Early Childhood Research Quarterly 55 (2021) 52–63

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

The Effects of Universal Preschool on Child and Adult Outcomes: A Review of Recent Evidence from Europe with Implications for the United States☆

David M. Blau

The Ohio State University, United States

ARTICLE INFO

Article history:
Received 12 May 2020
Received in revised form 15 October 2020
Accepted 31 October 2020

Keywords:
universal preschool
targeted preschool
early childhood education
preschool policy

ABSTRACT

Public preschool programs can be *universal*, open to all age-eligible children, or *targeted*, with eligibility limited to children from lower-income families. The effects on children of targeted programs have been intensively studied in the US, with results showing substantial beneficial impacts on child development and subsequent adult outcomes for disadvantaged children. However, there is little reliable evidence on the medium and long run effects of universal preschool programs in the US. This paper reviews studies from Europe that have exploited quasi-experimental variation to estimate the causal impact of universal public preschool eligibility and enrollment on child and adult outcomes. The evidence shows that these programs provide substantial short and long run benefits to disadvantaged children, but relatively modest benefits to more advantaged children. The implications of the European evidence for the issue of universal versus targeted programs in the US are discussed.

Contents

1. Introduction

Young children from disadvantaged families lag their more advantaged peers in developing cognitive and behavioral skills (Duncan & Magnuson, 2013). These socioeconomic gaps begin early in childhood and are attributable to differences in the home and external environments in which children are raised. The environment experienced early in life can have profound consequences for a child's wellbeing and success in childhood and throughout life. There is considerable public and scholarly interest in early interventions in the lives of disadvantaged children as a means of improving their developmental outcomes and prospects for success in life.

There are two main theoretical arguments in support of this approach, as opposed to later remedial interventions: (1) There are sensitive and critical periods in childhood during which investments in child development are relatively productive. These periods are concentrated in early childhood (Knudsen, Heckman, Cameron, & Shonkoff, 2006). (2) There are dynamic complementarities whereby investments in early childhood development make later investments more productive (Elango, García, Heckman, & Hojman, 2016). The benefits to children, families, and society from high-quality early interventions in the lives of disadvantaged chil-

☆ Helpful comments were provided by attendees at the Research Forum of the Crane Center for Early Childhood Research and Policy, Ohio State. I am also grateful to Michael Baker, Nabanita Datta Gupta, Bill Gormley, Andrea Ichino and coauthors, and Marianne Simonsen for useful comments. I am solely responsible for the content, including opinions.

E-mail address: blau.12@osu.edu

https://doi.org/10.1016/j.ecresq.2020.10.009
0885-2006/

dren have been shown to be very large relative to the costs (Elango et al., 2016). Hence public investment in early childhood education (ECE) is widely viewed as an effective and socially efficient way to improve the development prospects of disadvantaged children and thereby help to reduce social inequity.

Preschool programs supported by public funding come in two varieties: targeted and universal. Eligibility for targeted programs is restricted mainly to children from low-income families, while universal programs are open to all age-eligible children. Head Start is the leading example of a targeted program in the United States, with eligibility for the most part limited to children in families with income less than the poverty level. There is a substantial amount of experimental and quasi-experimental evidence that high-quality *targeted* preschool programs in the US have had large beneficial effects on the health, education, employment, earnings and other long run outcomes of enrollees. This evidence is discussed at length in several recent reviews (Baker, 2011; Elango et al., 2016; Council of Economic Advisors, 2016; Duncan & Magnuson, 2013; Hotz & Wiswall, 2019; Phillips et al., 2017).

These reviews also point out that evidence on the effects of high-quality *universal* preschool programs on child outcomes in the US is limited. A handful of states initiated and/or greatly expanded preschool programs in the 1990s, and such programs now exist in 44 states and the District of Columbia (Friedman-Krauss et al., 2019). There are three difficulties in evaluating the effects of state-level preschool programs on child outcomes. First, there is substantial variation across states in the quality and characteristics of the programs. About half are universal, available to any child in the state who meets the age requirement. Others are targeted to children from disadvantaged families. Some are rated as high-quality and others are not. This variability makes it difficult to define an appropriate set of treatment and comparison states for use in a quasi-experimental evaluation of the effects of universal programs in the states that have instituted such programs. Second, there is considerable variation in program quality within states, making it difficult to define exactly what the "treatment" is (Phillips et al., 2017). Third, most of these programs were initiated in the relatively recent past, limiting the amount of information that can be learned about their long run effects.

Another source of evidence on the effects of universal preschool programs is from studies conducted in several European countries, including Denmark, Germany, Italy, Norway, and Spain. These studies evaluate the impacts of universal national programs funded by central governments, which typically aim to impose and enforce uniform high-quality standards throughout a given country. There are differences across countries in the standards, and there is no direct evidence on the uniformity of quality within each country, but most of these programs would be considered high quality by US standards, as discussed below.

There were major reforms in several countries over the past four decades that significantly expanded funding for these preschool programs. Most of the funding is provided by national governments, but the programs are managed by local or regional governments. In countries with major reforms, there was often substantial variation across local jurisdictions in the speed at which they were able to expand the supply of preschools in response to the reforms. This provides the opportunity for quasi-experimental difference-in-differences (DD) research designs. The presumed relatively uniform quality of the national programs ensures a well-defined "treatment" and the slow-expanding municipalities provide a natural comparison group. In other countries, there was no specific reform that could be used as a quasi-experiment, but there was random or quasi-random rationing of slots at the local level, providing another useful source of variation. Several of these programs have been in place long enough to provide evidence on medium and long-run impacts.

This paper reviews evidence from Europe on the effects of universal preschool programs on child, adolescent, and adult outcomes. This evidence is informative for the US because there is strong popular support for such a program in the US (see Child Care Aware of America, 2018), but proponents often cite evidence from evaluations of *targeted* programs when arguing in favor of a *universal* program. Evidence from evaluations of targeted programs is of limited value in projecting the impact of a universal program. The early targeted demonstration programs in the US (Perry Preschool, Abecedarian, and others) were of exceptionally high quality and served extremely disadvantaged children without good alternative sources of care. It is doubtful that their effects can be extrapolated to more typical high-quality programs serving children from all families in a very different social and economic environment today (Baker, 2011). Evidence from evaluations of Head Start is also of limited value for inferring the impact of a universal program, since Head Start does not serve more advantaged children.

This paper builds on the reviews cited above, complementing them by incorporating evidence from several recently published articles that evaluate effects of universal public programs in the countries mentioned above. This review differs from previous reviews in focusing explicitly on results from recent studies of medium and longer run effects in European countries.

2. Methods

The programs discussed in the articles surveyed here are universal center-based preschool programs. Each article asserts that preschool quality in the country of interest is relatively high and uniform across centers. However, none of the articles presents direct evidence on quality as measured by an observation-based instrument. In some cases they support their assertions by describing the curriculum and standards governing child-staff ratio, group size, and teacher training. Several of the articles assert that the standards are rigorously enforced, resulting in the presumption of relatively uniform quality, especially compared with the US. In other cases they cite figures on expenditure per child to support the assertion of high quality. The absence of direct evidence on the level and variation of quality limits the usefulness of these studies for the purpose of drawing implications for the US. Nevertheless, as shown below, these programs have structural features consistent with high quality, and they operate under a uniform set of national standards.

The criteria for choosing studies for inclusion in this review were as follows. First, the methodology had to be quasi-experimental. Thus, the analysis had to include a comparison group of children who were not exposed (or were less likely to be exposed) to the program. Second, outcome measures beyond the ages of preschool enrollment had to be available. This is important in view of widely-documented evidence of fade out of cognitive gains but persistent social and economic benefits (Elango et al., 2016). Third, the program had to be universal, open to all age-eligible children. Fourth, the program had to be of high quality, as determined by descriptions in the articles, supplemented in a few cases with descriptions from other sources. Studies of the effects of child care subsidy programs in which the subsidy could be used for a variety of different types of child care are not included. Finally, the article had to be published recently (since 2015 in practice, although this was not imposed a priori) in a well-recognized scholarly journal. A number of otherwise interesting and informative studies were excluded based on one or more of these criteria. It should be noted that this selection process was not systematic in the sense of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; www.prisma-statement.org). This literature is small enough that standard literature searches turned up all of the

relevant studies and many others that ultimately were not included in the review.

Table 1 summarizes features of the preschool programs discussed in each article. The countries represented are Denmark, Germany (twice), Italy, Norway (twice), and Spain. There are two articles each on Germany and Norway because one article studies children enrolled as infants and toddlers (0-2) and the other studies children enrolled as preschoolers (3-6). In all, four of the articles study programs for infants and/or toddlers and the other three study programs for preschoolers. Four of the articles study the effects of a national program in specific regions or cities within a country, and the others study the entire country. In every case except Germany the preschools offer full day care. Public preschools in Germany offer four hours of care per day. The second-to-last column of Table 1 lists the typical type of care a child would receive if she is *not* enrolled in the public preschool system, as reported in the articles. This is an important piece of information for interpreting the estimated effects of universal public preschool, providing an answer to the question "compared to what?" "Home care" could be the mother, father, grandparent, or other family members in the child's home. Informal care refers to care by friends, relatives, and neighbors. It is notable that in no case is care in other centers outside of the public system a common alternative, in contrast to the US. In the case of Denmark, Datta Gupta and Simonsen (2016) study the impact of preschool compared to family day care, which is also publically funded and regulated. Children in home care are omitted from their analysis.

The last column in Table 1 displays the outcomes analyzed, the ages at which outcomes are measured, and the nature of the data used. The range of outcomes is wide, encompassing IQ, personality characteristics, school readiness, math and language achievement, grade retention, socioemotional status, and adult earnings and educational attainment. Five of the studies used data from administrative records, and the other two used data from surveys that included tests administered by professional psychologists or trained surveyors. Outcomes are measured at ages 6, 7, 8-14, 15, 16, and 30-40. The diversity of outcome measures and ages at evaluation across the studies is both a strength and a weakness for the purposes of this review. If the findings are qualitatively similar across outcomes and ages, this supports the robustness of the results. However, a uniform research design would ensure that any differences in results across countries are not a result of differences in program quality, outcome measures, or age at evaluation.

It is worth noting that most public preschool programs in the US enroll children aged 3-5, while several of the studies discussed here are of children aged 0-2, and the upper limit for preschool enrollment in Germany and Norway is age six. Nevertheless, it is useful to include studies of infants and toddlers in this review because proposals for universal or expanded care in the US typically cover ages 0-5. However, the inclusion of children enrolled at age six in one of the European studies (Havnes & Mogstad, 2015) makes it less comparable to the US than would be desirable.

Table 2 describes information about program quality, including federal regulations governing child-staff ratio, group size, and teacher training, the curriculum or philosophy of the approach to preschool care, and quotes or paraphrases from the articles about the relative uniformity of quality of care across preschools within the system. Among the countries studied, the maximum allowable child-staff ratio for infants and toddlers is 10:3, 7:2, 4:1, 5:1, and 6:1. At ages 3-6 the range is 16:2, 16:3, and 25:2. Maximum group size is 10 for infants and toddlers and 20 for preschoolers, if regulated at all. All of the countries except Germany require a four year college degree for lead teachers, with Germany requiring a vocational degree. Most countries require specialization in preschool education, and several require an apprenticeship. Several papers assert explicitly that the quality of care is uniform across preschools in the system, while others have little or nothing to say about this issue.

For comparison, the second and third to last rows of Table 2 report information about quality standards for the Georgia (GA) and Oklahoma (OK) Pre-Kindergarten programs, which were among the first universal state-supported public preschool programs in the US, and are considered to be of relatively high quality (Friedman-Krauss et al., 2019). The teacher qualification requirements in GA and OK are comparable to those in the European programs other than Germany, while the group size and child-staff ratio are higher than those for 3-5 year olds in most of the European cases, but comparable to Germany. There are no readily available succinct statements about curriculum and philosophy for GA and OK, so the table reports information derived from classroom surveys in those two states. Measures of classroom quality for GA and OK are relatively high but somewhat variable across preschools. There appears to be more variation in curriculum in GA and OK than in Europe, but this is based on small samples of classrooms from the two states. Unfortunately, there are no comparable classroom quality data for Europe.

The last row of Table 2 shows information about Head Start quality standards for classrooms with 4-5 year old children. This is useful because as discussed below some US proposals for universal preschool use Head Start standards as a basis for their quality standards. Head Start group size and child-staff-ratio standards are similar to those in GA and OK, but the teacher qualification standards are lower, as only 50% of Head Start lead teachers nationwide are required to have a four year college degree. Head Start philosophy is based on school readiness, broadly interpreted. Data from the Head Start Impact Study shown in the last column indicate considerable variation in structural quality measures, with a notable shortfall in teachers with four year degrees. 71% of classrooms received at least a "good" rating on the Early Childhood Environment Rating Scale ECERS (5 on a 1-7 scale), but only 65% met the child-staff ratio standard.

As noted above, the absence of measures of quality for European preschool settings other than structural measures is a drawback for the purposes of this study. It would be more compelling to argue that there are useful lessons for the US from studies of universal preschool in Europe if measures of classroom and instructional quality in Europe showed that quality is relatively high and uniform. It is widely believed that preschools in Europe are of higher and more uniform quality than in the US, but this belief may be based mainly on the perception that Europe places more importance on preschool, given its much wider availability than in the US. Based on the information in Table 2, we can be reasonably confident in arguing that preschool quality in Europe is higher on average than in the US and is likely more uniform within countries than it is within states in the US, but how much higher and how much more uniform cannot be determined. This is an important caveat to using these studies to draw implications for the US.

The European preschool programs discussed here are universal, but in all cases there has been considerable rationing of slots due to inadequate supply as the programs were implemented and expanded. This provides the basis for the research designs used to estimate the impact of the availability of or enrollment in preschool on child outcomes. There are two general approaches. The first uses "quasi-experiments" in which a federal reform led to a large increase in child care supply, but with substantial variation across municipalities in the rate of growth of supply. In the second approach, there was not a specific reform but the supply of child care was insufficient to serve all eligible children whose families applied for a slot. This approach exploits the fact that slots in oversubscribed centers were allocated by a random or quasi-random method. In all cases, the authors devote considerable effort to documenting the extent to which the variation used to identify and

Table 1
Summary of studies reviewed.

Study	Location, enrollment years, population studied	Treatment age and intensity	Counterfactual care condition	Outcomes analyzed, age at which outcomes are measured, type of data
Fort et al. (2020)	Bologna, Italy, 2001-05, two-earner couples	0-2, full & half day	Home	IQ and personality traits, 8-14, survey (including testing)
Cornellisen et al. (2018)	Lower Saxony, Germany 1994-2006, all families	3, half day	Home	School readiness, 6, administrative
Felfe and Lalive (2018)	Schleswig-Holstein, Germany, 2009-15, married couples	0-2, half day	Home	School readiness, 6, administrative
Felfe et al. (2015)	Spain, 2000-09, all families	3, full day	Family and informal	PISA reading test score and grade retention, 15, administrative
Havnes and Mogstad (2015)	Norway, 1976-83, married couples	3-6, full day	Informal	Earnings and years of schooling, 30-40, administrative
Drange and Havnes (2019)	Oslo, Norway, 2005-10, all families	1-2, full day	Parental and informal	Language and math achievement, 7, administrative
Datta Gupta and Simonsen (2016)	Denmark, 1996-97, ethnic Danish families	2, full day	Family day care	Academic high school track and GPA, 16, administrative

Notes: PISA = Programme for International Student Assessment. GPA = Grade Point Average.

estimate treatment effects is "as good as random." Examples of each of the methodological approaches are briefly described next.

The quasi-experimental approach is well-illustrated by the Havnes and Mogstad (2015) article on Norway. Subsidized child care was scarce in the late 1960s and early 1970s even as labor force participation of mothers was increasing rapidly. A 1975 reform (the Kindergarten Act) significantly increased federal funding for subsidized child care centers for 3-6 year old children. Funding approximately tripled in two years, and the number of available slots increased rapidly. The national coverage rate (slots per age-eligible child) increased from .10 in 1975 to .40 in 1984 and .63 by 1995. Local municipalities were responsible for building and operating new centers or expanding existing centers with subsidies from the federal government. There was considerable variation in the rate of expansion across municipalities. This was partly by design: municipalities with the greatest need received the highest subsidies. To exploit this variation, Havnes & Mogstad ordered the municipalities from lowest to highest according to the percentage point increase in the coverage rate from 1976 to 1979. They split the sample at the median and designated municipalities in the upper half of the distribution as the "treatment" units and those in the lower half as the "comparison" units. The time trends in the coverage rate for the two groups are similar in level and slope before the reform, and then there is a sharp upward divergence for the treatment units, while the comparison units have a much more gradual upward trend. The trends become parallel again after 1979.

This quasi-experiment is the basis for a Difference-in-Differences (DD) analysis: compare the average outcome of interest for cohorts of children in the treated municipalities who could have benefited from the expansion (born 1973-76, the "post" cohorts) to the average outcome for cohorts born too early to benefit (born 1967-69, the "pre" cohorts), and then compare this difference to the corresponding pre-post cohort difference in the comparison municipalities, in which the increase in the coverage rate was much slower. The pre-post comparison accounts for any permanent unobserved differences across municipalities that could be correlated with average child outcomes, by differencing within municipalities. The second difference accounts for any potentially confounding common national trends that could have influenced child outcomes in all municipalities. The authors analyze the determinants of the child care expansion rate across municipalities, and find no evidence suggestive of selection bias. This of course does not mean there was no selection bias: there could have been unobserved factors at work that the authors could not account for.

Other articles using variations on the DD approach include Cornelissen, Dustmann, Raute, and Schoenberg (2018), Felfe and Lalive (2018), and Felfe, Nollenberger, and Rodríguez-Planas (2015). Datta Gupta and Simonsen (2016) exploit differences across municipalities in whether they guarantee access to a preschool slot rather than guaranteeing a slot in either preschool or family day care. They rely on a cross section sample, which requires stronger assumptions than in the other studies. Specifically, the assumption is that selection of municipalities according to whether they guarantee access to preschool must be independent of child outcomes conditional on observable variables.

To illustrate the quasi-random rationing approach, consider the analysis of the Bologna Daycare System (BDS) in Italy by Fort, Ichino, and Zanellan (2020). Applicants to the BDS submit a list of centers ranked in order of their preferences. The system administration assigns each applicant to a priority group based on factors such as marital status and special needs. Within each priority group, applicants are ranked according to a "Family Affluence Index" (FAI) that is an increasing function of income and wealth. Applicants are ranked by FAI within priority groups, and are assigned to their most preferred center in increasing order of FAI, starting from the lowest-FAI applicant, until all slots in a given center are filled. Remaining applicants are then assigned to one of their lower-ranked alternatives or may not receive a slot. The threshold value of FAI that determines who is admitted and who is denied admission is center-and-year-specific, depending on the center's age-specific vacancy rate in a given year and other difficult-to-predict idiosyncratic factors.

There were no statistically significant differences within priority groups in pre-treatment household characteristics between households above and below the FAI admission threshold, for households relatively close to the threshold. These balancing tests support the "as good as random" assumption. This setup serves as the basis for a Regression Discontinuity (RD) analysis in which outcomes for children who were offered a slot in their most preferred center (because their FAI was below that center's threshold) are compared to outcomes of children whose FAI was above the threshold of their most preferred center, controlling for a continuous function of the FAI. Another version of this approach exploits a lottery used in Oslo to allocate slots for infants and toddlers in oversubscribed centers (Drange & Havnes, 2019).

Table 2
Information about preschool quality in studies reviewed.

	Max CSR	Max GS	Required lead teacher qualifications	Curriculum and educational goals	Uniformity of quality across centers
Fort et al. (2020), Bologna	4:1 at age 0 6:1 at ages 1-2	10[a]	3 or 4 year university degree[a]	"Essential elements of educational provision: personal relations between peers and with adults; enhancement of playing in all its forms and expressions; emphasis on productive making, and direct contact experiences with nature, things, materials, social environment, culture." [a]	Guidelines for programs are set at the central level with little autonomy left to the different facilities. Standards concerning goals and daily planning of educational activities, and the number of teachers and square meters per child are strictly enforced.
Cornellisen et al. (2018), Lower Saxony	25:2 Median observed: 9.4:1	25[a]	2-year state-certified vocational program, followed by a 1-year internship as a child care teacher.	Personal and emotional development, social skills, cognitive abilities and positive attitudes toward learning, physical development, creative development, and language and communication skills.	"Overall, these standards lead to a relatively homogeneous child care environment compared to, for example, the United States."
Felfe and Lalive (2018), Schleswig-Holstein	5:1	10	Two years of theoretical training and at least two years of practice in a care center.	Develop children's analytical, language, and motor skills. Center staff engage children in playful activities, such as circle play, reading, painting, or physical activities	"Care centers tend to comply with these regulations: over the period under study, groups accommodated, on average, 10.1 children, the average child-staff ratio was approximately 3:1, and 61.9% of the employed staff had a degree in ECC education."
Felfe et al. (2015), Spain	NA	20	College degree in pedagogy.	Emphasis on play-based education, group play, and learning through experiencing the environment. Child masters her body and understands movement possibilities. Interacts with others in a variety of contexts and modes. Forms good relationships with adults and peers and understands that people have different needs, views, cultures, and beliefs. Personal autonomy in the child's usual activities.	The reform provided federal provisions for the first time in Spain regarding educational content, group size, and staff skill composition for children 3 to 5 years old.
Havnes and Mogstad (2015), Norway	16:2 (1 asst.) or 16:3 (2 assts.)	NR[a]	College degree, including supervised practice in a formal child care institution.	Children should develop social, language, and physical skills mainly through play and informal learning integrated with the day-to-day social interaction between children and staff, and combined with some age-specific activities.	Subsidized child care was quite homogenous with respect to observable characteristics such as group size, staff-child ratio, and employee training.
Drange and Havnes (2019), Norway	10:3 (teacher plus two assistants	10	4-year degree, including supervised practice in a formal childcare institution.	Same as for 3-6 year olds in Norway	Same as for 3-6 year olds in Norway

Table 2 (*Continued*)

	Max CSR	Max GS	Required lead teacher qualifications	Curriculum and educational goals	Uniformity of quality across centers
Datta Gupta and Simonsen (2016), Denmark	7:2 (teacher plus one asst.)	NR[a]	University degree in teaching, specializing in young children.	Development of personal, linguistic, and physical skills. Develop understanding of nature and culture. Focus on socialization rather than on a basic skills curriculum.	Municipality provides care, yet the specific details are decided by the institutions.
Georgia, 4 year olds	10:1	20	BA in ECE	The state has approved a number of curricula for use in Georgia Pre-K classes. The most frequently used was High/Scope, followed by Creative and High Reach.[b]	Mean scores on Assessment profile (range 0-48): 35.2 (coefficient of variation .14). Mean score on ECERS (range 1-7): 4.7 (cv .14).[b]
Oklahoma, 3-4 year olds	11:1	22	BA in ECE	Tulsa's pre-K programs . . . exceed their counterparts in other states in both amount and quality of instruction to which young children are exposed. Tulsa pre-K teachers devoted substantially more time to virtually every academic activity as compared to teachers in other states. [c]	There was substantial variation across classrooms in scores on classroom assessments. Coefficients of variation ranged from .11 to .29.[c]
Head Start, 4-5 year olds	10:1	20	At least 50% of lead teachers must have a BA in ECE or CD.[d]	The goal of Head Start is school readiness. Essential ingredients are physical, cognitive, emotional, and social development. Programs must establish school readiness goals in approaches to learning, social and emotional development, language and literacy, cognition, and perceptual, motor, and physical development.[e]	Classrooms with four year olds: 29% of teachers had a BA or higher. 31% had AA as highest degree. 65% had CSR that meets APHA standard. 71% had ECERS-R total score at least 5 (on a 1-7 scale). 41% of teachers received 25 hours of training in past year.[f]

Notes: Unless otherwise noted, data are from the articles. CSR = child staff ratio. GS = group size. NR = not regulated. NA = not available. HSD = high school diploma. BA = bachelor's degree. ECE = early childhood education. ECC = early child care. CDA = child development associate. APHA = American Public Health Association. CD = child development.

[a] Item not reported in the article. Source: OECD (2006).

[b] Source: Henry et al. (2004), p. 7.

[c] Source: Phillips, Gormley, and Lowenstein, (2009), pp. 219, 225-6.

[d] Department of Health and Human Services, Administration for Children and Families, Head Start Program Performance Standards, 45 CFR Chapter XIII RIN 0970-AC63.

[e] eclkc.ohs.acf.hhs.gov/school-readiness/article/head-start-approach-school-readiness-overview.

[f] U.S. Department of Health and Human Services, Administration for Children and Families (January 2010). Head Start Impact Study. final report. Washington, DC. Page 3-5.

3. Results

3.1. Average effects

Table 3 presents estimates of treatment effects from each article. The third column shows that six of the seven articles report statistically significant (at the 5% level) beneficial treatment effects on at least one outcome. Six report beneficial and statistically insignificant effects for at least one outcome as well. One reports a statistically significant *adverse* effect on one outcome. The terms beneficial and adverse are used instead of positive and negative, because for some of the outcomes a higher score indicates a worse outcome. The statistically significant beneficial effects are for motor skills at school entry for children enrolled in preschool at ages 1-2 in Germany, the Programme for International Student Assessment (PISA) reading test score at age 15 for children enrolled at age three in Spain, completed years of schooling for children enrolled at ages 3-6 in Norway, language development at age seven for children enrolled at ages 0-2 years in Norway, and enrollment in the academic high school track and Grade Point Average (GPA) at age 16 for children enrolled in preschool at age two in Denmark, relative to children enrolled in family day care. The negative statistically significant effect is for IQ at ages 8-14 for children enrolled at ages 0-2 in the Bologna study.

The Bologna study reports the proportional effect of an additional month of enrollment in child care caused by receiving an offer from the most preferred center. The estimated effect of -.005 on IQ indicates a .5% reduction in IQ per additional month in care. The average increase in center care caused by receiving an offer from the most preferred center was six months, so the Average Treatment Effect (ATE) estimate implies an IQ reduction of 3.0%, equivalent to three points on the IQ scale. The personality effects in the Bologna study are all adverse and of a similar order of magnitude as the IQ effect, but are statistically insignificant. The counterfactual form of child care in this case is almost exclusively home care by parents or grandparents. The authors note that Bologna is a relatively high-income city, and the typical home environment is likely to be relatively good in developmental terms. Thus, preschool for infants and toddlers tends to replace high-quality home care with a better adult-child ratio. This may explain the adverse effects of preschool in this setting.

In the German case, the ATE of child care exposure at ages 3-6 is an increase in an index of school readiness of about six percentage points (pp) at age six, on a base of 91%, not significantly different from zero (Cornellisen et al., 2018). Child care at ages 0-2 in Germany yields beneficial effects on each of the three components of the age-six school readiness index - language, motor skills, and socioemotional development - of which the effect for motor skills is relatively large (.26 SD) and statistically significant (Felfe & Lalive, 2018). The counterfactual type of care in this case is also mainly home care.

In the Norwegian studies, Havnes and Mogstad (2015) find that child care at 3-6 has a precisely estimated effect of zero on adult earnings and a small positive and statistically significant effect of .07 on completed years of schooling. Drange and Havnes (2019) find that child care at ages 0-2 has positive effects on language and math skills at age 7, with the effect on language significantly different from zero. The counterfactual is parental and informal care.

In Spain, Felfe et al. (2015) report that living in a high-expansion-rate municipality at age 3 had a positive and statistically significant Intent-To-Treat (ITT) effect of .15 SD on the age-15 PISA reading score, compared to the counterfactual alternative of family and informal care. There was also a decline in the probability of grade retention of about 3 pp, statistically significant at the 10% level in the primary grades. The ITT effect measures the effect of living in a high-growth municipality, rather than the effect of enrollment, which is measured by the ATE. This study estimates ITT effects because they lack data on which children were actually enrolled in preschool.

In Denmark, enrollment in preschool at age two caused a statistically significant increase of 10 pp at age 16 on pursuing the academic high school track. There were also positive effects on GPA in Danish (.23 SD) and math (.10 SD), with the former statistically significant (Datta Gupta & Simonsen, 2016).

To summarize, the average effects of universal public preschool on child outcomes across the seven studies are generally beneficial. The beneficial effects range from very small (.07 increase in years of schooling) to moderate (.10–.16 SD increase in math and reading achievement) to fairly large (26 pp decrease in the incidence of motor skill problems, and a 10 pp increase in academic high school enrollment track). In some cases these effects are long lasting (PISA reading score at age 15, academic high school track at age 16, and completed years of schooling).

3.2. Effects by socioeconomic status

The last column of Table 3 reports evidence on heterogeneity in the treatment effects across groups defined by various measures of socioeconomic status (SES). Six of the seven articles report at least one outcome for which the estimated treatment effects are more beneficial (or less harmful) for lower-SES than for higher-SES children (the exception is Felfe & Lalive, 2018). In five of these cases at least one of the differences in effects by SES is itself significantly different from zero (not shown in the table; see the table note for details). In one case, the effects are of similar magnitude across SES groups and both estimates are statistically significant (Felfe & Lalive, motor skills). In several cases the effects are of similar magnitude across SES groups but are statistically insignificant, as are their differences.

Four of the seven articles report effects disaggregated by maternal or parental education. In three of these cases, there is at least one instance of a more beneficial or less harmful effect for children of lower educated parents, with two of these differences statistically significant. In two cases, there is at least one outcome for which the effects are very similar across different levels of maternal education. Three articles report effects by family income or affluence. In all three cases, the effects are more beneficial or less harmful for children in lower-income or less-affluent families, and several of the differences are themselves statistically significant. One article reports effects separately for immigrants and natives and finds similar effects for the two groups. Another article finds more beneficial effects for ethnic Turks than for ethnic Germans (Cornellisen et al., 2018), a difference that is statistically significant.

In other results not shown in the table, Havnes and Mogstad (2015) report estimates of treatment effects according to a child's location in the (untreated) distribution of the outcome measure (quantile treatment effects). They report a positive treatment effect on adult earnings at the bottom of the earnings distribution, which declines slightly from the 10^{th} to the 60^{th} percentile of earnings, and then drops sharply, becoming negative beyond the 80^{th} percentile. Thus, children who would be in the lower part of the distribution in the absence of treatment receive the largest benefit from the program. Three of the studies report treatment effects across the entire distribution of the propensity score (the probability of treatment conditional on observables). This approach estimates Marginal Treatment Effects (MTE), which can differ for each child as a function of the propensity score. This is useful because it illustrates how the treatment effect varies by the estimated likelihood of treatment. Most of the articles report that more disadvantaged children, as measured by parental education, income, etc., are less likely to enroll in public preschool. And more disadvantaged children tend to benefit more from treatment. The MTE distribution

Table 3
Summary of results of studies reviewed.

Study, location, treatment intensity, age at treatment, and age at outcome	Source of identification, estimation method, parameters estimated	Estimated Effects (standard errors) and units of measurement	Heterogeneity of effects	
Fort et al.,	Rationing in over-subscribed centers	SD per month of child care	Less affluent	More affluent
Bologna		IQ: **-.005 (.002)**	-.003 (.003)	**-.009 (.003)**
		Openness: -.004 (.003)	.001 (.007)	**-.014 (.005)**
Increase of 6 months of care, 4-36 months,	RD, IV	Conscientious: .000 (.004)	.007 (.006)	-.001 (.006)
8-14 years	ATE	Extraversion: -.006 (.004)	-.011 (.007)	-.006 (.005)
		Agreeableness: -.004 (.003)	.003 (.006)	**-.012 (.006)**
		Neuroticism: .002 (.003)	-.005 (.006)	.009 (.005)
Cornellisen et al.,	VMRE, 1992 reform, DD, MTE, ATE	School readiness probability, ATE: .059 (.072)	TT: -.051 (.082) TUT: **.173 (.085)** Positive effect for ethnic minorities was .117 (.035) pp larger than for ethnic Germans.	
Germany, 21 pp increase in coverage rate, 3, 6				

Felfe & LaLive, Germany, 20 pp increase in cover-age rate, 1-2, 6	VMRE, 2005 reform	Fraction without a problem in:	Maternal education		Nativity	
	DD		LTC	CG	Immigrant	Native
	MTE, ATE	Language: .09 (.07)	.10 (.07)	.07 (.08)	.15(.09)	.07 (.07)
		Motor development: **.26 (.06)**	**.24 (.06)**	**.28 (.06)**	**.22(.08)**	**.26 (.06)**
		Socioemotional devel.: .06 (.06)	.05 (.06)	.08 (.07)	.07(.07)	.06 (.06)

Table 3 (*Continued*)

Felfe et al., Spain, 26 pp increase in coverage rate, 3, 15	VMRE, 1991-2 reform, DD, ITT	PISA Reading **.15 (.07)** SD PISA Math .05 (.05) Fell behind grade in: primary: -.027 (.014) pp secondary: -.032 (.036)			Which parents have a secondary degree		
					Neither	At least one	
			Reading		**.17 (.06)**	.11 (.07)	
			Math		.04 (.07)	.03 (.05)	
			Falling behind primary		-.04 (.03)	-.01 (.01)	
			Falling behind secondary		-.04 (.06)	-.02 (.03)	
Havnes & Mogstad, Norway, 18 pp increase in coverage rate; 3-6, 30-40	VMRE, 1975 reform DD, QTE, ITT	Average annual earnings (000NK): .33 (1.60). mean: 362. Years of schooling: **.07 (.02)**			Parental family income		
					Low	Mid	High
			Earnings		**9.3 (3.3)**	−1.8 (3.4)	−8.1 (4.6)
			Schooling		**.24 (.04)**	**.08 (.04)**	.02 (.04)
Drange & Havnes, Norway, lottery offer, 0-2, 7	Random assignment in over-subscribed centers, OLS, IV ITT	Achievement scores: Language: **.16 (.05)** SD Math: **.11 (.05)**		Parental Education			
				Low	High	Low	High
			Language	**.24 (.12)**	.08 (.07)	**.26 (.12)**	.08 (.07)
			Math	.11 (.11)	.04 (.08)	**.26 (.11)**	.01 (.08)
Datta Gupta & Simonsen, Denmark, 42 pp increase in center enrollment, 2,16	Compare municipalities with & without access to GAPS IV, LATE	Academic high school track: **.105 (.023)**. mean: .65 (PP) GPA (SD): Danish: **.23 (.04)** Math: .10 (.07)			Maternal education:		
					≤HSG	>HSG	
			Academic high school track:		**.18 (.03)**	**.06 (.02)**	
			Danish GPA:		**.32 (.06)**	**.17 (.04)**	
			Math GPA:		.15 (.09)	.06 (.06)	

Sources: Fort et al.: Tables 6, 9. Cornellisen et al.: Table 7 (see the article for derivation of TT and TUT estimates from MTE estimates). Felfe and LaLive: Table 6 (moderate reform). Felfe et al.: Tables 4, 8. Havnes and Mogstad: Table 2. Drange and Havnes: Tables 5, 6. Datta Gupta and Simonsen: Table 4.

Notes: The table reports parameter estimates and standard errors from each study. Figures in **bold** are statistically significant at the 5% level. The alternative to preschool care is home or informal care in all cases except Datta Gupta and Simonsen. In their analysis the alternative to preschool care is care in a regulated Family Day Care Home and the sample excludes children cared for at home.

The text discusses the statistical significance of the difference in treatment effect estimates across SES groups. The standard errors of the difference in estimates across SES groups were not reported in four of the papers. These were computed by the author, assuming no covariance in the parameter estimates. The calculations are not reported here. The results of the calculations indicate that there were statistically significant differences in treatment effects across SES groups for at least one outcome in the following cases (level of significance in parentheses): Fort et al. (IQ, Openness, Agreeableness, Neuroticism, all at the 10% level), Cornellisen et al. (5%), Havnes and Mogstad (5%), Datta Gupta and Simonsen (academic high school track, 5%, and Danish GPA, 10%).

Abbreviations: VMRE = variation across municipalities in rate of expansion of child care coverage. RD = regression discontinuity. DD = difference in differences. IV = instrumental variables. OLS = ordinary least squares. PP = probability points. SD = standard deviations.

ATE = Average Treatment Effect. ITT = Intent to Treat. TT = Treatment on the Treated, derived from MTE. TUT = Treatment on the untreated.

MTE = Marginal Treatment Effects. QTE = quantile treatment effects. HSG = high school graduate. LTC = Less than College. CG = college graduate. GAPS = guaranteed access to preschool (GAPS). NK = Norwegian Kroner.

allows similar comparisons to be made based on *unobserved* factors that influence take up and treatment effects. In most cases, a similar pattern is found: children who for unobserved reasons (conditional on observables) would benefit the most from treatment are the least likely to enroll.

To summarize, the evidence from Europe indicates that universal preschool programs for the most part have larger beneficial effects on disadvantaged children than on more advantaged children. The treatment effects for disadvantaged children are beneficial in most cases and are never adverse and statistically significant, while the treatment effects for more advantaged children are beneficial and statistically significant in only four of the 22 estimates in the table.

4. Comparison to Evaluations of Universal Preschool in the US

There has been a considerable amount of research evaluating the effects of state-funded universal preschool in the US, but much of it is not as informative as the evidence from Europe due to limited opportunities for research designs that are well-suited to estimating causal effects (Phillips et al., 2017). The most common research design used to study the impact of state-funded universal preschool in the US is to compare outcomes of children who were eligible for a program in a given year, thanks to having birthdates before the eligibility cutoff date imposed by the state, to outcomes of children whose birthdates fell after the cutoff date. For example, in Oklahoma a child must turn four by September 1 in order to enroll in public preschool for four-year olds. This approach is implemented using an RD design, which allows for a discontinuous change in outcomes at the eligibility cutoff date while controlling for a smooth function of birthdate or age (Fitzpatrick, 2008; Gormley, Gayer, Phillips, & Dawson, 2005; Weiland & Yoshikawa, 2013). This approach can provide credible causal estimates of short run impacts at end of the preschool year. However, the comparison group in these studies becomes eligible for treatment in the following year, and it is unlikely that longer run effects of receiving the same treatment one year apart can be detected with much precision. To my knowledge, no study has used this approach to study outcomes beyond the end of preschool attendance. This is a significant limitation in view of the well-documented fade out of short run effects.

Another approach, sometimes combined with the birthdate eligibility cutoff, is a more traditional DD design, comparing outcomes of children in states with and without state-funded preschools, before and after the implementation of the state program. However, there is considerable heterogeneity across states in preschool program quality standards (Friedman-Krauss et al., 2019). Thus the treatment effect estimated from cross-state comparisons is difficult to interpret, as it reflects the weighted average quality of state programs in the sample, rather than the quality of a specific state program. This issue can be avoided by using only one or two states known to have high-quality standards as the treatment group. This is the approach used by Cascio and Schanzenbach (2013), with Georgia and Oklahoma as the treatment states. However, the problem reappears in another guise when choosing a group of comparison states: should they be states with no preschool program or all other states, including those with lower-quality programs and targeted programs? The same problem of interpretation arises depending on the choice of comparison states.

Cascio and Schanzenbach (2013) provide the most credible estimates using this approach, and it is useful to compare their results to those from the European studies. They use all states other than GA and OK as a comparison group in a DD analysis comparing outcomes of children in GA and OK before and after the introduction of public preschool to pre-post outcomes for the same birth cohorts in other states. Outcomes are measured by fourth and eighth grade scores on the National Assessment of Educational Progress (NAEP). They estimate effects separately for low-income (eligible for free or reduced-price lunch) and other children. Their main findings are reproduced in Table 4, in the form of ITT effects. They show positive and statistically significant effects on reading and math test scores for low-income children in fourth grade, which partially or fully fade out by eighth grade. There are no positive and statistically significant effects for higher-income children, and three of the four estimated effects are actually negative. Translating the fourth and eighth grade effects on math for low-income children into SD units yields effect sizes of .09 and .06, respectively.

Table 4
Results from Cascio & Schanzenbach's (2013) study of the effects of Georgia and Oklahoma Pre-K programs on NAEP test scores.

	Family Income	
Fourth grade NAEP scores	Low	High
Reading	**3.1 (1.2)**	−1.6 (.8)
Math	**3.1 (1.1)**	.9 (.7)
Eighth grade NAEP scores		
Reading	.8 (1.9)	-.8 (2.3)
Math	2.1 (1.1)	−1.3 (.8)

Source: Cascio and Schanzenbach (2013), Tables 6 and 7, column 5. National Assessment of Educational Progress (NAEP) test score units. The NAEP low income eighth grade math coefficient of 2.1 is equivalent to .06 SD. The treatment states are GA and OK. Control states are all others. Low and high income are defined by eligibility for free or reduced price school lunch.
Notes: Standard errors are in parentheses. **Bold** indicates a statistically significant treatment effect at the 5% level. The study reports Intent to Treat (ITT) effects.

There is no universal nationwide preschool program in the US, but during World War II the Lanham Act made federal subsidies available to all states to facilitate employment of mothers of children aged 0-12. Herbst (2017) used a DD approach to exploit variation across states in federal funding per child to identify the impact of subsidized child care. This is a methodologically sound approach, similar to several of the European studies. The results showed long run beneficial ITT effects on earnings and education, with the earnings effects concentrated in the lower part of the earnings distribution. However, anecdotal evidence suggested that the quality of care was highly variable across states (Herbst, 2017, p.527), so it is not clear how relevant these findings are for understanding the impact of universal care. The Lanham Act was intended first and foremost to provide care while mothers worked, with child development a secondary concern.

In sum, the best evidence on the effects of universal preschool programs in the US is consistent with the evidence from Europe: generally beneficial effects for children from low-income families and much smaller or no effects for children from higher-income families.

5. Discussion

To illustrate the policy implications of these findings for the US, it is useful to consider two specific US proposals for public preschool that have recently captured attention, one for a universal program and the other for a targeted program. The proposals provide concrete examples, including cost estimates, and are used here purely for illustrative purposes. This discussion is not intended to provide support for either program or for one over the other. Senator Elizabeth Warren's Universal Child Care and Learning Act (Warren, 2019) would establish a network of locally-run child care and early learning centers and family day care homes. Federal funds would be provided only to programs that meet the current standards used

by Head Start and the US military child care program. There would be a sliding scale fee, with no payment required for families with income less than twice the poverty level, and no family required to pay a fee greater than 7% of family income. Teacher compensation would be linked to the pay of equivalently credentialed local public school teachers, thus dealing with the perennial problem of low pay and high staff turnover among preschool teachers. Head Start would be folded into the new program. The net cost to the federal government is estimated to be $70 billion per year.

The other prominent proposal is a bill introduced in Congress in 2017 and reintroduced in 2019 known as the Child Care for Working Families (CCWF) Act (116th Congress, 2019). This is a proposal for a targeted program, providing eligibility to families with income up to twice the State Median Income (SMI). There would be no parental fee for families with income less than 75% of SMI, and no eligible family would be required to pay a fee of more than 7% of family income. Note that despite the use of the phrase "working families" in the title of the Act, eligibility is defined by a child's age and the family's income, and is not limited to children of working parents (116th Congress, 2019, p.67). The proposal sets aside funding for quality improvement, and ties reimbursement rates to program quality as measured by classroom observations. Head Start would be expanded to full day care for the school year and funds are set aside for quality improvement for Head Start. Funding of $20 billion in Fiscal Year (FY) 2020 was proposed, growing to $30 billion in FY21, and $40 billion in FY22. This proposal was endorsed by several candidates for the 2020 Democratic presidential nomination (Kamala Harris, Beto O'Rourke, and Amy Klobuchar).

An important question for public policy is what the $70 billion annual expenditure in the Warren proposal would produce in terms of improved child outcomes beyond the annual average of $30 billion in the CCWF Act. This is an oversimplified comparison, since the quality of care could differ across proposals and there are other relevant differences as well. Nevertheless, the question is legitimate even if we cannot provide more than a speculative answer here. Given the evidence that advantaged children gain relatively little on average in terms of developmental progress from universal public preschool, thanks to the relatively high quality of the alternatives available to them, the additional $40 billion would likely result in modest gains, if any, in child development.

The additional $40 billion in the Warren proposal is to a first approximation an upper middle class subsidy. Median household income in the US in 2018 was $63,179 (Semega, Kollar, Creamer, & Mohanty, 2019), so the proposed CCWF Act's income eligibility limit of twice state median income is $126,358, approximately the 78th percentile of the US household income distribution. Thus, part of the additional $40 billion in the Warren proposal would provide subsidies to families in the top 22% of the income distribution. There may be justifications for such subsidies, but based on the evidence discussed in this review it is unlikely that they could be rationalized as a socially efficient investment of public funds for the purpose of improving child outcomes, even if there are some small developmental benefits to the most advantaged children.

This conclusion, which is conditional on the assumption of equivalent program quality in the two proposals, is highly speculative, and is intended to highlight the tradeoff between universal and targeted programs. If some of the funds in the Warren proposal were used to increase the quality of preschool beyond the level provided by the CCWF Act, there would likely be some additional developmental benefits to disadvantaged children, which would increase the social efficiency of the proposal.

The findings discussed here support two general conclusions. First, evidence from quasi-experimental studies in Europe on the average causal effects of universal preschool on child outcomes show that these effects are generally beneficial, but are neutral in some cases and occasionally harmful. Second, most of the studies find larger beneficial effects for disadvantaged children and more modest or no effects for more advantaged children. Disadvantage is defined in a variety of ways across the studies, based on income, parental education, race and ethnicity, language spoken at home, and other metrics. These differences in program effects across socioeconomic groups are likely due in part to systematic differences in the nature of the alternative care arrangements available to children who do not have access to a public preschool. Preschool provides a boost to children who otherwise would be in a less enriching environment, while children from more advantaged families are more likely to be exposed to a developmentally stimulating environment even in the absence of access to high-quality preschool.

A key question is whether this evidence is informative for the US. A recent "consensus statement" by prominent scholars and policy analysts (Phillips et al., 2017) notes that US public preschool programs are highly variable in characteristics and quality both across and within states, and much of the research that evaluates their effects has methodological problems. Thus they argue that we should not expect definitive evidence from evaluations of these programs. The evidence from Europe discussed here is based on stronger research designs than have been possible in the US, thanks to national quality standards and funding, variation across municipalities in the rate of expansion in response to major reforms, and random or quasi-random rationing. Although the settings are very different from the US, the European studies can be thought of as providing potential best-case scenario evidence on the expected impacts of a universal high quality program in the US with well-enforced federal standards.

A reasonable conclusion based on the evidence discussed in this paper and in the reviews cited above is that the case for public investment in high-quality universal preschool programs *for the purpose of improving child outcomes* is not very compelling on either social efficiency grounds (rate of return on public investment) or equity grounds, in contrast to the case for programs targeted at disadvantaged children. Disadvantaged children benefit from enrollment in high-quality preschool provided through a universal program, but the impacts on children from more advantaged backgrounds are usually smaller and in some cases non-existent or harmful.

If the small benefits received by advantaged children do not justify the costs of providing the services to them, why should the government fund universal rather than targeted programs? One common answer is that these programs also serve as a child care subsidy, facilitating labor force participation by parents of young children. This is a valid point, although there is no employment requirement in most universal preschool programs, nor in the Warren and CCWF proposals. However, to the extent that such programs serve as a child care subsidy to relatively affluent families, they are not well-targeted. In 2011 families with monthly income of $4500 or more who paid for child care spent 6.7% of family income on child care, compared to 13.3% for families in the $,000–$4499 range, 18.8% for families in the $1500–$2999, and 39.6% for families with income less than $1500 (Laughlin, 2013). Another common answer is that universal programs are more popular than programs targeted at disadvantaged families and therefore garner more political support. A third answer is that part of the effect of universal preschool for disadvantaged children could operate through exposure to more advantaged peers (Cascio, 2017; Gormley et al., 2017). However, universal programs do not necessarily result in greater mixture of children by SES *within* classrooms.

Drawing policy implications for the US from studies of European preschool programs is subject to an important caveat. The impact of universal preschool may depend on the mix of other public programs and services available to families of young children. In addition to heavily subsidized preschool, European countries pro-

vide more generous paid family leave and health insurance than in the US, and other important aspects of the social, economic, and policy environment differ as well. Children in Europe may receive smaller benefits from preschool if they have already benefitted in developmental terms from more time at home with parents during infancy, better and more regular health care, and other social safety net programs. Alternatively, the benefits from high-quality preschool may be enhanced by exposure to other public programs if these programs are complements to rather than substitutes for the services provided by high-quality preschool. For example, food security may be a necessary condition for gaining the potential benefits from a high-quality preschool. In either case, results in Europe may not generalize to a very different social and economic context in the US. Thus inferences for the US from the European experience are only suggestive.

There are two other important limitations of this study that are worth emphasizing. First, as discussed previously, the European studies reviewed here do not have measures of process quality or teacher instructional and emotional support. The structural features of European preschools indicate that the necessary conditions for high quality are present – low ratios and group size, and strong educational requirements. These features are comparable to or better than those of the GA and OK universal preschool programs (for 3-5 year olds), which score well on assessments of classroom quality on average (Gormley et al., 2005; Henry et al., 2004). However, in the absence of measures of classroom and instructional quality in Europe, we cannot be confident that assertions of high and uniform quality in the articles are correct. This underlines the speculative nature of implications drawn from these studies for the US.

Second, one of the main strengths of the European studies – measurement of medium and long terms effects – could also be a disadvantage if the social and economic circumstances in present-day Europe are substantially different from those that prevailed when the children in these studies were enrolled in preschool. This is unlikely to be a significant concern for most of the studies, which measure effects at ages 6-16, but could be relevant for the Havnes and Mogstad (2015) study of outcomes in Norway at ages 30-40. This is a generic concern for studies of long run effects of preschool (for example, see Bailey, Sun, & Timpe, 2018, and Ludwig & Miller, 2007 concerning Head Start; and Elango et al., 2016, concerning Perry Preschool and Abecdarian). Nevertheless, in view of evidence of fadeout of initial cognitive effects but persistent social and economic benefits, understanding long run impacts is crucial.

References

Baker, M. (2011). Universal early childhood interventions, what is the evidence base? *Canadian Journal of Economics*, *44*(4), 1069–1105. http://dx.doi.org/10.1111/j.1540-5982.2011.01668.x

Bailey, M. J., Sun, S., & Timpe, B. (2018). *Prep school for poor kids, the long-run impacts of Head Start on human capital and economic self-sufficiency, mimeo.* University of Michigan., www-personal.umich.edu>Bailey_Sun_Timpe

Cascio, E. U. (2017). *Does universal preschool hit the target? Program access and preschool impacts* Working Paper 23215, March. www.nber.org/papers/w23215. National Bureau of Economic Research.

Cascio, E. U., & Schanzenbach, D. W. (2013). *High-quality preschool education*. pp. 127–178. Brookings Papers on Economic Activity, Fall.

Child Care Aware of America. (2018). *Igniting possibilities, promoting innovation.* www.childcareaware.org

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

Council of Economic Advisors. (2016). Inequality in early childhood and effective public policy interventions. Chapter 4. In *Economic Report of the President*. pp. 153–206. obamawhitehouse.archives.gov/administration/eop/cea/economic-report-of-the-President/2016

Datta Gupta, N., & Simonsen, M. (2016). Academic performance and type of early child care. *Economics of Education Review*, *53*, 217–229. http://dx.doi.org/10.1016/j.econedurev.2016.03.013

Drange, N., & Havnes, T. (2019). Early childcare and cognitive development, evidence from an assignment lottery. *Journal of Labor Economics*, *37*(2), 581–620. http://dx.doi.org/10.1086/700193

Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *Journal of Economic Perspectives*, *27*(2), 109–132. http://dx.doi.org/10.1257/jep.27.2.109

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2016). Early childhood education. In Robert A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States* (Vol. 2) (pp. 235–297). University of Chicago Press.

Felfe, C., & Lalive, R. (2018). Does early child care affect children's development? *Journal of Public Economics*, *159*, 33–53. http://dx.doi.org/10.1016/j.jpubeco.2018.01.014

Felfe, C., Nollenberger, N., & Rodríguez-Planas, N. (2015). Can't buy mommy's love? Universal childcare and children's long-term cognitive development. *Journal of Population Economics*, *28*, 393–422. http://dx.doi.org/10.1007/s00148-014-0532-x

Fitzpatrick, M. D. (2008). Starting school at four, the effect of universal Pre-Kindergarten on children's academic achievement. *The B.E. Journal of Economic Analysis & Policy, Advances*, *8*(1) http://dx.doi.org/10.2202/1935-1682.1897. Article 46

Fort, M., Ichino, A., & Zanellan, G. (2020). Cognitive and non-cognitive costs of daycare for children in advantaged families. *Journal of Political Economy*, *128*(1), 158–205. http://dx.doi.org/10.1086/704075

Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & DiCrecchio, N. (2019). *The state of preschool 2018*. Rutgers Graduate School of Education. nieer.org/state-preschool-yearbooks/2018-2

Gormley, W. T., Jr., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal Pre-K on cognitive development. *Developmental Psychology*, *41*(6), 872–884. http://dx.doi.org/10.1037/0012-1649.41.6.872

Gormley, W., et al. (2017). Universal vs. targeted Pre-Kindergarten, reflections for policymakers. In D. Phillips (Ed.), *The current state of scientific knowledge on prekindergarten effects* (pp. 51–56). Brookings & Duke. www.brookings.edu/research/puzzling-it-out-the-current-state-of-scientific-knowledge-on-pre-kindergarten-effects/

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

Henry, G. T., Ponder, B. D., Rickman, D. K., Mashburn, A. J., Henderson, L. W., & Gordon, C. S. (2004). *An evaluation of the implementation of Georgia's Pre-K program, report of the findings from the Georgia Early Childhood Study (2002–2003)* December. https://ies.ed.gov/ncee/wwc/study/29347. Andrew Young School of Policy Studies, Georgia State University.

Herbst, C. M. (2017). Universal child care, maternal employment, and children's long-run outcomes, evidence from the US Lanham Act of 1940. *Journal of Labor Economics*, *35*(2), 519–564. http://dx.doi.org/10.1086/689478

Hotz, V. J., & Wiswall, M. (2019). Child care and child care policy, existing policies, their effects, and reforms. *ANNALS of the American Academy of Political and Social Science*, http://dx.doi.org/10.1177/0002716219884078, published online November 6

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Science*, *103*(27), 10155–10162. http://dx.doi.org/10.1073/pnas.0600888103

Laughlin, Lynda. (2013). *Who's minding the kids? Child care arrangements: Spring 2011* Household Economic Studies P70-135, April. https://www.census.gov/library/publications/2013/demo/p70-135.html. US Census Bureau.

Ludwig, J., & Miller, D. L. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics*, *22*(1), 159–208. http://dx.doi.org/10.1162/qjec.122.1.159

OECD. (2006). *Starting Strong II, Early Childhood Care and Education, Paris.* www.oecd.org/education/school/startingstrongiiearlychildhoodeducationandcare.htm

Phillips, D. A., Gormley, W. T., & Lowenstein, A. E. (2009). Inside the pre-kindergarten door, classroom climate and instructional time allocation in Tulsa's pre-K programs. *Early Childhood Research Quarterly*, *24*, 213–228. http://dx.doi.org/10.1016/j.ecresq.2009.05.002

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., et al. (2017). *Consensus statement from the prekindergarten task force. The current state of scientific knowledge on prekindergarten effects*. Brookings Institution and Duke Center for Child and Family Policy. www.brookings.edu/research/puzzling-it-out-the-current-state-of-scientific-knowledge-on-pre-kindergarten-effects/

Semega, J., Kollar, M., Creamer, J., & Mohanty, A. (2019). *Income and Poverty in the United States, 2018. Current Population Report P60-266* September. www.census.gov/library/visualizations/2019/demo/p60-266.html.

Warren, Senator Elizabeth. (2019). *Universal Child Care and Early Learning Act* February. https://www.warren.senate.gov/newsroom/press-releases/warren-unveils-universal-child-care-and-early-learning-proposal.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130. http://dx.doi.org/10.1111/cdev.12099

116TH CONGRESS 1ST SESSION H. R. 1364. (2019). *An Act To amend the Child Care and Development Block Grant Act of 1990 and the Head Start Act to promote child care and early learning, and for other purposes.* www.congress.gov/bill/116th-congress/house-bill/1364/text