfogel, 2016). These have prompted policy makers to put improving access to ECE for all families on the policy agenda in recent years at the local, state, and federal levels. A comprehensive picture of ECE access is vital to understand where the highest need for child care is and what the attributes of child care programs are that policymakers should focus on.

Access to child care is a multi-dimensional concept including the physical availability, cost, and quality of child care programs, and how well they serve families' needs (Friese et al., 2017). Common indicators to measure ECE access are physical availability, i.e. the number of programs in a given geographic area (Cochi Ficano, 2006; Gordon & Chase-Lansdale, 2001; Malik et al., 2018), and enrollment in ECE (Bainbridge et al., 2005; Magnuson & Waldfogel, 2016). These indicators show that lower-income families and those living in rural areas have fewer programs available for them and they are also less likely to use center-based care. These indicators focus on specific attributes of access and point towards the same conclusion: low-income areas are the most in need of public investments. Distance-based indicators on the other hand suggest that in some areas higher-income families travel more for child care and urban areas have lower access (Davis et al., 2019; National Survey of Early Care and Education Project Team, 2016a).

The current paper contributes to our understanding of ECE access by income. I use a novel indicator to examine ECE access: distance traveled for child care. By using fine-grained geographic mobility data, I document trends in how far families travel to access care in the United States and in the state of Illinois. I also examine differences by income across the whole income distribution. This new indicator provides a data-driven method to assess the size of the relevant child care market and allows additional insights into current patterns of care. Furthermore, I revisit the physical availability indicator by neighborhood income in Illinois. Given investments in public pre-kindergarten and early care programs over the last two decades (Friedman-Krauss et al., 2020), reexamining this indicator is useful to assess where future investments in early care should be targeted to.

Literature review

Indicators of ECE access

Several indicators have been used to measure ECE access. Enrollment in ECE is one of the most popular ones. Lower-income children under the age of 5 are less likely to be enrolled in center-based care than higher-income children (Laughlin, 2013). Magnuson & Waldfogel (2016) show that the difference in preschool enrollment for 3- and 4-year-olds between the top and bottom income quintiles are persistent. Enrollment was around 30% for children from the lowest income categories, while around 50% of children from the highest income families were enrolled in child care in the 1970s. In the 21^{st} century, there is still a 20 percentage point difference between the top and bottom quintiles, with 60% of 4-year-olds enrolled in pre-school in the lowest income category versus 80% in the highest (Magnuson & Waldfogel, 2016).

Other commonly used indicators of access are related to the physical availability of child care programs. This is most often measured by the number of programs in a given geographic area, the licensed capacity in child care programs, or the number of licensed child care workers (Friese et al., 2017; Gordon & Chase-Lansdale, 2001). One specific indicator in this category is the child care desert concept. Researchers at the Center for American Progress define a child care desert as a census tract that has either no licensed child care options, or has a ratio of more than three young children for every licensed child care slot (Malik & Hamm, 2017). According to this measure, 51% of Americans live in neighborhoods that can be characterized as child care deserts. Families living in rural areas and Hispanic/Latino families are the most likely to live in a child care desert (Malik et al., 2018). Supply indicators relying on the number of child care workers in a particular area similarly show that rural and mixed-income communities have the lowest availability of child care (Cochi Ficano, 2006; Gordon & Chase-Lansdale, 2001).

These widely used access indicators paint a bleak picture about availability of and enrollment in ECE for low-income and rural families. Research on ECE access is not conclusive, though. Other papers show slight differences in access in the opposite direction by income and urbanicity. Using the number of children for each licensed slot as another indicator Davis et al. (2019) find that families in Minnesota have access to about half a slot on average. Families in large urban areas of the state have somewhat lower access to ECE programs than families who live in more rural and smaller urban areas. Their analyses also indicate small but statistically significant differences in access to slots by income: the lowest-income families have somewhat higher supply of child care than middle- and high-income families (Davis et al., 2019). Nationally representative survey data shows that lower-income families travel shorter distances to access child care for 3-5-year-old children than higher-income families (2.7 miles vs. 4.6 miles, NSECE Project Team, 2016a). Papers use different indicators and different geographic areas over which they compute those indicators, which likely matters for conclusions about access.

Defining the child care market

When parents decide on which (if any) child care program to choose, they have to take into account both the physical availability of care, as well as their own preferences for child care programs. Many families may not look for non-parental care regardless of the options available for them. The ones who do may prioritize different factors and make decisions based on a combination of location, quality, cost, and other characteristics like hours of operations or language spoken by caregivers (Carlin et al., 2019; Herbst et al., 2020; Johnson et al., 2017). I conceptualize the geographic area where the programs that families consider are located as the child care market. Supply and demand jointly determine how families make choices given the market they face. One relevant question for measuring access is the size of this market, as many indicators of access like the number of programs or the number of licensed slots are often computed over a pre-defined geographic area (Friese et al., 2017). Existing studies determined the size of the market using area-based and distance-based measures.

Area-based measures often take pre-defined administrative areas like states (E. U. Cascio & Schanzenbach, 2013), counties (Bassok et al., 2016), or census tracts (Malik et al., 2018), and use these as approximations of the market. As another type of area-based measure, Brown (2019) creates a grid of tessellated hexagons for New York City, and analyzes how child care supply changes within one such hexagon. Gordon & Chase-Lansdale (2001) use two levels of geography, zip codes and counties, as well as a 30-mile radius around a zip code. They posit that counties capture the market best. Driving distance is another popular method to define markets, Davis et al (2019) and Malik et al (2020) both consider a 20-minute drive. These area-based measures of the child care market are not based on where parents look for care, but on pre-defined geographic areas that seem to be relevant for families or for which data are available.

The National Survey of Early Care and Education (NSECE) Project Team (2016b) uses a data-driven method to get at the size of the market relevant for families. They rely on data from the 2012 NSECE survey and assess how far child care programs that children attend are located from families' residences. Based on these data they define a two-mile radius around families' census tracts as the market.

Current study

In the current study, I analyze how far families travel to access child care and look at differences by income. I follow a similar data-driven method to that of the NSECE, but revise their estimates using a larger sample and a more fine-grained dataset. The study contributes to the literature on several fronts. First, it defines a new indicator of access. Then, I use this new indicator to define the market that is relevant for families. Existing research often fails to explain why they chose the pre-specified geographic area in their measures. I provide evidence that the size of the pre-specified geographic area matters for conclusions about access. I also compare the size of the market I compute using the data-driven approach to two pre-specified geographic areas: Census Block Groups and Zip Code Tabulation Areas. It is important to understand which of these geographic areas approximates the market best as demographic indicators from nationally representative surveys are only available for these areas. Lastly, I use this data-driven approach of the relevant market size to recalculate existing indicators of ECE supply, which gives novel insights into ECE access.

Method

Data

I used SafeGraph's Geographic Patterns data for September 2019 (SafeGraph, 2023). I chose this specific year and month because it provided a snapshot of the child care landscape in the beginning of the most recent school year not affected by the COVID-19 pandemic. SafeGraph provides geographic mobility data from about 45 million smartphones across the United States. The company partners with smartphone application providers who obtain opt-in consent from their users, and then collect anonymous latitude and longitude information from the devices. SafeGraph then aggregates this location information and connects it to points of interest (POI), which are businesses or other establishments that people visit during a particular month, including child care programs. They work with a panel of mobile devices, so the same mobile devices are tracked over time, until people opt out of using the mobile applications that are collecting their location data. SafeGraph assigns business names to the longitude/latitude information. They also provide North American Industry Classification System (NAICS) codes for each POI. They use the business names as the indications of a category, and they also use other open-source information and machine learning techniques to assign the correct NAICS code to a business (SafeGraph, n.d.). Code 624410 is associated with Child Day Care Services, which I used to identify child care programs in the dataset. One observation in the dataset is one child care program.

Sample of devices. SafeGraph did not provide individual-level information like name or address, they only provided the number of devices that were tracked in a Census Block Group (CBG). Based on the number of tracked devices, I described the geographic representativeness of the sample. In September 2019, 32,466,856 devices were tracked across the United States, and 1,340,121 devices were tracked in the state of Illinois. An estimated 322,004,832 people lived in the U.S. during this month, and 12,851,612 people lived in Illinois. If the sampling were geographically representative, the ratio of the devices tracked in Illinois over the devices tracked in all of the U.S. should be very similar to the ratio of all people in Illinois over all people in the U.S. The first ratio was 0.0413, the second was 0.0399. Geographic representativeness was less precise with levels of geography below the state-level and no claims could be made about representativeness with regards to other demographic characteristics.

Sample of child care programs. My analytic sample contained child care programs that people with tracked devices visited. In the month of September 2019 there were n = 106,916 child care programs in the U.S. sample, and there were n = 2,131 child care programs in the state of Illinois.

In order to characterize the representativeness of these child care programs, the universe of all child care programs would be necessary. The universe of child care programs for the U.S. was not readily available, thus as a case study I examined child care programs in the state of Illinois, as well. Illinois is a state with a relatively large population, it provided a list of all licensed child care programs, and had an active child care policy environment. I downloaded the registry of all licensed child care programs from the website of the Illinois Department of Child and Family Services for December 2020. There were 9,663 licensed child care programs in the state of Illinois, of which 2,930 (30%) were child care centers and 6,733 (70%) were family day care homes. Child care centers had an average capacity of 81 slots, family day care homes had an average capacity of 1.6 slots. 85% of centers provided care for children under the age of 3, 93% of centers provided care for children between the ages of 3 and 5, and 86% of centers provided care for children above age 5.

My analytic sample of child care programs in Illinois (n = 2,131) contained 75% of licensed child care centers and virtually no other child care programs. Child care centers in my analytic sample had higher capacity than child care centers not in the sample. The centers in sample had 84.8 slots on average compared to 72.8 slots in centers not in the sample (t-stat = -5.31). Child care centers in and out of the analytic sample were equally likely to provide care for children under the age of 3 (85% of centers provided such care). More centers provided care for 3- to 5-year-olds in the sample (94%) than centers not in the analytic sample (89%, t-stat = -4.51). Centers in the analytic sample were also more likely to provide care for children above age 5 (87% of centers in the sample versus 81% of centers not in the sample, t-state = -3.97).

Measures

The outcome of interest was the median distance visitors traveled for child care. If 100 people visited the child care program in a month, the median distance these 100 visitors traveled between their homes and the child care program was reported as an aggregate measure at the child care program level. It is important to note that while the majority of the visitors were most likely parents bringing their children to child care, it was not possible to distinguish between teachers, staff, cleaners, or other suppliers who visited to the establishment for reasons other than to access child care.

To proxy for visitors' demographic information like income or urbanicity, I used demographics for the geographic area where they resided. These data were downloaded from the

National Historical Geographic Information System (Manson et al., 2021) and were 5-year estimates from the American Community Survey for the years 2015-2019 aggregated to the level of the geographic area. Two levels of pre-specified geographic areas were used in my analyses: Census Block Groups (CBG) and Zip Code Tabulation Areas (ZCTA).

To measure supply or physical availability of child care in the state of Illinois, I constructed two indicators: (1) the number of licensed child care programs in a pre-specified geographic area, and (2) licensed capacity in a given pre-specified geographic area, which was the number of licensed child care slots per child under the age of 5. More details about these measures are available in the Appendix.

Analytic strategy

The main analytic approach was descriptive statistical analysis that compared means and medians of access indicators. I used Ordinary Least Squares regressions to estimate the correlation coefficients. I conducted analyses of distance traveled by income both for the U.S. sample and for the Illinois sample. In sensitivity tests by urbanicity and in the calculation of supply measures I only included child care programs from Illinois as these analyses required a complete set of zip codes and the universe of child care programs, which were not available for the U.S. All statistical analyses were executed in Stata (version 17).

Results

Distance traveled to access child care

During September 2019, visitors traveled on average 5.26 miles (st.dev = 30.90) for child care programs in the United States (n = 101,758 with non-missing distance information). 50% of child care programs received visitors who traveled fewer than 3.70 miles. The average distance traveled to child care in Illinois was 4.08 miles (st.dev = 4.37), and the median was 3.24 miles (n = 1,810 with non-missing distance information). There were 8 programs in Illinois that received visitors from more than 20 miles away. These observations were excluded from the analyses. Appendix Figure 1 shows histograms for distance traveled.

Distance traveled by income

The relationship between distance traveled and income followed an inverted U-shaped pattern. In the U.S. sample at the left end of the income distribution, visitors who lived in the lowest-income Census Block Groups (CBGs) traveled on average 4 miles to access child care. Similarly, at the right end of the distribution, visitors who lived in the highest-income CBGs traveled around 4 miles for child care. As income increased distance traveled increased, as well, up to the 70th percentile of the distribution, peaking at about 4.8 miles. Figure 1, Panel A illustrates this relationship. The figure shows binned scatter plots grouping income into deciles and plotting the mean distance traveled in each decile.

The pattern between distance traveled and income followed a similar inverted U-shape for the Illinois sample. Distance traveled increased with income up to the 90th income percentile. However, people who visited from the highest-income areas traveled about as much as people who lived in areas below the 40th percentile of the income distribution. Figure 1, Panel B illustrates this relationship. To check whether the relationship was only linearly increasing or approximated the inverted U-shape pattern of the sample for the U.S. more closely, I also plotted income and distance traveled using 15 and 20 bins. As I increased the number of bins, the right end of the income distribution increasingly showed the downward trend in distance traveled (see Appendix Figure 2).

Regression results supported the graphical evidence of the relationship between distance traveled and income (see regression coefficients in Table 1). The coefficient on median income in the linear Ordinary Least Squares regression on distance traveled was not significant in the sample including all U.S. child care programs (column 1). In regressions including both linear and quadratic terms for median income (column 2), both coefficients were statistically significantly different from zero (with $p < 0.001$), which supported the existence of an inverted U-shaped relation between the two variables. For the sample of Illinois child care programs, the coefficient on median income in the univariate regression with only the linear term was significantly different from zero (column 3), which reflected the fact that distance traveled increased with income up to the 90th percentile in Illinois. However, the coefficients on both the linear ($\beta_1 = 0.426$ with $p < 0.001$) and the quadratic term ($\beta_2 = -0.020$ with $p < 0.001$) were statistically significantly different from zero in the multivariate regression (column 4), which supported the existence of the inverted U-shape relation in the Illinois sample, too.

Sensitivity tests by urbanicity

The relationship between income and distance traveled can be confounded by the fact that rural areas are more scattered, so people who live in rural areas are more likely to travel longer distances than people who live in urban areas. Urban areas were more affluent in Illinois, the correlation coefficient between median income and the percentage of the population who lived in urban areas was 0.15. This could lead to biased coefficients in the regressions with only income as the explanatory variable. Controlling for urbanicity in analyses estimating the correlation between distance traveled and income in Illinois did not change the conclusions from the univariate analyses. The coefficients on median income ($\beta_1 = 0.401$ with $p < 0.001$) and its square term ($\beta_2 = -0.018$ with $p < 0.001$) in multivariate OLS regressions including urbanicity as a control variable were similar in magnitude and in significance to the regression coefficients where urbanicity was not included as a control variable (see Table 1, column 5).

Physical availability of child care

I calculated physical availability indicators for the state of Illinois using two levels of pre-specific geographic areas. Census block groups (CBGs) are the second smallest level of geography defined by the Census Bureau. They cover areas with 600-3000 inhabitants. Illinois had 9,691 CBGs. The average size of a CBG was 5.8 square miles, but the distribution had a long right tail. The median CBG was 0.31 square miles, and the CBG at the 75th percentile of the distribution was 1.4 square miles. CBGs in urban areas were much smaller than in rural areas, the mean CBG size was 0.6 square miles in urban areas and 11 in rural areas. Zip Code Tabulation Areas are zip codes created by the Census Bureau to make tractable, aerial representation of postal zip codes. ZCTAs are two levels bigger than CBGs, census tracts are in between. There were 1,383 ZCTAs in Illinois. ZCTAs were on average 40.5 square miles. 25% of ZCTAs were below 7.7 square miles, and 25% were larger than 58 square miles. Urban ZCTAs were smaller than rural ones, the mean size was 9.3 square miles in urban areas, and 51 in rural. Chicago ZCTAs were on average 4 square miles.

Of the 9,691 CBGs, 859 (8.9%) had fewer than 10 children under the age of 5. I included in analyses only the 8,832 CBGs that had at least 10 children under the age of 5. A little more than half (53%) of these CBGs had at least one child care program. There were 1.92 programs (st.dev = 1.42) per CBG on average in CBGs with at least one child care program, and there was on average 1.00 (st.dev = 1.81) licensed child care slot per child. However, the licensed capacity indicator had a long right tail, the median CBG had only 0.375 slots perc child, or about 2.7

children for each slot. 43% of children under the age of 5 lived in CBGs with no child care programs.

When working with ZCTAs, I included ZCTAs with at least 50 children under age 5 in my analyses. Of the 1,383 ZCTAs in Illinois 450 (32.5%) had fewer than 50 children under age 5. 82.6% of the 933 ZCTAs with at least 50 children under age 5 had at least one child care program. There were 12.32 programs (st. dev = 19.41) on average in these ZCTAs, and on average 0.39 (st. dev = 0.33) licensed slots per child. The median ZCTA had 0.32 slots per child, or about 3.1 children for each slot. Only 1.2% of children under the age of 5 lived in ZCTAs without any child care program.

There was a negative correlation between the number of child care programs and median household income in the geographic area both for CBGs ($\rho = -0.095$) and for ZCTAs ($\rho = -0.037$). CBGs with the lowest household income had on average 1.2 child care programs, CBGs in the middle of the income distribution had about 1.0 child care program, and CBGs with the highest median income had about 0.8 child care programs on average. This relationship was driven by family day care programs: there were 0.8 programs in the lowest income CBGs and the number of family day cares decreased gradually until the highest income decile where there were fewer than 0.25 programs. There were between 0.25 and 0.5 centers per CBG across the income distribution. The lowest and the highest income deciles had somewhat more centers than the middle of the income distribution. Figure 2, panels A and B illustrate this relationship with binned scatter plots for number of programs by type and by income at the CBG level.

Patterns at the ZCTA level were more pronounced than they were at the CBG level. The two lowest-income deciles had the most programs, about 25 and 13 on average. ZCTAs in the middle of the income distribution had fewer programs, between 5-10. The 3 highest-income deciles had on average 10 programs. This relationship was driven by family child care programs. The two lowest-income deciles had on average 18 and 9 family child care programs, and the rest of the income deciles had about 5. The lowest income decile had the fewest, about 3. The number of child care centers followed a U-shaped pattern. ZCTAs both at the lowest and at the highest end of the income distribution had more than 5 child care centers, and ZCTAs at the middle of the income distribution had between 1-3. Figure 2, Panels C and D show binned scatter plots for the number of programs by type and by income at the ZCTA level.

Table 2 shows regression results for the association between the number of programs and median income both at the CBG level (Panel A) and at the ZCTA level (Panel B). As median income increased the number of programs decreased by about 0.02 for each $10,000 at the CBG level and by 4 at the ZCTA level. A quadratic functional form approximates the relationship better than a linear one. Adding urbanicity and the number of children under 5 as control variables did not change the results. Appendix Table 1 shows similar regressions results for the number of centers and the number of family day care homes as outcome variables.

The number of slots per children under 5 showed a U-shaped pattern both at the CBG and at the ZCTA level. In the lowest-income CBGs there were on average 0.7 slots per child under age 5. The middle of the income distribution had fewer slots, about 0.5 on average. CBGs at the 60th and 70th income percentile had the fewest slots, around 0.4 per child. The highest-income decile CBGs had about 0.6 slots. The lowest-income ZCTAs had about 0.4 slots per child. ZCTAs in the middle of the income distribution had between 0.2-0.3 slots per child. ZCTAs in the highest-income decile had about 0.5 slots. Appendix Figure 6 shows the binned scatter plots that depict these relationships.

For CBGs, the univariate regression coefficient of capacity on the linear term on income was not significant. Including both the linear and quadratic term for income in regressions approximates the U-shaped relationship apparent on the graphs. The coefficient on the linear term was negative (β_1 = -0.04 with $p < 0.01$) and the coefficient on the quadratic term was positive (β_2 = 0.002 with $p < 0.01$) in multivariate regressions. Both terms remained significantly different from zero in regressions including urbanicity as a control variable. Regressions at the ZCTA level were qualitatively similar. Appendix Table 2 shows regressions results for both sets of regressions.

Discussion

In this paper I analyzed access to child care using a novel indicator: distance traveled for care based on geographic mobility data. I found that families traveled on average 5.26 miles across the United States, and 4 miles in the state of Illinois. The medians were 3.7 and 3.2 miles, respectively. Distance traveled by income followed an inverted U-shaped pattern: families residing in lower-income areas traveled less than families in higher-income areas, but families in the highest income areas traveled similar distances to the lowest-income families. This pattern was not explained by urbanicity. These findings were consistent with earlier distance traveled estimates using the National Survey of Early Care and Education. The NSECE Project Team (2016a) showed a gradient by income: distance traveled increased across income categories. My analyses painted a more nuanced picture. While distance traveled by income increased up to about the 70th percentile of the income distribution, people who resided in the highest-income areas traveled shorter distances.

The estimates of distance traveled can inform our understanding of the relevant child care market families face. It is common in the literature to approximate the market using pre-defined geographic areas, like states (E. U. Cascio & Schanzenbach, 2013), counties (Bassok et al., 2016;

Gordon & Chase-Lansdale, 2001), or census tracts (Malik et al., 2018). Using pre-defined geographic areas might not accurately depict families' child care options as parents can easily cross boundaries of these areas to access care (Davis et al., 2019). These administratively defined areas have advantages, though. Most importantly, many demographic variables from nationally representative surveys are only available at these pre-defined geographic levels. My estimates of distances of 3-5 miles of travel indicate that geographic areas below Zip Code Tabulation Areas (ZCTA) such as census tracts or census block groups are too small to approximate well the child care options that families likely consider.

Supply indicators of access like the number of child care programs or the licensed capacity per pre-school aged child are most often calculated over one of the pre-defined geographic levels (Friese et al., 2017). These indicators are sensitive to the level of geographic area they are calculated for. For example, Malik et al. (2018) reports that 57.7% of families live in a child care desert in Illinois. They use census tracts as the level of geography. Using ZCTAs as the level of aggregation, which corresponds better to the distances families travel to access care, the same indicator suggests that 44.3% of families live in a child care desert.

To provide a more comprehensive picture of child care supply in Illinois, I analyzed the number of licensed programs and the number of child care slots per under-5 children by income. While some papers find that rural and mixed-income areas have lower availability of child care (Cochi Ficano, 2006; Gordon & Chase-Lansdale, 2001; Malik et al., 2018), others show small differences and in the opposite direction: rural and lower-income areas have somewhat more child care options (Davis et al., 2019; NSECE Project Team, 2016b). I described these supply measures across the whole income distribution controlling for urbanicity. I found that the highest- and lowest-income areas had more child care programs and higher capacity than areas in the middle of the income distribution. This finding might be somewhat unexpected given the narrative around disparities in access by income (Magnuson & Waldfogel, 2016; Malik et al., 2020), but can be explained by federal, state, and local efforts to invest more in early childhood education over the last few decades. Publicly funded pre-kindergarten programs have been on the rise and enrollment in these programs have increased from 14% in 2002 to 34% by 2019 (Friedman-Krauss et al., 2020). Many of these pre-k programs were targeted towards families who live at 185% of the federal poverty line or who are otherwise disadvantaged. The Child Care Development Block Grants, which provide federal funding to states to invest in child care for low-income families (Herbst & Tekin, 2010) are another set of policies that could have contributed to the current supply patterns.

Analyzing only supply patterns cannot paint a comprehensive picture about families' child care decisions. Outcome indicators like distance traveled or enrollment in child care are determined by both the supply of programs and families' preferences for child care. While the distances families travel or whether they enroll their children in child care are observable in datasets, these descriptive analyses do not answer the question of why families made those decisions. For example, several papers have documented that lower-income children are less likely to be enrolled in center-based care than higher-income children (Laughlin, 2013; Magnuson & Waldfogel, 2016). One explanation for this pattern could be related to the supply-side: there are not enough programs in lower-income areas. While analyses of indicators of supply showed that programs are available in many settings, it is possible that these are not the types of programs that families are looking for. These programs may have low quality, may be too expensive, may not operate during hours when families would need care, or may have other attributes that make them undesirable (Forry et al., 2013; Henly & Adams, 2018). Using different indicators of access both from the supply and from the demand side would help researchers, practitioners, and policy makers to better understand families' child care realities.

Limitations

This study provided novel insights into ECE access using distance traveled as an indicator and by reevaluating supply across the whole income distribution in the state of Illinois. There are important dimensions of access this paper did not consider. I did not have data about the cost or the quality of the child care programs that families have in their vicinity. While my findings indicated that families in the lowest and highest income categories traveled similar distances and had similar supply of child care available for them, it is very likely that these programs differ in terms of quality and/or cost. A more comprehensive access indicator would have to incorporate these other omitted dimensions, too. I used data on licensed child care programs, and many families, especially marginalized families, tend to use non-licensed child care arrangements (L. Schochet, 2019; Snyder et al., 2005). My analyses could not account for relative-care or other informal arrangements, thus may not provide a comprehensive picture of access.

Descriptive analyses such as the ones in this paper are helpful to understand the current child care landscape that families face and to point towards potential levers for intervention to increase access. However, they do not provide causal evidence that any of those interventions would indeed lead to higher access. Causal research designs would have to rely on exogeneous changes in either the supply- or demand-side and analyze how parental choice changes following those shocks. A supply-side change could be the entry of new child care programs to the market. Analyzing such a shock would require a panel dataset of child care programs, but these datasets are difficult to obtain and to harmonize over time (Bassok et al., 2016; Brown, 2019). A demand-side intervention could be a policy change that increases families' disposable income for child care, such as an increase in child care subsidies. Analyzing such a policy change would require a longitudinal dataset of detailed parental child care choices, but to my knowledge such a dataset does not exist with regular frequency and a large enough sample-size.

Conclusion

Using different indicators of ECE access, distance traveled and supply of child care, I showed that families in the lowest and highest income areas have better access to ECE than families in the middle of the income distribution. These indicators help us better understand certain dimensions of access and bring attention to the fact that next to continued support for the most disadvantaged communities policy efforts should also focus on middle-income families. Yet, a stronger focus on parental preferences within these contexts is needed. Future work should bring together supply- and demand-side factors to make more informed investments into early care and education programs that serve families' needs.

Chapter 2 – Tables

Table 2.1: Regression results of distance traveled on income

	(1) US	(2) US	(3) Illinois	(4) Illinois	(5) Illinois
Median income ('0000 $)	0.0761	0.571***	0.0645**	0.426***	0.401***
	(0.0445)	(0.117)	(0.0213)	(0.0846)	(0.0818)
Median income (squared)		-0.0309***		-0.0203***	-0.0176***
		(0.00607)		(0.00458)	(0.00440)
Urbanicity					-0.0200***
					(0.00335)
Constant	4.759***	3.043***	3.426***	2.021***	3.705***
	(0.232)	(0.425)	(0.169)	(0.358)	(0.466)
Observations	101,657	101,657	1,895	1,895	1,895
R^2	0.000	0.000	0.005	0.014	0.039

Table shows coefficients from Ordinary Least Squares regressions at the child care program level. Robust standard errors are in parentheses. Median income is measured as median household income in visitors' home census block groups. Urbanicity is measured as the percentage of people in visitors' census block groups who live in an urban area. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.2: Number of child care programs and median household income in Illinois

Panel A: Census Block Groups

	(1)	(2)	(3)
Median income ('0000 \$)	-0.0109***	-0.0226***	-0.0213***
	(0.00141)	(0.00460)	(0.00456)
Median income (squared)		0.000591**	0.000427
		(0.000221)	(0.000219)
Urbanicity			0.00134***
			(0.000162)
Under-5 population			0.000813***
			(0.0000643)
Constant	0.604***	0.649***	0.472***
	(0.0114)	(0.0203)	(0.0247)
Observations	8623	8623	8623
R^2	0.007	0.008	0.032

Panel B: Zip Code Tabulation Areas

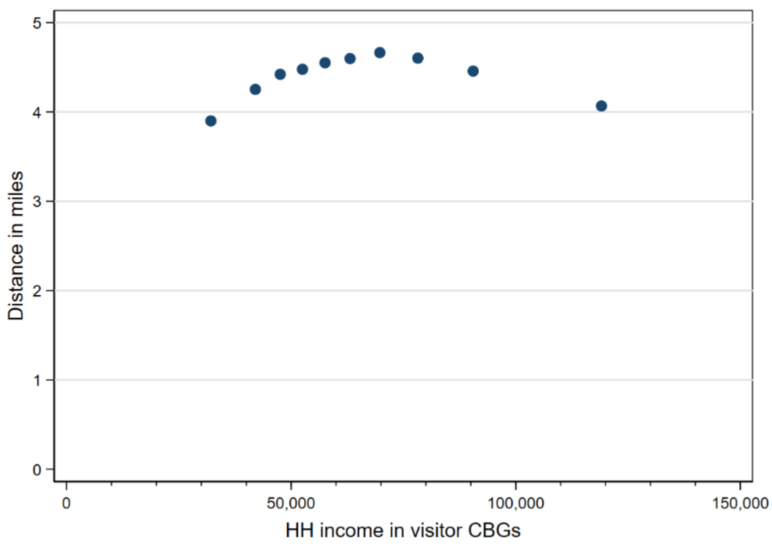
	(1)	(2)	(3)
Median income ('0000 $)	-1.093***	-4.215***	-4.301***
	(0.265)	(1.106)	(0.653)
Median income (squared)		0.168**	0.156***
		(0.0529)	(0.0289)
Urbanicity			0.00284
			(0.0113)
Under-5 population			0.0127***
			(0.00105)
Constant	19.81***	32.17***	20.97***
	(2.337)	(5.301)	(2.845)
Observations	770	770	770
R^2	0.022	0.041	0.613

Table shows coefficients from Ordinary Least Squares regressions at the level of Census Block Groups (Panel A) and Zip Code Tabulation Areas (Panel B). Illinois provides a registry of all licensed programs through the website of the Illinois Department of Child and Family Services. The list of licensed programs in December 2020 is used to generate the outcome variable in the above table. Median income is measured as median household income in the geographic area where visitors reside. Urbanicity is measured as the percentage of people who live in an urban area. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

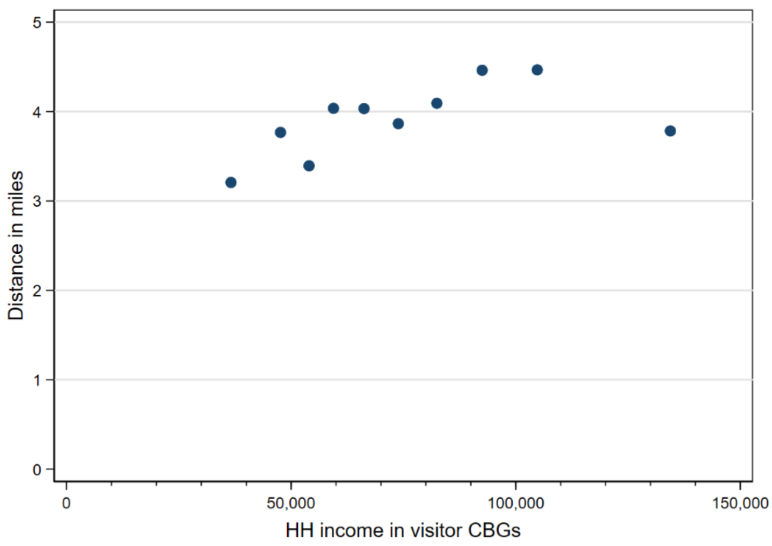
Chapter 2 – Figures

Figure 2.1: Distance Traveled and Income

Panel A: United States



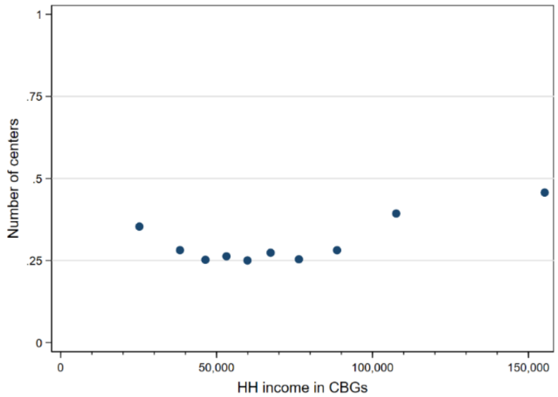
Panel B: Illinois



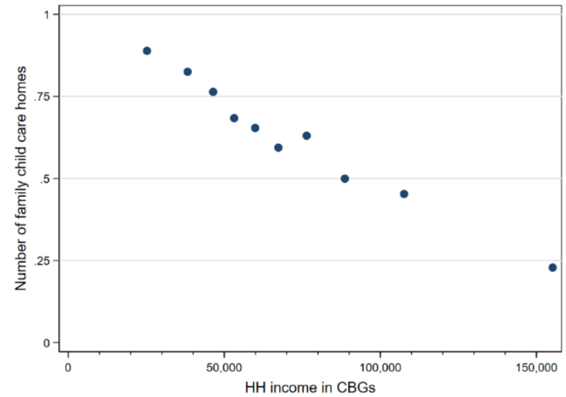
Notes: Figures present SafeGraph data for September 2019 and 5-year estimates from the American Community Survey for 2015-2019. Both panels show binned scatter plots with 10 bins of distance traveled to access child care by median household income in visitors' home Census Block Groups.

Figure 2.2: Number of programs by income in Illinois

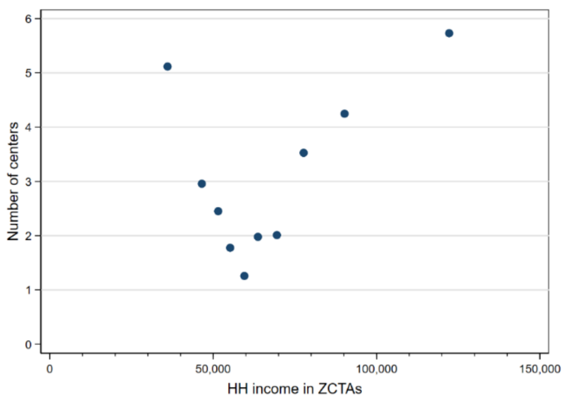
Panel A: Number of child care centers by CBG



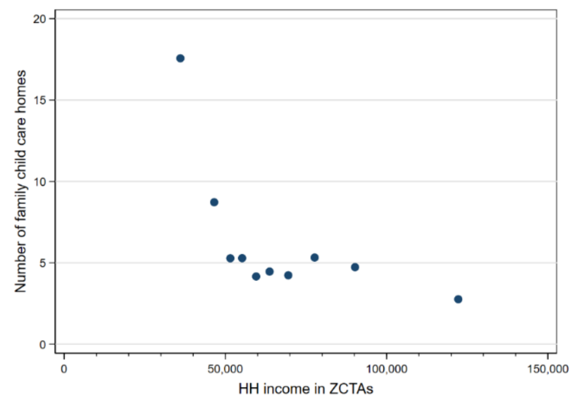
Panel B: Number of family child care programs by CBG



Panel C: Number of child care centers by ZCTA



Panel D: Number of family child care programs by ZCTA



Notes: Figure shows the number of child care centers (Panels A and C) and the number of family child care programs (Panels B and D) in the state of Illinois by median household income in the Census Block Group (CBG) (Panels A and B) and Zip Code Tabulation Area (ZCTA) (Panels C and D) where the program is located. Illinois provides a registry of all licensed programs through the website of the Illinois Department of Child and Family Services. The list of licensed programs in December 2020 was used to generate the above graphs. Median household income at the CBG level was downloaded from the National Historical Geographic Information System (Manson et al., 2021) and were 5-year estimates from the American Community Survey for the years 2015-2019.

Chapter 2 – Appendix

Additional information on measures

Distance traveled

The main outcome of interest in the paper, distance traveled, is a variable that is readily available in the SafeGraph Patterns dataset. They provide the median distance that all visitors traveled between their homes and the child care center they are visiting. If a child care program had 100 visitors during the month of September 2019, the dataset contains the median distance of these 100 visitors as an aggregated variable.

As no individual demographic information is available about visitors other than the Census Block Group where they reside, I proxy for visitors' demographic information using demographic information for their home CBG. Technically, their home CBG is determined as the CBG where the tracked mobile device resides during the night.

I use the following method to determine median household income of child care program visitors. For each child care center, SafeGraph provides a list of all CBGs where visitors are coming from with the count of visitors (reported as 4 if the number is 4 or below). Following with the above example, if of the 100 visitors 25 are coming from CBG-1, 50 from CBG-2, and another 25 from CBG-3, this information is available in the dataset. For each CBG, I merge in the median household income from the ACS 5-year estimates for that CBG. As visitors are coming from several CBGs, there are multiple observations for each child care program. To aggregate these to only one observation per child care program, I take the weighted average of the median income across all visitor CBGs, weighted by the number of visitors in each CBG. I use this weighted average as the "visitors' household income" measure for the specific child care program.

Demographic Information for Geographic Areas

I downloaded demographic information from the National Historical Geographic Information System (NHGIS) website at the CBG and at the ZCTA level (Manson et al., 2021). These are 5-year estimates from the American Community Survey for the years 2015-2019 aggregated to the CBG or the ZCTA level. The primary demographic variables of interest are total population, population under the age of 5, median household income, and urbanicity. Urban/rural distinction is only available from the decennial censuses, and the latest at the time of the analyses was from 2010. The Census Bureau defines urban areas as “a continuously built-up area with a population of 50,000 or more“ and rural areas as “not urban” (Census Bureau, 1994). They provide the number of people who live in an urban area. I divide this number by the total population to receive a percentage of the population who live in an urban area.

The sample of child care centers in the SafeGraph data contains the CBG where the child care program is located. For all other child care programs in the registry of Illinois child care programs that do not have this information, addresses have been geocoded and linked to CBGs using ArcGIS. This allows me to perform CBG-level analyses.

Zip Code Tabulation Areas are zip codes created by the Census Bureau to make tractable, aerial representation of postal zip codes. ZCTAs and postal zip codes align very closely. Postal zip codes are available for child care programs both in the SafeGraph sample and in the Illinois registry data. I used a ZCTA to zip code crosswalk from the Uniform Data System (UDS) Mapper created by the American Academy of Family Physicians and available from their website https://udsmapper.org/zip-code-to-zcta-crosswalk/ (last accessed on 11/9/2022).

Chapter 2 – Appendix Tables

Appendix Table 2.1: Number of child care programs by type and by household income in Illinois

Panel A: Census Block Groups

	(1) Centers	(2) Centers	(3) Centers	(4) Family CC	(5) Family CC	(6) Family CC
Median income ('0000 $)	0.0121***	-0.00583	-0.00413	-0.0481***	-0.0672***	-0.0697***
	(0.00210)	(0.00670)	(0.00659)	(0.00247)	(0.00911)	(0.00889)
Median income (squared)		0.000909**	0.000693*		0.000967**	0.000865*
		(0.000338)	(0.000333)		(0.000367)	(0.000360)
Urbanicity			0.00176***			0.00160***
			(0.000147)			(0.000298)
Under-5 population			0.00108***			0.00262***
			(0.000138)			(0.000287)
Constant	0.219***	0.288***	0.0553	0.967***	1.041***	0.710***
	(0.0156)	(0.0286)	(0.0315)	(0.0249)	(0.0456)	(0.0507)
Observations	8623	8623	8623	8623	8623	8623
R^2	0.005	0.006	0.034	0.026	0.026	0.060

Panel B: Zip Code Tabulation Areas

	(1) Centers	(2) Centers	(3) Centers	(4) Family CC	(5) Family CC	(6) Family CC
Median income ('0000 $)	0.193*	-0.300	-0.324	-1.252***	-3.813***	-3.865***
	(0.0757)	(0.306)	(0.173)	(0.199)	(0.782)	(0.538)
Median income (squared)		0.0265	0.0224*		0.138***	0.130***
		(0.0161)	(0.00878)		(0.0360)	(0.0231)
Urbanicity			0.00553			0.00373
			(0.00317)			(0.0103)
Under-5 population			0.00346***			0.00775***
			(0.000279)			(0.000895)
Constant	2.434***	4.385**	1.079	16.16***	26.30***	19.37***
	(0.574)	(1.333)	(0.735)	(1.748)	(3.842)	(2.368)
Observations	770	770	770	770	770	770
R^2	0.010	0.016	0.652	0.054	0.078	0.483

Notes: Appendix Table 1 shows coefficients from Ordinary Least Squares regressions at the level of Census Block Groups (Panel A) and Zip Code Tabulation Areas (Panel B). Illinois provides a registry of all licensed programs through the website of the Illinois Department of Child and Family Services. The list of licensed programs in December 2020 is used to generate the outcome variables in the above table. Median income is measured as median household income in the geographic area where visitors reside. Urbanicity is measured as the percentage of people who live in an urban area. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix Table 2.2: Regression results for capacity on income

Panel A: Census Block Groups

	(1)	(2)	(3)
Median income ('0000 $)	0.00375	-0.0405**	-0.0304*
	(0.00448)	(0.0147)	(0.0146)
Median income (squared)		0.00224**	0.00160*
		(0.000706)	(0.000705)
Urbanicity			0.00405***
			(0.000300)
Constant	0.494***	0.665***	0.310***
	(0.0351)	(0.0659)	(0.0682)
Observations	8623	8623	8623
R^2	0.000	0.002	0.011

Panel B: Zip Code Tabulation Areas

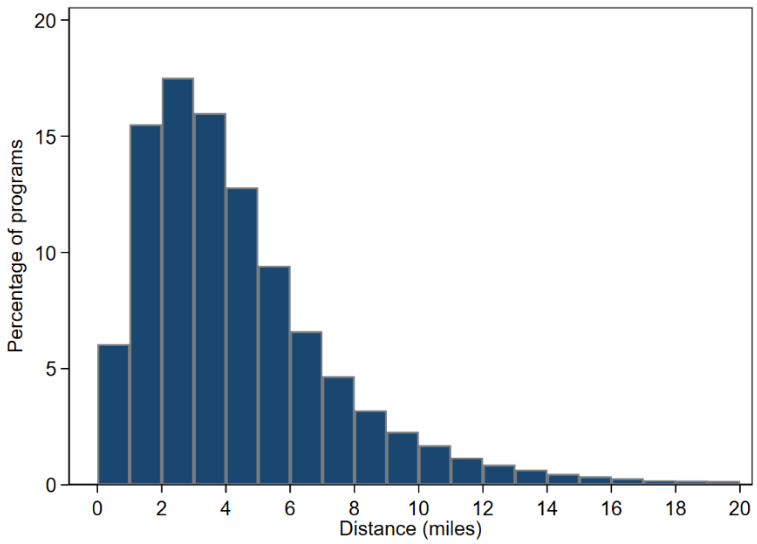
	(1)	(2)	(3)
Median income ('0000 $)	0.0170*	-0.0529*	-0.0531*
	(0.00726)	(0.0228)	(0.0218)
Median income (squared)		0.00376**	0.00353**
		(0.00141)	(0.00136)
Urbanicity			0.00132***
			(0.000263)
Constant	0.269***	0.546***	0.476***
	(0.0486)	(0.0869)	(0.0867)
Observations	770	770	770
R^2	0.019	0.052	0.079

Notes: Table shows coefficients from Ordinary Least Squares regressions at the level of Census Block Groups (Panel A) and Zip Code Tabulation Areas (Panel B). Illinois provides a registry of all licensed programs through the website of the Illinois Department of Child and Family Services. The list of licensed programs in December 2020 is used to generate the outcome variables in the above table. Median household income, urbanicity, and the number of children under the age of 5 are available from the National Historical Geographic Information System (Manson et al., 2021) and are 5-year estimates from the American Community Survey for the years 2015-2019. Capacity is slots per child, defined as the number of licensed seats divided by the number of children under the age of 5. Median income is measured as median household income in the geographic area where visitors reside. Urbanicity is measured as the percentage of people who live in an urban area. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

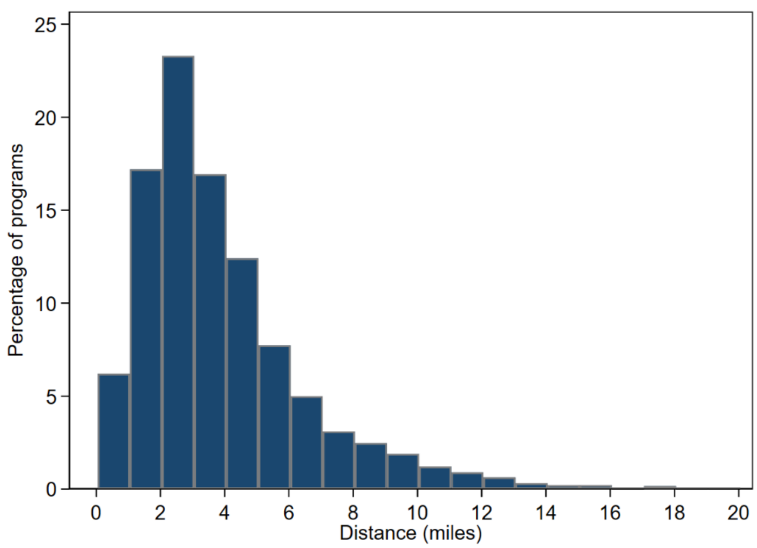
Chapter 2 – Appendix Figures

Appendix Figure 2.1: Histogram of distance traveled for child care

Panel A: United States



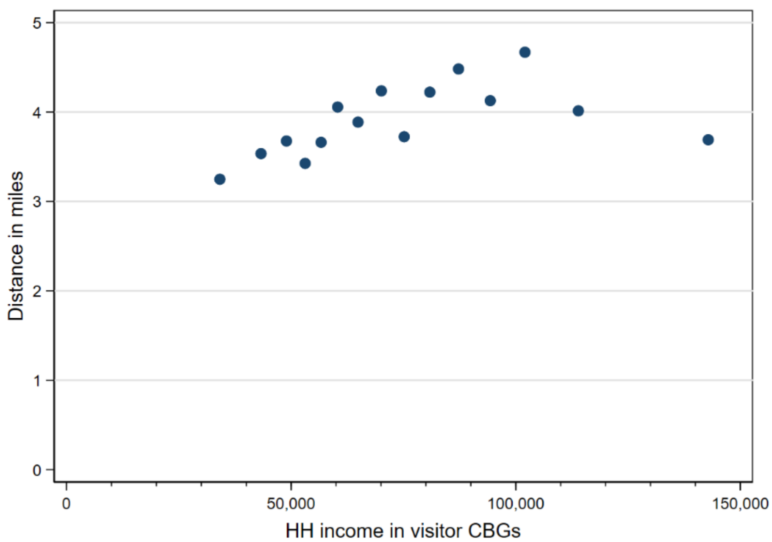
Panel B: Illinois



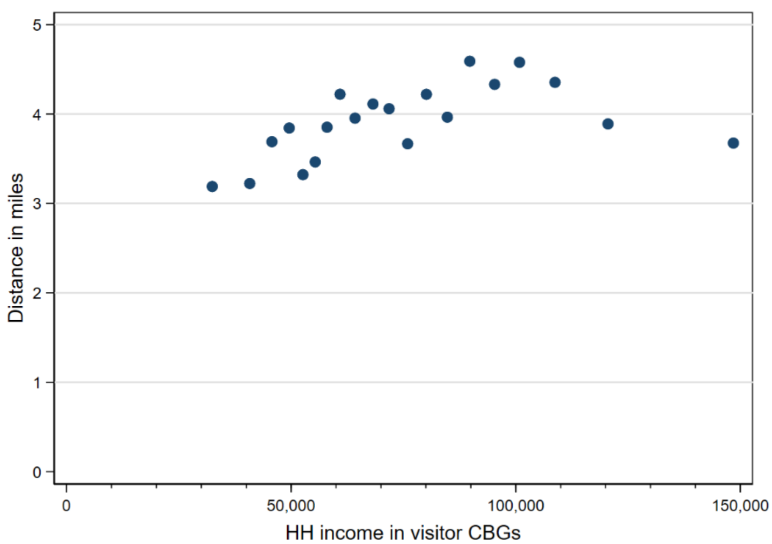
Notes: Figures show the distribution of the median distance visitors traveled to access child care programs. The dataset is at the child care program level. For each child care program the median distance that all visitors in the month of September 2019 traveled to that child care program is reported. For example, the data show that 15% of U.S. child care programs received visitors from between 1-2 miles away. In Illinois, almost 25% of programs received visitors from 2-3 miles away.

Appendix Figure 2.2: Distance traveled and median household income in Illinois, robustness check

Panel A: Binned scatter plot with 15 bins



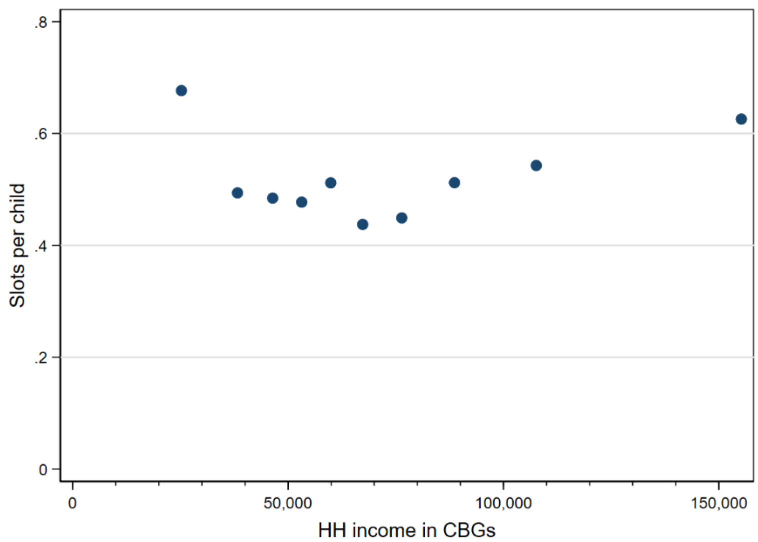
Panel B: Binned scatter plot with 20 bins



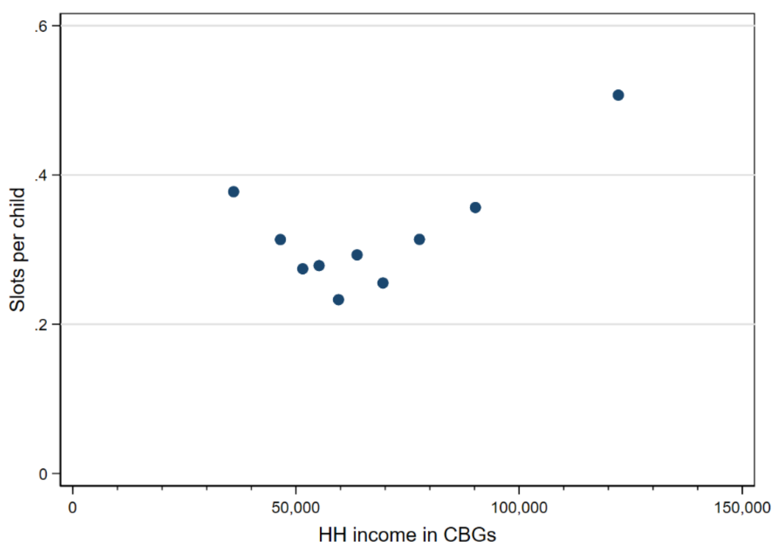
Notes: Figure presents SafeGraph data for September 2019 and 5-year estimates from the American Community Survey for 2015-2019. Both panels show binned scatter plots of distance traveled to access child care by median household income in visitors' home Census Block Groups in the state of Illinois. Panel A uses 15 bins and Panel B uses 20 bins.

Appendix Figure 2.3: Child care capacity and income

Panel A: Census Block Groups



Panel B: Zip Code Tabulation Areas



Notes: Figure shows child care capacity at the CBG (Panel A) and at the ZCTA (Panel B) level by median household income in the geographic area where the program is located. Illinois provides a registry of all licensed programs through the website of the Illinois Department of Child and Family Services. The list of licensed programs in December 2020 is used to generate the above graphs. The registry lists the licensed capacity at each program. Median household income and the number of children under the age of 5 are available from the National Historical Geographic Information System (Manson et al., 2021) and are 5-year estimates from the American Community Survey for the years 2015-2019. Slots per child are defined as the number of licensed seats divided by the number of children under the age of 5.

3) Chapter 3 – Revisiting Family Stress Theory: A Case Study from the Head Start Impact Study

Introduction

Parents in the United States face a complex early childhood education (ECE) market. Currently the U.S. spends 0.3 percent of its Gross Domestic Product (GDP) on early childhood education and care from public sources, which is one of the lowest among member countries of the Organisation for Economic Co-operation and Development (OECD), where the overall average is 0.7 percent (OECD, 2021). As there is no universal, federally funded early childhood education system in the U.S., parents have to find individual solutions for child care. Nearly half of all families report difficulty finding care (L. Schochet, 2019). Even when families find child care, the financial burden associated with these arrangements is high: middle-class working families spend on average 14% of their monthly income on child care (Malik, 2019).

The difficulties associated with finding and paying for early care and education are significant for families (Malik, 2019; L. Schochet, 2019), especially for low-income families. These difficulties may become repeated stressors in parents' lives. Social and psychological stressors were shown to trigger the body's stress response systems (Adam et al., 2017), and repeated stress exposure leads to adverse health outcomes (Sapolsky, 2004). Public provision of early care for free could ease some of the logistical and financial burdens families face and could promote better parental health.

Head Start is one publicly funded preschool program which provides high-quality early education services for free for eligible families. It has been available for low-income families since 1965 (Vinovskis, 2008) and in 2019, Head Start programs around the U.S. served about 1 million children and their families (Administration for Children and Families, 2019). Next to promoting school readiness, Head Start supports family well-being by providing nutritional, health, and social services, as well. The benefits of Head Start on children's outcomes have been extensively documented (Barnett, 1995; Currie, 2001; Deming, 2009; Ludwig & Miller, 2007; Ludwig & Phillips, 2008; Morris et al., 2018; Puma et al., 2010a; Yoshikawa et al., 2013), but research has focused less on how the program influences outcomes for parents. Head Start likely increases parental education and maternal labor force participation (Sabol & Chase-Lansdale, 2015; Wikle & Wilson, 2021) and has an influence on parenting practices (Gelber & Isen, 2013; Padilla, 2020; Puma et al., 2010a). This paper aims to add to our understanding on how Head Start promotes parental well-being using data from the Head Start Impact Study.

The Head Start Impact Study is a nationally representative, longitudinal study that assigned eligible families of newly entering 3- and 4-year-olds to receiving Head Start services or to a non-recipient control group. It was a multi-site randomized control trial with over 380 Head Start centers participating (Puma et al., 2010a). In this paper, next to documenting the average effect of offering Head Start on parental well-being, we also explore heterogeneity in these effects by program characteristics, including family-centered services and supports in a program. Investigating treatment impact heterogeneity has become an important tool in education effectiveness research to understand "under what circumstances" and "for whom" interventions such as Head Start work best (Reardon & Stuart, 2017).

Head Start Participation and Parental Well-being

Offering families Head Start services could lead to better health outcomes for parents through reducing the stressors associated with finding and paying for good quality child care. Parents who look for non-parental care prioritize quality, safety, and cost in child care programs

(Forry et al., 2013). Head Start programs are high-quality, they are subject to accreditation, and follow evidence-based curricula (see for example the Improving Head Start for School Readiness Act of 2007). These measures make sure that the programs provide enriching early experiences for children. When Head Start is available for families it can be a straightforward choice for parents, which makes the search process shorter. As Head Start is provided to eligible families for free, it helps with the cost of child care, too. If the time and monetary considerations are significant stressors for families, removing these stressors from their lives could influence their well-being.

Insights from stress biology help to explain how removing stressors from families' lives could promote better parental well-being. Stressors, including social and psychological stressors, initiate the body's stress response system (Sapolsky, 2004). Psychological factors related to socio-economic status, financial hardship, unemployment, relationships, or parenting tend to frequently activate the stress system that has evolved to help us survive short-term crises (Sapolsky, 2004; Schetter & Dolbier, 2011). These too frequent activations can lead to dysregulated stress hormones, which alters cardiovascular, metabolic and immune functions (McEwen, 2003). Research has also linked exposure to stress hormones to brain function, cognition, and mental health (Lupien et al., 2009). If the provision of Head Start is able to reduce the frequency and severity of stressors related to costs and logistical concerns families face when looking for and attending child care, existing literature on stress biology suggests that parental well-being can improve.

Next to the direct channel of Head Start reducing stressors associated with finding and paying for high-quality child care, there are also indirect channels that may influence parental well-being. One such indirect channel could be related to increased maternal labor force participation, which might lead to work-family conflict, which could be associated with worse parental well-being. Head Start, next to providing a high-quality early education setting for children, also acts as a work-support program. Availability of early child care, either through increased access or lower costs, is predicted to increase parental employment (Blau & Currie, 2006). Findings suggest that the effect of early care availability on maternal labor force participation is positive but the magnitude varies by context (Morrissey, 2017b). In the United States, public kindergarten enrollment around children's 5th birthday increases single mothers' labor supply (E. U. Cascio, 2009; Gelbach, 2002). Universal pre-kindergarten programs in Georgia and Oklahoma did not change maternal labor force participation on average (Fitzpatrick, 2010), but once they disaggregate the effect by education Cascio & Schanzenbach (2013) show that while the effects are zero for mothers with at least some college education, universal pre-k helped lower-educated mothers to return to the labor market. Head Start does not seem to increase employment among parents who are unemployed when the program is offered to the families (Sabol & Chase-Lansdale, 2015), but it does increase full-time employment of single mothers with one young child (Wikle & Wilson, 2021).

When mothers re-enter the labor force or increase their hours worked, work-family conflict can arise. Work-family conflict refers to the difficulty of balancing both employment-related and caregiving responsibilities (Greenhaus & Beutell, 2019). Working longer hours, lack of control over one's schedule, job-related stress, or lack of family support are examples of factors that contribute to work-family conflict (Byron, 2005; Frone et al., 1997). Work-family conflict is associated with lower life, marital, and family satisfaction, and it significantly increases psychological strain and both work- and family-related stress (Allen et al., 2000). As repeated psychological stressors are related to health and well-being, work-family conflict can be a mediator in the relationship between parental employment and well-being. Baker et al. (2008) find that in

Canada, following an expansion of public child care, families seem to have become more stressed. Mothers reported more hostile and less consistent parenting, worse mental health, and lower relationship satisfaction. Fathers' self-reported health also declined (Baker et al., 2008). Chatterji et al. (2013) find that among employed mothers of 6-month-old children more hours worked predicts more depressive symptoms and lower self-reported overall health. Qualitative evidence also supports the conclusion that the dual role of working and parenting is a significant stressor for mothers, especially in a country like the United States with low levels of social support (Collins, 2019).

However, a higher likelihood of being employed or more hours worked could also mean higher income for families. Family stress theory posits that instability in families' lives including economic hardship or unstable work environments can lead to parental stress and negatively impact mental well-being, which can in turn negatively influence parent-child interactions (Conger & Elder, 1994; Yeung et al., 2002). Higher likelihood of being employed, increased income, and potentially more stable work arrangements related to having reliable child care can mitigate these effects. Higher family income has been causally linked to better child outcomes, as well (Brooks-Gunn & Duncan, 1997). Prior research did not find a statistically significant average effect of Head Start on parental earnings, but parents of children who enrolled in Head Start at age 3 tend to have higher earnings 2-3 years after enrollment (O. N. Schochet & Padilla, 2021).

Head Start has the potential to overcome the likely negative effects of parental employment on well-being by its focus on engaging and supporting parents, too. Head Start program design rests on the idea that caregivers and early education programs should work together to provide the best possible environment for the developing child (US DHHS ACF OHS, 2018). Many programs provide education- and employment-related services for parents (Aikens et al., 2017; Sabol &

Chase-Lansdale, 2015) and these work best to promote parental and child well-being when they incorporate a dual developmental framework (Chase-Lansdale & Brooks-Gunn, 2014; Sabol et al., 2021). This framework recognizes that children and parents mutually influence each other's development, and next to considering only one generation at a time, it incorporates the bidirectional relationship between children and adults and its contribution to developmental pathways (Sabol et al., 2021). Career*Advance* is one example of a two-generation program that combines Head Start services for children with workforce development training for parents (Chase-Lansdale et al., 2019). Parents can further their education and potentially achieve better labor market outcomes while their children are enrolled in Head Start. One important feature of the program is the alignment of training and child care hours and the additional supports parents receive for child care (Chase-Lansdale et al., 2019). This suggests that when we examine heterogeneous treatment effects, centers that have a better alignment of child and parent-support services could have a larger impact on parental well-being.

Prior evidence on the Head Start Impact Study

Our study uses data from Head Start Impact Study (HSIS), a nationally representative randomized control trial from 2002. There is a large body of research examining the effects of offering and enrolling in Head Start on child outcomes by analyzing data from the HSIS. Children who were offered Head Start were more ready for school one year after randomization based on measures of cognitive tests (Puma et al., 2010a). The average effects of higher scores in language and literacy and social-emotional outcomes faded out by first grade (Puma et al., 2010a). However, Head Start is not a uniform service. Researchers have found variation in program effects by counterfactual child care arrangements, child characteristics, and geographic location (Morris et al., 2018). For example, using moderation analyses, studies have found that dual-language learners, children who reside in nonurban settings, and children whose mothers were themselves enrolled in Head Start gained more on average than other children (Bitler et al., 2014; Chor, 2018; Puma et al., 2010a).

Moving beyond individual and family characteristics that moderate the effects of access to Head Start on child outcomes, Bloom & Weiland (2015) use a novel Hierarchical Linear Modeling technique, the fixed intercept, random coefficient approach, to quantify variation in treatment effects across Head Start centers. They find that there is a statistically significant variation in effect size across centers on language and literacy measures.

While we have a relatively good understanding about the effects of Head Start on children's cognitive and socio-emotional outcomes, fewer studies have looked at the effects of access to Head Start on parents. Researchers have analyzed data from the HSIS to look at parenting behaviors. Parents are more likely to engage in cognitively stimulating behaviors, like reading to their children, and are less likely to spank children as a disciplinary measure (Gelber & Isen, 2013; Padilla, 2020; Puma et al., 2010a). Padilla (2020) found significant variation across Head Start centers in their success at promoting beneficial parent-child interactions. While all centers were effective at increasing parental reading time, they were not as successful in influencing disciplinary behaviors.

Evidence on parental employment outside the home and earnings as a result of access to Head Start also exists. Head Start increased short-run labor supply of single mothers (Wikle & Wilson, 2021), but it did not increase maternal employment on average (Sabol & Chase-Lansdale, 2015). Access to Head Start increased maternal education, especially for those mothers who already had some college education (Sabol & Chase-Lansdale, 2015).

To our knowledge, no published studies have looked at the effect of access to Head Start on parental well-being. While some evidence exists that child care in the form of government provided pre-school has negative influence on parental stress and well-being (Baker et al., 2008), Head Start with its two-generation focus might have different influences on parental well-being. Next to quantifying the average effects of Head Start on parental well-being, we are also analyzing treatment impact variation using the fixed-intercept, random coefficient method (Sabol et al., 2022) to see whether some centers are more effective at increasing parental well-being than others. Thus, we provide both theoretical contributions to the literature on early care and education and parental well-being, as well as use novel methodologies to advance our understanding about impact variation.

Current study

This study uses data from the Head Start Impact Study, which was a multi-site randomized control trial with data collection running between 2002-2006 and a follow-up data collection in 2008 when children were in grade 3. The study design used oversubscription, and at each center randomized families into two groups: one which were offered to enroll in Head Start and one control group (Puma et al., 2010a, 2012). This design makes it possible to estimate not just the overall average treatment effects pooling information on all centers, but also to analyze variation in treatment impact by center.

The study contributes to the literature on several fronts. First, we examine average intent-to-treat effects of offering Head Start on parental mental well-being, an outcome that has not been rigorously evaluated. We use an intent-to-treat analysis to shed light on average effects. Second, we use recent methodologies specifically designed to analyze variation in treatment effects to understand variation in effect size across centers. Lastly, we test specific features of Head Start centers chosen based on developmental theory to see which, if any, are associated with impact variation.

Methods

Data and Sample

The nationally representative HSIS successfully randomized 4,442 children into treatment and control groups in 378 centers. About 60% of the total sample was offered Head Start services. The program included newly entering 3- and 4-year-olds in two cohorts (Puma et al., 2010a). We restrict the full sample at baseline the following ways. We use parent interviews to get information about their health and well-being, so we drop 19.5% of the observations with missing data (n=865) in the 2002 parent interview. We keep observations only if the respondent was the biological, adoptive, or stepmother, or the father. Overall, we keep 3,436 observations. We run the analyses only for mothers, as well, thus we exclude the n=152 fathers.

When we run our analyses, we further restrict the sample the following way for each outcome (following Bloom & Weiland (2015) and O. N. Schochet & Padilla (2021)). We first drop observations where the outcome variable is missing. We then calculate for each center the number of treatment and control group members, and the percentage of compliers. If a center has no compliers, we drop all observations within the center. Number of observations for each outcome thus vary.

Measures

Mental well-being. We draw measures of well-being from parent interviews at each follow-up time point, from the spring parent interviews for years 2003, 2004, 2005, 2006 and at the time of the third-grade follow-up, in 2008. We use the fall 2002 parent interview items as the baseline.

To measure mental well-being, we use a set of questions that are drawn from the questionnaire of the Center for Epidemiologic Studies Depression Scale (CES-D). These questions concern how often parents during the past week (1) were bothered by things that usually don't bother them; (2) did not feel like eating, their appetite was poor; (3) felt depressed; (4) felt that they could not shake off the blues, even with help from their family and friends; (5) had trouble keeping their mind on what they were doing; (6) felt that everything they did was an effort; (7) felt fearful; (8) their sleep was restless; (9) they talked less than usual; (10) felt lonely; (11) felt sad; (12) they could not get "going". They could choose from the answers (1) rarely or never; (2) some or little; (3) occasionally or moderately; (4) most or all of the time. We follow the HSIS recommendation in how to represent these items with a single number. The responses are first rescaled so that "refused", "don't know", or "not asked" answers are assigned missing values. Then, values between 0-3 are assigned to the categorical answers the following way. Rarely or never is 0, some or little is 1, occasionally or moderately is 2, and most or all of the time is 3. Then, we sum these values for each variable for each person, and then collapse into four categories. Values between 0-4 are assigned 1 (no depressive symptoms), values between 5-9 are assigned 2 (mild depressive symptoms), values between 10-14 are assigned 3 (moderate depressive symptoms), and values between 15-36 are assigned 4 (severe depressive symptoms). If a parent did not respond to at least 3 questions of the 12 the overall variable is set to missing.

Figure 1 shows raw, unadjusted means for treatment and control groups members over time for the two outcomes. Overall health seems to be getting worse over time and control groups means are above treatment groups means. While overall depression scores are getting lower over time, there seems to be no different pattern between the two groups.

Treatment status. In models where we estimate average and heterogeneous effects of offering Head Start services on parental well-being, we use the indicator of whether a child was assigned to the Head Start group or to the control. Thus, we estimate intent-to-treat effects.

Sample characteristics. Baseline sample characteristics are reported in Tables 1 and 2 for the three-, and four-year-old cohorts by treatment and control group status. There are some characteristics that are different between the two groups at baseline for the three-year-old cohort, but given the large number of characteristics examined, these could be due to chance. The treatment group has a higher percentage of parents who show mild depressive symptoms (26.5% vs. 22.3% in the control group, t = -2.03), more control group members report to sleep well at night (34.15% vs. 29.3% in the treatment group, t = 2.2), and more children have special needs in the treatment group (13% vs. 8.7% in the control group t = -2.83). The one characteristic that is concerning is the different rate of compliance with the random assignment in the two groups. While 89.7% complied with the random assignment in the treatment group, 85.7% complied in the control group (t = -2.59). The same pattern was observed in the full sample, too, this difference is not due to the sample restrictions. There are no statistically significant differences between treatment and control groups for the four-year-old sample.

Parents are on average 29 years old, about a third of them are working full-time and half are not working for pay. 75-78% of parents have at most a high school diploma and about a quarter gave birth to their first child before their 18th birthday. There are on average 2 adults and 2.6 children in a household. In terms of race/ethnicity parents were asked to choose the race/ethnicity category that best described them. They could choose more than one category. For example, 40% of people who identified as white also reported to be Hispanic/Latino (n=689). In terms of race, the majority of the sample reported to be either white or Black. Of the 20% of the sample who are neither white nor Black (n=661), 86% (n=570) reported that they are from “another race” and did not list a specific race/ethnicity.

In terms of income, parents were asked to report their monthly household income. For a third of the respondents who did not know what their income was (n=1,031), categorical responses for their monthly income ranges were provided to choose from. The categories were (1) less than \$250; (2) between \$251 and \$500; (3) between \$501 and \$1,000; (4) between \$1,001 and \$1,500; (5) between \$1,501 and \$2,000; (6) between \$2,001 and \$2,500; (7) over \$2,500. We use a random value falling in the respective range to have a continuous measure of income for all respondents who provided an answer. For category (7) we use the upper bound of \$5,000 as the 99th percentile of the reported continuous monthly income is \$5,000. This way, we have continuous income values for n=3,073 respondents. Average monthly household income is \$1,400 for the three-year-old cohort and \$1,500 for the four-year-old cohort.

About half of the sample reports to have no depressive symptoms at baseline, and a third of the sample reports to sleep well. Between 50-58% of people report to have excellent or very good health at baseline.

Program characteristics. Children in the sample are either enrolled in Head Start or another program of their choice (if they are in the control group or if they did not end up enrolling in Head Start). Interviews were conducted with program directors to gather information about the programs themselves. We use some of this information to generate variables to analyze whether any program characteristics predict treatment impact variation. The interviews were conducted at several time points, and we use the earliest available one, which is spring 2003, as these describe families’ experiences in the first year of the program.

Program characteristics are reported at the child level. We use the following procedure to generate center-level variables. In the center director survey the data are at the child level. There is one indicator for the Head Start center where the child initially applied and where the randomization happened (“centerID”). There is another indicator for the actual program where the child ended up enrolling in ("careID"). Center directors answer questions about the program where the child is enrolled (answers are at the "careID" level). However, there is no crosswalk between these two variables, so it is not always clear whether the answers apply to the Head Start center or to another program where the child enrolled. Even for treatment group compliers there are a few cases of where the “centerID” and the “careID” are not unique pairs, which means that more than one “careID” can be associated with a specific Head Start center. To overcome this issue, we assign the most frequent “careID” associated with a “centerID” as a unique match, and then take the average across treatment group compliers to generate the variables of interest for the Head Start centers.

We use the following variables to describe centers. First, we generate measures of program structure. These variables include indicators for whether the center provides full-time care, part-time care, or offers extended hours; the number of hours a center is open; and center capacity.

To measure family-centered services and supports in a program, we use the list of services offered to family members other than children. These services are: (1) income assistance, including welfare, SSI, or unemployment insurance; (2) help with medical care; (3) food and nutrition assistance, including food stamps and WIC; (4) help with housing; (5) help with utilities (water, heat, electric, telephone); (6) adult education/literacy; (7) job training and employment assistance; (8) alcohol or drug abuse treatment or counseling; (9) family counseling or mental health services; (10) help dealing with family violence; (11) foster care payments; (12) any other service. We sort these items into two domains that can be related to improving family and child outcomes based on prior research (Sabol et al., 2018). These two domains are (1) family support services (items 1, 2, 3, 4, 5, 8, 9, 10, 11); and (2) education, career, and assets (items 6 and 7). We created a measure of breadth: how many of these 2 domains centers provide services from; and a measure of depth: how many sub-services they provide regardless of domain (out of the possible 11).

Lastly, we generate variables that describe family demographics in each center. We take the average across all treatment compliers for these measures. We use family income; race/ethnicity (percentage white, percentage Black, percentage Hispanic/Latino); and percentage of children who are English language learners. We infer that a child is an English language learner if the language spoken in the home is Spanish.

Analytic plan

The set-up of the Head Start Impact Study, the randomization of families within centers, makes it possible to study not just the average effect of offering Head Start services, but also the variation in the average treatment effect across Head Start centers. Since there are both treatment and control families in each center, we can treat each center as its own mini-randomized control trial, and estimate treatment impacts for each (Sabol et al., 2022). We use a fixed intercept, random coefficient (FIRC) model, which is a hierarchical linear modeling technique (HLM) where an indicator variable is included for each center (this is the fixed intercept) and random effects are used to estimate the site-specific treatment impact (Bloom et al., 2017; Bloom & Weiland, 2015). Including center-fixed effects eliminates bias that arises due to the fact that there is variation across Head Start centers in the proportion of families who are assigned to treatment or control conditions (Bloom & Weiland, 2015). In order to have sufficient power to detect variations in treatment effects we need to have enough treatment and control group members in each center (Sabol et al., 2022). For the three-year-old cohort there are on average 10.4 children per center, and for the four-year-old cohort there are on average 10.7 children in each center. We also investigate whether specific characteristics of centers (e.g. whether they offer family support services or extended care hours) predict differences in treatment impact.

We use a two-level model that nests parents (level 1) in Head Start centers (level 2) where they initially applied for child care. More specifically, we estimate the following model to assess average treatment effect and variation in this effect for both cohorts separately:

Level 1: Parents

$$Y_{ij} = \alpha_j + \beta_j T_{ij} + \sum_{m=1}^{M} \pi_m W_{mij} + v_{ij}$$

Level 2: Head Start centers

$$\alpha_j = \alpha_j$$

$$\beta_j = \beta_0 + u_j, u_j \sim N(0, \tau_\beta^2)$$

At level 1, Y_{ij} represents the outcomes of interest for parent i in Head Start center j; α_j indexes a series of indicator variables for each Head Start center j; T_{ij} is an indicator variable for the random assignment status; W_{mij} is the matrix of covariates at baseline (indicator variables for whether the parent is female, white, Black, Hispanic, or of another race/ethnicity, whether the parent is foreign born, imputed family income, indicator variables for the highest level of education is less than high school, high school diploma, some college, or at least a bachelor's degree, whether the parent is married, whether they are in good health, and baseline CES-D score in 2002). v_{ij} is the error term which varies independently across parents and has a mean of zero and separate variances for treatment (τ_T^2) and control group (τ_C^2) members. These separate individual variances across treatment group status are important because the it is possible that the Head Start program effects vary across individuals within a center, in which case outcome variance for treatment group members would be different than outcome variance for control group members (Bloom et al., 2017). At level 2, β_j is the random slope that represents the average treatment effect in Head Start center j, which is the linear combination of β_0, the mean of all the center-specific mean effects, and a random error u_j, which varies across sites and has a mean of 0 and variance of τ_β^2.

Our estimands of interest are β_0, the average treatment effect across all centers and τ_β^2, the standard deviation of the distribution of the site-specific treatment effects. Larger values of τ_β^2 would mean a greater variation in the effect of offering Head Start services to parents across centers (Sabol et al., 2022).

We also test whether the estimated variation in treatment effects is significantly different from 0 with specifying the null hypothesis that $\tau_\beta^2 = 0$. We compute a Q-statistic, which is used in meta-analysis to test whether cross-study impact variation is zero (Hedges & Olkin, 2014). We follow the method outlined in Bloom et al. (2017) to calculate the Q-statistic, which approximates a chi-square distribution with j-1 degrees of freedom, which we can use to test statistical significance.

We conduct the analyses separately for the three-year-old and the four-year-old cohort, because initial random assignment was done separately for the two cohorts and baseline demographic characteristics are different between the two groups (Puma et al., 2010b).

To assess whether center characteristics predict treatment impact variation, in a second set of analyses we include interaction terms between the treatment variable and the center characteristic in the Level 2 (Head Start center) equations. Center-level characteristics are indexed with P_j in the below equation.

Level 1: Parents

$$Y_{ij} = \alpha_j + \beta_j T_{ij} + \sum_{m=1}^{M} \pi_m W_{mij} + \upsilon_{ij}$$

Level 2: Head Start centers

$$\alpha_j = \alpha_j$$

$$\beta_j = \beta_0 + \beta_1 P_j + u_j, u_j \sim N(0, \tau_\beta^2)$$

For each of the center-level characteristics (full-time care, part-day program, number of hours open in a week, capacity, whether the center is at full capacity, breadth of services, depth of services, average family income, percentage white families, percentage Black families, percentage Hispanic families, percentage families of other race, percentage of children who are English language learners) we run separate regressions.

Results

Effects of offering Head Start on mental health

Results on the average treatment effects of access to Head Start on parental mental health outcomes are presented in Table 2. Random assignment to Head Start reduced mental health by 0.08 points on a 4-point scale in 2003 in the three-year-old cohort (one year after being offered Head Start), however, this result was not statistically significantly different from zero (p=0.141). Similarly, at each follow-up time (in 2004, 2005, 2006, and 2008) the point estimates were negative and not statistically significant. For the four-year-old cohort, the magnitude of the coefficients was very small and they were all statistically non-significant. For example, the estimated coefficient was -0.008 one year after being enrolled in Head Start, with a p-value of 0.89.

Outcomes did not vary significantly across centers. While the standard deviation of the distribution of the site-specific treatment effect ($\sqrt{\tau_\beta^2}$) was 0.23 for the three-year-old cohort one year after randomization, it was not statistically significantly different from zero (Q-statistic=250.58, p=0.978). Even though their variation is not significantly different from zero, we computed and graphed the site-specific treatment effects. These site-specific treatment effects were computed using the random effects estimates produced by the regression specified in the Analytic plan section. These random effects estimates were then adjusted (multiplied) by the estimated variance of the predicted random effects. Figure 2, panel A shows the histogram of the site-specific treatment effects. The histogram shows that the large majority of the centers had a predicted zero effect on parental mental health. While some centers reached an estimated 1-point reduction or a 1-point increase in the depression scale measure, this may have just been due to estimation error. For each follow-up year, τ_β^2's were less than 0.001, which indicated no variation in effects across centers. It was not possible to produce histograms of these effects as they are all very close to zero.

For the four-year-old cohort, the standard deviation of the site-specific treatment effect one year after randomization was 0.12, and 0.192 two years after randomization, but these were not statistically significantly different from zero. Figure 2, panels B and C show histograms for the site-specific treatment effects. In 2005 and 2008 τ_β^2 was smaller than 0.001.

Variation by center characteristics

To predict treatment impact based on center characteristics, we used a set of variables describing structural aspects of Head Start centers (indicator variables for whether the center provides full-time care, whether the center has a part-day program, the number of hours the center is open in a week, capacity, whether the center is at full capacity); measures of family engagement and support services (breadth of services – how many domains they provides services from where the domains are (1) family support services and (2) education, career, and assets; and depth of services – the number of domains they provide services for from a total of 11). We also used a set of family demographic characteristics at the center (average family income, percentage white families, percentage Black families, percentage Hispanic families, percentage families of other race, percentage of children who are English language learners). We found that none of these center characteristics predicted parental mental health outcomes. Table 3 presents regression coefficients on the interaction term between the treatment variable and the center characteristic under consideration. None of them are statistically significantly different from zero.

Discussion

A large body of research examines the effects of Head Start on child outcomes, but we know much less about how Head Start contributes to the well-being of other family members, like parents. This study aimed to inform our understanding about whether offering families Head Start could ameliorate parental mental health. We found no statistically significant average treatment effect of offering Head Start on parental depression. Furthermore, we found that there was no variation in treatment impacts across Head Start centers. No centers contributed to parental mental health more than others either in the positive or the negative direction. Overall, we also found that no center characteristic considered in this study moderated treatment impacts on parental mental health outcomes.

There are several potential reasons for the overall null results of our study, which provide avenues for future research. The first one could be theoretical. Ecological systems theory posits that families' contexts, including early childhood education centers and the larger policy environment around child and family support, matters for human development (Bronfenbrenner & Morris, 2006). Family stress theory posits that instability in families' lives including economic hardship or lack of adequate child care can lead to parental stress and negatively impact mental well-being, which can in turn negatively influence parent-child interactions (Conger & Elder, 1994; Yeung et al., 2002). Stress biology provides the biological underpinnings of how repeated social stressors can "get under the skin" and influence health and well-being (Adam et al., 2017; Sapolsky, 2004; Seeman et al., 2010). Based on these theories we hypothesized that providing free, high-quality child care to low-income families is going to ameliorate parental mental health. Potentially, stressors around child care are relatively small in families' lives, or they may not be persistent. Future research could build theories that incorporate smaller or not persistent stressors and how they might contribute to parental mental well-being.

A second potential explanation for the null findings could be measurement-related. We used the only available parental mental health measure in the Head Start Impact Study, the Center for Epidemiologic Studies Depression Scale (CES-D), which is a clinical depression measure collected once per year. Items that are more suitable to measure short-term changes in parental stress levels could be more appropriate. Annual scales may not be able to pick up changes in parental stress levels related to child care. More frequent data collection would be a useful addition in future studies about the effects of child care on parental mental health.

Third, in this paper we only consider intent-to-treat effects. The counterfactual child care arrangements of participants of the Head Start Impact Study may have been similar in terms of mitigating parental stress around child care concerns. 60% of control-group children were also enrolled in center-based care (Puma et al., 2010a). Exploring treatment-on-the-treated might shed light on some aspects of Head Start programs that can contribute to parental mental health.

Conclusion

We explored whether providing free, high-quality early childhood education improves parental mental health. We did not find any effects on average on parental mental health. Our study's strength is not just the theoretical contribution on how early childhood education programs can contribute to the well-being of family members other than children. We also use robust and novel methods to test whether there is variation in treatment impact across child care centers. Our estimates from the random coefficient, fixed intercept (FIRC) approach show no statistically significant differences in impacts across centers. Furthermore, using moderation analyses, we also explore whether center characteristics, like structural factors and family support services provided, predict variation in treatment effects. We find that none of these characteristics predict variation.

These null results may not mean that early childhood education cannot contribute to the well-being of family members other than children. The study uses data from the Head Start Impact Study, and the sample includes low-income and largely marginalized members of the society. Economic hardship and other stressors may be very salient in families' lives and access to child care may not be enough support to ameliorate parental mental health.

Chapter 3 – Tables

Table 3.1: Summary statistics and balance test

A. Three-year-old cohort

	Treatment Mean (sd)	Control Mean (sd)	Difference t-stat/p-val
Complier	89.69	85.73	-2.59**
	(30.42)	(35.00)	0.01
Parent female	95.18	96.43	1.29
	(21.43)	(18.56)	0.20
Hispanic/Latino	33.44	32.33	-0.50
	(47.20)	(46.81)	0.62
White	44.58	48.56	1.68
	(49.73)	(50.02)	0.09
Black/African American	38.27	35.20	-1.33
	(48.62)	(47.79)	0.18
Another race	16.48	14.66	-1.05
	(37.12)	(35.39)	0.29
Not born in USA	26.73	28.57	0.87
	(44.27)	(45.21)	0.38
Works full-time	34.64	33.09	-0.68
	(47.60)	(47.09)	0.49
Works part-time	16.28	16.48	0.11
	(36.93)	(37.12)	0.91
Does not work for pay	49.00	50.36	0.57
	(50.01)	(50.03)	0.57
Less than high school	32.03	35.43	1.52
	(46.68)	(47.86)	0.13
High school diploma	43.04	40.29	-1.17
	(49.53)	(49.08)	0.24
Some college	21.27	20.57	-0.36
	(40.94)	(40.45)	0.72
College or more	3.67	3.71	0.05
	(18.81)	(18.92)	0.96
Parent age	28.89	28.57	-1.08
	(6.25)	(6.03)	0.28
Income	1423.78	1401.21	-0.45
	(1004.20)	(976.17)	0.65
Parents married	43.19	45.06	0.79
	(49.55)	(49.79)	0.43
Age at first birth <18	24.54	24.71	0.08
	(43.05)	(43.17)	0.93

Number of children	2.63	2.61	-0.27
	(1.32)	(1.27)	0.79
Number of adults	2.04	2.11	1.50
	(0.97)	(1.05)	0.13
Number of adult earners	1.6	1.66	1.61
	(0.72)	(0.86)	0.11
CES-D score (0-36)	5.96	5.64	-1.10
	(6.29)	(6.15)	0.27
Not depressed	51.71	55.97	1.79
	(49.99)	(49.68)	0.07
Mild depression	26.48	22.30	-2.03*
	(44.14)	(41.66)	0.04
Moderate depression	12.70	13.09	0.25
	(33.31)	(33.76)	0.80
Severe depression	9.11	8.63	-0.35
	(28.78)	(28.11)	0.73
Sleeps well	29.29	34.15	2.20*
	(45.53)	(47.45)	0.03
Health excellent/very good	58.78	58.00	-0.33
	(49.24)	(49.39)	0.74
Child female	51.54	51.07	-0.20
	(50.00)	(50.02)	0.84
Child is English language learner	22.98	22.42	-0.27
	(42.09)	(41.74)	0.79
Child has special needs	12.97	8.70	-2.83**
	(33.61)	(28.21)	0.00
Observations	1203	701	1904

B. Four-year-old cohort

	Treatment Mean (sd)	Control Mean (sd)	Difference t-stat /p-val
Complier	85.70	87.63	1.07
	(35.03)	(32.95)	0.29
Parent female	95.62	95.30	-0.29
	(20.48)	(21.19)	0.77
Hispanic/Latino	42.42	43.11	0.26
	(49.45)	(49.57)	0.79
White	54.84	56.16	0.50
	(49.79)	(49.66)	0.62
Black/African American	22.84	22.71	-0.06
	(42.00)	(41.93)	0.95
Another race	20.63	20.60	-0.02
	(40.49)	(40.48)	0.99
Not born in USA	38.87	38.74	-0.05
	(48.77)	(48.76)	0.96
Works full-time	33.33	32.98	-0.14
	(47.17)	(47.06)	0.89
Works part-time	13.90	16.06	1.15
	(34.61)	(36.74)	0.25
Does not work for pay	52.77	50.87	-0.72
	(49.95)	(50.04)	0.47
Less than high school	41.00	42.66	0.63
	(49.21)	(49.50)	0.53
High school diploma	37.03	36.71	-0.12
	(48.31)	(48.24)	0.90
Some college	17.26	16.61	-0.33
	(37.81)	(37.25)	0.74
College or more	4.71	4.02	-0.63
	(21.19)	(19.66)	0.53
Parent age	29.69	29.58	-0.33
	(6.29)	(6.22)	0.74
Income	1495.33	1509.94	0.26
	(965.53)	(1074.29)	0.80
Parents married	46.27	49.13	1.08
	(49.89)	(50.04)	0.28
Age at first birth <18	24.32	25.61	0.57
	(42.92)	(43.69)	0.57
Number of children	2.68	2.65	-0.37
	(1.30)	(1.28)	0.71
Number of adults	2.15	2.15	0.15

	(1.06)	(1.00)	0.88
Number of adult ear-ners	1.62	1.60	-0.30
	(0.77)	(0.77)	0.76
CES-D score (0-36)	6.32	6.02	-0.85
	(6.87)	(6.36)	0.40
Not depressed	50.21	50.87	0.25
	(50.03)	(50.04)	0.80
Mild depression	26.26	26.57	0.13
	(44.03)	(44.21)	0.89
Moderate depression	13.34	11.89	-0.82
	(34.02)	(32.39)	0.41
Severe depression	10.19	10.66	0.29
	(30.27)	(30.89)	0.77
Sleeps well	32.67	32.52	-0.06
	(46.92)	(46.89)	0.95
Health excellent/very good	55.50	52.18	-1.26
	(49.72)	(50.00)	0.21
Child female	48.96	49.48	0.20
	(50.02)	(50.04)	0.84
Child is English language learner	33.78	32.4	-0.54
	(47.32)	(46.84)	0.59
Child has special needs	14.20	12.02	-1.21
	(34.92)	(32.55)	0.23
Observations	958	574	1532

Table 3.2: FIRC model estimates of the effect of offering Head Start services to families on parental mental health

Panel A: Three-year-old cohort

	CES 2003	**CES 2004**	**CES 2005**	**CES 2006**	**CES 2008**
ATE (s.e.) p-value	-0.084 (0.057) 0.141	-0.057 (0.055) 0.304	-0.0087 (0.051) 0.865	-0.07 (0.050) 0.166	-0.021 (0.052) 0.689
$\sqrt{\widehat{\tau_\beta^2}}$	0.23	<0.001	<0.001	<0.001	<0.001
Q-stat	250.58	219.01	244.38	193.62	190.33
p-value of Q-stat	0.978	0.999	0.994	0.999	0.999
N individuals N centers	1463 290	1454 292	1465 295	1443 287	1390 287

Panel B: Four-year-old cohort

	CES 2003	**CES 2004**	**CES 2005**	**CES 2008**
ATE (s.e.) p-value	-0.008 (0.056) 0.89	0.016 (0.062) 0.794	0.06 (0.059) 0.312	0.059 (0.06) 0.33
$\sqrt{\widehat{\tau_\beta^2}}$	0.12	0.195	<0.001	<0.001
Q-stat	127.95	190.12	140.16	134.72
p-value of Q-stat	1	0.999	1	1
N individuals N centers	1218 257	1172 257	1190 255	1131 250

Notes: Table shows estimates of average treatment effects and the standard deviation of the site-specific treatment effects using the fixed-intercept, random coefficient method. Each column represents a separate regression. Sample includes parents who had non-missing CES-D depression scores at baseline in 2002, and non-missing CES-D depression scores in the indicated follow-up year. Centers where compliance was zero are excluded.

Table 3.3: Impact of access to Head Start on parental mental health moderated by center characteristics

Outcome variable: CES-D 2003

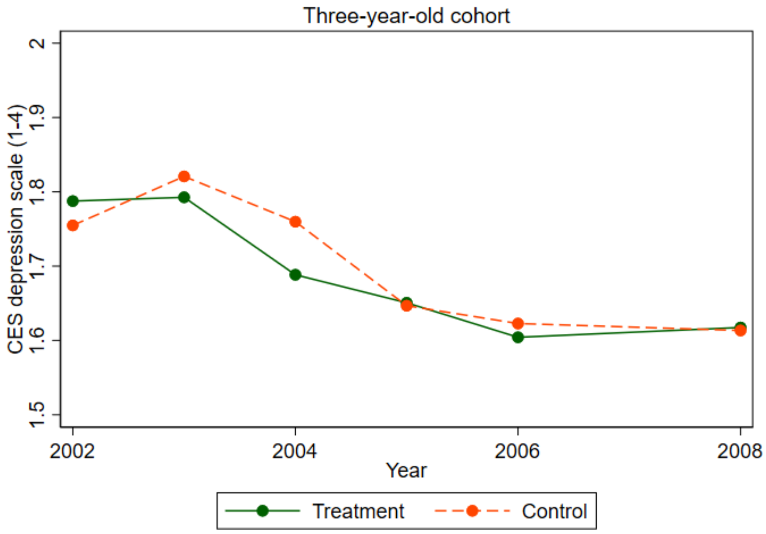
	Three-year-old cohort	Four-year-old cohort
Structural factors		
Treatment*Full-day program	-0.147 (0.122)	0.037 (0.115)
Treatment*Part-day program	0.042 (0.122)	0.180 (0.135)
Treatment*Center hours	-0.005 (0.005)	0.001 (0.004)
Treatment*Capacity	-0.001 (0.001)	-0.008 (0.001)
Treatment*Center at full capacity	0.117 (0.119)	-0.027 (0.116)
Family support services		
Treatment*Breadth	-0.006 (0.016)	0.005 (0.015)
Treatment*Depth	-0.094 (0.104)	0.099 (0.095)
Family demographic characteristics		
Treatment*Average family income	0.000 (0.000)	0.000 (0.000)
Treatment*Percent Black	-0.049 (0.142)	-0.276 (0.168)
Treatment*Percent white	0.081 (0.145)	0.249 (0.152)
Treatment*Percent Hispanic	-0.058 (0.149)	0.104 (0.142)
Treatment*Percent other race/ethnicity	-0.138 (0.240)	-0.103 (0.207)
Treatment*English language learners	-0.137 (0.187)	0.054 (0.160)

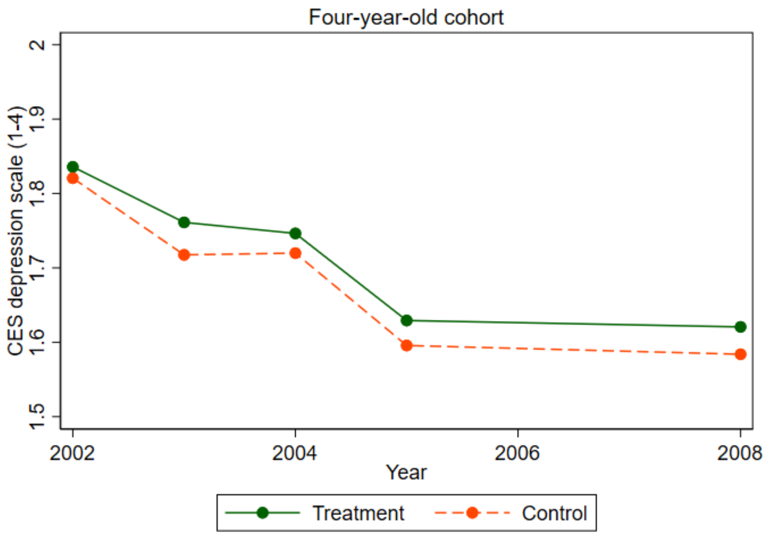
Notes: Each cell represents the coefficient on the interaction term, which is the result of estimating a separate fixed intercepts, random coefficient (FIRC) model. All models control for parent demographic variables outlined in the Methods section. Parentheses contain standard errors.

To measure family-centered services and supports in a program, we use the list of 11 potential services offered to family members other than children. We sort these 11 items into two domains. Breadth is defined as the number of domains (out of 2) that are covered by a center, and depth is the number of items covered (out of 11).

Chapter 3 – Figures

Figure 3.1: Graphs of outcomes over time for Treatment and Control group members, unadjusted

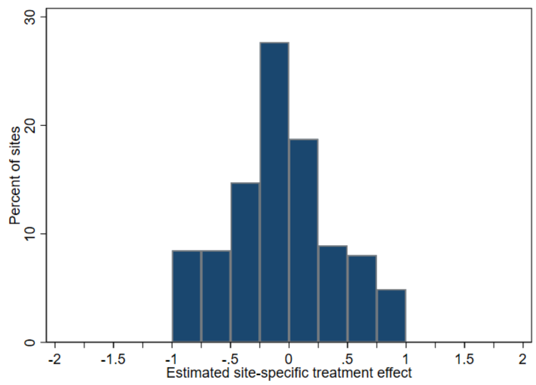




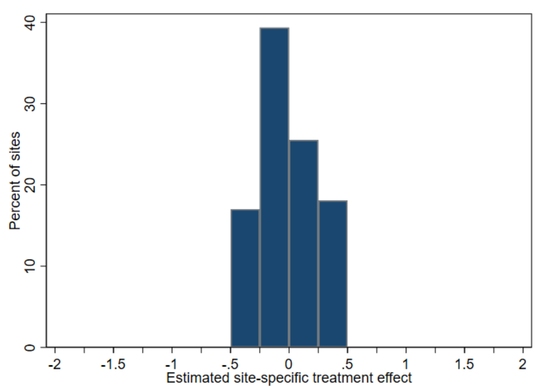
Notes: These graphs show the unadjusted means of the CES depression scale over time for both cohorts across treatment status. The scale is constructed the following way: parental categorical responses for individual items are assigned values (1->0, 2->1, 3->2, 4->3), which are then summed and collapsed into 4 categories: 0-4 = 1, 5-9 = 2, 10-14 = 3, 15-36 = 4. The graphs show the mean across all respondents.

Figure 3.2: Adjusted site-specific treatment effect estimates of offering Head Start on mental health

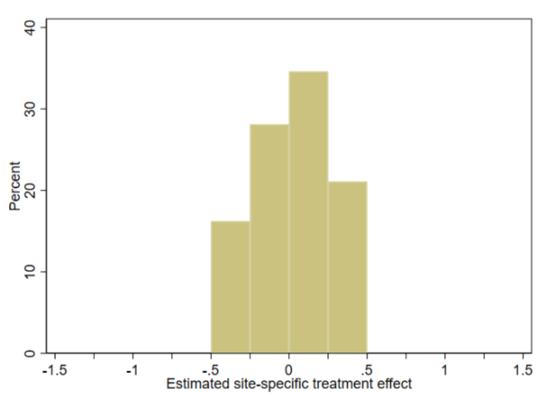
Panel A: Three-year-old cohort, one year after randomization



Panel B: Four-year-old cohort, one year after randomization



Panel C: Four-year-old cohort, two years after randomization



Notes: Histograms show adjusted empirical Bayes estimates of random treatment effects predicted from fixed intercept, random coefficient models for each outcome separately.

Conclusion

This dissertation explored two primary policies that are important tools for societies to support families with young children: parental leave, and early care and education. While there are large bodies of research on the average effects of these policies across a wide range of countries (for reviews see e.g. E. Cascio, 2021; Morrissey, 2017c; Rossin-Slater, 2018), we know much less about how these policies operate within countries by characteristics of families and child care programs. In this dissertation I explored paid leave and early care and education policies in two countries, Germany and the United States, and examined how these policies influence families from different socio-economic backgrounds. The results of these analyses can inform future public investments into families with young children.

Study 1 of this dissertation documented maternal wage dynamics around childbirth in Germany. I documented a 53% decrease in mothers' monthly wages one year after childbirth compared to their earnings in the year right before childbirth. These low post-birth wages were persistent, they were still 23% less 10 years after childbirth than what they were before having children. Women with high and low education levels have similar dynamics, although women with low education experience a smaller drop in their earnings in percentage terms than women with high education. A parental leave policy change in 2007 aimed to increase female labor force attachment (Huebener et al., 2016). It affected high- and low-income women differently by design. Lower-income women were worse off under the new policy as their length of transfer receipt was cut in half. Higher-income women were better off as they received higher transfers for a longer time period. I found that the policy had no effect on high-income mothers' post-birth wages or on the length of leave-taking. They seem to have behaved similarly on the labor market under both regimes. Lower-income women responded to the policy change: they returned to work earlier and had higher post-birth earnings under the new regime.

Even if the response of women with low education was in line with the policy incentives, it is not clear whether the policy made families better or worse off. The decision of how much market work, housework, and child care to provide after children arrive depends on many factors including personal preferences, social norms, as well as institutional, financial, and time constraints. Constraints are easier to measure as it is possible to map the policy environment, see income information in administrative datasets, and there are only 24 hours during a day. Parental preferences about staying home versus returning to the labor force, especially by socio-economic background, are not salient. An imperfect measure of preferences are items in surveys that ask participants about values. Musick et al (2020) report that 59% of German women without a college education agree with the statement that pre-school-aged children likely suffer when their parents work, while only 39% of women do. This indicates that mothers with lower levels of education might have stronger preferences to stay home when their children are young. This fact, along with the reduction of transfer-receipt to lower-income families, may not off-set the increase in transfer amount and the gains in income that result from higher post-birth wages. The policy likely made lower-income families worse off. As higher-income families did not change their behavior as a response to the policy change, but they received on average higher transfers for a longer time period, the policy change most likely made these families better off. Thus, the policy likely increased inequalities between families.

Parents are only able to work outside the home if they have non-parental child care available. Access to early care and education is a concept that often comes up in policy discussions and increasing access to early care and education has been on the policy agenda in many countries, including the United States. Access is a multi-dimensional concept measured by different indicators (Friese et al., 2017). In the second study of this dissertation, I investigated access by using two different indicators: distance traveled for care and the number of child care programs within a geographic area. I showed that families living in the lowest- and the highest-income areas have better access to child care than families living in areas in the middle of the income distribution. These findings bring attention to the fact that next to continued support for the most disadvantaged communities policy efforts should also focus on middle-income families. To illustrate this point further, consider child care costs. Families with a monthly household income less than 200 percent of the poverty line pay 35% of their income on child care on average, but middle-class families also pay 10-14%, which is almost twice as high as deemed "affordable" by the U.S. Department of Health and Human Services (Malik, 2019). Thus, investments into the child care sector should also target middle-class families.

Early care and education is one of the most important contexts for families with young children. Next to the potential of providing a conducive environment for child development, it also contributes to the well-being of the whole family (Phillips et al., 2017; Teti et al., 2017) and is an important part of the social safety net (Small, 2006b). Even when parents have access to child care, everyday logistics and the financial burden associated with paying for it can be significant for families (Malik, 2019; L. Schochet, 2019). Public provision of child care might alleviate some of these stressors and contribute to family well-being. Study 3 of this dissertation tested this hypothesis by using family stress theory and a novel empirical method to explore whether providing free, high-quality pre-school improves parental mental health. We used data from the Head Start Impact Study and did not find any effects on average on parental mental health. We also found no statistically significant differences in impacts of offering Head Start to families on parental mental health across centers. Furthermore, using moderation analyses, we also explored whether center characteristics, like structural factors and family support services provided, predicted variation in treatment effects. None of these characteristics predicted treatment effects.

These null results do not necessarily mean that early childhood education cannot contribute to the well-being of family members other than children. The analytic sample in the study included low-income and marginalized members of the American society. Economic hardship and other stressors may be very salient in families' lives and access to child care may not be enough support to ameliorate parental mental health. Continued investments into early care and education and a focus on aspects of programs that parents need stays a vital policy tool to help families and children to reach their full potential.

References

Adam, E. K., Quinn, M. E., Tavernier, R., McQuillan, M. T., Dahlke, K. A., & Gilbert, K. E. (2017). Diurnal cortisol slopes and mental and physical health outcomes: A systematic review and meta-analysis. *Psychoneuroendocrinology*, *83*(May), 25–41. https://doi.org/10.1016/j.psyneuen.2017.05.018

Administration for Children and Families. (2019). *Head Start Program Facts: Fiscal Year 2019*.

Aguilar-Gomez, S., Arceo-Gomez, E., & De la Cruz Toledo, E. (2019). Inside the black box of child penalties: Unpaid work and household structure. In *SSRN* (Available at SSRN 3497089, Issue 2019). https://doi.org/10.2139/ssrn.3497089

Aikens, N., Cavadel, E., Hartog, J., Hurwitz, F., Knas, E., Schochet, O., Malone, L., & Tarullo, L. (2017). Building Family Partnerships: Family Engagement Findings from the Head Start FACES Study. In *OPRE Report 2017-102. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services*. https://www.proquest.com/scholarly-journals/discerns-special-education-teachers-about-access/docview/2477168620/se-2?accountid=17260%0Ahttp://lenketjener.uit.no/?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Aed

Allen, T. D., Herst, D. E. L., Bruck, C. S., & Sutton, M. (2000). Consequences associated with work-to-family conflict: a review and agenda for future research. *Journal of Occupational Health Psychology*, *5*(2), 278–308. https://doi.org/10.1037/1076-8998.5.2.278

Andresen, M. E., & Nix, E. (2021). *What Causes the Child Penalty? Evidence from Adopting and Same-Sex Couples* (Accepted for Publication by Journal of Labor Economics).

Andresen, M. E., & Nix, E. (2022). Can the Child Penalty Be Reduced? Evaluating Multiple Policy Interventions. *Working Paper*.

Bainbridge, J., Meyers, M. K., Tanaka, S., & Waldfogel, J. (2005). Who gets an early education? Family income and the enrollment of three- to five-year-olds from 1968 to 2000. *Social Science Quarterly*, *86*(3), 724–745. https://doi.org/10.1111/j.0038-4941.2005.00326.x

Baker, M., Gruber, J., & Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, *116*(4), 709–745. https://doi.org/10.1086/591908

Bana, S., Bedard, K., & Rossin-Slater, M. (2018). Trends and Disparities in Leave Use under California's Paid Family Leave Program. *AEA Papers and Proceedings*, *108*(May), 388–391.

Barnett, W. S. (1995). Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes. *The Future of Children*, *5*(3), 25–50. https://doi.org/10.2307/1602367

Bassok, D., Miller, L. C., & Galdo, E. (2016). The effects of universal state pre-kindergarten on the child care sector: The case of Florida's voluntary pre-kindergarten program. *Economics of Education Review*, *53*, 87–98. https://doi.org/10.1016/j.econedurev.2016.05.004

Berge, P. vom, Frodermann, C., Graf, T., Grießemer, S., Kaimer, S., Köhler, M., Lehnert, C., Oertel, M., Schmucker, A., Schneider, A., & Seth, S. (2021). *Weakly anonymous Version of the Sample of Integrated Labour Market Biographies (SIAB) – Version 7519 v1*. Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). https://doi.org/10.5164/IAB.SIAB7519.de.en.v1

Bergemann, A., & Riphahn, R. T. (2022). Maternal employment effects of paid parental leave. In *Journal of Population Economics* (Issue 0123456789). Springer Berlin Heidelberg. https://doi.org/10.1007/s00148-021-00878-7

Bertrand, M., Goldin, C., & Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics, 2*(3), 228–255. https://doi.org/10.1257/app.2.3.228

Bitler, M., Hoynes, H. W., & Domina, T. (2014). Experimental Evidence on Distributional Effects of Head Start. In *National Bureau of Economic Research Working Paper Series* (No. 20434; NBER Working Paper Series). https://doi.org/10.3386/w20434

Bitler, M., Hoynes, H. W., & Whitmore Schanzenbach, D. (2020). The Social Safety Net in the Wake of COVID-19. In *NBER Working Paper* (No. 27796; NBER Working Paper Series). http://www.nber.org/papers/w27796

Blau, D., & Currie, J. (2006). Pre-School, Day care, and After-school care: Who's minding the kids? In *Handbook of the Economics of Education* (Vol. 2, Issue 06). https://doi.org/10.1016/S1574-0692(06)02020-4

Bloom, H. S., Raudenbush, S. W., Weiss, M. J., & Porter, K. (2017). Using Multisite Experiments to Study Cross-Site Variation in Treatment Effects: A Hybrid Approach With Fixed Intercepts and a Random Treatment Coefficient. *Journal of Research on Educational Effectiveness*, *10*(4), 817–842. https://doi.org/10.1080/19345747.2016.1264518

Bloom, H. S., & Weiland, C. (2015). Quantifying Variation in Head Start Effects on Young Children's Cognitive and Socio-Emotional Skills Using Data from the National Head Start Impact Study. *SSRN Electronic Journal*, *March*. https://doi.org/10.2139/ssrn.2594430

Bronfenbrenner, U., & Morris, P. A. (2006). The Bioecological Model of Human Development. In R. M. Lerner (Ed.), *Theoretical models of human development. Handbook of child psychology* (6th ed., pp. 793–828). Hoboken, NJ: Wiley.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *Future of Children*, *7*(2), 55–71. https://doi.org/10.2307/1602387

Brooks-Gunn, J., Markman-Pithers, L., & Rouse, C. E. (2016). Starting early: Introducing the issue. *Future of Children*, *26*(2), 3–20. https://doi.org/10.1353/foc.2016.0009

Brown, J. (2019). *Does Public Pre-K Have Unintended Consequences on the Child Care Market for Infants and Toddlers?* https://doi.org/10.2139/ssrn.3360616

Buckles, K. S., & Hungerman, D. M. (2013). Season of Birth and Later Outcomes: Old Questions, New Answers. *The Review of Economics and Statistics*, *95*(3), 711–724.

Bünning, M. (2015). What happens after the "Daddy Months"? Fathers' involvement in paid work, childcare, and housework after taking parental leave in Germany. *European*

Sociological Review, *31*(6), 738–748. https://doi.org/10.1093/esr/jcv072

Byron, K. (2005). A meta-analytic review of work – family conflict and its antecedents. *Journal of Vocational Behvaior*, *67*, 169–198. https://doi.org/10.1016/j.jvb.2004.08.009

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). Rdrobust: Software for regression-discontinuity designs. *Stata Journal*, *17*(2), 372–404. https://doi.org/10.1177/1536867x1701700208

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. *Stata Journal*, *14*(4), 909–946. https://doi.org/10.1177/1536867x1401400413

Carlin, C., Davis, E. E., Krafft, C., & Tout, K. (2019). Parental preferences and patterns of child care use among low-income families: A Bayesian analysis. *Children and Youth Services Review*, *99*(February), 172–185. https://doi.org/10.1016/j.childyouth.2019.02.006

Cascio, E. (2021). Early Childhood Education in the United States: What, When, Where, Who, How, and Why. In *NBER Working Paper Series*.

Cascio, E. U. (2009). Maternal labor supply and the introduction of kindergartens into American Public Schools. *Journal of Human Resources*, *44*(1), 140–170. https://doi.org/10.3368/jhr.44.1.140

Cascio, E. U. (2021). Early Childhood Education in the United States: What, When, Where, Who, How, and Why. In *NBER Working Paper Series* (No. 28722).

Cascio, E. U., & Schanzenbach, D. (2013). The Impacts of Expanding Access to High-Quality Preschool Education. *Brookings Papers on Economic Activity*, *FALL 2013*, 127–178. https://doi.org/10.1353/eca.2013.0012

Census Bureau. (1994). *Geographic Areas Reference Manual*.

Chase-Lansdale, P. L., & Brooks-Gunn, J. (2014). Two-Generation Programs in the Twenty First Century. *The Future of Children*, *24*(1), 13–39.

Chase-Lansdale, P. L., Sabol, T. J., Sommer, T. E., Chor, E., Cooperman, A. W., Brooks-Gunn, J., Yoshikawa, H., King, C., & Morris, A. (2019). Effects of a two-generation human capital program on low-income parents' education, employment, and psychological wellbeing. *Journal of Family Psychology*, *33*(4), 433–443. https://doi.org/10.1037/fam0000517

Chatterji, P., Markowitz, S., & Brooks-gunn, J. (2013). Effects of early maternal employment on maternal health and well-being. *Journal of Population Economics*, *26*, 285–301. https://doi.org/10.1007/s00148-012-0437-5

Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2017). *Cradle to Kindergarten. A New Plan to Combat Inequality*. Russell Sage Foundation. https://doi.org/10.2105/ajph.38.9.1289-b

Chhaochharia, V., Ghosh, S., Niessen-Ruenzi, A., & Schneider, C. (2020). Child Care Provision and Women's Careers in Firms. In *SSRN Electronic Journal* (No. 2943427). https://doi.org/10.2139/ssrn.2943427

Chor, E. (2018). Multigenerational Head Start Participation: An Unexpected Marker of Progress. *Child Development*, *89*(1), 264–279. https://doi.org/10.1111/cdev.12673

Cochi Ficano, C. K. (2006). Child-care market mechanisms: Does policy affect the quantity of care? *Social Service Review*, *80*(3), 453–484. https://doi.org/10.1086/505447

Collins, C. (2019). *Making Motherhood Work: How Women Manage Careers and Caregiving*. Princeton University Press.

Conger, R. D., & Elder, G. H. (1994). *Families in troubled times: Adapting to change in rural America*. New York: Aldine de Gruyter.

Cortes, P., & Pan, J. (2020). Children and the remaining gender gaps in the labor market. In *National Bureau of Economic Research Working Paper Series* (No. 27980).

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives*, *15*(2), 213–238. https://doi.org/10.1257/jep.15.2.213

Currie, J., & Schwandt, H. (2013). Within-mother analysis of seasonal patterns in health at birth. *Proceedings of the National Academy of Sciences*, *110*(30), 12265–12270. https://doi.org/10.1073/pnas.1307582110

Dauth, W., & Eppelsheimer, J. (2020). Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide. *Journal for Labour Market Research*, *54*(1). https://doi.org/10.1186/s12651-020-00275-9

Davis, E. E., Lee, W. F., & Sojourner, A. (2019). Family-centered measures of access to early care and education. *Early Childhood Research Quarterly*, *47*, 472–486. https://doi.org/10.1016/j.ecresq.2018.08.001

Del Boca, D., Flinn, C., & Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies*, *81*(1), 137–185. https://doi.org/10.1093/restud/rdt026

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, *1*(3), 111–134. https://doi.org/10.1257/app.1.3.111

Dorner, M., Heining, J., Jacobebbinghaus, P., & Seth, S. (2010). The Sample of Integrated Labour Market Biographies. In *Schmollers Jahrbuch* (Vol. 130, Issue 4). https://doi.org/10.3790/schm.130.4.599

Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *The Journal of Economic Perspectives*, *27*(2), 109–131.

Ehlert, N. (2008). *Elterngeld als Teil nachhaltiger Familienpolitik.*

Elder, G. H., & Caspi, A. (1988). Economic Stress in Lives: Developmental Perspectives. *Journal of Social Issues*, *44*(4), 25–45. https://doi.org/10.1111/j.1540-4560.1988.tb02090.x

Fitzpatrick, M. D. (2010). Preschoolers Enrolled and Mothers at Work? The Effects of Universal PreKindergarten. *Journal of Labor Economics*, *28*(1). https://doi.org/10.2139/ssrn.1416029

Forry, N. D., Tout, K., Rothenberg, L., Sandstrom, H., & Vesely, C. (2013). *Child Care*

Decision-Making Literature Review (Issue December).

Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & Gardiner, B. A. (2020). *The State of Pre-School*. http://nieer.org/wp-content/uploads/2020/04/YB2019_Full_Report.pdf

Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & Gardiner, B. A. (2021). The State of Preschool 2020. In *National Institute for Early Education Research*. https://cutt.ly/oiQl6kg

Friese, S., Lin, V.-K., Forry, N., & Tout, K. (2017). *Defining and measuring access to high-quality early care and education (ECE): A guidebook for policymakers and researchers*. *February*, OPRE Report #2017-08. https://www.acf.hhs.gov/opre/research.

Frone, M. R., Yardley, J. K., & Markel, K. S. (1997). Developing and Testing an Integrative Model of the Work–Family Interface. *Journal of Vocational Behavior*, *167*(50), 145–167.

Gelbach, J. B. (2002). Public Schooling for Young Children and Maternal Labor Supply. *American Economic Review*, *92*(1), 307–322. https://doi.org/10.1257/000282802760015748

Gelber, A., & Isen, A. (2013). Children's schooling and parents' behavior: Evidence from the Head Start Impact Study. *Journal of Public Economics*, *101*(1), 25–38. https://doi.org/10.1016/j.jpubeco.2013.02.005

Ginja, R., Jans, J., & Karimi, A. (2020). Parental Leave Benefits, Household Labor Supply, and Children's Long-Run Outcomes. *Journal of Labor Economics*, *38*(1).

Gordon, R. A., & Chase-Lansdale, P. L. (2001). Availability of child care in the United States: A description and analysis of data sources. *Demography*, *38*(2), 299–316. https://doi.org/10.2307/3088307

Greenhaus, J. H., & Beutell, N. J. (2019). Sources of Conflict between Work and Family Roles. *The Academy of Management Review*, *10*(1), 76–88.

Haas, L., & Hwang, C. P. (2008). The impact of taking parental leave on fathers' Participation in childcare and relationships with children: Lessons from Sweden. *Community, Work and Family*, *11*(1), 85–104. https://doi.org/10.1080/13668800701785346

Heckman, J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, *312*(June), 1900–1902. https://doi.org/10.1016/j.adolescence.2005.09.001

Hedges, L., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.

Henly, J. R., & Adams, G. (2018). *Increasing Access to Quality Child Care for Four Priority Populations Challenges and Opportunities with CCDBG Reauthorization* (Issue October). https://www.urban.org/research/publication/increasing-access-quality-child-care-four-priority-populations

Herbst, C. M., Desouza, K. C., Al-Ashri, S., Srivatsav Kandala, S., Khullar, M., & Bajaj, V. (2020). What do parents value in a child care provider? Evidence from Yelp consumer reviews. *Early Childhood Research Quarterly*, *51*, 288–306. https://doi.org/10.1016/j.ecresq.2019.12.008

Herbst, C. M., & Tekin, E. (2010). Child care subsidies and child development. *Economics of Education Review*, *29*(4), 618–638. https://doi.org/10.1016/j.econedurev.2010.01.002

Huber, K. (2019). Changes in parental leave and young children's non-cognitive skills. *Review of Economics of the Household*, *17*(1), 89–119. https://doi.org/10.1007/s11150-017-9380-2

Huebener, M., Kuehnle, D., & Spiess, C. K. (2019). Parental leave policies and socio-economic gaps in child development : Evidence from a substantial benefit reform using administrative data. *Labour Economics*, *61*(August), 101754. https://doi.org/10.1016/j.labeco.2019.101754

Huebener, M., Müller, K., Spieß, C. K., & Wrohlich, K. (2016). The Parental Leave Benefit: A Key Family Policy Measure, One Decade Later. *DIW Economic Bulletin*, *49*, 571–578.

Johnson, A. D., Padilla, C., & Votruba-Drzal, E. (2017). Predictors of Public Early Care and Education Use among Children of Low-Income Immigrants. *Child Youth Services Review*, *73*, 24–36. https://doi.org/10.1016/j.childyouth.2016.11.024.Predictors

Jürges, H. (2017). Financial incentives, timing of births, and infant health: a closer look into the delivery room. *European Journal of Health Economics*, *18*, 195–208. https://doi.org/10.1007/s10198-016-0766-5

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child Penalties across Countries: Evidence and Explanations. *AEA Papers and Proceedings*, *109*, 122–126. https://doi.org/10.1257/pandp.20191078

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2020). Do family policies reduce gender inequality? Evidence from 60 years of policy experimentation. In *NBER Working Paper Series* (Working Paper 28082; NBER Working Paper Series).

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, *11*(4), 181–209. https://doi.org/10.1257/app.20180010

Kluve, J., & Tamm, M. (2013). Parental leave regulations, mothers' labor force attachment and fathers' childcare involvement: Evidence from a natural experiment. *Journal of Population Economics*, *26*(3), 983–1005. https://doi.org/10.1007/s00148-012-0404-1

Kotsadam, A., & Finseraas, H. (2011). The state intervenes in the battle of the sexes : Causal effects of paternity leave. *Social Science Research*, *40*(6), 1611–1622. https://doi.org/10.1016/j.ssresearch.2011.06.011

Lalive, R., Schlosser, A., Steinhauer, A., & Zweimüller, J. (2014). Parental leave and mothers' careers: The relative importance of job protection and cash benefits. *Review of Economic Studies*, *81*(1), 219–265. https://doi.org/10.1093/restud/rdt028

Lalive, R., & Zweimüller, J. (2009). How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments. *Quarterly Journal of Economics*, *124*(3), 1363–1402.

Laughlin, L. (2013). Who's Minding the Kids ? Child Care Arrangements : Spring 2011 Household Economic Studies. *Current Population Reports, U.S. Census Bureau*, *2009*(April), P70-135.

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281

Ludwig, J., & Miller, D. L. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics*, *122*(1), 159–208. https://doi.org/10.1162/qjec.122.1.159

Ludwig, J., & Phillips, D. A. (2008). Long-term effects of head start on low-income children. *Annals of the New York Academy of Sciences*, *1136*, 257–268. https://doi.org/10.1196/annals.1425.005

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, *10*(6), 434–445. https://doi.org/10.1038/nrn2639

Magnuson, K., & Waldfogel, J. (2016). Trends in Income-Related Gaps in Enrollment in Early Childhood Education. *AERA Open*, *2*(2), 233285841664893. https://doi.org/10.1177/2332858416648933

Malik, R. (2019). *Working Families Are Spending Big Money on Child Care*. https://cdn.americanprogress.org/content/uploads/2019/06/19074131/Working-Families-SpendingBRIEF.pdf?_ga=2.61444580.977246131.1619902055-1350100502.1618780341

Malik, R., Hamm, K., Lee, W. F., Davis, E. E., & Sojourner, A. (2020). *The Coronavirus Will Make Child Care Deserts Worse and Exacerbate Inequality*. www.childcaredeserts.org

Malik, R., Hamm, K., Schochet, L., Novoa, C., Workman, S., & Jessen-Howard, S. (2018). America's Child Care Deserts in 2018. In *Center for American Progress* (Issue December).

Manson, S., Schroeder, J., Van Riper, D., Kugler, T., & Ruggles, S. (2021). *IPUMS National Historical Geographic Information System: Version 16.0 [dataset]*. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V16.0

McEwen, B. S. (2003). Interacting mediators of allostasis and allostatic load: towards an understanding of resilience in aging. *Metabolism*, *52*(10), 10–16. https://doi.org/10.1053/s0026-0495(03)00295-6

Morris, P. A., Connors, M., Friedman-Krauss, A., McCoy, D. C., Weiland, C., Feller, A., Page, L., Bloom, H., & Yoshikawa, H. (2018). New Findings on Impact Variation From the Head Start Impact Study: Informing the Scale-Up of Early Childhood Programs. *AERA Open*, *4*(2), 233285841876928. https://doi.org/10.1177/2332858418769287

Morrissey, T. W. (2017a). Child care and parent labor force participation: a review of the research literature. *Review of Economics of the Household*, *15*(1), 1–24. https://doi.org/10.1007/s11150-016-9331-3

Morrissey, T. W. (2017b). Child care and parent labor force participation: a review of the research literature. *Review of Economics of the Household*, *15*(1), 1–24. https://doi.org/10.1007/s11150-016-9331-3

Morrissey, T. W. (2017c). Child care and parent labor force participation: a review of the research literature. *Review of Economics of the Household*, *15*(1), 1–24.

https://doi.org/10.1007/s11150-016-9331-3

Müller, D., Filser, A., & Frodermann, C. (2022). *Update: Identifying mothers in administrative data*.

Musick, K., Bea, M. D., & Gonalons-Pons, P. (2020). His and Her Earnings Following Parenthood in the United States, Germany, and the United Kingdom. *American Sociological Review*, *85*(4), 639–674. https://doi.org/10.1177/0003122420934430

National Survey of Early Care and Education Project Team. (2016a). *Fact Sheet: How Far are Early Care and Education Arrangements from Children's Homes?* (Issue OPRE Report No. 2016-10). https://www.acf.hhs.gov/opre/project/national-survey-early-care-and-education-nsece-2012

National Survey of Early Care and Education Project Team. (2016b). *Households' Geographic Access to Center-based Early Care and Education: Estimates and Methodology from the National Survey of Early Care and Education* (OPRE Report #2016-08; Washington DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.). http://www.acf.hhs.gov/sites/default/files/opre/hh_geoaccessto_cb_ece_toopre_042916_b508.pdf

Nepomnyaschy, L., & Waldfogel, J. (2007). Paternity leave and fathers' involvement with their young children. *Community, Work and Family*, *10*(4), 427–453. https://doi.org/10.1080/13668800701575077

Neugart, M., & Ohlsson, H. (2013). Economic incentives and the timing of births: Evidence from the German parental benefit reform of 2007. *Journal of Population Economics*, *26*, 87–108. https://doi.org/10.1007/s00148-012-0420-1

OECD. (2019). *Education at a Glance 2019: OECD Indicators*. OECD Publishing, Paris. https://doi.org/https://doi.org/10.1787/f8d7880d-en

OECD. (2020). *OECD Family database: Maternal employment*. OECD - Social Policy Division - Directorate of Employment, Labour and Social Affairs.

OECD. (2021). *OECD Family Database: Public policies for families and children*. https://www.oecd.org/els/family/database.htm

OECD. (2022). *Family benefits public spending (indicator)*. https://doi.org/10.1787/8e8b3273-en

Olivetti, C., & Petrongolo, B. (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics*, *8*(1), 405–434. https://doi.org/10.1146/annurev-economics-080614-115329

Olivetti, C., & Petrongolo, B. (2017). The economic consequences of family policies: Lessons from a century of legislation in high-income countries. *Journal of Economic Perspectives*, *31*(1), 205–230. https://doi.org/10.1257/jep.31.1.205

Padilla, C. M. (2020). Beyond Just the Average Effect: Variation in Head Start Treatment Effects on Parenting Behavior. *AERA Open*, *6*(4), 233285842096969. https://doi.org/10.1177/2332858420969691

Persson, P., & Rossin-Slater, M. (2021). When Dad Can Stay Home: Fathers' Workplace Flexibility and Maternal Health. *SSRN Electronic Journal*, *1752203*. https://doi.org/10.2139/ssrn.3401154

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., Duncan, G. J., Dynarski, M., Magnuson, K., & Weiland, C. (2017). *Puzzling It Out: The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf

Puma, M., Bell, S., Cook, R., & Heid, C. (2010a). *Head Start Impact Study. Final Report.* http://files.eric.ed.gov/fulltext/ED507845.pdf

Puma, M., Bell, S., Cook, R., & Heid, C. (2010b). *Head Start Impact Study. Technical Report.*

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). Third Grade Follow-up to the Head Start Impact Study Final Report. In *OPRE Report # 2012-45* (Issue October).

Reardon, S. F., & Stuart, E. A. (2017). Editors' Introduction: Theme Issue on Variation in Treatment Effects. *Journal of Research on Educational Effectiveness*, *10*(4), 671–674. https://doi.org/10.1080/19345747.2017.1386037

Rossin-Slater, M. (2018). Maternity and family leave policy. In S. L. Averett, L. M. Argys, & S. D. Hoffman (Eds.), *The Oxford Handbook of Women and the Economy*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190628963.013.23

Rossin-Slater, M., & Uniat, L. (2019). Paid Family Leave Policies and Population Health. *Health Affairs Health Policy Brief*, *march*. https://doi.org/10.1377/hpb20190301.484936

Sabol, T. J., & Chase-Lansdale, P. L. (2015). The Influence of Low-Income Children's Participation in Head Start on Their Parents' Education and Employment. *Journal of Policy Analysis and Management*, *34*(1), 136–161. https://doi.org/10.1002/pam.21799

Sabol, T. J., Eckrich Sommer, T., Chase-Lansdale, P. L., & Brooks-Gunn, J. (2021). Intergenerational Economic Mobility for Low-Income Parents and Their Children: A Dual Developmental Science Framework. *Annual Review of Psychology*, *72*(20). https://doi.org/10.2307/1418941

Sabol, T. J., McCoy, D., Gonzalez, K., Miratrix, L., Hedges, L., Spybrook, J. K., & Weiland, C. (2022). Exploring treatment impact heterogeneity across sites: Challenges and opportunities for early childhood researchers. *Early Childhood Research Quarterly*, *58*, 14–26. https://doi.org/10.1016/j.ecresq.2021.07.005

Sabol, T. J., Sommer, T. E., Sanchez, A., & Busby, A. K. (2018). A New Approach to Defining and Measuring Family Engagement in Early Childhood Education Programs. *AERA Open*, *4*(3), 233285841878590. https://doi.org/10.1177/2332858418785904

Sandler, D., & Szembrot, N. (2019). Maternal Labor Dynamics: Participation, Earnings, and Employer Changes. In *CES Working Papers*.

Sapolsky, R. M. (2004). *Why zebras don't get ulcers: The acclaimed guide to stress, stress-*

related diseases, and coping. St. Martin's Press, New York.

Schetter, C. D., & Dolbier, C. (2011). Resilience in the Context of Chronic Stress and Health in Adults. *Social and Personality Psychology Compass*, *5*(9), 634–652. https://doi.org/10.1111/j.1751-9004.2011.00379.x

Schochet, L. (2019). The Child Care Crisis Is Keeping Women Out of the Workforce. In *Center for American Progress* (Issue March). https://cdn.americanprogress.org/content/uploads/2019/03/19103744/ECPP-ChildCare-Crisis-report-2.pdf

Schochet, O. N., & Padilla, C. M. (2021). Children Learning and Parents Earning: Exploring the Average and Heterogeneous Effects of Head Start on Parental Earnings. *Journal of Research on Educational Effectiveness*, *0*(0), 1–32. https://doi.org/10.1080/19345747.2021.2005202

Schönberg, U., & Ludsteck, J. (2014). Expansions in maternity leave coverage and mothers' labor market outcomes after childbirth. *Journal of Labor Economics*, *32*(3), 469–505. https://doi.org/10.1086/675078

Seeman, T., Epel, E., Gruenewald, T., Karlamangla, A., & Mcewen, B. S. (2010). Socio-economic differentials in peripheral biology: Cumulative allostatic load. *Annals of the New York Academy of Sciences*, *1186*, 223–239. https://doi.org/10.1111/j.1749-6632.2009.05341.x

Small, M. L. (2006a). Neighborhood institutions as resource brokers: Childcare centers, interorganizational ties, and resource access among the poor. *Social Problems*, *53*(2), 274–292. https://doi.org/10.1525/sp.2006.53.2.274

Small, M. L. (2006b). Neighborhood Institutions as Resource Brokers: Childcare Centers, Interorganizational Ties, and Resource Access among the Poor. *Social Problems*, *53*(2), 274–292. https://doi.org/10.1525/sp.2006.53.2.274

Snyder, K., Dore, T., & Adelman, S. (2005). *Use of Relative Care by Working Parents* (Issue 23). https://www.urban.org/sites/default/files/publication/51546/311161-Use-of-Relative-Care-by-Working-Parents.PDF

Tamm, M. (2013). The impact of a large parental leave benefit reform on the timing of birth around the day of implementation. *Oxford Bulletin of Economics and Statistics*, *75*(4), 585–601. https://doi.org/10.1111/j.1468-0084.2012.00707.x

Tamm, M. (2019). Fathers' parental leave-taking, childcare involvement and labor market participation. *Labour Economics*, *59*(April), 184–197. https://doi.org/10.1016/j.labeco.2019.04.007

Tanaka, S., & Waldfogel, J. (2007). Effects of parental leave and work hours on fathers' involvement with their babies. *Community, Work and Family*, *10*(4), 409–426. https://doi.org/10.1080/13668800701575069

Teti, D. M., Cole, P. M., Cabrera, N., Goodman, S. H., & McLoyd, V. C. (2017). Supporting Parents: How Six Decades of Parenting Research Can Inform Policy and Best Practice. *Social Policy Report*, *30*(5), 1–34. https://doi.org/10.1002/j.2379-3988.2017.tb00090.x

Torun, H., & Tumen, S. (2017). The empirical content of season-of-birth effects: An investigation with Turkish data. *Demographic Research*, *37*(57), 1825–1860. https://doi.org/10.4054/DemRes.2017.37.57

U.S. Department of Health and Human Services Administration for Children and Families Office of Head Start National Center on Parent Family and Community Engagement. (2018). *Head Start Parent, Family, and Community Engagement Framework*.

Vinovskis, M. A. (2008). *The birth of Head Start: Preschool education policies in the Kennedy and Johnson administrations.* Chicago, IL: University of Chicago Press.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills. *Child Development*, *84*(6), 2112–2130.

Wikle, J., & Wilson, R. (2021). *Access to Head Start and Maternal Labor Supply: Experimental and Quasi- Experimental Evidence*.

Yeung, W. J., Linver, M. R., & Brooks-Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child Development*, *73*(6), 1861–1879. https://doi.org/10.1111/1467-8624.t01-1-00511

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L., Gormley, W. T., Ludwig, J., Magnuson, K., Phillips, D., & Zaslow, M. (2013). *Investing in our future: The evidence base on preschool education.*