Aalto-yliopisto
Aalto-universitetet
Aalto University

Master's Programme in Economics

Impacts of Universal Early Education on Children's Outcomes

A Review of Empirical Economics Literature

Minka Melin

Master's thesis
2023

Copyright ©2023 Minka Melin

Aalto-yliopisto
Aalto-universitetet
Aalto University

Author Minka Melin

Title of thesis Impacts of Universal Early Education on Children's Outcomes

Programme Master's Programme in Economics

Thesis supervisor Prof. Kristiina Huttunen

Date 23.12.2023 **Number of pages** 59+2 **Language** English

Abstract

Early childhood education is high on the policy agenda in many countries, with growing interest in making subsidized child care universally accessible to children from all backgrounds. Proponents see universal early education as an equitable means to promote child development and support parental employment, but the costs are high and may exceed the benefits for advantaged children. This thesis reviews empirical economics literature aiming to estimate causal impacts of universal child care programs on children's outcomes.

Findings from programs in the United States, Canada, Norway, Denmark, and Germany indicate that universal child care can substantially affect children's development and long-term socioeconomic success, either in the positive or negative direction. Significant results are found for behavioral and cognitive development, health, educational attainment, earnings, and crime. Variation in the effects across programs could be explained by differences in children's background characteristics, in program type and quality, and in the counterfactual modes of care. Overall, universal early education appears to benefit disadvantaged children the most, pointing to its potentially equalizing role. Moreover, high-quality center-based care appears to show greater promise in improving children's long-run prospects than family day care, especially when it displaces informal non-parental care as opposed to home care.

Keywords universal early childhood education, child care, child development, human skill formation, labor economics

Aalto-yliopisto
Aalto-universitetet
Aalto University

Tekijä Minka Melin

Työn nimi Universaalin varhaiskasvatuksen vaikutukset lasten tulemiin

Koulutusohjelma Taloustieteen maisteriohjelma

Vastuuopettaja/valvoja Prof. Kristiina Huttunen

Päivämäärä 23.12.2023 **Sivumäärä** 59+2 **Kieli** Englanti

Tiivistelmä

Varhaiskasvatus on nouseva prioriteetti monissa kehittyneissä maissa, ja kasvava mielenkiinto kohdistuu julkisilla varoilla tuetun päivähoidon laajentamiseen universaaliksi, eli suunnatuksi kaikille lapsille perheen sosioekonomiseen taustaan katsomatta. Universaalin varhaiskasvatuksen puolestapuhujat näkevät sen tasa-arvoa edistävänä tapana edistää lasten kehitystä ja tukea vanhempien työssäkäyntiä, mutta vasta-argumenttina painavat korkeat kustannukset, jotka saattavat ylittää hyödyt parempiosaisten lasten osalta. Tämä tutkielma pyrkii kirjallisuuskatsauksella selvittämään, mitä taloustieteellinen empiirinen tutkimus kertoo universaalien varhaiskasvatusohjelmien vaikutuksista lasten lopputulemiin.

Löydökset Yhdysvalloista, Kanadasta, Norjasta, Tanskasta ja Saksasta osoittavat, että universaali varhaiskasvatus voi merkittävästi vaikuttaa lasten kehitykseen ja pitkän aikavälin sosioekonomiseen menestykseen, joko negatiivisesti tai positiivisesti. Merkittäviä vaikutuksia havaitaan behavioraaliseen ja kognitiiviseen kehitykseen, terveyteen, koulutukseen, tuloihin sekä rikollisuuteen. Vaihtelu eri ohjelmien vaikutuksissa voi selittyä eroilla lasten taustoissa, hoidon laadussa ja tyypissä sekä vaihtoehtoisessa lastenhoitomuodossa. Yleisesti ottaen heikommassa asemassa olevat lapset vaikuttavat hyötyvän osallistumisesta eniten, mikä viittaa universaalin lastenhoidon tasapäistävään potentiaaliin. Lisäksi korkealaatuinen päiväkotimuotoinen varhaiskasvatus vaikuttaa perhepäivähoitoa lupaavammalta keinolta tukea lasten pitkän aikavälin menestymisen edellytyksiä, erityisesti sen korvatessa epämuodollisia hoitojärjestelyjä kotihoidon sijaan.

Avainsanat universaali varhaiskasvatus, lastenhoito, lasten kehitys, taitojen muodostuminen, työn taloustiede

Table of contents

1 Introduction 7
2 Theoretical Framework 11
3 Review of Empirical Literature 15
3.1 Impacts of Boston's Public Preschool Program 16
3.1.1 Institutional Setting 16
3.1.2 Data and Outcomes Measured 18
3.1.3 Identification Strategy 19
3.1.4 Results 20
3.2 Impacts of Quebec's Universal Child Care Program 22
3.2.1 Institutional Setting 23
3.2.2 Data and Outcomes Measured 25
3.2.3 Identification Strategies 26
3.2.4 Short-Term Results 28
3.2.5 Long-Term Results 29
3.2.6 Distributional Results 30
3.3 Impacts of Norway's Universal Child Care Reform 32
3.3.1 Institutional Setting 32
3.3.2 Data and Outcomes Measured 33
3.3.3 Identification Strategies 34
3.3.4 Results 37
3.4 Impacts of Early Child Care in Norway 40
3.4.1 Institutional Setting 40
3.4.2 Data and Outcomes Measured 41
3.4.3 Identification Strategy 42
3.4.4 Results 43
3.5 Impacts of Denmark's Universal Preschool and Family Day Care 45
3.5.1 Institutional Setting 45
3.5.2 Data and Outcomes Measured 47
3.5.3 Identification Strategy 48
3.5.4 Results 51
3.6 Impacts of Germany's Universal Child Care Program 52
3.6.1 Institutional Setting 53
3.6.2 Data and Outcomes Measured 54

3.6.3 Identification Strategy 55
3.6.4 Results 56
4 Conclusions 60
References 66

1 Introduction

The very first years of children's lives critically influence many later-life outcomes. A vast multidisciplinary literature documents that multiple cognitive and noncognitive abilities play a vital role in explaining how well individuals succeed in life; these skills are both inherited and created over the life cycle, determined by the interplay of genes and environments. The influence of early childhood conditions is particularly potent, and skill gaps across different socioeconomic groups emerge already before schooling age (Heckman & Mosso, 2014). For a growing body of children across the world, the early environment comprises not only the home but also formal child care. Public interest in early childhood education has spread widely over the recent decade or so, inspired by accumulating evidence of benefits to children's human capital development. At the same time, women's rising labor market participation has contributed to a growing demand for out-of-home child care arrangements. Indeed, investment in early education programs has been an increasing priority in many EU member countries (European Commission et al., 2018). In fact, the European Council has set a target of reaching a 96 percent early education participation rate for children aged three years or above by 2030, and the EU-27 average was already nearly 93 percent in 2019 (Chircop & Claros Gimeno, 2021). Policymakers' attention has been increasingly drawn to a universal model of early education, where eligibility to publicly subsidized care extends to all children regardless of socioeconomic background. While prevalent in Scandinavia, many developed countries are considering shifting from income-targeted to widely accessible early education programs (Havnes & Mogstad, 2015). Yet, evidence on the effects of universal child care is mixed and rather sparse, leaving many advocates to rely on findings from studies on smaller-scale programs targeted at disadvantaged children instead. This thesis aims to shed light on the impacts of universal early education on children's outcomes by reviewing the economics literature currently available.

While the literature on the effects of early childhood education spans multiple academic fields such as psychology and neuroscience, the topic has attracted considerable attention within economics in its own right. On the theoretical side, economists have focused on the technology of human skill formation (Heckman & Mosso, 2014). Cunha et al. (2006) incorporate the existence of sensitive and critical periods of child development (limited time windows in life where certain skills are malleable) into a model of multi-staged childhood, where human capital investments exhibit self-productivity and dynamic complementarity.

That is, skills acquired earlier promote skill attainment in later stages and improve the productivity of later investments. The dynamic nature of human capital accumulation explains why investing early has the highest returns and no equity-efficiency tradeoff, as opposed to later investments: once skills have become settled, the economic returns are highest for investing in the most capable rather than disadvantaged individuals. Clearly, early childhood education is then important both on the grounds of fairness and economic efficiency.

Consistent with theory, a substantial empirical literature has established the striking positive influence of small-scale "model" programs targeted at particularly disadvantaged children in the United States in the 1960s and 70s, such as the Perry Preschool and Abecedarian programs. As reviewed by Duncan & Magnuson (2013), empirical studies document that these intensive interventions improved children's IQ and achievement test scores in the short run and had long-lasting impacts on later outcomes like educational achievement, teen parenthood, employment, earnings, and crime, which were reflected in remarkable benefit-cost ratios of these programs. An adjacent branch of research explores the impacts of large-scale targeted programs. Scholars have mainly focused on Head Start, the largest early education program in the US that began in 1965 as part of the War on Poverty and presently enrolls more than a million low-income children each year. The evidence on the effects of Head Start remains inconclusive, with a randomized evaluation from 2002 (the Head Start Impact Study) pointing to initial cognitive gains that dissipated quickly, but many quasi-experimental studies indicating longer-run improvements in socioeconomic outcomes (Cascio & Schanzenbach, 2013; Bailey et al., 2021). Overall, the pattern of results from the literature on programs targeted at low-income children can be summarized as a medium-term fadeout of strong immediate gains in cognitive development, followed by significant beneficial impacts on adult outcomes (Cascio & Schanzenbach, 2013). The perplexing finding of disappearing effects on cognitive test scores accompanied by persistent impacts on later-life outcomes has led researchers to speculate that the long-term effects operate through "noncognitive" channels, by which economists typically mean personality traits and behavioral attributes (Duncan & Magnuson, 2013). For instance, Heckman et al. (2021) demonstrate that lasting improvements in personality skills played a significant role in the long-run effects of the Perry Preschool program. Nevertheless, findings from targeted early interventions offer only limited insight when it comes to the effectiveness of more widely accessible programs. Uncovering the impacts of universal early education requires analyzing such programs operating at scale, precisely because they serve a much larger and more

diverse population of children (Cascio, 2015). Compared to the bulk of evidence on income-targeted programs, the body of research on the impacts of universal child care is much smaller but growing.

In this thesis, I seek to determine what the empirical literature currently says about the causal impacts of universally accessible early education on children's outcomes. Specifically, I narrow the focus to papers published in reputable economics journals. The question about how universal early education affects children's development and later-life outcomes holds clear policy relevance, not least because such programs are heavily subsidized. While proponents argue that universal child care can generate equitable gains to children's development, the costs may exceed the benefits for children from more advantaged backgrounds (Cascio, 2015). By synthesizing findings on the impacts of these programs on children, I aim to contribute an important piece of the puzzle to guide optimal policy. The nine articles within the scope of this review provide insight into five universal child care regimes spanning North America and Europe, in Boston, Quebec, Norway, Denmark, and Germany.

Overall, the evidence indicates that universal early education can significantly influence children's behavioral and cognitive development, health, educational attainment, earnings, and criminality. The effects range from positive to negative and vary across different groups of children. First, Gray-Lobe et al. (2023) report that participation in Boston's public preschool program had significant positive impacts on children's long-term educational outcomes, like college enrollment, high school graduation, and adolescent disciplinary behavior. By contrast, Baker et al. (2008, 2019) demonstrate that the large-scale expansion of subsidized child care in the Canadian province of Quebec led to immediate deterioration in children's noncognitive development (namely, increased anxiety and aggression) and health, with the behavioral impacts persisting into elementary school years. Exposed cohorts also had poorer health and higher rates of crime in adolescence. However, Kottelenberg & Lehrer (2017) uncover that the negative mean effect hides the fact that some children did benefit from the Quebec reform, particularly disadvantaged children in single-parent households. Meanwhile, Havnes & Mogstad (2011, 2015) discover that the introduction of universal child care in Norway positively impacted children's long-run educational and labor market outcomes, with equalizing effects on the earnings distribution. Analyzing how attending Norway's public child care affects younger toddlers in more recent cohorts, Drange & Havnes (2019) report positive effects on language and mathematics

performance at school start. As for Denmark, Gupta & Simonsen (2010) document that center-based preschool led to similar behavioral outcomes at school starting age as home care, but family day care had a significant negative impact. Finally, Cornelissen et al. (2018) find that disadvantaged children with most to gain from Germany's universal child care in terms of school readiness are the least likely to enroll.

As can be seen, the effects of universal child care vary drastically across programs, from decidedly negative in Quebec to highly positive in Boston and Norway. Moreover, there is variation in the impacts across individuals within the same program. Understanding what explains the differences and who benefits from universal child care would be of great interest to policymakers and parents alike. Based on synthesis of the evidence, the effects of a given policy appear to depend on the children's background, the type and quality of the program, and the counterfactual mode of care. In general, the programs tend to be more beneficial for disadvantaged children, such as those from single-parent, low-income or minority families, which points to an equalizing role of universal child care. Furthermore, high-quality center-based care seems to outperform family day care, and the effects may be more positive when the programs displace informal non-parental care as opposed to parental home care. Nevertheless, the evidence base is yet too small to reach strong verdicts. While the heterogeneity in the results caution against generalizing findings from single universal programs, let alone income-targeted ones, policymakers should note that publicly subsidized child care can significantly affect children's development, health, educational attainment, earnings, and crime. Indeed, universally accessible early education has the potential to influence the life trajectories of large numbers of individuals, at best contributing to greater long-term success and wellbeing.

2 Theoretical Framework

Recent theoretical work in the economics of human development and social mobility focuses on modeling how skills develop over the life cycle (Heckman & Mosso, 2014). Cunha & Heckman (2007) put forward a model of skill formation that captures the dynamic complementarities of human capital investments, proposing that "skills beget skills" through multiplier effects. Notably, the model highlights the importance of early investment: investing in early childhood years has the highest returns and escapes the equity-efficiency trade-off that characterizes later investments.

Building upon a foundation of multidisciplinary empirical evidence, Cunha & Heckman (2007) formalize important features of child development into a simple economic model. First, the authors recognize that a multitude of skills contribute to socioeconomic success, and both cognitive and noncognitive abilities matter. These skills are both inherited and created, influenced by genes and environmental factors alike in an interactive manner. Moreover, gaps in these abilities between individuals and socioeconomic groups emerge at early ages, with high correlations between child skill levels and family background factors. For example, children from disadvantaged family backgrounds are exposed to a much more limited vocabulary and subjected to parenting styles less conducive to learning and development (Heckman & Mosso, 2014). Indeed, Cunha et al. (2006) emphasize that when it comes to gaps in attaining higher education, "the empirically important market failure" is not a parental credit constraint at the time of college enrollment but rather "the inability of children to buy their parents or the lifetime resources that parents provide" (p. 703). While explaining educational choices to a large extent, permanent family income acts as a proxy for the greater investment in skills that wealthier families tend to provide for their children (Cunha et al., 2006).

Second, skills differ in their plasticity over the course of life. The notion of sensitive periods of child development refer to limited time frames in human life where experiences alter and stabilize certain neural pathways; in critical periods, such experiences are essential for typical development. When the windows of opportunity for the development of given abilities are missed, remediation can be very costly and full recovery impossible. Moreover, empirical literature on interventions aimed at disadvantaged children points to low economic returns to remedial investments in adolescence, but high returns to early interventions. Late investments have higher returns to the most able rather than disadvantaged individuals.

Indeed, Cunha & Heckman (2007) argue that from an economic perspective, the definition of critical and sensitive periods should consider not only technical possibilities but also the costs and returns of remediation of early deficits. The authors introduce the concepts of "self-productivity" and "complementarity" to represent the dynamic nature of human capital accumulation, where abilities attained in earlier life stages promote later skill development. That is, self-productivity conveys the idea that skills persist over time and reinforce one another. Complementarity means that early human capital investment improves the productivity of later investments, but also needs to be followed by subsequent investment in order to be productive. Given these points, Cunha & Heckman (2007) develop a skill formation technology that accounts for multiple stages of childhood, where inputs at different periods are complements, and where investment exhibits self-productivity.

In Cunha & Heckman's (2007) formulation, agents are said to possess a vector of skills at each age. These abilities are then used with various weights in different occupations and other tasks of adult life. A multistage technology governs the production of these skills, such that investments and abilities in each life stage produce outputs (levels of the skills) in the next. Specifically, vector θ_t represents the skill stocks at each stage t, with θ_1 signifying the initial conditions a child is born with. The production function for skills is defined recursively as $\theta_{t+1} = f_t(h, \theta_t, I_t)$, for $t = 1,2, \dots, T$, where h denotes parental characteristics and I_t investments in the child's skills. The authors assume that f_t is strictly increasing and concave in I_t, and twice continuously differentiable. Furthermore, the stock of skills can be rewritten as a function of all past investments: $\theta_{t+1} = m_t(h, \theta_1, I_1, \dots, I_t), t = 1, \dots, T$. The technologies and the used inputs may be different in each stage. Moreover, the stages may differ in their effectiveness in producing some skills, and the productivity of the inputs may vary across stages. Sensitive periods are incorporated into the model as stages that are more productive in generating certain skills; if only a single stage is effective in producing a given skill, it is a critical period. To be specific, period t^* is a critical period for θ_{t+1} when investments in θ_{t+1} are productive only in that period, as captured by the assumption:

$$\frac{\partial \theta_{t+1}}{\partial I_s} = \frac{\partial m_t(h, \theta_1, I_1, \dots, I_t)}{\partial I_s} = 0 \text{ for all } h, \theta_1, I_1, \dots, I_t, s \neq t^*,$$

but

$$\frac{\partial \theta_{t+1}}{\partial I_{t^*}} = \frac{\partial m_t(h, \theta_1, I_1, \dots, I_t)}{\partial I_{t^*}} > 0 \text{ for some } h, \theta_1, I_1, \dots, I_t.$$

Moreover, period t^* is sensitive relative to period s when the productivity of investment is higher in stage t^* than in another stage $s \neq t^*$, at the same level of inputs:

$$\frac{\partial \theta_{t+1}}{\partial I_s}\bigg| h = \bar{h}, \theta_1 = \theta, I_1 = i_1, \dots, I_t = i_t \quad < \quad \frac{\partial \theta_{t+1}}{\partial I_{t^*}}\bigg| h = \bar{h}, \theta_1 = \theta, I_1 = i_1, \dots, I_t = i_t$$

The concept of self-productivity is captured by $\frac{\partial f_t(h,\theta_t,I_t)}{\partial \theta_t} > 0$, meaning that a higher skill level in one period elevates the skill stock in the next one. Furthermore, dynamic complementary arises when $\frac{\partial^2 f_t(h,\theta_t,I_t)}{\partial \theta_t \partial I_t'} > 0$, which implies that skills acquired in the prior stage increase the productivity of subsequent investment.

Crucially, the structure of the technology dictates whether later remediation of early deficits in investments is possible. For simplicity, the authors suppose that the life cycle consists of only two stages of childhood and a third, final adulthood stage. In this case, the adult skill stock h' $(= \theta_3)$ is determined by the function of parental characteristics, initial conditions, and childhood investments in two periods: $h' = m_2(h, \theta_1, I_1, I_2)$. For now, θ_1, I_1, I_2 are assumed to be scalars. The critical question is whether inputs at childhood stages 1 and 2, or early and late investments, are substitutes or complements. To illustrate a case where the adult skill stock does not depend on the timing of childhood investments, the technology is defined as $h' = m_2(h, \theta_1, \gamma I_1 + (1-\gamma) I_2)$, where $\gamma = \frac{1}{2}$. Here, investments I_1 and I_2 are perfect substitutes, and adult abilities are determined only by the total amount of investment. In contrast, the opposite situation of perfect complementarity would be characterized by a technology $h' = m_2(h, \theta_1, \min\{I_1, I_2\})$. Now the temporal distribution of investments is critical: it only pays off to invest in both periods. More generally, skill formation can be formulated as a standard CES technology: $h' = m_2\left(h, \theta_1, \left[\gamma (I_1)^{\phi} + (1-\gamma)(I_2)^{\phi}\right]^{1/\phi}\right)$, for $\phi \leq 1$ and $0 \leq \gamma \leq 1$. The authors refer to parameter γ as a "skill multiplier," which captures how early investment I_1 not only increases h' directly (self-productivity) but also raises the productivity of later investment I_2 by increasing θ_2 (direct complementarity). The elasticity of substitution $\frac{1}{1-\phi}$ measures how easy it is to substitute between first and second period investments. Furthermore, the parameter ϕ determines how substitutable early and late investments are in producing human capital: that is, how easy it is to compensate later for low levels of early investments and skills. Small ϕ signifies that later investment I_2 cannot easily remediate deficiencies in early investment I_1, but also that high

early investment needs to be followed by high late investment for the early investment to pay off. As argued by the authors, the empirical evidence supports the complementarity of inputs in different stages of childhood.

The dynamic nature of human skill formation has important implications for the optimal distribution of human capital investments over the life cycle and across individuals. Summarizing the central finding from the empirical literature, Cunha et al. (2006) maintain that the rate of return to an equal investment is higher at younger than later ages, other things being equal. Optimal investment equates the marginal rate of return with the opportunity cost of funds in each period and for each person. The skill multiplier emphasizes the importance of early investment: self-productivity and complementarity of the skill formation technology imply that later investment cannot compensate for prior deficits, meaning that economically efficient investments perpetuate early inequalities stemming from disadvantaged childhood environments. This equity-efficiency tradeoff is not present at earlier ages. (Cunha et al., 2006). In conclusion, it is clearly important to invest in children's human capital when they are young, so as to provide a basis for the multiplier process where "skills beget skills." Public investments in early education therefore have a strong empirical and theoretical backing.

3 Review of Empirical Literature

This section reviews the empirical literature on the effects of universal early education, organized by the program under study. For each program, I discuss the institutional setting (including the program's quality and counterfactual modes of care), the data and outcomes measured, the quasi-experimental approach, and the results. Table 1 summarizes the papers.

Table 1. Summary of the reviewed papers

Paper	Program	Age at program	Age at follow-up	Identification strategy	Results
Gray-Lobe et al. (2023)	Boston, USA	Age 4	Up to early 20s	Randomized instrumental variable (IV) design, based on admission lotteries	Positive impacts on college enrollment, high school graduation, SAT taking, and adolescent disciplinary behavior; larger effects for boys. No effects on state achievement test performance
Baker et al. (2008)	Quebec, Canada	Ages 0-4	Ages 0-4	Difference-in-differences	Negative short-term effects on behavior (anxiety and aggression), motor and social development, and health. No impact on cognitive test performance
Baker et al. (2019)	Quebec, Canada	Ages 0-4	Up to age 20	Difference-in-differences	Persistence of the negative shock on childhood behavior into school years, negative effects on adolescent health, life satisfaction, and crime. Ambiguous impacts on standardized test performance
Kottelenberg & Lehrer (2017)	Quebec, Canada	Ages 0-4	Ages 0-5	Non-linear difference-in-differences (change-in-changes)	Negative short-term effects on motor and social development and cognitive test performance for children at the lower half of the outcome distribution in two-parent families; positive impacts for children in single-parent households

Havnes & Mogstad (2011)	Norway	Ages 3-6	Ages 30-33	Difference-in-differences	Positive impacts on educational attainment, earnings and welfare dependency, and delayed family formation; larger effects for girls and children with low-educated mothers
Havnes & Mogstad (2015)	Norway	Ages 3-6	Up to early 30s	Non-linear difference-in-differences (RIF-DID)	Equalizing impacts on earnings, with positive effects in the lower and middle parts of the distribution and negative at the top; most gains to children from low-income families
Drange & Havnes (2019)	Norway	Ages 1-2	Age 7	Randomized instrumental variable (IV) design, based on admission lotteries	Positive effects on language and mathematics test performance; larger gains to low-performing children
Gupta & Simonsen (2010)	Denmark	Age 3	Age 7	OLS conditional on observables, and IV design	Relative to home care, no impact of center-based preschool on behavioral development; negative impacts of family day care, driven by boys with low-educated mothers. Negative effects of longer hours
Cornelissen et al. (2018)	Germany	Ages 3-6	Age 6	Marginal treatment effects (MTEs), based on difference-in-difference-IV design	Reverse selection on gains, based on both observed and unobserved characteristics: larger positive effects on school readiness to children who are less likely to attend (e.g., disadvantaged minority children)

3.1 Impacts of Boston's Public Preschool Program

Gray-Lobe et al. (2023) document positive impacts of a large-scale public preschool program operating in Boston, Massachusetts, on children's long-term educational outcomes. In a randomized IV research design based on admission lotteries, the authors find that participation led to favorable effects on college enrollment, high school graduation, standardized test taking, and adolescent disciplinary behavior.

3.1.1 Institutional Setting

Gray-Lobe et al. (2023) study outcomes for cohorts who attended Boston's public preschool in the years 1997-2003, predating a post-2005 program reform. While Boston's universal preschool program ranks currently highly in quality metrics, its implementation during the study period was presumably of lower and more uneven quality. As argued by the authors, this should permit conclusions likely more relevant to large-scale public programs implemented elsewhere. While universal in the sense that families of all income levels are eligible to apply, a large share of the students comes from racial minority and low-income backgrounds. Moreover, the seats are rationed, which provides a basis for the paper's identification strategy.

The city's publicly funded preschool is operated by Boston Public Schools (BPS). Separate kindergarten programs called K0, K1, and K2 serve children aged three, four, and five, respectively. The authors focus specifically on K1 programs enrolling four-year-olds, a common age for public preschool entry. All Boston residents are eligible for attendance regardless of income, although seats are rationed and the enrolled population is relatively disadvantaged, with a large percentage of nonwhite and low-income children. In the studied sample, nearly three-quarters of the applicants were Black or Hispanic and 11% were bilingual Spanish speakers, and more than two-thirds were eligible for a free or reduced-price lunch.

BPS preschools hold high qualification standards for teachers, including a requirement for a bachelor's or master's degree. During the study period (1997-2003), curriculum and hiring decision were at the discretion of each school principal. The average class size was 19 students, with a range from 10 to 25. An estimated annual cost of the program was $13,000 per full-day student (in 2020 US dollars), a high investment compared to the typical state-funded program cost of $6,000. In comparison, the federal Head Start program costs around $11,000 per student.

Boston's preschool program evolved significantly during and after the period under study. In 1997, a decision to run down half-day K1 programs in order to offer full-day kindergarten for all 5-year-olds led to a sharp decrease in the number of K1 seats, which declined from around 2,500 to 900 by 1998. The number of seats remained on the same level throughout the study period, and both full-day and half-day K1 programs were available. After the study period beginning in 2005, the number of preschool seats expanded rapidly along with the implementation of a deliberately designed and centralized program model,

which was reflected on subsequent surveys documenting sustained quality improvement. The improvement record suggests that preschool quality during the study period was lower than later on. In recent years, Boston's public preschools score highly on quality evaluations. Gray-Lobe et al. (2023) argue that the potentially more heterogenous and less distinguished quality during the study period is likely more representative of preschool program implementations in large urban cities, potentially yielding more generalizable conclusions.

Based on administrative records on PBS, Head Start and private school enrollment and Census data on Boston's four-year-old population, the authors deduce that a significant majority of Boston four-year-olds were enrolled in some form of kindergarten during the study period. Furthermore, the expansion of BPS preschool seats co-occurs with decreases in the Head Start and private preschool enrollment shares. Consequently, private and Head Start centers are likely relevant counterfactuals for the city's public preschool attendees.

3.1.2 Data and Outcomes Measured

To assess the impact of preschool attendance on later educational attainment and college readiness, Gray-Lobe et al. (2023) combine preschool application records from Boston's public school district with administrative data on college outcomes as well as on earlier academic performance and behavior. The authors have obtained data covering all preschool applicants in seven admission cohorts between 1997 and 2003 from the BPS district. This application data includes demographical information together with the variables used in the preschool assignments, including students' school preference ranks, admission priorities, and randomized tie-breaking numbers. In addition, a complementary set of post-application data contains information on the school assignments and enrollment.

To measure student outcomes, the BPS applicant data is merged with additional sources providing information on college attendance and preparedness. Data on the key outcomes of college enrollment, college type, and college graduation come from the National Student Clearinghouse (NSC). The NSC records reportedly cover more than 90% of all undergraduate institutions in 2011 (the earliest applicable enrollment year in this study) in the US and 95% of those in Massachusetts. Gray-Lobe et al. (2023) construct outcome measures separately for attending college "on time" and "ever", depending on whether the individual achieved the outcome within a time window assuming typical academic progress. Furthermore, the authors observe the type of college (two-year or four-year, private or public,

Massachusetts), the number of semesters attended, and graduation. In addition to these postsecondary outcomes, the authors have access to the preschool applicants' earlier academic records from the Massachusetts Department of Elementary and Secondary Education (DESE). This administrative data allows them to measure the standardized test performance of public-school students across grades 3-8 and 10, SAT scores and high school graduation, and disciplinary outcomes like truancy, suspensions, and juvenile incarceration.

The analysis is restricted to applicants subject to random assignment to preschool, as described in the following section. The subsample whose slot offers were determined by random tiebreaks consists of 4,215 students, and their characteristics do not notably differ from those of the full applicant population.

3.1.3 Identification Strategy

Aiming to identify causal impacts of preschool attendance, Gray-Lobe et al. (2023) employ an instrumental variable strategy that exploits random variation in enrollment generated by admission lotteries. Namely, the identification relies on a random tie-breaking feature of Boston's centralized mechanism for school assignment. Applicants to Boston's public preschools rank the district's programs in their order of preference. Those who have siblings in a given program and live within walking distance are given a higher priority. Within priority groups, a random number assigned by the district is used to break ties. In essence, the mechanism allocates each applicant a single assignment, either a seat in a specific preschool program or no seat at all, based on a combination of applicant preferences, preschool priorities, and random tie-breaker numbers. This means that assignments between students who share the same preferences and priorities differ only due to the random tiebreaker.

However, few applicants are completely similar in their preferences and priorities. Given that the likelihood of an offer depends in practice on school-level cutoffs, Gray-Lobe et al. (2023) control for the assignment propensity score, or the probability of an offer conditional on the applicant's preferences and priorities. Rosenbaum and Rubin's (1983) propensity score theorem entails that if the offers are random (independent of outcomes) after controlling for preferences and priorities, they are also random conditional on the propensity score. The preschool assignment setting allows the authors to compute the propensity scores based on knowledge of the preferences, priorities, and the structure of the assignment

algorithm used by the school district. Gray-Lobe et al. (2023) calculate the probability of an offer to any Boston preschool by summing the propensity scores for each ranked program.

Preschool program assignment is then used as an instrument for enrollment, controlling for the assignment propensity score. Additionally, the authors add as covariates baseline demographic characteristics, namely race, sex, and bilingual Spanish status. They estimate the model using two-stage least squares, and the resulting coefficient can be understood as a weighted average of local average treatment effects (LATEs) for the compliers (who enroll in preschool due to receiving an offer) at each value of the propensity score. Notably, the LATE essentially corresponds to the effect of treatment on the treated, since few applicants can enroll without an offer.

As the validity of the research design rests on random assignment, the authors check whether the students who were and were not offered a slot are similar in predetermined dimensions. When the propensity score is controlled, the differences are not statistically significant except for gender and race, which are included as controls in the subsequent analysis. In contrast, the treatment and control group are significantly imbalanced in terms of several applicant, application and neighborhood characteristics when assignment risk is not controlled, which highlights the importance of adjusting for the propensity score. Indeed, students from different backgrounds likely apply to different programs. Another threat to validity would be nonrandom attrition between the treatment and control group. Reassuringly, the authors find that the NSC records contain information for about 99% of the non-offered students, signifying a very low attrition for their main postsecondary outcomes. However, there are small but statistically significant differences in attrition for earlier outcomes, with offer-receiving applicants more likely to be observed, perhaps because public preschool attendance leads to higher attachment to the public education system.

3.1.4 Results

Gray-Lobe et al.'s (2023) findings point to positive effects of preschool on a multifaceted set of educational outcomes. According to the 2SLS estimate, attendance in Boston's public preschools increases on-time college enrollment by 8.3 percentage points ($p < 0.01$). This corresponds to an 18 percent boost relative to the 46 percent on-time enrollment rate for students who were not offered a slot. The impact differs by college type, with statistically significant positive estimates for four-year institutions but no effect for two-year ones.

Moreover, the positive effect extends to attending college at any time. When late enrollment is included, 65 percent of non-offered applicants attend college, which enrollment increases by 5.4 percentage points ($p < 0.1$). The estimated impacts on college graduation are statistically insignificant, but the positive estimate for ever graduating is quantitatively large (5.2 percentage points, with a control rate of 33%).

Gray-Lobe et al. (2023) also examine outcomes prior to college enrollment as potential channels driving the postsecondary results. They find no discernable effects on grade repetition and special education classification, with a pattern of small and insignificant estimates throughout grades from kindergarten to 12th grade. Conversely, the results on high school graduation indicate that preschool attendance raises the probability of ever graduating high school by 6.0 percentage points ($p < 0.05$), a relative increase of 9.4% to the non-offered rate of 64%. Nevertheless, some caution in interpreting these estimates is warranted given that they rely on students' appearance in the Massachusetts public school database and the gap in attrition between the treatment and control groups. The authors also investigate the outcomes in the SAT, a standardized aptitude test widely used in college admissions. The effect on test taking is positive and statistically significant at 8.5 percentage points, while approximately two-thirds of non-offered preschool applicants take the SAT. Looking at SAT performance as probabilities of scoring above thresholds defined by quartiles of the state distribution, the results show that below one-quarter of preschool applicants score above the state medians. Preschool attendance increases the likelihood of clearing the bottom quartile of SAT performance in Reasoning (Math and Verbal) at 10% significance level. The effect is positive and statistically significant ($p < 0.01$) for scoring in the top quartile for math, a 5.7 percentage point increase to the control mean of 9.7 percent. However, students who do not take the test are included in the data coded as zeros. This means that the effects on thresholds conflate the extensive-margin impact on SAT-taking and the intensive-margin impact on scores. Since the estimated effects on average SAT scores in a sample of test-takers are imprecise and not statistically different from zero, the authors interpret that the performance effects likely reflect the increased probability of taking the SAT.

In addition to college preparatory outcomes, the paper considers the medium-term effects on performance in state standardized tests and disciplinary behavior. The first outcome of interest is the test scores in the Massachusetts Comprehensive Assessment System (MCAS) exams, administered in grades 3-8 and 10. The findings indicate that preschool attendance had no statistically significant impacts on achievement in elementary, middle, or high school.

Although no test score effects were detected, previous literature suggests that preschool can influence long-term outcomes through noncognitive channels. Gray-Lobe et al. (2023) measure impacts on disciplinary outcomes in middle and high school, namely suspensions, truancy, absences, and incarceration. Estimates are also reported on a summary index, scaled to have means zero and standard deviation one for non-offered students, with positive estimates signifying a decline in behavioral problems. No effect is observed on disciplinary outcomes in middle school, whereas preschool attendance improves the index significantly, by 0.17 standard deviations ($p < 0.01$) in high school. Specifically, preschool is estimated to marginally ($p < 0.1$) reduce the number of suspensions and significantly ($p < 0.05$) decrease juvenile incarceration by one percentage point (the rate for non-offered applicants is 1%).

Investigating differential effects for subgroups of students, the authors compare the estimates across subsamples formed based on sex, race, and income. Generally, preschool attendance has larger effects for boys than for girls. On-time college enrollment does increase for both sexes, although the girls' estimate is only marginally significant at 10% level ($p < 0.05$ for boys). Effects on four-year college going, SAT-taking, high school graduation, and the discipline index are significant for boys, whereas the corresponding estimates for girls are smaller and insignificant. In contrast, differences across race and income are mainly statistically insignificant.

To summarize, Gray-Lobe et al.'s (2023) analysis reveals that enrollment in Boston's public preschool improves postsecondary educational attainment for a relatively disadvantaged population. Students randomly assigned a slot in the program are more likely to attend college, take the SAT, graduate high school, and exhibit lower levels of disciplinary problems in adolescence. State standardized test scores, conversely, appear unaffected by preschool attendance. Taken together, the results point to the potential of universal public early education programs in boosting long-term educational outcomes, especially for boys. Across America's northern border, however, drastically different results are found for the universal child care policy in the Canadian province of Quebec.

3.2 Impacts of Quebec's Universal Child Care Program

Baker et al. (2008, 2019) study the introduction of universal child care in Quebec in the late 1990s, reporting sizeable negative effects on children's noncognitive outcomes. Based on a difference-in-differences comparison to the rest of Canada, their influential 2008 paper

documents an immediate adverse shock on the development of 0-4-year-olds, who are worse off in terms of anxiety and aggression, motor and social development, and health. Baker et al. (2019) extend the follow-up time frame, showing that the negative behavioral impacts persist to elementary school years. Furthermore, children exposed to the child care reform show poorer self-reported health and life satisfaction as well as higher rates of crime in adolescence. However, Kottelenberg & Lehrer's (2017) distributional results suggest that the average causal effects reported by Baker at al. (2008) disguise substantial heterogeneity in the policy impacts: some children do gain from the program, particularly those from single-parent households.

3.2.1 Institutional Setting

In what has become North America's largest modern universal early education program, the Canadian province of Quebec introduced heavily subsidized child care aimed for all children aged 0-4 in the late 1990s. Phased in over the period 1997-2000, this drastic policy change (part of a larger set known as Quebec Family Policy) offered each child a space at regulated care for a $5.00 daily contribution from parents. The program was implemented gradually by age, starting with 4-year-olds in 1997 and extending to all 3-year-olds in 1998, all 2-year-olds in 1999, and all children below age two in 2000. Two forms of subsidized slots were offered, namely center- and home-based. Children aged older than two typically enrolled in centers known as centres de la petite enfance (CPE), created out of existing nonprofit child care venues. Younger children usually entered home-based care, organized by a network of regulated providers affiliated with a local CPE. In addition to preschool, the Quebec Family Policy instituted voluntary full-time kindergarten and after-school care for 5-12-year-olds. Transition to the new system led to substantial expansion in the number of child care slots. New places were created in response to excess demand, and the number more than doubled between 1997 and 2005, from just below 80,000 to around 190,000.

The policy change also aimed to improve the quality of care. Qualification standards were raised for both CPE and home-based care: the share of staff required to have a postsecondary degree was raised to two-thirds from one-third, and training requirements for CPE-affiliated home-based care providers increased (24-45 hours, as well as annual 6 hours of professional development). Staff-to-child ratios decreased slightly for 4-5-year-olds, from 1:8 to 1:10. New higher wage policies were also implemented in the sector. However, an audit

study from 2005 observed that just above 60 percent of the public child care centers met the criteria for minimum quality, with around one-quarter placing in the range from good to excellent. Nevertheless, evidence by Baker et al. (2008) suggests that quality measures in Quebec evolved similarly relative to the rest of Canada over the study period. CPE quality was also higher than that of for-profit or unregulated child care in Quebec.

The Family Policy substantially lowered the price of child care. The program introduction in 1997 induced a steep rise in Quebec's subsidy rates for two-parent families, settling by 2002 at 80 percent, nearly three times higher compared to other Canadian provinces. Not all families were equally affected, however, with largest gains accruing to highest earners. The impact is only half as large for single parents, since significant subsidies were available to low-income families before the program's introduction. This fact coupled with coinciding changes to welfare and child benefit policies impacting single-parents motivates Baker et al. (2008) to limit the analysis to two-parent families. In contrast, the authors' 2019 paper samples children in all family types, acknowledging that by older ages of childhood, family forms may have changed from two- to single-parent and vice versa.

Baker et al. (2008) provide a comprehensive evaluation of the regime change, including effects on self-reported parental labor supply and use of various forms of child care. The results indicate that utilization of non-parental care rose substantially in Quebec. Specifically, the proportion of 0-4-year-olds in care grew by 14.6 percentage points relative to the rest of Canada, or more than one-third relative to the baseline of 41.5 percent in Quebec. Looking at changes in the types of child care arrangements, the authors find that it is mainly the use of institutional care that increased substantially, whereas home-based care (both own or another's home) remained the same. However, a closer examination reveals that the care arrangements within homes shifted notably to licensed caregivers (including the home-based care of CPE program) away from relatives or non-licensed providers. Maternal employment increased along with the rise in child care utilization, but only by about half as much (7.7 percentage points relative to other provinces, or 14.5 percent of the pre-policy rate in Quebec). The gap in the growth rates of employment and child care usage mainly reflects crowding out of unreported informal arrangements: around a third of the 14.6 percentage-point increase in child care use seems to represent a shift from informal (misreported as no child care usage) to formal care. Also, the share of women who do not work but use child care increased a little. Thus, the policy impacts likely reflect the consequences of moving children out of parental as well as other informal arrangements into more formal care.

3.2.2 Data and Outcomes Measured

All three papers utilize nationally representative survey data to study the developmental outcomes of young children aged four years and below. Furthermore, Baker et al. (2019) expand the analysis to cover longer-run effects on child development in elementary school years and on standardized test performance, health, and criminal behavior in adolescence.

Baker et al. (2008) and Kottelenberg & Lehrer (2017) rely primarily on the NLSCY, or National Longitudinal Survey of Children and Youth, a panel survey conducted biannually on a nationally representative sample of Canadian children. The NLSCY is likewise used by Baker et al. (2019), who also use additional data sources to study a richer set of children's outcomes extending further in time. The first NLSCY data collection cycle of 1994-1995 included a sample of children aged 0-11 that is followed longitudinally throughout the study's eight cycles until 2008-2009, with the inclusion of new 0-1-year-olds in each wave adding a cross-sectional component to the survey. As a result, around 2,000 children between ages 0 and 5 are available in each wave. Baker et al. (2019) and Kottelenberg & Lehrer (2017) use data from all cycles, whereas Baker et al. (2008) keep to the first five. All papers drop the transitional cycle 3 (1998-2000) from the analysis and consider the first two waves as the pre-period before program implementation.

Collected via household interviews, the NLSCY contains information on the child's development in the physical, cognitive, social, and behavioral domains, as well as on child care usage, parental labor supply, and family characteristics. Baker et al. (2008) examine a broad range of children's short-term outcomes. First, they consider behavioral summary scores for hyperactivity-inattention, general anxiety, separation anxiety, and physical aggressiveness and opposition. Second, they look at scores on motor and social development as well as on the Peabody Picture Vocabulary Test (PPVT), a cognitive measure; Kottelenberg & Lehrer's (2017) analysis covers these two outcomes. Finally, they estimate effects on health, including infections (nose/throat and ear), asthma, injuries, and an indicator for excellent health. The outcomes are measured at varying ages between 0 and 4: ages 2-3 for the behavioral outcomes, 0-3 for the motor and development score, 4 for the PPVT score, 0-2 for infections, and 0-4 for the other health indicators.

Baker et al. (2019) complement the above survey data with other sources documenting children's development over a longer time frame. To begin with, they pool the NLSCY data with the Survey of Young Canadians (SYC), a cross-sectional survey from 2010-2011 with

similar children's developmental content, thus adding one additional cycle 9 to their time series. This allows the authors to replicate their 2008 results using cycles 1-5 and extend the analysis to further years covering cycles 6-8 and the SYC. In their analysis of outcomes at ages 0-4, they focus on the same outcomes as Baker et al. (2008), whereas the available behavioral indices for ages 5-9 (hyperactivity, anxiety, aggression, indirect aggression, and pro-social behavior) are based on a separate set of age-appropriate questions.

In addition to the NLSCY/SYC data that provide information on young children, Baker et al. (2019) investigate test scores, health, and criminal activity of adolescents. For information on exam performance, they use data from national achievement tests as well as from OECD's Program for International Student Assessment, PISA. The national dataset comes from two successive programs, the School Achievement Indicators program (SAIP) and the following Pan Canadian Assessment Program (PCAP), both of which evaluate students aged 13 and 16 in the core subjects of math, reading, and science. Correspondingly, PISA assesses 15-year-olds around the world in the same subject areas. To analyze the health impacts on older children, Baker et al. (2019) obtain data from the Canadian Community Health Survey (CCHS), a biannual cross-sectional survey. The authors examine the self-assessed health, life satisfaction, and mental health of a sample of 12-20-year-olds. As the final additional outcome area, Baker et al. (2019) observe police-reported criminal behavior for 12-20-year-olds. The data from Statistics Canada's Uniform Crime Reporting Survey (UCRS) allows the authors to investigate separate rates for four crime categories (crimes against persons, crimes against property, "other criminal violations," and drug violations) in addition to an aggregate crime rate, including both incidents that did and did not lead to conviction. The most prevalent youth crimes under the "other" category are failures to appear in court and breaches of probation.

3.2.3 Identification Strategies

Baker et al. (2008, 2019) employ a difference-in-differences (DID) strategy to estimate the impact of Quebec's child care reform in comparison to other Canadian provinces. In the spirit of the same DID approach, Kottelenberg & Lehrer (2017) aim to identify distributional treatment effects of the program by using a non-linear change-in-changes (CIC) estimator. To elaborate on the DID model, Baker et al. (2008, 2019) identify the causal effect of the policy by comparing how the outcomes evolved from pre-reform to the post-reform period in

Quebec, relative to the change in the rest of Canada. The authors include year and province fixed effects, as well as a set of controls for child and parent characteristics. While the authors' two papers follow essentially the same strategy, they do adopt slightly different specifications motivated by partly different data sets.

The DID approach evaluates a reduced-form effect of the program on all children exposed to the reform, including those who do not attend subsidized child care. Baker et al. (2008) maintain that this intention-to-treat effect has the potential of capturing the policy's full impact on all care arrangements, peer externalities included. In order to arrive at the effect of treatment on the treated, the intention-to-treat estimates would need to be divided by the probability of treatment. However, deciding on the scaling parameter is not straightforward, as the probability depends on how the treatment is defined. While Baker et al. (2008) report different estimates scaled either by the impact on maternal labor supply (7.7%) or on child care utilization (14.6%), Baker et al. (2019) adhere to the reduced-form findings and focus on their sign and significance.

The central identifying assumption behind difference-in-differences is parallel time trends between Quebec and rest of Canada. In other words, any pre-existing trends in the outcome variables would pose a threat to causal inference. To alleviate concerns, Baker et al. (2008, 2019) check how the outcomes evolved before and after the policy change in Quebec and other provinces, finding no evidence of notably differential prior trajectories. Another potential short-coming is that any Quebec-specific shocks co-occurring with the Family Policy could bias the estimates. Finally, the reform itself should be exogenous to the children's outcomes. According to Baker et al. (2008), the Quebec Family Policy emerged after long public debate and not in reaction to any contemporary development, giving reason to believe in the policy's exogeneity.

In contrast to the linear DID strategy of Baker et al. (2008, 2019), Kottelenberg & Lehrer (2017) undertake a distributional approach that aims to uncover heterogeneity in the policy effects. Their specification uses the Athey and Imbens (2006) change-in-changes (CIC) estimator, which imitates DID estimation at each quantile of the potential outcome distribution. Namely, a counterfactual cumulative distribution function is constructed by using the observed pre-treatment outcomes in the treated group and the observed changes over time in the control group. This counterfactual distribution represents the imputed outcomes of the treatment group in the absence of treatment. The identifying assumptions are slightly different from ordinary difference-in-differences. Importantly, an individual's rank within a group is

assumed to stay constant over time and be the same under treatment and control conditions. For instance, a violation of this assumption could mean that any zero effects are due to individuals moving along the distribution and not to ineffective treatment. To indirectly test whether rank invariance holds, the authors check whether the distribution of observable population characteristics at different quantile ranges of the outcome distribution varies between Quebec and the rest of Canada. Their results are reassuring, revealing only little change in the demographic characteristics between the groups over time.

3.2.4 Short-Term Results

Baker et al.'s (2008) results indicate that the introduction of the subsidized child care program made young children significantly worse off in multiple domains of noncognitive development. First, behavioral outcomes for 2-3-year-olds worsened for all indices, with statistically significant estimates for generalized anxiety and aggression. Similarly, the result for motor and social skills shows a statistically significant drop that represents 1.65% of the mean in pre-program Quebec, or more than 10% of a standard deviation. The estimate on 4-year-olds' PPVT score is insignificant. Moreover, the results on health indicate heightened exposure to contagious illnesses. The odds of being in excellent health declines by 5.5 percentage points ($p < 0.01$) relative to the rest of Canada, and the likelihood of never having had a nose or throat infection at ages 0-2 decreases by 14 percentage points at 1% significance level (and a similarly significant effect of -5.7% for ear infections). On the other hand, no significant effects are found for asthma or injuries. Important to remember, these estimates represent intention-to-treat effects that, when scaled by the probability of treatment, would appear quite a bit larger. For example, the estimate for motor and social development that reflected a 1.65% decline relative to mean would translate into an effect on the treated between 11.3 and 21.4 percent (assuming that 7.7-14.6 percent of children are treated).

In addition to direct child outcomes, Baker et al. (2008) report negative impacts on parenting quality and parental wellbeing. Their DID-estimates mark statistically significant increases in hostile and aversive parenting along with a significant decline in consistent parenting, implying a deterioration in parent-child relationships. Furthermore, the self-assessed health measures for parents indicate a decline in fathers' odds for excellent health as well as an increase in mothers' depression. The effect on reported relationship satisfaction is also negative and statistically significant.

As can be seen, the introduction of Quebec's universal child care regime led to worse outcomes for small children. Children demonstrate greater anxiety and aggression, poorer motor and social skills, and worse health outcomes, in addition to being exposed to more hostile and less consistent parenting. However, the Baker et al. (2008) raise the possibility that the short-term negative results might simply reflect initial costs of socialization. If the policy only moved up the inevitable difficulties associated with transitioning into an environment with other children to an earlier age, the long-term impact might even be positive. Nevertheless, the authors' 2019 paper is inconsistent with this interpretation, instead demonstrating that the early adverse effects persisted into a longer time horizon. Moreover, Baker et al.'s (2019) replication of the short-term outcomes shows that the significant results on child care use and children's behavioral scores are robust to extending the sample to include single-parent families. That is, Baker et al. (2019) find similar statistically significant increases in anxiety and aggression but not in separation anxiety nor hyperactivity, although the estimated magnitudes are marginally smaller. The later paper does find a different result for the measure of cognitive development, the PPVT score: now measured for ages 4 and 5 (rather than just 4), the authors report a statistically significant decline.

3.2.5 Long-Term Results

Baker et al. (2019) demonstrate that the negative effects on children's early noncognitive development persist to elementary school years and are in some cases amplified. At ages 5-9, the impact on anxiety is more than doubled compared to that for 2-3-year-olds, at 28 percent of a standard deviation. The estimate for aggression (17 percent of a standard deviation) closely parallels that for the younger age group. Hyperactivity also increases (by 13 percent of a standard deviation), as does indirect aggression (by 19 percent of a standard deviation). In contrast to the statistically significant effects on the above behavioral indices, the estimate for pro-social behavior is very close to nil and imprecise. Analyzing impact heterogeneity in an unconditional quantile regression based on the recentered influence function (RIF), the authors find that the effects are larger for boys and for those with already higher levels of behavioral problems.

When it comes to cognitive outcomes in adolescence as measured by performance in standardized tests (PCAP/SAIP and PISA), the pattern of results is rather inconclusive. The estimates for the national tests indicate negative but statistically insignificant effects on math,

reading, and science scores. Conversely, the PISA results reveal a statistically significant improvement in math scores, with insignificant and smaller but positive estimates for reading and science. Overall, the program's long-run effect on cognitive ability remains ambiguous, but scores on reading and science point to a lack of impact.

The findings on the wellbeing of 12-20-year-olds indicate that the exposure to the Quebec program is connected to worse outcomes for self-reported health and overall life satisfaction, but not for self-reported mental health. The estimate for the health indicator signifies a worsening equal to 7.3 percent of a standard deviation, whereas the negative impact on life satisfaction (0.043) is smaller but also statistically significant. In contrast, the mental health estimate points to a positive direction but is very small and statistically insignificant.

Finally, Baker et al. (2019) investigate the impacts on criminal activity in adolescence by comparing the accusation and conviction rates in Quebec relative to the rest of Canada. The results suggest that exposure to the child care program leads to sizeable increases in youth crime. In particular, a graphical examination of year-of-birth cohorts reveals that crime rates rise with each additional year of program eligibility (ranging from 0 to 5), shrinking the gap between Quebec and the other provinces whose crime rates are higher. Quantified in regression estimates, the simple DID results indicate a statistically significant increase of 514 accusations per 100,000 children, or 27 percent relative to the overall mean of 1,872. The effect on convictions is similar in relative magnitude. Looking at crime types, the impacts are largest for property crime and for other criminal code violations, with lesser effects on crimes against persons or drug offenses. Gender differences observed for noncognitive child development are also evident for crime, with larger impacts for boys.

Taken together, the negative shock on the noncognitive development of small children first identified by Baker et al. (2008) is echoed by these later findings of lasting detrimental effects. Children exposed to Quebec's child care program show elevated levels of aggression, anxiety, and hyperactivity in elementary school age, together with poorer self-reported health and life satisfaction as well as higher rates of criminal activity in teenage years.

3.2.6 Distributional Results

In their evaluation of distributional effects on two short-term developmental outcomes, motor and social skills as well as PPVT scores, Kottelenberg & Lehrer's (2017) illustrate that there

is substantial variation in the impacts of Quebec's child care policy. Particularly, children from single-parent families at lower quantiles of the distribution seem to benefit, whereas children from two-parent households between the 10th and 50th quantiles are significantly worse off. Surprisingly, the policy does not generally seem to affect children in the top half of the distribution.

To elaborate, children from two-parent families show significant negative effects at the lower half of the distribution for motor and social development (at ages 0-3), whereas no statistically significant declines are observed for certain top percentiles. Thus, the negative estimate obtained by Baker et al. (2008) seems to be driven by the bottom 50% of the distribution for two-parent families. In contrast, children raised by single parents seem to benefit at nearly all percentiles, and especially so towards the lower end of the distribution. The distributional estimates for PPVT scores (ages 4-5) of children from two-parent families show a similar pattern to the one for motor and social skills, with statistically significant negative effects concentrated between the 5th and 40th percentiles.

The effects also differ across children's gender. Among girls, the effects on motor and social development are significantly negative in the bottom quartile, but positive in several percentiles in the top half of the distribution. A similar pattern is mirrored in the results for boys' PPVT scores, with declines in the bottom half percentiles but no effects between the 50th and 85th percentiles. No such evidence of heterogeneity is found for girls' PPVT scores or for boys' motor and social skills.

Overall, Kottelenberg & Lehrer (2017) paint a more nuanced picture of the child care policy's impacts, demonstrate that the negative DID estimates by Baker et al. (2008) fail to show that some children do benefit from the program, especially those in single-parent households. All results are consistent with Kottelenberg & Lehrer's (2017) finding that children from two-parent households in the bottom half of the outcome distribution are significantly worse off. To summarize the findings from all three papers, the Quebec Family Policy on average led to considerable immediate declines in children's noncognitive development that endure into the long run. Specifically, children exposed to the subsidized child care program show elevated anxiety and aggression, poorer health, and more youth criminal activity. The following section discusses a Norwegian policy reform, which instead reveals positive consequences to expansion of universal child care.

3.3 Impacts of Norway's Universal Child Care Reform

Havnes & Mogstad (2011, 2015) study the long-run impacts of a large-scale expansion of subsidized child care in Norway, discovering strong positive effects on children's educational and labor market outcomes in adulthood. Havnes & Mogstad (2011) adopt a difference-in-differences framework that exploits variation in child care coverage across time and municipalities in the immediate years following the 1975 reform. The authors' subsequent paper relies on the same identification but uses non-linear DID methods to estimate quantile treatment effects, revealing that the largest gains of the child care expansion accrue to children from low-income families and those towards the lower end of the earnings distribution.

3.3.1 Institutional Setting

The Norwegian child care system underwent a radical shift towards universally available subsidized care in the late 1970s. As a backdrop to the reform, formal child care in post-WWII Norway was heavily rationed, failing to meet the needs of a growing base of working mothers. In 1968, about 35 percent of surveyed women with children aged 3-6 expressed demand for formal child care, when child care coverage was below five percent. Out of this group just 34 percent reported using out-of-home care regularly, relying mostly on informal arrangements. Passed by the Norwegian parliament with broad political support, the Kindergarten Act of 1975 assigned the responsibility for child care to municipalities but included federal regulations and subsidies for both operating costs and investments in new facilities. Eligibility was extended to all children, and care places were allocated based on age and the length of time on a waiting list.

In the following years, subsidized child care expanded rapidly. Whereas the coverage rate for 3-6-year-olds was below 10 percent in 1975, it had sprouted to more than 28 percent by 1979. Over the period, the number of child care places more than doubled, increasing by 17,500 slots. During the study period, the child care coverage for younger children aged 1-2 remained nearly absent. The authors focus their analysis on the early expansion of care for children aged 3-6, arguing that it plausibly reflects a supply side shock of easing constraints rather an increase in local demand.

The Kindergarten Act required child care centers, whether run by public institutions or private organizations, to fulfil federal provisions on educational content, group size, staff qualifications, and physical environment. The reform mandated that every formal care institution be run by a college-educated preschool teacher, in addition to having at least one educated preschool teacher for every 16 children aged 3-6. The teachers were usually accompanied by one or two assistants, a position without educational requirements. The educational approach was rooted in the social pedagogy tradition typical to Nordic countries, mainly relying on play and informal learning. Overall, the child care quality appears on par to other large-scale child care programs in different countries, but significantly lower than in many targeted programs. The average ratio of staff to children was about 1:8 and the average annual expenditure per child approximately 6600 USD. In both measures, Norway falls between Canada and Denmark, whose universal child care programs had staff-child ratios equal to 1:12 and 1:7, respectively.

The relevant counterfactual mode of care seems to be informal non-parental arrangements, given that the program didn't noticeably increase maternal labor supply. Therefore, the estimated treatment should be interpreted as moving children from informal care (e.g. relatives or non-licensed caregivers) into subsidized care.

To ease potential concerns about factors confounding the treatment effects, the authors point out that the child care reform didn't coincide with other significant policy changes or trend breaks. Moreover, the quality of child care does not appear to vary with the extent of the coverage expansion. The authors report a lack of significant differences between treatment or control municipalities, either in observable quality measures or public spending per student, which is consistent with Norway's known unified public school system rooted in egalitarian principles.

3.3.2 Data and Outcomes Measured

Havnes & Mogstad (2011, 2015) utilize longitudinal administrative data sets from Statistics Norway to estimate effects on long-term educational and labor market outcomes. Renowned for its coverage and reliability, Norwegian register data facilitates linking individual demographic and socioeconomic information via unique identifiers. Moreover, merging the demographic files documenting the total number of children by age and residence with

administrative data on all formal child care institutions allows the authors to construct a time series of annual child care coverage for each of Norway's 414 municipalities.

The sample of Havnes & Mogstad (2011) comprises children born in 1967-1976. Since the child care expansion began in 1976, the sample is divided into "post-reform cohorts" born in 1973-1976, "phase-in cohorts" born in 1970-1972, and "pre-reform cohorts" born in 1967-1969. The adult outcomes are measured in 2006, when the treated children are aged 30-33. In comparison, Havnes & Mogstad (2015) focus on the pre-reform cohorts born in 1967-1969 and the post-reform cohorts born in 1973-1976, excluding the partially treated cohorts. Their outcome measurement of earnings spans the years 2006-2009. Both papers restrict the sample to children of married mothers (based on marital status prior to the reform, in 1975). The final samples consist of 499,026 (2011) and 341,170 (2015) children.

The two papers examine a slightly different set of adult socioeconomic outcomes. Havnes & Mogstad (2011) consider educational attainment, earnings and welfare dependency, and family formation. Specifically, educational attainment is measured in terms of years of schooling completed. For example, indicators for having attended college or being a high school dropout are determined by having obtained at least 13 and at most 11 years of education, respectively. The measures for labor market attachment and welfare dependency are based on the basic amount thresholds of the Norwegian Social Security Scheme, with one basic amount corresponding to around 10,500 USD in 2006. Impacts on earnings, consisting of both wages and self-employment income, are estimated as probabilities of being a low, average, high, or top earner, determined by the basic amount thresholds (with low earners earning at most two basic amounts, and respective minimums of four, eight, and twelve basic amounts for the higher brackets). Correspondingly, being on welfare signifies receiving more than one basic amount in public cash transfers. The family formation outcomes refer to having children and cohabitation. By contrast, the main outcome of interest in Havnes & Mogstad (2015) is average earnings over the 2006-2009 period, including both wages and income from self-employment. As mediating channels, the authors also look at educational attainment, measured as years of schooling, as well as cognitive ability test scores. The cognitive test scores are available only for males, who are tested upon entering compulsory military service, typically at ages 18-19.

3.3.3 Identification Strategies

Havnes & Mogstad (2011) estimate the impacts of child care expansion by adopting a difference-in-differences approach based on regional variation in the size of the supply shock. That is, the authors compare the adult outcomes for children aged 3-6 before and after the reform between treatment municipalities where child care coverage increased a lot and control municipalities with little or no expansion. Norwegian municipalities are categorized according to their percentage point increase in the child care coverage rates in the expansion period of 1976-1979, split at the median such that the upper half forms the treatment group and the lower half the control group. Graphical evidence shows that coverage in the two groups developed almost in parallel prior to the reform, but kinks sharply in treatment municipalities afterwards. Havnes & Mogstad (2015) follow their earlier paper in the identification strategy but apply non-linear DID-methods to estimate quantile treatment effects.

The causal parameter of interest in the regression model of Havnes & Mogstad (2011) estimates the mean effect of additional care spaces on children in post-reform cohorts who live in the treatment municipalities. Both papers capture the reduced form effects of the reform, encompassing the influence of changes in both formal and informal care as well as spillover effects on children not enrolled in care. Like Baker at al. (2008), the authors scale these intention-to-treat (ITT) effects with the probability of treatment to evaluate the impact of treatment-on-the-treated (TT). That is, they divide the ITT estimates by the increase in child care coverage in treatment municipalities relative to the control group (0.1785 in the 2011 estimation), thus yielding the treatment effect per child care place. Havnes & Mogstad (2015) also report TT estimates scaled by the increase in child care expenditure, giving effects per dollar spent.

The DID model relies on the identifying assumption of parallel trends, meaning that the outcomes in 3-6-year-olds in pre- and post-reform cohorts should have developed similarly in treatment and control municipalities in the absence of reform. While the DID approach controls for unobserved differences between children born in different cohorts as well as between children living in treatment and comparison municipalities, one might worry about bias arising from differentially evolving population characteristics in the treatment and control regions. For instance, if a determinant of children's outcomes, such as parental education, changed over time differently in the two groups, the estimates would be biased. To alleviate such concerns over omitted variable bias, Havnes & Mogstad (2011, 2015) control for a set of child and parental characteristics in addition to including municipality-specific

fixed effects. Furthermore, Havnes & Mogstad (2011) run two placebo tests in order to test whether the common trends assumption holds. The first pretends that the child care expansion occurred in the pre-reform period, whereas the second tests whether the reform shows an impact on adult height (measured for males before their military service), known to be primarily determined by genetics in developed countries (hence, unaffected by the reform) and positively correlated with higher education and earnings. The insignificant results from these placebo test support the notion that the estimated effects reflect true policy impacts rather than differential time trends. Moreover, a graphical examination of treatment and comparison group trends before and after the reform shows congruent development in the pre-reform period and a remarkable departure afterwards.

As further specification checks, Havnes & Mogstad (2011) and (2015) interact cohort fixed effects with observed pre-reform municipality characteristics. Both papers also address the threat that coinciding secular changes between urban and rural areas are driving the effects, estimating additional models with either dropping the three largest cities from the analysis (2011) or including separate time trends for the largest cities (2015). The authors also test robustness to different exact definitions of treatment and the expansion period. Specifically, both papers explore dividing the sample at the 33rd and 67th percentiles of care coverage growth rather than at the median, excluding the children from in-between municipalities from the analysis. Additionally, they test postponing the expansion period by one year to 1977-1979. Their results turn out to be consistent across different treatment definitions.

An additional threat to identification arises from two potential sources of selection bias. First, municipalities might select into increasing child care coverage based on the developmental gains of care, with child care potentially expanding the most in regions where it has higher impacts. At the same time, there might be selection of municipalities into child care expansion depending on underlying trends in the child outcomes. For example, if municipalities expand subsidized child care in order to counteract a negative trend in children's development, the estimates would be downwards biased. Conversely, expansion in response to a positive trend would lead to an upward bias. This concern motivates the authors to examine the determinants of child care expansion across municipalities. The treatment and control municipalities are spread out across Norway, covering both urban and rural areas. The baseline specifications of both papers define five of the ten largest (by the number of children) cities as treatment and the other five as control municipalities. Reassuringly, a

comparison of treatment and control municipality characteristics reveals that both are comparable in terms of local public expenditure and income as well as political and demographic composition. An important difference does stand out: child care expanded most strongly in municipalities where the ratio of formal child care coverage to maternal employment was lowest. This is to be expected, as municipalities with low child care coverage were not only granted higher federal subsidies but also were more likely to face greater political pressure to alleviate the rationing of care.

Because selective migration of families could also bias the estimates, Havnes & Mogstad (2011) exclude children from families who relocate between treatment and control municipalities during the expansion period. They also control for moving between municipalities within the treatment/control region, in addition to reporting robust results from a model where children's municipality of birth rather than residence defines their treatment status.

Relying on the above identification approach, Havnes & Mogstad (2015) estimate non-linear econometric models aiming to uncover distributional impacts of the child care reform, in addition to the mean impacts given by the standard DID. Namely, the estimated quantile treatment effects (QTEs) record how children in various parts of the outcome distribution were affected. The QTEs are identified by comparing the treatment group's observed cumulative distribution function of the outcome to the counterfactual distribution, that is, what would have happened to the treatment group in the absence of the child care expansion. Non-linear DID methods construct this counterfactual by using the observed change in the comparison group's outcome distribution. In contrast to Kottelenberg & Lehrer (2017) who use a change-in-changes estimator, Havnes & Mogstad (2015) adopt a method called recentered influence function (RIF) regression. The identifying assumption behind the RIF-DID estimate is that the change in population shares from pre-treatment to post-treatment around a given outcome level should be the same in the treatment and control groups, in the absence of treatment. Reassuringly, the authors find no significant changes over time in the differences in child and parent characteristics between the treatment and control municipalities, giving no reason to suspect unobserved compositional changes.

3.3.4 Results

Havnes & Mogstad (2011, 2015) observe substantial positive effects of the child care reform on children's educational attainment and earnings in adulthood. The author's 2011 paper reports that children exposed to the child care expansion completed more years of education, were less likely to be low earners or dependent on welfare, and more likely to delay family formation. Given that the effects are driven by girls and children with low-educated mothers, the authors argue that subsidized child care contributed to intergenerational mobility and closing of the gender wage gap. Havnes & Mogstad (2015) corroborate the findings of equalizing effects on earnings, as their quantile treatment effect estimates indicate largest positive impacts in the lower parts of the earnings distribution, with negative effects at the top. Moreover, children from low-income families were found to be the main beneficiaries of the reform.

Havnes & Mogstad (2011) report statistically significant ($p < 0.01$) estimates on educational attainment, revealing an ITT effect of added 0.06 years of schooling per child, or a TT effect of 0.35 years. Overall, the additional 17,500 child care slots generated around 6,200 years of education. Furthermore, child care decreased the likelihood of dropping out of high school by almost 6 percentage points and increased the probability of attending college by nearly 7 percentage points.

The 2011 results on adult earnings and welfare dependency suggest that subsidized child care had an equalizing effect on the earnings distribution. Whereas the reform reduced the likelihood of having low earnings by 3.6 percentage points and increased the chances of being at least an average earner by 5.1 percentage points, the effects on high and top earnings go in the negative direction (3.4 and 2.2 percentage point declines in probabilities, respectively). However, the impacts on high and top earners do not pass all robustness checks. Correspondingly, the probability of being on welfare declined by about 5 percentage points in response to the child care expansion, a sizeable effect relative to the pre-reform treatment group mean of around 16 percent.

In line with the results on greater educational attainment, Havnes & Mogstad (2011) also find that the child care reform led to delayed family formation. Specifically, treated children have a nearly 8 percentage points lower likelihood of having a child at the time of measurement, and almost 3.5 percentage points higher likelihood of being single without a child, both significant at the 1% level. In contrast, the probability of being a single parent is unaffected by the reform. Although it is unclear whether these effects at ages 30-33 endure

over time, delayed childbearing could ultimately reduce the family size of the impacted children, as female fertility weakens at ages beyond 30.

Subsample analysis conducted by Havnes & Mogstad (2011) shows that there's variation in the effects across children's sex and mothers' education level. In particular, the positive effects on earnings are mainly driven by girls, while the educational gains occur mostly to children from low-educated mothers. These findings lead the authors to conclude that the child care reform "leveled the playing field" by increasing intergenerational mobility and diminishing the gender wage gap.

The results from Havnes & Mogstad's (2015) QTE estimations present a similar picture of equalizing effect on earnings, with positive effects in the lower and middle parts of the earnings distribution of treated children as adults, and negative at the top. In detail, the effect remains zero or positive for all percentiles up to the 82nd with a peak at the 11th percentile, where the QTE estimate indicates a $1.27 increase in earnings per dollar spent (or $8403 per child care place). In contrast, the estimate for the mean impact is small and insignificant at $0.05 per dollar spent, or $311 per child care slot. To further illustrate the equalizing effect of the reform, the authors calculate that the Gini coefficient in the observed earnings distribution of the post-reform cohorts in treatment municipalities was reduced by 2.9 percent compared to that of the counterfactual distribution.

Havnes & Mogstad (2015) also demonstrate that the impacts on earnings vary by family income, with benefits accruing to individuals from low-income homes. By estimating local linear regressions of children's earnings on the log of family income separately for pre- and post-reform cohorts in treatment and control municipalities, the authors show that the income gains of the child care reform decline consistently in family income, with losses to upper-class children. The differential impacts by parental income also translate into positive effects on social mobility, with a standard DID model estimating that the reform reduced intergenerational income elasticity by 2.5 percentage points.

In line with the results on earnings and their 2011 findings on educational attainment, Havnes & Mogstad (2015) show that the child care reform had a positive mean effect on years of schooling that mainly stems from children with low family income. Moreover, the RIF-DID estimates for years of education reveal that the increase in educational attainment is largely driven by higher likelihoods of completing high school and attending college. In contrast, the results on cognitive test scores appear to be unaffected by the reform, with the QTE effects being close to nil and precise enough to rule out meaningful impacts.

To summarize, the results from the two papers point to positive and equalizing effects of the child care expansion on long-run earnings and educational attainment. The mean impacts on years of schooling and labor market attachment are sizeable, while subgroup and distributional analysis indicates that the gains accrue particularly to those in the lower parts of the earnings distribution and from low-income families. The following section continues to analyze Norway's universal child care system but shifts the focus to the very early care of 1-2-year-old children in the city of Oslo in more recent times.

3.4 Impacts of Early Child Care in Norway

Drange & Havnes (2019) examine the cognitive effects of child care enrollment of toddlers aged 1-2 via randomized admission lotteries used by the Norwegian capital Oslo. The results show significant improvements in language and mathematics performance at age 7, both in terms of mean scores and probabilities of exceeding nationally defined thresholds for low performance, especially for underperforming children.

3.4.1 Institutional Setting

While the previous section on Norway's universal child care system focused on the institutional details of the program's introduction in the 1970s, the description here focuses on the more contemporary organization of child care. Specifically, Drange & Havnes (2019) investigate care for children younger than 3 years old in Oslo. The majority of child care slots in Norway are allocated annually in centralized allocation rounds, where parents can apply for a place in up to seven institutions. The authors' identification strategy relies on oversubscription and random seat offers generated by the assignment mechanism.

In practice, most Norwegian parents can stay at home with their newborn for approximate one year. Fully wage-compensated parental leave lasted 43 weeks (or 53 weeks with 80% compensation) in 2005, which was extended to 44 (54) weeks in 2006. Immediately on top of that, each parent is entitled to a year of unpaid leave. In 2004, Norway's child care enrollment rate of children younger than age 3 was 44 percent. This is significantly lower than in several other countries, such as Denmark's 83 percent, but comparable to the US rate of 38 percent at the time.

As described earlier, Norwegian child care is strongly regulated and governed by a social pedagogy tradition. Regulations necessitate that there should be at minimum one educated preschool teacher per 10 children below age 3, who is typically accompanied with two assistants. In the studied sample, the average child care institution enrolls 14 children aged 0-2, with around 3 adults (including one teacher) per 10 children. About 16 percent of children come from an immigrant background. In Oslo, about 60 percent of child care institutions are publicly operated, while the rest are private. Both public and private centers are subject to municipal supervision and eligible to federal government subsidies covering around 80 percent of the costs. The maximum monthly contribution from parents has been set at around 2,500 NOK for a full-time slot since 2003.

Based on survey data on stated demand and actual utilization of child care, parental care seems to be the predominant alternative mode of care for Norwegian toddlers. About 70 percent of parents express demand for either full time or part time child care, whereas only a third actually enroll their children in care. Moreover, 56 percent take care of their children themselves, but only 17 percent prefer this arrangement. On the other hand, receiving a lottery appears to modestly increase mothers' full-time employment but has little impact on the employment margin or on paternal labor supply. This suggests that parents may have access to informal care arrangements.

3.4.2 Data and Outcomes Measured

The authors have obtained administrative data on cohorts born in 2004-2006, including all child care applications, lottery offers and enrollments as well as language and mathematics test scores at age 7. Register data from the city government of Oslo are complemented with individual-level demographic records available from Statistics Norway.

The outcome of interest is performance in mathematics and Norwegian language tests conducted at the beginning of first grade. The tests are designed on the national level and aimed at identifying poorly performing children for whom schools should allot supportive resources. The grading scale ranges from 0 to 105 in language and 0 to 50 in math. Given the purpose of detecting problems, the test score distributions are skewed, with approximately 10 and 15 percent of children scoring at the ceiling in language and mathematics, respectively. In the analysis, the authors construct two outcome measures from each exam. First, the scores are normalized to have mean zero and standard deviation one. Additionally, an unweighted

average of the two test scores are considered as a summary measure of cognitive development. Second, dummy variables are defined for performance below nationally determined thresholds, set for individual test parts with the intention of identifying the bottom 15-20 percent of children. The dummy gets the value one if a child scores below the threshold in one or more test parts.

The authors' final sample consists of 2,888 children, whose parents applied for a care slot in the calendar year the child turns one. Analysis of baseline characteristics reveals that such parents are on average more highly educated and higher earners than average parents of a one-year-old, as can be expected.

3.4.3 Identification Strategy

Drange & Havnes (2019) exploit randomized assignments to early child care in Oslo, using lottery offers as an instrument for starting child care early. Crucially, the assignment lottery generates random variation in early child care enrollment, which allows the authors to circumvent selection bias.

Recall that placement into Norway's child care in determined in annual centralized allocation rounds, in which parents may apply up to seven child care centers, both public and private. Based on the received applications, municipalities generate lists of non-priority applicants to each care establishment. Priority may be awarded to children with disabilities or siblings in the same institution. The authors exclude children receiving priority placement (24 percent of the original sample) as well as those born after September 1 who are not included in the main allocation round. Moreover, since lists for private centers are transferred to the institutions themselves that also manage their own admissions based on fuller set of applicant information, children who have ranked a private institution first on their application are also excluded from analysis due to endogeneity issues.

The assignment mechanism to public institutions works as follows: a computer randomizes the order of children on the full list of applicants to each center, and available seats are then allocated according to the random rank on the list. Parents are free to either accept or reject the offer; rejected offers are transmitted to the highest-ranking applicant who did not already receive one from any institution. Once an offer has been received, the child is removed from the list to other public institutions applied, except for the center ranked first (for which the child may retain their place on the list). This implies that individual ranking of

institutions does not influence the likelihood of receiving an offer. Rather, the probability of winning a lottery offer from a particular institution depends on the extent of oversubscription. To account for the fact that children apply to different centers with varying numbers of applicants and available slots, Drange & Havnes (2019) control for lottery-specific risk set, or a full set of dummies for each institution by year. Moreover, applying to a greater number of institutions increases the odds of being offered a slot. Potential bias-inducing correlation between the number of institutions applied for and children's developmental outcomes could arise from parental preferences for child care being connected to family background characteristics. Indeed, parents who apply to more institutions are found to be better educated, to earn more, and be less likely of immigrant background. Consequently, the authors control for the number and identity of institutions on the application list.

Overall, the identification strategy relies on oversubscription of child care slots as well as successful randomization of the offers. The relevancy of lottery offers is confirmed by strong and widespread oversubscription to public child care in Oslo: there are on average about 25 applicants to each slot, or 4.6 applicants to the first choice (where each child is counted only once). Correspondingly, only 29 percent of the children in the sample are offered a place in the lottery, and the offers are shown to have a sizeable effect on child care starting age. Randomization appears to have succeeded as well, given that children who did and did not receive an offer are shown to be well-balanced in observable background characteristics. However, a threat to validity could arise from strategic application behavior. That is, applicants may not base the rankings on their true preferences but rather on maximizing their chances of receiving an offer (or an offer from a particular institution). Listing institutions that are expected to have lower oversubscription rates, or ranking such institutions higher on the list, would count as examples of such behavior. However, the authors point out that controlling for the institution identities should account for the former behavior, and that oversubscription rates are high for most centers and not directly observable to families.

3.4.4 Results

The results suggest that enrolling early in child care has positive effects on children's cognitive development at age 7. To begin with, a simple comparison shows that children who received a lottery offer score on average about 10 percent of a standard deviation higher than

children who did not receive an offer. In both language and mathematics, children with offers are about 2 percentage points less likely to score below the low performance threshold. Among children who did not win offers, 12 percent fall under the threshold in language and 6 percent do so in math. Moreover, regression analysis with controls for lottery risk and sociodemographic characteristics yields similar estimates. The effect of receiving a lottery offer on the average test performance is approximately 13 percent of a standard deviation, driven by improvements of about 16 and 11 percent of a standard deviation in language and math, respectively. Relative to the control group's mean, these effects imply an approximately 30 percent decline in the probability of scoring below the limits for low performance. To put these effects into context, the authors point out that the impact on the language test corresponds to the performance gap between children from high- and low-income families, or about half the gap between children from high- and low-educated parents. The math estimate corresponds to around one-third of the gap between high and low family incomes, or about one-quarter of the gap between parental education levels. Hence, receiving an offer for a public child care place appears to have a substantial effect on children's language and mathematical development.

Examining heterogeneity in the treatment effects between subsamples formed by gender and socioeconomic background, the authors find a pattern of two to three times greater gains in underperforming groups, although the point estimates are imprecise and lack statistical significance. For example, the gap in language performance between children from families with low and high education would be halved when both are enrolled. Moreover, results of analysis where dependent variables are indicators for scoring above 19 quantiles of the test score distribution suggest substantial improvements at the lower part of the distribution, with modest and statistically insignificant effects upwards from the middle. Of course, it should be kept in mind that the tests are designed to capture variation specifically at the low end of the distribution, not at the top.

Drange & Havnes (2019) aim to identify potential mechanisms behind the reduced-form effects on test performance, concluding that an earlier starting age is the likeliest driver. Specifically, the authors consider three possible channels: receiving a lottery offer can either allow the child to enter child care at an earlier age or to attend a higher-quality institution, or it can induce changes in the family environment (via parental income and labor supply). While lottery offers strongly predict child care attendance in the year of application, many non-offer-receiving children also enter early, and eventually nearly all are enrolled in care

before school age. That being said, receiving an offer does seem to lower the age of entrance to child care. Offer-receiving children start child care when they are about 15 months old on average, whereas the average starting age among children who are not randomly offered a slot is 19 months. Among children whose entry is delayed, about two-thirds wait at least one year before attending. Based on an IV model where lottery offer instruments for the age at first enrollment, it appears that starting child care 1 month earlier causes a statistically significant test score improvement (0.05 and 0.03 standard deviations for language and math, respectively). In contrast, little support is found for the other two channels. Children who received lottery offers do not seem to attend higher-quality institutions, and impacts on parental labor market outcomes are very modest.

In short, early child care enrollment at ages 1-2 is found to improve children's performance in language and mathematics tests at age 7, both in terms of the mean score and the probability of scoring below a nationally set threshold for low achievement. Especially underperforming children appear to benefit, and the effects seem to be driven by lowering of the child care starting age. The next subsection discussing the effects of another Nordic universal child care program, in Denmark.

3.5 Impacts of Denmark's Universal Preschool and Family Day Care

Gupta & Simonsen (2010) investigate the noncognitive effects of participation in Denmark's universal child care, adopting an OLS approach as well as an instrumental variable strategy that exploits municipality-level variation in policies guaranteeing access to preschool. Relative to home care, center-based preschool appears to make no difference in children's behavioral skills at age seven, whereas family day care has a significant negative impact. The superiority of preschool over family day care seems to apply especially to boys with lower levels of education. Moreover, the length of time spent in non-parental care matters, with weekly hours exceeding 30 leading to worse behavioral outcomes.

3.5.1 Institutional Setting

Publicly subsidized child care in Denmark is organized both as center-based preschool and family day care, with preschool being the dominant mode of care for 3-year-olds. Gupta &

Simonsen (2010) study the effects of participation in either mode of universal child care, focusing on children who were aged three in 1999. Overall, the majority of child care in Denmark is publicly provided, organized by the country's 271 municipalities. Municipalities offer center-based care as nurseries for 0-2-year-olds, preschools for 3-6-year-olds, and after-school programs for school-aged children, in addition to family day care organized within private homes. All children are eligible regardless of their parents' income or employment status – in fact, excluding certain groups of children is against the law.

Each municipality must provide enough places to cover local needs but can freely decide on how the slots are distributed across the different types of care. Importantly for the paper's identification strategy, some municipalities offer all children a right to a preschool slot, whereas others are not obligated to accommodate parental preferences between preschool and family day care. Care spaces are allocated based on applications filed by parents, with the first open slot offered to the child highest on the waiting list. The application process is the same for preschool and family day care, and seniority can be determined by either date of birth or date of application, depending on the municipality. Parents may either accept the received offer or decline and possibly reapply later. Once enrolled in care, a child no longer has the right to apply for an alternative slot.

The quality of Danish public child care is high in international comparison, indicated by measures of per capita expenditure, staff qualification requirements, and staff-to-child ratios. For instance, the average staff-to-children ratio in preschools was 1:7, the lowest among all OECD countries. The average annual expenditure per slot in center-based preschool for 3-year-olds was around 8000 USD, similar to the yearly cost of Canada's universal child care program but notably higher than the American Head Start, which amounts to around 5000 USD per year. By contrast, family day care in Denmark costs around 10,000 USD, supposedly due to the lower number of children per staff (1:5). Federal law regulates municipal child care on a general level, requiring institutions to supply care, education, and opportunities for play. Moreover, Danish early education is child-centered and focused on socialization rather than formal instruction. Municipalities are legally obligated to monitor the institutions closely (preschools and family day care alike), in terms of educational content, safety, and hygiene.

The care settings differ in significant ways between preschool and family day care. The average preschool, possibly integrated with nursery centers, services about 60 children in smaller groups of around 20, with a staff comprising around 9 permanent teachers together

with assistants and other personnel. Preschool teachers are required to have a post-secondary degree in teaching specialized in young children. In contrast, family day care is organized in the homes of the carers who are employed directly by the municipality. The maximum number of children in each home is five. Unlike preschool teachers, family day carers are not required to hold teaching degrees but are offered short, 3-week-long vocational courses. A comparison of average educational level of staff between preschools (non-teaching roles included) and family day care confirms the higher qualification level of preschool personnel. As for expenses to families, municipalities set the prices once a year for each type of care. Parental contributions depend on income and the number of siblings enrolled in public child care, with the maximum yearly price amounting to 33% of total costs; this is higher for family day care (3500 USD) than preschool (2600 USD). Families with an annual income above 60,000 USD (about 60% of parents) pay the full subsidized price, while families who earn below 20,000 USD in a year (about 1% of parents) can enroll their children for free.

Unlike the other empirical papers discussed, Gupta & Simonsen (2010) have well-defined counterfactuals for the given form of care. Namely, they estimate impacts of preschool and family day care relative to home care, as well as of preschool relative to family day care. The authors' descriptive analysis on the take-up of different modes of child care reveals that approximately 15 percent of children are cared for at home, overrepresented by children with immigrant parents and a greater number of siblings. The majority of children attend municipality-run preschools (around two-thirds of all 3-year-olds), whereas 16 percent attend subsidized family day care. The remaining four percent of children are either in private care (3%) and cared for in other arrangements (1%); this minor fraction is ignored by the authors. On average, children spend 30 hours weekly in outside-of-home care.

3.5.2 Data and Outcomes Measured

Gupta & Simonsen (2010) base their analysis on a panel survey dataset known as the Danish Longitudinal Survey of Children (DALSC), which contains information on children's outcomes, forms of care, and family background. The data follows around 6,000 children born in 1995 in three surveys, completed by the primary caregivers (mothers 99% of the time) when the children were aged 0.5, 3.5, and 7.5 years. At the time of the study, three waves were available to the authors, from 1996, 1999 and 2003. Thus, children's care enrollment is measured at age 3.5 in 1999 and the outcomes at age 7.5 in 2003. Gupta & Simonsen (2010)

merge the survey data with Danish administrative register data providing detailed socioeconomic information on the parents, including their educational attainment, labor market status, work hours, and income.

The authors' main outcome of interest is a standard behavioral measure called Strength and Difficulties Questionnaire (SDQ), which assesses behavioral problems in the domains of emotional development, conduct, hyperactivity and inattention, and peer relationships. The questionnaire forms a discrete interval scale from 0 to 40, with higher scores indicating more problems. As mentioned, the outcome measurements are based on parental responses when the children were seven years old and just began the first grade in school.

3.5.3 Identification Strategy

Gupta & Simonsen (2010) estimate the effects of public child care on children's behavioral outcomes via OLS regression, conditioning on a comprehensive set of observable child, parent, and municipality characteristics. Furthermore, they employ an instrumental variable (IV) strategy that exploits exogenous variation in preschool enrollment stemming from the fact that only some municipalities offer guaranteed access to center-based care. Finally, they examine the marginal effects of hours spent in non-parental care.

First, Gupta & Simonsen (2010) estimate linear regression models aiming to identify the effects of being enrolled in municipality-provided child care relative to home care. The authors address the fundamental issue of non-random selection into public child care by conditioning on a rich set of observables, which have been recognized in the literature as correlates of children's outcomes and the choice of care. In particular, child development has been modeled as a function of the type and intensity of current and past care, purchased inputs, and exogenous shocks, while demand for child care is known to be influenced by maternal employment and the costs of care. Consequently, Gupta & Simonsen (2010) use the panel data on the type and intensity of the chosen mode of care, in addition to municipality-level information on preschool quality as measured by the average teacher-to-children ratio. They also add detailed controls for parental income and labor market outcomes as proxies for purchased inputs, maternal employment, and child care costs. Furthermore, to cover the blanket category of "production shocks," the authors control for several characteristics of the children (e.g. birth weight, gender, disabilities), parents (e.g. age, education level, immigrant status, whether the mother breastfed the child or had post-partum depression), and

municipalities (e.g. unemployment level, number of immigrants, proportion of households with children).

Whether the resulting estimates can be interpreted causally depends on whether the covariates sufficiently control for differences between the groups enrolled in different forms of care. A particular concern pointed out by Gupta & Simonsen (2010) is that unobserved child and parent ability might confound the estimates. For example, if high-ability mothers have a greater likelihood of having high-ability children and are more likely to work and use formal child care, the impacts of non-parental care would be biased upwards. On the other hand, the bias could go downwards if high-ability parents have stronger preferences for quality in their children and hence decide to personally stimulate their development at home. The authors argue that their extensive set of controls should adequately capture parental ability and preferences (including prior measures for occupational status, hourly wage, and leave length). Nevertheless, the selection into a particular type of care is presumably non-random. For instance, parents who strongly prefer a given form of care and are capable of waiting for a preferred slot likely differ from parents who accept a non-preferred one. This leads the authors to exploit variation in access to center-based care that is unrelated to the children's outcomes.

Besides the OLS approach, the authors pursue an IV strategy exploiting plausibly exogenous municipality-level variation in the take-up of preschool. Recall that while all Danish municipalities are obligated to supply enough care slots to meet local demand, they are free to decide on the allocation of slots between preschool and family day care. Importantly, certain municipalities offer guaranteed access to preschool (GAPS). GAPS means that all children have the right to a preschool slot within the municipality. Assuming that parents on average prefer preschool (with its highly qualified teachers) over family day care, the policy can be expected to increase the take-up of preschool, which leads Gupta & Simonsen (2010) to use is as an instrument. The authors reason that the optimal decision for municipal governments is to exactly match the demand for preschool: offering too many slots is costly, whereas offering too few risks political discontent. Therefore, most variation in GAPS provision should arise from differences in demand (such as cohort size differences), conditional on municipality characteristics like the available funds.

To be a valid instrument, GAPS should affect preschool take-up but have no impact on children's developmental outcomes by itself. The authors demonstrate that the instrument significantly increases pre-school enrollment as expected. While the municipality's decision

to offer GAPS is not a parental choice, parents might be expected to relocate selectively. Thus, Gupta & Simonsen (2010) investigate selection into GAPS-municipalities by estimating a probit regression conditioning on observables. The results indicate that selection bias is not a clear issue: only a few coefficients are significant at the 5% significance level. The number of preschool teachers per 100 children is negatively associated with the policy, suggesting a downwards bias on the estimates as far as the indicator reflects unobserved quality of child care. Furthermore, mothers with short tertiary degrees are 4 percentage points more likely to live in GAPS-municipalities, whereas those working full time in 1996 are 7 percentage points less likely to do so, which mitigates concerns over high-ability parents selectively moving to GAPS-municipalities. An additional worry is that local child care policies likely correlate with other municipality characteristics that could influence families' settlement decisions and the municipality's general capacities to offer care services, which motivates the authors to control for a set of municipality-specific variables.

A final identifying assumption is monotonicity: the instrument must affect parents' child care choices in one direction only. Because the authors exclude children in home care from the analysis and instead estimate effects of preschool relative to family day care, they need an extended version of monotonicity. Namely, children in home care are assumed to stay at home regardless of whether GAPS is offered, and children in publicly provided child care are assumed to be in public care with or without GAPS. Moreover, children in preschool in the absence of GAPS must be in preschool also under GAPS, and children in family day care under GAPS must be in family day care also in the absence of GAPS. The observations that more children attend preschool under GAPS and that a similar share of children are enrolled in homecare in non-GAPS and GAPS-municipalities tentatively support the monotonicity assumption. Nevertheless, if high-ability parents preferring preschool would choose home care over family day care, the estimates would be biased upwards.

If the identifying assumptions hold, the IV estimation will yield a local average treatment effect, or LATE. In other words, the estimate captures the effect of preschool enrollment on compliers only, or those children who attend only because they live in a municipality that guarantees preschool access – that is, those truly constrained by the limited availability of slots. The parents of compliers would either be indifferent between the types of care or weakly prefer a slot in family day care over declining a non-preferred offer.
Finally, Gupta & Simonsen (2010) study the impact of treatment intensity, or hours spent in care. Namely, they estimate the intensive-margin effects of increasing time in a given type of

public care by a small amount. They form six categories of weekly hours in care: 10h or less, 10-20 h, 20-30h, 30-40h, 40-50h, and above 50h. The identifying assumption is that whether children spend plus or minus 10 hours per week in non-parental care is random, conditional on observables.

3.5.4 Results

Gupta & Simonsen's (2010) OLS results suggest that attending preschool does not lead to significantly different outcomes from home care, but family day care is associated with a greater incidence of behavioral problems. To elaborate, the estimate for preschool is 0.4 points in the Strengths and Difficulties Questionnaire (SDQ), a statistically insignificant and small rise relative to the mean score of 6.6. In contrast, attending family day care significantly increases the index by 1.8 points. The effect of family day care is comparable to the difference in average SDQ between children of highly educated mothers and children whose mothers have at most a high school degree.

Separate estimations for different subpopulations indicate that the effects depend on children's gender and parental education level. Notably, the negative impact of family day care is only significant for boys, specifically those whose mothers have low levels of education. Attending family day care raises the SDQ index by 2.1 points ($p < 0.05$) for boys with mothers who have completed at most high school, and by 1.5 points for boys whose mothers have a vocational degree (although only at the 10% significance level). In contrast, preschool leads to similar outcomes as home care for all subgroups of children.

Moreover, preschool appears to outperform family day care in children's behavioral outcomes. In specifications estimating the effects of preschool relative to family day care by excluding children in home care from the sample, preschool is shown to benefit children in the whole sample, and especially boys whose mothers have lower levels of education. However, an exception is girls with highly educated mothers: the behavior of this group deteriorates when enrolled in preschool, reflected in a statistically significant increase of 1.4 SDQ points.

The IV results on the effects of preschool relative to family day care largely corroborate the OLS conclusions. Preschool enrollment has a positive impact on the whole sample (significant only at the 10% level). This effect is again driven by boys of mothers with vocational degrees, with a statistically significant estimate of -7.4 SDQ points. The estimate

for girls of mothers with higher education is still positive but no longer significant. Overall, the IV parameters are greater than the OLS estimates, but the standard errors are also larger. The noisiness from the smaller sample sizes makes it difficult to draw conclusions on heterogeneous treatment effects.

The impact of treatment intensity seems to be negligible at low hours (below 30h per week), but negative at longer hours in care. Increasing the time in care from 0-10 to 10-20 weekly hours and from 10-20h to 20-30h does not have significant effects on SDQ points, in preschool and family day care alike. Conversely, the effect turns negative when the hours are increased further to 30-40h and 40-50h, with statistically significant estimates for preschool.

Taken together, Gupta & Simonsen's (2010) results indicate that enrollment at age three in Denmark's municipality-provided preschool leads to similar noncognitive outcomes at age seven as home care, independent of children's gender or their mother's educational attainment. Conversely, attending family day care has a significant negative impact on behavioral problems, driven by boys whose mothers have low levels of education. Children overall seem to be better off attending preschool, with its more qualified staff and larger group sizes, than family day care, although the opposite may be true for girls with highly educated mothers. Finally, increasing the time in non-parental beyond 30 weekly hours leads to worse behavioral outcomes. The article covered next looks into the heterogeneity in treatment effects in the context of Germany's universal early education program.

3.6 Impacts of Germany's Universal Child Care Program

Cornelissen et al. (2018) estimate marginal treatment effects (MTE) of Germany's universal child care program, shedding light on which children benefit from participation in terms of pediatrician-assessed school readiness at age six. Exploiting a policy reform from the 1990s that led to a large expansion of subsidized half-day care staggered across municipalities, the authors document that children who gain the most from attending child care are the ones least likely to enroll, based on both observed and unobserved characteristics. Particularly, children from immigrant backgrounds are more resistant to enrollment but experience higher returns from attendance, such that child care appears to help minority children fully catch up with their more advantaged native German peers. Given that the greater treatment effects to disadvantaged high-resistant children are driven by their worse outcomes in the absence of

treatment and not by differences in the treated state, universal child care appears to play an equalizing role.

3.6.1 Institutional Setting

The German public child care system guarantees each 3-6-year-old a place in a half-day program, instituted by a reform in the 1990s that led to substantial expansion of care. Prior to the policy change, provision of formal care was severely rationed, with available slots allocated primarily to older children and those whose mothers were in the labor force, and secondarily based on waiting times. In 1992, the federal government implemented a legal mandate for each municipality to provide every child between ages 3 and 6 a subsidized 4-hour care slot by the beginning of 1996. However, the mandated deadline proved unattainable despite generous state subsidies made available for municipalities. In particular, the state government in Lower Saxony (West German region from where the authors have obtained data) permitted exceptions until the end of 1998. Between 1992 and 2002, around 11,000 new child care seats were created for 3-6-year-olds in the 80 municipalities comprising the study's sample, an increase of nearly 40 percent. The expansion mostly increased early enrollment at age 3 (that is, full 3-year-long attendance), and reduced the share of children enrolling for only 2 years starting at age 4. Over the study period from 1994 to 2002, the 3-year-attendance rate increased from 41 to 67 percent. Consequently, the authors' baseline specification defines treatment as early or 3-year attendance of child care.

Formal child care in Germany is virtually fully publicly provided and strongly subsidized, with parental fees covering on average only around 10 percent of the overall costs. Stringent nationwide standards govern the care quality, resulting in relatively homogenous care environments. Moreover, the pedagogical approach leans mostly towards informal and play-oriented learning. The maximum ratio of students to staff is set at 25 children per two teachers, and teachers are required to hold state certifications for completing a 2-year vocational program followed by a year-long internship. In international comparison, Germany falls in the middle category on quality metrics: the 12.5:1 student-teacher ratio is higher than UK's 8:1 and US Head Start's 10:1, but lower than France's 25:1. The estimated annual expenditure in 2002 was 4,998 USD, in line with other continental European programs but below Head Start's 7,200 USD. Descriptive statistics on the studied sample reveal a median

child-to-staff ratio of 9.4, and an average share of university-educated teachers of 9 percent, and a male staff share of 2 percent.

In the German context, home care by parents or grandparents forms the nearly exclusive alternative to universal child care. Among 3-year-olds, the share of children in public care was around 41 percent in 1994, rising to 76 percent in 2000. Less than 2 percent of families resort to informal arrangements, or care by nannies or other nonrelatives. Moreover, maternal labor force participation falls beneath child care attendance rates at 31 percent in 1994 and 39 percent in 2000, implying that many children whose mothers do not work are also enrolled in formal care.

3.6.2 Data and Outcomes Measured

Cornelissen et al. (2018) have access to administrative records from Weser-Ems, a large West-German region in the state of Lower Saxony, which include school readiness results from compulsory exams conducted by licensed pediatricians on the full school entry-aged population during 1994-2006. In addition to the school entrance examination data, the authors have collected information on the number of child care seats in each municipality and year from regional youth welfare offices. These sources are supplemented with municipality-level yearly data on local sociodemographic characteristics and care quality indicators obtained from Lower Saxony's statistical office or calculated from social security records.

The studied outcome measures are derived from the pediatrician-administered school readiness exam results, which provide standardized information on children's development and health. Early in the year of elementary school entry, all German children undergo a mandatory 45-minute examination consisting of an interview and a battery of tests with the aim of evaluating their school preparedness and identifying any needs for preventive interventions. The main outcome variable of Cornelissen et al. (2018) is a dummy for the pediatrician's recommendation for school entry in the fall, equal to one for positive assessments and zero when a child is held back a year. Official guidelines suggest postponed school entry when delays in physical, cognitive, or emotional development are deemed too severe to be remedied by any therapeutic or special-needs measures by the start of school. Importantly, the pediatrician's recommendations are nearly always followed, and subsequent deferred entry into the labor market translates into significant lifetime earnings losses. Further outcomes considered include motor skill problems, logarithm of the body mass index (BMI),

overweight (per official German pediatric guidelines, BMI above the 90th percentile of the age- and gender-specific BMI distribution), and need for compensatory sport (prescribed in cases of postural or coordination problems, lack of muscular tension, overweight due to lack of physical exercise, or psychosomatic developmental issues).

Restricting the sample to include only municipalities with available data on care slots yields a baseline sample of 135,906 children in 80 municipalities. Of these children, 51 percent were enrolled in child care for at least 3 years. An important demographic characteristic of interest to the authors besides gender is minority status; around 12 percent of the sampled children were of immigrant descent. The two largest groups of minority children are ethnic Germans from the former Soviet Union whose parents immigrated to Germany mainly following the collapse of Eastern European communist regimes and children of Turkish ancestry, respectively forming approximately 35 and 30 percent of the immigrant children in the sample. Children of Turkish origin come from lower educated families and are less likely to speak German at home compared to those from the former Soviet Union. Regarding the outcomes, 91 percent of all children are deemed to be ready for on-time school entry, 85 percent show no motor skill deficiencies, 82 percent have no need for compensatory sport, and only 8 percent are classified as overweight.

3.6.3 Identification Strategy

For identification, the authors rely on Germany's plausibly exogenous child care expansion, which varied across time and municipalities. Specifically, the authors estimate marginal treatment effects (MTE), based on a difference-in-difference-IV setting where the local child care coverage rate three years prior to school entry instruments for early child care attendance. In the MTE framework, heterogeneity in the individual treatment effects (differences between the potential outcomes in the treated and untreated states) may arise from both observed and unobserved characteristics. Unlike quantile treatment effects such as those estimated by Havnes & Mogstad (2015) and Kottelenberg & Lehrer (2017) that identify differential treatment effects on the individual level only under the rank invariance assumption, marginal treatment effects permit unobserved gains from treatment to be correlated with unobserved characteristics driving selection into enrollment. Importantly, the authors observe individual child care attendance status, allowing them to distinguish between differences in the take-up of care and the effects of the care itself. Moreover, the MTE approach makes it possible for

standard IV estimates like average treatment effect (ATE), effect of treatment on the treated (TT), and effect of treatment on the untreated (TUT) to be computed. Since the slot expansion was staggered across municipalities, generating variation in the care availability not only across areas but also across cohorts, the authors are able to adopt a stronger identification strategy than what is typical in the MTE education literature.

The validity of the instrument first requires that the municipality's child care coverage rate creates variation in the probability of treatment after controlling for observed child covariates as well as municipality and cohort dummies. This existence of the first stage is clearly observed in the data. Secondly, the timing and intensity of the care expansion should be as good as random, again conditional on municipality and cohort effects. The authors demonstrate that that increase in the care supply is strongly predicted by lower baseline coverage rates (as expected, reflecting higher state subsidies and political pressure driving greater expansion of care availability), but not by baseline municipality characteristics save for one. Moreover, the inclusion of municipality dummies accounts for time-constant differences in municipality characteristics. The timing of expansion is similarly argued to be plausibly exogenous, given that none of the individual and municipality-level characteristics measured in the previous period are statistically significantly correlated with the instrument. Furthermore, exclusion restriction necessitates that the local child care coverage rate must not affect the child outcomes directly but only via the enrollment decision. Additional potential threats to identification raised by the authors include crowding out of other public expenditure or reductions in household income and worsening of care quality due to the child care expansion, which could negatively affect child outcomes. Reassuringly, income taxes and social benefits are regulated at the federal, not municipal, level, and controlling for observed child care quality measures changes the results little. Finally, the concern over endogenous family migration is mitigated by the findings that below five percent of families moved across municipalities in the two years leading up to the examination and that the mobility rates are uncorrelated with changes in the local child care availability.

3.6.4 Results

Cornelissen et al. (2018) uncover notable heterogeneity in the enrollment propensity and returns to care attendance, with respect to both observed and unobserved child characteristics. Namely, the results point to a pattern of reverse selection on gains, where the greatest benefits

accrue to disadvantaged children who are the least likely to enroll in care. Regarding observable child characteristics, the authors find strong differences between native German and minority children: children of immigrant ancestry are 20-30 percentage points less likely to enroll in public child care, but experience around 12 percentage points greater treatment effect on the main outcome of school readiness. Importantly, untreated minority children have about 12 percentage points lower likelihood of being assessed ready to start school on time than majority children, which implies that early child care attendance can close the inter-group gap. These findings hold separately for children from both major minority groups, but children of Turkish origin are initially more disadvantaged and thus see a larger catch-up than children descending from the former Soviet Union. Gender differences in the propensity of care uptake are small, with girls slightly more likely to enroll, but a similar equalizing pattern emerges in the treatment effects: 3-year-attendance helps poorer-performing boys fully catch up to girls.

Correspondingly, the authors document evidence of reverse selection on gains based on unobservables. That is, children with higher resistance to early enrollment on the basis of unobservable characteristics appear to benefit more from attendance, and those most likely to attend gain the least. For the top 40 percent of children most likely to enroll in care for at least three years, the effects on school readiness are negative but not statistically significant. Conversely, the 30 percent of children with the lowest probability of attendance see positive and statistically significant returns to early attendance. Deriving summary treatment effect measures, the authors report a statistically insignificant ATE of 0.059, which represents a 5.9 percentage-point increase in the likelihood of being recommended for on-time school entry for the randomly chosen child. The estimated TT effect suggests that the treatment reduces the probability of a recommendation for school readiness by 5 percentage points for the average treated child, but this estimate similarly lacks statistical significance. By contrast, the TUT effect is sizeable and statistically significant at the 5 percent level, revealing that the average untreated child would be over 17 percentage points more likely to be immediately ready for school by attending child care for at least 3 years.

Results on the additional examination outcomes reveal a similar pattern as documented above for the overall school readiness measure, with untreated children benefiting the most in terms of physical health. Selection into treatment based on unobserved gains is not statistically significant for motor skills nor compensatory sport requirement, whereas reverse selection for overweight and log BMI is marginally significant at the 10 and 15 percent level,

respectively. As for the summary treatment effects, the ATE and TUT estimates for compensatory sport indicate large and statistically significant benefits to child care attendance for the average and currently untreated child. Moreover, enrollment appears to increase BMI and the risk of overweight for the average and treated child, but the effects are imprecise and not statistically significantly different from zero.

Cornelissen et al. (2018) investigate whether the larger treatment effects on school readiness among children with high versus low resistance to enrollment are explained by differential outcomes in the untreated or treated states. It turns out that the greater gains to high-resistance children are entirely driven by their poorer outcomes in the untreated state, and that the gap in school readiness between the groups nearly disappears in the treated state. Thus, universal child care acts as an equalizer. Illustrating how families with low resistance to enrolling their children in care differ from high-resistance families, the authors present evidence that low-resistance parents are more likely to have attended routine medical postnatal checkups as well as to be college-educated. Therefore, it appears that children with high resistance may come from less advantaged family backgrounds.

The results beg the question of why disadvantaged children with much to gain aren't enrolled in care more often, and why advantaged children attend even in the absence of benefits. One potential explanation for the high care utilization by advantaged families could lie in parental preferences favoring labor market participation or leisure time at the expense of their children's development. Furthermore, the enrollment decisions made by the child care centers in the case of excess demand presumably favored advantaged parents who are more likely to be employed and better informed about the admission process. Even after enough child care slots were created to serve all applicants, the participation rate of disadvantaged minority children remains lower than that of advantaged majority children, possibly due to lesser knowledge about the benefits of or more critical attitudes towards formal child care rooted in cultural or religious reasons. The costs of public care may also be relatively higher for disadvantaged families. Given that the German child care program appears to disproportionately subsidize advantaged families while failing to reach disadvantaged and minority children with more to gain from attendance, the authors suggest that optimal policy would strive to attract high-resistance children into care. Such measures could involve offering child care for free for disadvantaged families and reducing the subsidies for well-off families, and informational and culturally sensitive campaigns to boost the enrollment of resistant children. Moreover, child care programs targeted at disadvantaged and minority

children appear likely to outperform universal programs in terms of benefits and cost-effectiveness.

In summary, Germany's universal child care program is shown to generate larger benefits in school readiness to disadvantaged and minority children who also have the highest resistance to enrolling in care. Particularly, early child care attendance by children of immigrant origin can nearly eliminate the 12-percentage-point disparity relative to native German children who are better prepared for on-time school entry. The pattern of reverse selection on gains is also apparent for unobserved child characteristics, and the larger treatment effects for high-resistant children are driven by their worse outcomes in the untreated state. Indeed, the results indicate that there are potentially high returns to be claimed from attracting resistant children into public child care.

4 Conclusions

This paper has reviewed empirical economics papers that aim to evaluate the impacts of large-scale, universally accessible early education programs on children's outcomes. Articles to be covered have been selected based on their publication in a highly ranking economics journal and their focus on universal programs. On the whole, evidence on programs in the US, Canada, Norway, Denmark, and Germany demonstrates that universal child care can have significant effects on children's development and long-term socioeconomic performance, but the effects can go in either direction. To begin with, Gray-Lobe et al. (2023) exploit randomized admission lotteries into Boston's public preschool to show that enrollment increased postsecondary educational attainment and improved college-preparatory outcomes. Based on difference-in-differences findings, the introduction of large-scale universal child care led to immediate and persistent negative impacts on children's noncognitive development, health, and youth crime in Quebec (Baker et al., 2008, 2019), but significant gains in educational attainment and earnings in Norway (Havnes & Mogstad, 2011, 2015). Drange & Havnes's (2019) lottery-based research on more contemporary effects of the Norwegian program on a younger age group reveals positive effects on language and mathematical performance at school start. In Denmark, attending center-based care made no difference in behavioral development relative to home care, but the effect of family day care was detrimental, as Gupta & Simonsen (2010) demonstrate in OLS and IV models. Finally, marginal treatment effects estimates by Cornelissen et al. (2018) reveal that Germany's public child care generates larger gains in school readiness for disadvantaged children who are at the same time less likely to attend. While the results are clearly far from unified, certain noteworthy patterns emerge from the evidence.

First, it remains unclear whether cognitive ability is affected. For many of the programs under study, performance in standardized tests does not significantly change. Attendance in Boston's preschool had no discernable effect on state standardized test scores in elementary, middle, or high school, and the partly positive impact on SAT scores (an aptitude test used in college admissions) was likely driven by the increased likelihood of taking the test (Gray-Lobe et al., 2023). Similarly, Baker et al. (2019) reach ambiguous findings on the impacts of Quebec's reform on adolescents' scores on national achievement tests and the international PISA examinations, covering core subjects of math, reading, and science. In addition to standardized testing, papers studying the Quebec program considered an earlier measure of

cognitive development, the Peabody Picture Vocabulary Test. Results by Baker et al. (2008) suggest a lack of impact on 4-year-olds, although Baker et al. (2019) and Kottelenberg & Lehrer (2017) do find significant effects on a sample consisting of children aged 4 and 5; nevertheless, the measurements reflect only immediate effects. Indicating a lack of long-run impact, Havnes & Mogstad (2015) show that the Norwegian universal child care program did not influence performance on the cognitive tests conducted for males upon starting compulsory military service, typically at 18-19 years of age. On the other hand, attending Norway's child care at age 1-2 significantly improved the performance on language and mathematical tests at age 7, especially for underperforming students (Drange & Havnes, 2019). Similarly, Cornelissen et al. (2018) identify significant gains on German children's school readiness, a comprehensive measure that includes consideration of cognitive deficits. Overall, the ambiguity in short-term results and the absence of significant long-run effects on cognitive development is consistent with the body of evidence on income-targeted programs, which documents a fade-out of gains in IQ and achievement tests.

While universal early education programs may not significantly impact children's cognitive abilities, they can have significant effects on children's noncognitive development. Particularly, the programs in Quebec and Denmark were found to impact children's scores on indices comprising parent-reported measures of emotional, behavioral, and social development. Children exposed to the child care reform in Quebec exhibit statistically significant increases in anxiety and aggression at ages 2-3, which prevail into ages 5-9 (Baker et al., 2008, 2019). As for Denmark, Gupta and Simonsen (2010) discover that while the outcomes for 7-year-olds are similar between children in subsidized center-based care and home care, attending family day care leads to a greater incidence of behavioral problems. It is likely that the lasting influence of the child care programs on later-life outcomes operates through these noncognitive skills.

All the papers evaluating longer run socioeconomic outcomes find significant effects in multiple domains, including educational attainment, earnings, and crime. First, participation in Boston's public preschool increased college enrollment and boosted the likelihood of graduating high school (Gray-Lobe et al., 2023). Norway's child care reform likewise produced a substantial increase in years of schooling (Havnes & Mogstad, 2011). Furthermore, the Norwegian program exerted a positive and equalizing influence on labor market outcomes, reducing welfare dependency and increasing the earnings of those low in the income distribution (Havnes & Mogstad, 2011, 2015). As for crime, the programs in

Boston and Quebec had significant effects on criminal activity in adolescence. Whereas Boston's program improved disciplinary behavior in high school, specifically reducing suspensions and juvenile incarceration (Gray-Lobe et al., 2023), the child care reform in Quebec substantially increased rates of youth crime (Baker et al., 2019). In addition to the above markers of socioeconomic success, universal early education also seems to matter for children's health and wellbeing. The immediate negative shock of the Quebec program on children's health was mirrored as poorer self-reported health and life satisfaction in adolescent years (Baker et al., 2019). Moreover, attending public child care in Germany had an impact on children's physical health at school starting age, as reflected in body mass index, overweight, and need for compensatory sport (Cornelissen et al., 2018). In short, it strongly appears that universal early education shapes children's later prospects in multidimensional ways.

As can be seen, the studies on different child care programs reach quite disparate conclusions. While it is difficult to quantitatively compare the results from studies using different outcome measures, the considerably negative findings from Quebec stand in stark contrast to the positive impacts of many other programs. What could explain such wide variation in the effects? Cascio (2015) points out that theoretically the impacts of early education programs can be expected to depend on children's background, the program's quality, and the counterfactual care alternative. To begin with, the impacts of formal child care are predicted to be larger for children from less privileged backgrounds, whose home environments are presumably less conducive to learning and development. Notably, Boston's public preschool program served a relatively disadvantaged population with a high share of low-income and racial minority students, which could contribute to its positive influence. Program quality could also play a role in the result pattern. For example, the impacts of high-quality center-based care of the Nordic countries were positive in Norway and neutral in Denmark, whereas Quebec's program quality was largely judged to fall short of requirements (although it didn't deviate from the comparison group). Moreover, the counterfactual mode of care seems to matter, as does the type of the universal care. Whereas the introduction of universal child care in Norway mainly displaced informal non-parental care, in Quebec it drew children not only from such arrangements but also from parental care. This would imply that positive effects are more likely when the counterfactual consists of informal care by non-relatives as opposed to home care by parents. However, the results are inconclusive on this front: for example, the predominant alternative for public child care for German children and

Norwegian 1-2-year-olds is apparently home care, and both programs proved beneficial (in Germany's case, for enrollment-resistant disadvantaged children). Furthermore, the results from Denmark indicate that center-based preschool was indistinguishable from home care, but family day care led to significantly poorer outcomes. Therefore, center-based care with its highly educated teaching staff would appear to be superior to home-based care provided by licensed caregivers. This interpretation is supported by the fact that the subsidized care in Quebec comprised regulated home care providers in addition to child care centers. Finally, differences between the identification strategies should be kept in mind. The papers on the introduction of universal child care in Norway and Quebec identify reduced form impacts of the reforms, including spillover effects on children who did not attend care, while the rest of the articles observe individual treatment status and study the effects of participating in the programs.

Besides differences in the average effects across studies, not everyone is equally affected within the same program. Importantly, disadvantaged and low-performing children appear to benefit the most from universal early education, pointing to the equalizing potential of widely accessible programs. Cornelissen et al. (2018) present evidence of heterogeneity in the treatment effects as well as in the enrollment propensity of children in Germany, identifying a pattern of reverse selection on gains where disadvantaged children less likely to enroll benefit more from public child care. In particular, attendance appears to help minority children fully catch up to native German children in school readiness. Since the higher gains to disadvantaged children seem to be driven by their poorer outcomes in the untreated state, universal child care appears to play an equalizing role by providing a more homogeneous learning environment. Correspondingly, Havnes & Mogstad (2011, 2015) find that Norway's universal child care reform led to a more equal income distribution, with children from low-income and low-educated families being the main beneficiaries. Analysis by Drange & Havnes (2019) also suggests that attending Norwegian early child care boosts especially the test scores of weakly performing children. Moreover, Kottelenberg & Lehrer (2017) reveal that children from single-parent families low in the outcome distribution benefited from the Quebec reform despite significant losses for children raised in two-parent households. In addition to differences across family socioeconomic characteristics, subsample analyses indicate that the effects may vary depending on child gender. For example, Gray-Lobe et al. (2023) and Baker et al. (2019) report larger effects for boys, whereas Havnes & Mogstad (2011) find that the gains in earnings are driven by girls. While the picture on gender

differences remains inconclusive, the evidence quite consistently implies that the benefits from universal child care tend to be higher for children from disadvantaged backgrounds, such as minority, low-income, low-educated, or single-parent families.

The observed heterogeneity in the effects of universal early education programs has important ramifications for optimal policy. Furthermore, it may help reconcile the disparity between findings on universally accessible and targeted programs, as Cornelissen et al. (2018) argue. As can be seen, the evidence on universal programs is mixed with effects ranging from highly beneficial to detrimental, which contrasts with the literature on income-targeted programs that unanimously documents a positive influence on children's outcomes. Given that the model interventions like Perry Preschool served highly disadvantaged children, their high returns are in line with the notion that formal child care may yield larger benefits to children from less advantaged backgrounds than to children from well-off families. In terms of policy implications, Cornelissen et al. (2018) point out that public child care programs aimed at disadvantaged and minority children are likely to be more advantageous and cost-effective than universal programs, and that there are potentially high returns to be claimed from attracting hard-to-reach children into care.

Taken together, universal early education appears to lead to more equal child outcomes, helping especially underperforming and disadvantaged children. Furthermore, the results tentatively suggest that universal child care programs may be more likely to impact children's outcomes positively when they are center-based and of high quality, and when they displace informal care arrangements rather than parental home care. Conversely, moving children from the care of parents to licensed non-relative caregivers might predict more negative effects. Nevertheless, the evidence is yet too scant to draw firm conclusions. Policymakers should thus beware that findings from individual studies do not necessarily generalize to all contexts, and that ultimately each program may need to be analyzed directly to determine its effectiveness.

Further research is also needed to relate the benefits to costs as well as to gain a deeper understanding of the mechanisms behind the impacts. A comprehensive cost-benefit analysis of the programs would require evaluating the effects on parental labor supply in addition to children's outcomes. Another important area for continued inquiry is the role of child care starting age. Most of the economic evidence explores the effects of early education on preschool-aged children, upwards of age three, while the impact of formal care on the youngest age group of 0-2-year-olds remains an understudied question that has for the most

part been left outside the scope of this review. Worth noting in this regard is a paper by Fort et al. (2020), which reports negative IQ and personality effects of day care at age 0-2 in the affluent Italian city of Bologna, exploiting admission thresholds in a regression discontinuity design. Moreover, we do not yet know enough about how specific child care quality measures and other characteristics like pedagogical orientation contribute to the outcomes. Future work could also consider the drivers behind treatment effect heterogeneity, such as peer effects between children from disadvantaged and advantaged backgrounds.

Even with these limitations in mind, the literature strongly indicates that universal child care can have significant impacts on children's development and success later in life. In the case of beneficial programs, short-term benefits on children's behavioral development tend to persist and translate into substantial gains in adult outcomes ranging from greater educational attainment and earnings to reduced crime. Indeed, universal early education has the potential to generate important benefits to children that extend to society at large.

References

Bailey, M. J., Sun, S., & Timpe, B. (2021). Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency. *The American Economic Review, 111*(12), 3963–4001. https://doi.org/10.1257/aer.20181801

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

Baker, M., Gruber, J., & Milligan, K. (2019). The Long-Run Impacts of a Universal Child Care Program. *American Economic Journal: Economic Policy, 11*(3), 1–26. https://doi.org/10.1257/pol.20170603

Cascio, E. U., & Schanzenbach, D. W. (2013). The impacts of expanding access to high-quality preschool education. *Brookings Papers on Economic Activity, 2013*(Fall), 127–192. https://doi.org/10.1353/eca.2013.0012

Cascio, E. (2015). The promises and pitfalls of universal early education. IZA World of Labor 2015: 116. https://doi.org/10.15185/izawol.116

Chircop, D. & Claros Gimeno, E. (2021) *Participation in early education and care*. European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)696210

Cornelissen, T., Dustmann, C., Raute, A., & Schönberg, U. (2018). Who benefits from universal child care?: Estimating marginal returns to early child care attendance. *The Journal of Political Economy, 126*(6), 2356–2409. https://doi.org/10.1086/699979

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *The American Economic Review, 97*(2), 31–47. https://doi.org/10.1257/aer.97.2.31

Cunha, F., Heckman, J., Lochner, L., & Masterov, D. (2006). Chapter 12: Interpreting the Evidence on Life Cycle Skill Formation. In *Handbook of the Economics of Education* (Vol. 1, pp. 697–812). Elsevier B.V. https://doi.org/10.1016/S1574-0692(06)01012-9

Drange, & Havnes, T. (2019). Early Childcare and Cognitive Development: Evidence from an Assignment Lottery. *Journal of Labor Economics, 37*(2), 581–620. https://doi.org/10.1086/700193

Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *The Journal of Economic Perspectives, 27*(2), 109–131. https://doi.org/10.1257/jep.27.2.109

European Commission, Directorate-General for Education, Youth, Sport and Culture, Lenaerts, K., Vandenbroeck, M., Beblavý, M. (2018). *Benefits of early childhood education and care and the conditions for obtaining them*, Publications Office. https://data.europa.eu/doi/10.2766/20810

Fort, M., Ichino, A., & Zanella, G. (2020). Cognitive and Noncognitive Costs of Day Care at Age 0–2 for Children in Advantaged Families. *The Journal of Political Economy, 128*(1), 158–205. https://doi.org/10.1086/704075

Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2023). The Long-Term Effects of Universal Preschool in Boston. *The Quarterly Journal of Economics, 138*(1), 363–411. https://doi.org/10.1093/qje/qjac036

Gupta, N. D. & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics, 94*(1), 30–43. https://doi.org/10.1016/j.jpubeco.2009.10.001

Havnes, T. & Mogstad, M. (2011). No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy, 3*(2), 97–129. https://doi.org/10.1257/pol.3.2.97

Havnes, T. & Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of Public Economics*, 127, 100–114. https://doi.org/10.1016/j.jpubeco.2014.04.007

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *The American Economic Review, 103*(6), 2052–2086. https://doi.org/10.1257/aer.103.6.2052

Heckman, J. & Mosso, S. (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics, 6*(1), 689–733. https://doi.org/10.1146/annurev-economics-080213-040753

Kottelenberg, M. & Lehrer, S. F. (2017). Targeted or universal coverage?: Assessing heterogeneity in the effects of universal child care. *Journal of Labor Economics, 35*(3), 609–653. https://doi.org/10.1086/690652