VIDEN-TIL-VELFÆRD

Notat

Universal Preschool Programs and Long-Term Child Outcomes

A Systematic Review

Jens Dietrichson, Ida L. Kristiansen & Bjørn C. V. Nielsen

Universal Preschool Programs and Long-Term Child Outcomes – A Systematic Review

Universal Preschool Programs and Long-Term Child Outcomes: A Systematic Review*

Jens Dietrichson, Ida L. Kristiansen & Bjørn C. V. Nielsen

VIVE – The Danish Center for Social Science Research

Abstract

This systematic review included 25 studies using natural experiments to estimate the effects of universal preschool programs for children aged 0-6 years on child outcomes measured from third grade to adulthood. Studies comparing preschool with parental, family, or other informal modes of care showed mixed effects on test scores, and on measures related to health, well-being, and behavior. All estimates for outcomes related to adequate primary and secondary school progression, years of schooling, highest degree completed, employment, and earnings indicated beneficial average effects. Three of the included studies calculated benefits-to-costs ratios and found ratios clearly above one. Effects tended to be more beneficial for children with low socioeconomic status, though there were examples of the opposite pattern. Effects were not consistently different for boys or girls. Few studies compared two alternative types of universal preschool programs in terms of long-term outcomes.

Table of contents

1 Introduction

A large body of literature shows that the early childhood environment has a strong impact on long-term child outcomes, including educational attainment, earnings, health, and well-being (e.g., Almond, Currie, & Duque, 2017; Black et al., 2017). Many children spend a substantial share of their early childhood in *preschool programs*; that is, they receive formal pre-primary education and care in facilities outside of their homes. Figure 1.1 shows that the share of children enrolled in preschools has been increasing in recent decades in both developed and developing countries. Public spending on preschools in the OECD countries averaged just over 0.7 percent of GDP, and private child care expenditures were 15 percent of net family income on average (OECD, 2016, 2017). The importance of the early childhood environment for child development and the resources devoted to preschool make the effect of preschool programs an important issue for families and policy makers alike.

Resource intensive and high quality preschool programs targeting highly disadvantaged children and families, such as the Abecedarian and Perry Preschool projects, substantially improve long-term child outcomes (e.g., Campbell et al., 2014; Gertler et al., 2014; Heckman, Pinto, & Savelyev, 2013; Reynolds & Temple, 2008); often with highly beneficial rates of return (e.g.,García, Heckman, Leaf, & Prados, 2016; Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Reynolds & Ou, 2011). Broader, but still targeted, programs, such as Head Start in the United States (US), also have long-term beneficial effects (e.g., Carneiro & Ginja, 2014; Currie & Thomas, 1995; Deming, 2009; Kline & Walters, 2016; Ludwig & Miller, 2007; McCoy et al., 2017; Rossin-Slater & Wüst, 2017). However, the demand for preschool is likely to come from all sorts of families, not just the disadvantaged. Therefore, the results from targeted programs are not sufficient to answer the question of whether and in what form – targeted or universal – governments should support preschool programs. To answer that question, long-term evidence from universal programs including a more general population of children is needed.

If the development of skills is a cumulative, dynamic, and self-reinforcing process, then the early childhood is an especially important period, and investment during this period can have a high rate of return later in life (Cunha & Heckman, 2007). Although universal preschool can be seen as an investment in child development, the effects of such programs are theoretically ambiguous and depend not only on the quality of the program itself, but also on the quality of the counterfactual mode of care. That is, the care the child would have received, if the child had not attended a universal preschool program.

Attachment theory assigns an important role to the connection between the primary caregiver and the child (Flaherty & Sadler, 2011). High quality adult-child interactions and caregiving is the strongest predictor of children's skill development (NICHD Early Child Care Research Network, 2002), and is perhaps the most important aspect of preschool quality (e.g., Barnett, 2011; Sabol, Hong, Pianta, & Burchinal, 2013). Multiple caregivers, as in a preschool setting, may damage the attachment between the primary caregiver and the child, thereby harming the child's development (Belsky, 2001). Moreover, reduced one-to-one adult interaction in preschools compared to parental or family care may be harmful more generally, especially at a young age. However, the counterfactual mode of care to universal preschool is not necessarily one-to-one high quality parental care for all children. Siblings may compete for attention (Bradley & Corwyn, 2002), and insecure attachments to parents may be compensated for by secure attachments to preschool teachers (Goossens & van Ijzendoorn, 1990). Children may further attend low quality informal out-of-home care, if universal preschool is not available.

Figure 1.1 The average share of children enrolled in pre-primary education in OECD countries and developing countries
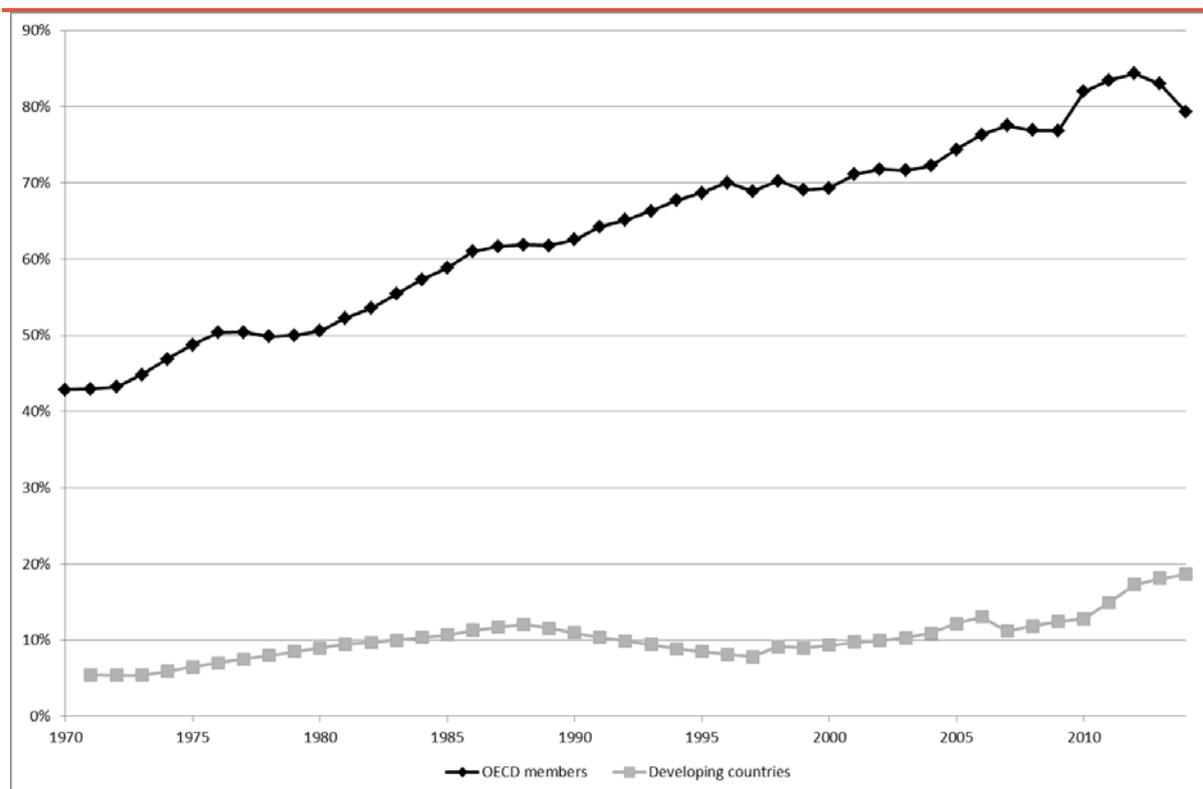
Preschool programs may give the parents, and especially the mother, better labor market opportunities. Because universal programs are often heavily subsidized, they redistribute resources from other tax payers to families with preschool children. Household income may therefore rise. Families with more financial resources can invest more in child development (Elango, García, Heckman, & Hojman, 2015). Furthermore, while child health is likely to be negatively affected in the short term by attending preschool due to the increased risk of infection, the hygiene hypothesis states that such infections may strengthen the immune system and thus have long-term health benefits (Strachan, 1989). Similarly, socializing with other children and adults may have short-term harmful effects, but be beneficial in the long run (Baker, Gruber, & Milligan, 2008). The latter two hypotheses underline the importance of examining long-term outcomes.

Furthermore, the effects may be heterogeneous in terms of SES and gender, as the quality of the counterfactual mode of care or the quality of the preschool program may differ among groups of children, and initial skill levels may matter for how much children benefit from education and care.

Regarding the counterfactual mode of care, parents with higher income or education may give their children a better home environment, and may live in neighborhoods that are more conducive to educational achievement and job market success (e.g., Björklund & Salvanes, 2011; Bradley & Corwyn, 2002; Hart & Risley, 2003). The quality of the counterfactual mode of care may also differ according to gender, as, for example, the home environment seems less stimulating for boys in the US (Bertrand & Pan, 2013) and for girls in many low and middle income countries (Costa, da Silva, & Victora, 2017).

In some contexts, the quality of the same preschool program may differ among groups of children. For example, being exposed to high SES peers may have beneficial effects for low SES

children (Cascio, 2017; Henry & Rickman, 2007), and the quality of the adult-child interaction may depend on whether the teacher and the child are of the same gender (Bauchmüller, Gørtz, & Rasmussen, 2014).

For these reasons, we may expect universal preschool to have more beneficial effects for low SES children, while gender differences should be context dependent. However, if skills produced at one stage of childhood raise the productivity of investment in subsequent stages – if child development exhibits dynamic complementarities (Cunha & Heckman, 2007) – this may cause heterogeneous effects. If we see the difference in quality between universal preschool and the counterfactual mode of care as an investment, then dynamic complementarities may imply that, all else equal, if the preschool program has better quality than the counterfactual mode of care, then the effects would be most beneficial for children with an initial skill advantage. If the program has a lower quality, the effects would be most harmful for children with an initial skill advantage. That is, all else equal, the absolute magnitude of the effects would be largest for children who start preschool with the highest level of skills.

Which groups of children are likely to have an initial skill advantage? If high SES parents provide care of a higher quality, as discussed above, high SES children are likely to develop a skill advantage early on. However, it is unclear whether the dynamic complementarities are strong enough to offset the other reasons why low SES children are generally expected to gain more from preschool. Because a number of studies indicate that girls develop faster than boys in domains like vocabulary and socio-emotional skills, girls are more likely to have an initial advantage (Magnuson et al., 2016), at least in countries without substantial gender bias against girls.

Summing up this discussion, there are few clear-cut predictions regarding any of our research questions. Several recent reviews included analyses of the long-term effects of universal preschool programs (e.g., Almond et al., 2017; Baker, 2011; Cascio, 2015; Elango et al., 2015; Melhuish et al., 2015; Phillips et al., 2017; Ruhm & Waldfogel, 2012; van Huizen & Plantenga, 2015; Waldfogel, 2015). The key message from these reviews is that the evidence is mixed for the general population of children, and that universal preschool has more beneficial effects for children that are in some way disadvantaged (often in terms of SES). However, few reviews included more than a handful of studies with adulthood outcomes, and most of them did not provide analyses of separate outcome categories. The results regarding gender differences were scarce and not consistently in favor of either boys or girls (see also Magnuson et al., 2016, for similar results in a review of mainly targeted programs).

1.1   The Present Study

We reviewed the literature on the effects of universal preschool programs on child outcomes from third grade to adulthood. Included studies compared attending universal preschool programs to parental, family, and other informal modes of care, or alternative universal preschool programs. In countries and states where a substantial share of the children already attend universal preschool programs, the important question for parents may not be whether to place their children in preschool or not, but which type of preschool they should choose (Datta Gupta & Simonsen, 2012). We were therefore interested in comparing the long-term outcomes of alternative types of universal preschool programs, for example in terms of their ownership (e.g., private/public or for-profit/non-profit) or pedagogical approaches (e.g., Reggio Emilia/Montessori). For both study types, we examined whether there were differences in the effects according to socioeconomic status (SES) and gender.

We used systematic review methods to maximize our chances of finding all relevant studies and to increase the transparency of our analyses and conclusions. Included studies used natural experiments to obtain a credible identification of the effects of universal preschool programs. Just

including a variable measuring preschool attendance or exposure to a universal preschool program would likely yield biased estimates, as families and children differ in terms of, potentially unobserved, characteristics that influence the attendance decision, where to live, and child outcomes. In successful randomized and natural experiments, the assignment of treatment is unrelated to both observed and unobserved family and child characteristics, and they thus avoid this type of bias. We focused on these research designs for this reason, but found no randomized experiments. Included outcomes were not limited in any other way than measurement timing, and we analyzed the following outcome categories: health, well-being, and behavior; test scores and school grades; primary and secondary school progression; years of schooling and highest grade completed; employment and earnings; and benefit-cost analyses (BCA). Our research questions were:

1. What are the effects of universal preschool programs compared to parental, family, or other informal modes of care on child outcomes measured from third grade to adulthood?
2. What are the effects of alternative types of universal preschool programs on child outcomes measured from third grade to adulthood?
3. Are the effects different for a) children from families with high and low SES, and b) boys and girls?

2   Method

This section outlines the inclusion criteria and how we located and analyzed relevant studies.

2.1   Inclusion Criteria

We included studies that had at least one estimate of the effect of a program that met all seven of the following inclusion criteria.

Primary empirical studies: We excluded reviews, comments on research, and theoretical papers from the analysis.

Preschool programs: Included studies examined preschool programs; that is, formal out-of-home education and care that children attend before they start primary school. In most countries (e.g., the US, Denmark, and Sweden), Kindergarten (or preschool class or grade 0) is a part of primary school, and studies that exclusively examined kindergarten in such countries were excluded. Some studies met parts of this criterion: Haimovich Paz (2015) examined the early kindergarten movement in the US, which operated at a time when kindergarten was not a regular part of primary school and the content of the program was more like contemporary preschool programs than primary school. The participants in the program studied by Herbst (2017) were both preschool and school children. We included both these studies, as omitting them seemed more likely to bias the results than including them.

Universal programs: The preschool programs should be universal; that is, not targeted at a specific group of children. Studies of programs targeting selected groups, such as Head Start, were excluded. This criterion did not imply that all or even a large share of children in an area had to attend preschool, only that the program under study should be open to children from the general population.

Long-term child outcomes: We included studies reporting child outcomes in third grade or later. We included all types of long-term child outcome measures, such as grade point averages (GPA), standardized academic tests, measures of health and well-being, behavior, social skills, school attendance, labor market outcomes, and crime rates. Studies reporting only parental or family (including sibling) outcomes were excluded.

Type of comparisons: Included studies compared outcomes between children attending or being more exposed to formal preschool programs and children in modes of family or informal care (e.g., care by parents, relatives, or nannies). We also included studies that compared groups of children receiving care and education in alternative types of preschool programs. Type of preschool could be defined in terms of, for example, the ownership status of preschool (private/public) or the pedagogical approach. We excluded studies of interventions in existing preschools, where part of a preschool program was changed for some children (e.g., a changed staff-to-child ratio), for example, or where preschool teachers or managers got professional development.

Country, period, publication status, and language: We did not restrict inclusion by country, time period studied, or publication status of the study (i.e., we included studies not published in scientific journals). However, we limited the search period backwards in time to 1980 and included only studies written in a language that at least two members of the research team understood (Danish, English, German, Norwegian, and Swedish).

Estimation methods: We included studies that estimated the effects of preschool programs by comparing a treatment group to a control group, or two alternative treatments against each other, and where the assignment of treatment was made by randomization or some form of natural experiment (although we found no randomized field experiments). In a natural experiment, the assignment of treatment occurred through some form of "natural" (or administrative) process, which

was outside the control of researchers and attempted to mimic the assignment in a randomized experiment, in the sense that the assignment was unrelated to observable and unobservable characteristics of the participating children (see e.g., Cascio, 2015; Ruhm & Waldfogel, 2012; van Huizen & Plantenga, 2015, for reviews using a similar criterion).

To further illustrate how we applied the criteria, we provide examples of excluded studies in the appendix along with a explanation of why they were excluded.

2.2   Search Strategy, Screening, and Coding

We searched the following electronic databases for relevant studies: EconLit, ERIC, PsycINFO, Academic Search, Teacher Reference Center and SocIndex. All searches were performed in EB-SCO-host in November 2017 and were limited to 1980-2018. We present search documentation for all databases in the appendix. In addition to the search of electronic databases, we used the reference lists of included studies and the associated reviews mentioned in the introduction for citation tracking.

We screened unique identified records from the electronic databases using the title and the abstract to exclude irrelevant records. We first piloted the inclusion criteria on 100 studies, until we reached at least a 95 percent agreement between all three screeners (the first two authors and a research assistant). We obtained and screened records that we did not exclude in the first level screening in full text. At least two screeners performed both levels of screening for each study independently. In the case of differences in the assessment, a third screener decided. The first and the second author extracted information from included studies about, for instance, the preschool program, the estimation method, and the effect estimates. We resolved discrepancies by discussion, and it was possible to reach a consensus in all cases.

2.3   Analysis

In the analysis, we used the estimates from the specification designated as the preferred one by the studies, as long as this specification met our inclusion criteria. If a study did not indicate a preferred specification, we used the one with the lowest risk of bias according to our assessment. If there were effects using the same outcome measure but estimated at different ages, we reported the estimate for the oldest children. Some included studies examined the same programs and used (partly) overlapping samples. When they also reported the same outcome measures, we only included one in the analysis to avoid double-counting. We chose the study that provided the most information (e.g., had a larger sample) or had the lowest risk of bias. The section *Included Estimates* in the appendix contains a detailed motivation for each of these cases.

In studies that did not have access to data for individual preschool attendance, we reported intention-to-treat (ITT) estimates of living in an area that was (more) exposed to the universal preschool program. Some of these studies also reported treatment-on-the-treated (TOT) estimates, calculated by scaling the ITT estimate with an estimate of the probability of being treated. To be unbiased, TOT estimates require that the scaling-up of preschool programs did not change the type of children attending or the preschool quality, and that there were no spillover effects on children that grew up in a treated area but did not attend preschool (see e.g., Baker, Gruber, & Milligan, 2015; Havnes & Mogstad, 2011; and van Huizen, Plantenga, & Dumhs, 2017, for discussions). As it was unclear whether these assumptions were met, we reported the ITT estimates.

To make the effect estimates as comparable as possible across studies, we calculated effect sizes for the studies that contained sufficient information. Effect sizes were of three types: for continuous outcome measures without an easily interpretable scale, such as standardized test scores,

we calculated Cohen's *d* by dividing the effect estimate by the standard deviation in the treatment and control groups. For dichotomous outcome measures, we reported the absolute effects in percentage points and the relative effects, calculated as the increase or decrease in percent and using the sample mean as the base rate.

As the estimates from studies that compared two alternative universal preschool programs were not fully comparable to the estimates from studies in which children in preschool are compared to children that are not in any type of formal care, we discussed them separately.

We reported average effects for the general population of children and heterogeneity over SES and gender and used the full sample means as the base rate for the relative effects in the heterogeneity analysis also (separate means for high/low SES children, or boys and girls were not reported in most studies). Statistically significant estimates ($p < 0.05$), as reported by the studies, are shown in bold in the tables.

Although we standardized the effect estimates, the definitions of outcome variables and the measures used by the studies were different for nearly all outcome types, and there were few studies that studied exactly the same outcomes. Furthermore, the included studies examined very different preschool programs in terms of program features, age of attending children, the studied period, and the broader study context. We therefore refrained from using meta-analysis to synthesize the results, as we believed such an analysis risked downplaying these differences.

Not performing a meta-analysis precluded a formal analysis of the consistency of effects across studies. If there is a stochastic component in effect estimates – due to sampling variance, for instance – we should expect estimates to vary across studies, even if the true effect of the evaluated programs was the same. Evaluating whether the results in the literature are "mixed" or not by counting negative, null, and positive effects may therefore be misleading (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009; Higgins & Green, 2011). Our discussion should be read with this caveat in mind.

We will describe which outcome measures were used in more detail for each outcome type in section 3 below. Note that because of the procedures described above, the effect sizes and relative effects we report may differ from the ones reported in the studies.

3 Results

This section presents the results of the search and screening process, a discussion of the risk of bias and quality of inference in the included studies, and the analysis of the effects of universal preschool programs on long-term child outcomes. The analysis of effects is divided into the following outcome categories: health, well-being, and behavior; test scores and school grades; primary and secondary school progression; years of schooling and highest grade completed; and employment and earnings. The results section concludes with an analysis of the three benefit-cost analyses (BCAs) in the literature and of the studies that compared two or more universal programs. The BCAs provided a natural context for discussing the magnitude of the effects across studies, so we postponed this discussion until that section. A broader discussion of the results follows in section 4.

3.1 Results of the Search and Screening Process

The search of the electronic databases yielded 1,516 unique records, and we found an additional 86 records from the citation tracking. After excluding irrelevant studies based on information in the title and abstract, we screened 145 studies in full text. Of these, 25 met the inclusion criteria. In some cases, we used information from earlier versions of studies, if, for instance, they included outcomes that were not covered in the published/latest version. Such cases were counted as one study. A flowchart of the search and screening process can be seen in the appendix, where we have also included a detailed description of each study in Appendix Table 1.

The main characteristics of the included studies are summarized in Table 3.1. We found 22 studies that compared children attending or being more exposed to universal preschool programs and children in modes of family or informal care. We found only two comparisons of preschool types, contained in three studies. The studies covered a relatively broad range of countries: there were studies from 14 countries and four continents, and 20 examined programs in developed countries and five in developing countries. Most studies were published in a journal (six were not), and were relatively new (eight studies were dated before 2012). The studied periods were wide-ranging, but a majority of studies (14) examined a program that children attended during the period 1981-1999. There were fewer studies that included very young children: 9 studies included participants that were between 0 and 2 years of age, and 23 included 3-6-year-olds (some included both age categories). The research designs included difference-in-differences (DID; 17 studies), instrumental variables (IV; six studies), and sibling fixed effects (two studies).

Table 3.1   Descriptive statistics of the 25 included studies

Variables	N	%
Type of study		
Preschool/no preschool/more or less preschool	22	88
Type comparison	3	12
Country		
Developing	5	20
Developed	20	80
Continent		
Europe	13	52
North America	8	32
South America	3	12
Africa	1	4
Publication status		
Published in scientific journal/books	19	76
Not published in scientific journal/books	6	24
Publication period		
-2012	8	32
2013-2018	17	68
Studied period		
-1960	4	16
1961-1980	4	16
1981-1999	14	56
2000-	3	12
Age of participants		
0-2	9	36
3-6	23	92
Study design		
Difference-in-differences	17	60
Instrumental variables	6	28
Sibling/family fixed effects	2	8
Outcomes		
Health, well-being and behavior	6	24
Test scores and school grades	10	40

Variables	N	%
Primary and secondary school progression	9	36
Years of schooling and highest grade completed	6	24
Employment and earnings	6	24
Benefits-costs analysis	3	12

Note: Not all categories sum to 25 because some studies covered more than one category, e.g., included both 0-2 and 3-6-year-olds. In these cases, they were counted in all covered categories. Studied period refers to the period in which the preschool program started.

3.2 Risk of Bias and the Quality of Inference

All included studies used some form of natural experiment to estimate the effects of universal preschool programs. However, the research designs differed in the type of natural experiments used, the type of effect they aimed to estimate, in the assumptions underlying the identification of causal effects, and in the estimation and inference techniques used (Appendix Table 1 includes a brief description of the research designs). Below, we first discuss the main risk of bias in each type of research design. Second, we discuss the quality of the statistical inference (i.e., getting standard errors and p-values right). We conclude with an overall assessment of the direction of bias and whether the statistical significance of the results was likely to be over- or understated.

We did not exclude any study on account of having too high risk of bias or too much inferential problems. The intention of this section is to present the most important objections, as we see them, to the claim that these studies estimate the causal effects of universal preschool programs, before we present the results of the review.

3.2.1 Risk of Bias

The most common research design, used in some form by 17 studies, exploited expansions of universal preschool programs that created variation over time (between cohorts of children) and groups (often defined by an area, such as a municipality, state, or city) in how much children were exposed to the programs. The control groups in these studies were often in informal or parental care, but in most cases at least some children in the control group also attended a formal preschool program.

The studies typically lacked information on which children attended preschool and used a DID design to estimate an ITT effect of being more exposed to the preschool program. An ITT estimate has the advantage of capturing the full effects of the program, including any peer effects on children in treated areas that did not attend preschool (e.g., Cascio & Schanzenbach, 2013; Havnes & Mogstad, 2011). As the control group was in many cases not a no-treatment control, only not as exposed, the absolute magnitudes of the estimates are smaller than a contrast between a treatment and no-treatment control group would have been. That is, beneficial effects would be more beneficial and harmful effects more harmful in the latter type of contrast.

The main assumption needed for DID designs to estimate the causal effects is that the trends of the outcome variable would have been parallel, had the treatment group not been more exposed to the preschool program (e.g., Abadie, 2005). The most serious risk of bias in the included studies is that several studies included few areas (seven studies have less than 20 areas, see Appendix Table 1). In the most extreme case, only one area was treated. In the case of one treated area, the treatment effect will be confounded by any idiosyncratic trend or shock affecting the outcome variable differently in the treated area compared to the control areas, even if the shock is completely random (including few areas in the estimation also makes inference more problematic, which we return to below). This risk of bias decreases, the more treated areas there are, as posi-

tive and negative shocks will be more likely to cancel each other out. However, the direction of such bias is difficult to sign and there were both beneficial and harmful effects among the studies with few treated areas.

Six studies used some form of IV design to estimate the effects of attending preschool. Just including a variable measuring preschool attendance would likely yield biased estimates as families and children differ in terms of characteristics that influence both the attendance decision and child outcomes. The IV designs attempted to solve this problem by using a two-stage least squares estimation procedure. In the first step, attendance is predicted by a set of variables, at least one of which (the instrument) was assumed to 1) exert a substantial influence on attendance, and 2) only affect child outcomes through its influence on attendance. The included studies used either thresholds in the admission system that determined whether a child was offered a preschool slot or variants of differences in the preschool supply created by municipal guarantees, historical differences, or similar preschool expansions to those used in the DID designs.

The IV studies all have access to data on preschool attendance and estimated variants of a local average treatment effect (LATE). A LATE is the effect for the so-called "compliers"; that is, the children who would not have attended preschool, if they had not been influenced by the instrument (Imbens & Angrist, 1994). A first problem with the IV designs is that this group is not readily observable and may not be representative of the larger population of interest. LATE estimates are therefore not easily comparable across studies, as the compliers change from context to context.

The instruments used in the IV designs seemed to be strong enough according to the information contained in the studies (i.e., they met condition 1) above). However, it was hard to rule out correlation with child outcomes through other channels than preschool attendance for all instruments used. Historical and geographical differences in the supply of preschools may be correlated with other unobserved determinants of child outcomes (e.g., the value placed by families on having an education or school quality), and admission rules may compare families with different characteristics when samples include children who are not directly at the cut-off created by the rule. Signing this bias across the IV designs was difficult, however, and there were IV designs showing both beneficial and harmful effects.

Two studies employed family or sibling fixed effects, both in the context of expansions of access to universal preschool. The research design uses variation in preschool attendance among siblings to estimate a treatment-on-the-treated (TOT) effect. The sibling fixed effects control for all influences that affect the siblings in the same way, so if the attendance differences between siblings was only driven by access preschools, for instance, this design may recover the causal effect. A problem is that expanding preschools often means that access increases over time and therefore tends to affect younger rather than older siblings. The effects may therefore be confounded by birth-order effects, which tend to favor older siblings (e.g., Black, Grönqvist, & Öckert, 2017). Both studies control for birth-order effects to mitigate these problems. More generally, parental investments in education and care may be correlated with the decision to send one child and not the other(s) to preschool. The sign of the bias, if any, was therefore again uncertain, but both studies employing this design showed beneficial effects.

3.2.2   Quality of Inference

For a number of reasons, the standard errors and p-values reported in the included studies were more likely downward than upward biased. Most studies reported multiple outcomes but only two adjusted for multiple hypothesis testing (Heckman et al., 2017; Lebihan, Haeck, & Merrigan, 2017). Treatment was often assigned on the area level, which means that the standard errors needed to be adjusted for the clustering of children in areas. However, standard methods for cluster-robust variance estimation often underestimate the standard errors when there are few clusters or the

number of children per cluster differs a lot among clusters (e.g., Cameron & Miller, 2015; Mackinnon & Webb, 2017). Few included studies used methods that have been found to work better in these cases (like the wild-cluster bootstrap of Cameron, Gelbach, & Miller, 2008). Furthermore, Mackinnon and Webb (2017) found that even these methods may yield poor results, when the number of treated units is very small. Lastly, Young (2017) found that IV designs tend to produce too small standard errors and p-values when standard inference methods were used.

3.2.3 Overall Assessment

The claim that included studies estimated the causal effects of universal preschool programs has several caveats. However, although individual studies may be biased we found few indications that the estimates were systematically biased toward showing either beneficial or harmful effects. In light of the inference problems mentioned above, the included studies seem more likely to overstate the statistical significance of their findings than understate it.

A problem that pertains to the whole literature rather than the individual studies is the issue of publication bias; that is, the tendency that statistically significant results are more likely to be published than null findings. For this reason, we included unpublished reports. However, Franco, Malhotra, and Simonovits (2014) found results indicating that publication bias in the social sciences was more driven by researchers not writing up null results than journals not publishing them. It is difficult to tell whether this is the case for the literature we reviewed, but as we shall see in the following results sections, there were plenty of examples of insignificant results reported in our included studies.

3.3 Health, Well-being, and Behavior

Table 3.2 displays the estimated effects of universal preschool programs on measures related to health, well-being, and behavior. We included personality measures, family formation, and crime, as personality is closely related to behavior, family formation is a type of behavior, and crime is a measure of anti-social behavior. Personality traits and family formation are not clear-cut measures of beneficial or harmful effects, although some are related to other more unambiguous measures (for instance, conscientiousness is positively associated with earnings and health; Almlund, Duckworth, Heckman, & Kautz, 2011). Four studies reported estimates on measures of problem behavior or personality traits, three studies reported measures related to health, healthy behaviors, and well-being, and two studies reported the effects on outcomes related to crime. The outcomes were measured when the children where between 8 and 59 years old. It was not possible to convert all estimates to a common effect size measure. Whenever there was sufficient information, we calculated either effect sizes in standard deviations or percentage points and percent. The estimates were calculated so that a positive (negative) sign implied an increase (decrease) of the behavior/health/trait measured. For example, a negative sign on an estimate of overall health implied decreased health and therefore a harmful effect.

For all three subcategories (behavior/personality, health and well-being, and crime) the estimates were mixed. There were examples of beneficial and harmful effects in all subcategories. Furthermore, most estimates were insignificant. Few studies reported heterogeneity over SES or gender, and there was no clear pattern in either category.

Table 3.2 Health, well-being, and behavior

(1) Study	(2) Age/grade	(3) Type of effect	(4) Average effect	(5) SES	(6) Gender
Baker et al. (2015)	12-20 years	ITT effect of being more exposed on stress, quality of life, and being accused of a crime	Effect sizes in standard deviations or percentage points and % Stress: 0.094 Quality of life: **-0.36** Crime: **0.30** (3.7%)	Not reported	Not reported for stress and quality of life Crime *Girls*: **0.17** (2.1%) *Boys*: **0.43** (5.3%)
Berlinski et al. (2009)	3rd grade	ITT effect (uptake close to 1 though) of new preschool places per child on teacher ratings of behavior	Effect in percentage points and % Attention: **12** (13%) Effort: **21** (24%) Discipline: 11 (15%) Participation: **17** (20%)	Not reported	Not reported
Fort et al. (2018)	8-14 years (mean 10.7)	LATE of 1 extra month of preschool on the (log of) openness (O), conscientiousness (C), extraversion (E), agreeableness (A), and neuroticism (N)	Effects in % O: -0.4% C: -0.0% E: -0.6% A: -0.4% N: 0.2%	*Lower income* O: 0.1% C: 0.7% E: -1.1% A: 0.3% N: -0.5% *Higher income* O: **-1.4**% C: -0.1% E: -0.6% A: **-1.2**% N: 0.9%	*Girls* O: -0.5 % C: 0.3% E: **-1.2**% A: -0.3% N: 0.0% *Boys* O: -0.3% C: 0.1% E: -0.3% A: -0.4% N: 0.2%
Havnes & Mogstad (2011)	Age 30-39	ITT effect on being more exposed to preschool on the probability of being a parent, single without children, and single with children.	Effects in percentage points and % Parent (P) **-1.4** (1.8%) Single without children (S) **0.62** (4.4%) Single with children (SC) -0.04 (-0.48%)	*Mother high school/no high school* *No high school* P: **-1.1** (-1.4%) S: **0.56** (4.0%) SC: 0.12 (1.4%) High school P: **-1.3** (-1.6%) S: 0.23 (1.7%) SC: **-0.61** (-7.3%)	*Girls* P: **-2.04** (-2.5%) S: **0.95** (6.8%) SC: -0.31 (-3.7%) *Boys* P: **-0.87** (-1.1%) S: 0.31 (2.2%) SC: 0.21 (2.6%)
Herbst (2017)	44-59 years	ITT effect of $100 more in spending on the probability of a work-limiting disability	Effects in percentage points and % -0.3 (-4.8%)	Not reported	Not reported
Lebihan et al. (2017)	Behavior: 8-9 years Health and well-being: 12-14 years	ITT effect of being more exposed on behavior, health, and well-being	Effect sizes in standard deviations *Behavior* Hyperactivity (H): 0.074 Anxiety (A): **0.21** Physical aggression (P): 0.10 Indirect aggression (I): 0.094	*Mothers, post-secondary schooling:* *Without* H: -0.026 A: **0.30** P: 0.067 I: **0.27**	Not reported

(1) Study	(2) Age/grade	(3) Type of effect	(4) Average effect	(5) SES	(6) Gender
			Health and well-being Overall health: -0.04 Had asthma attack: 0.005 Mental health: -0.037 Belonging: -0.023 Life satisfaction: -0.098 Drank alcohol: -0.018 Doesn't smoke: **0.080**	*With* H: 0.12 A: **0.19** P: 0.11 I: **-0.005** Not reported for health and well-being	
Smith (2015)	18-19 years	ITT effect of being more exposed on the probability of being charged with a crime	Effects in percentage points and % *Black* Felonies: -2.8 (-17%) Misdemeanors: -5.7 (-32%) *White* Felonies: -0.6 (-20%) Misdemeanors: 0.9 (18%)	Not reported	Not reported

Note: Whenever there was sufficient information, we calculated either effect sizes in standard deviations or percentage points and percent. The estimates were calculated so that a positive (negative) sign implied an increase (decrease) of the behavior/health/trait measured. For example, a negative sign on an estimate of overall health implied decreased health and therefore a harmful effect. Statistically significant effects ($p < 0.05$), as reported by the studies, are shown in bold. Type of SES heterogeneity is shown in italics in column 5.

3.4   Test Scores and School Grades

The effect sizes in Table 3.3 are based on standardized tests of science, mathematics, and literacy, combinations of the latter two subjects, broader tests of cognitive skills and IQ, or school grades. The earliest tests were performed when children were around 8 years and the latest when they were 18-20 years old. Although tests of educational achievement and school grades measure different skills from those that an IQ test measures, achievement and IQ tests are significantly correlated (e.g., Borghans, Golsteyn, Heckman, & Humphries, 2016). Furthermore, motivation and incentives to perform well are important for all tests (e.g., Kautz, Heckman, Diris, ter Weel, & Borghans, 2014), which is another reason to believe that standardized achievement tests and IQ capture overlapping skills. Grades are likely to capture other skills than IQ to an even higher degree. We therefore analyzed these outcomes together.

Most studies included in Table 3.3 reported beta-coefficients with the scores standardized to have mean zero and standard deviation equal to one. We reported effect sizes calculated in this way whenever possible in the table (some studies lack information), but it should be noted that the standardization procedure differed between studies (e.g., some were standardized by grade, site, or year, and some by the overall standard deviation). The studies did not contain sufficient information to enable a uniform way of standardization, so we kept the results as reported in the studies. As the standardization may affect the effect size, variation in this procedure may be one reason for the variation in effect sizes between studies. Positive estimates imply a beneficial effect.

The effects of universal preschool programs for the general population of children on test scores and school grades were mixed, in the sense that Table 3.3 contains significant beneficial and harmful effects, as well as insignificant estimates. The ITT effects range from large harmful effect sizes to large beneficial effect sizes (-0.23 to 0.26, both extremes are from Baker et al., 2015, and both are for math tests). The two examples of studies having access to attendance data

were likewise mixed, one showing significant beneficial effects (Bietenbeck, Ericsson, & Wamalva, 2017) and one significant harmful effects (Fort, Ichino, & Zanella, 2018).

Most studies reporting heterogeneity over SES found more beneficial/less harmful effects for children from families with low SES, and no study found a consistent opposite pattern. The absolute magnitude of effects was larger for girls in all studies reporting heterogeneity over gender. However, most gender differences were small and the differences not significant in any study performing such a test.

Table 3.3   Effects on test scores and school grades

(1) Study	(2) Age/grade	(3) Type of effect	(4) Average effect	(5) SES	(6) Gender
Baker et al. (2015)	13-16 years	ITT effect of being more exposed on math, reading, and science test scores	Math SAIP/PCAP: -0.23 PISA: **0.26** Reading SAIP/PCAP: -0.074 PISA: 0.074 Science SAIP/PCAP: -0.042 PISA: -0.032	Not reported	Not reported
Berlinski et al. (2009)	3rd grade	ITT effect of new preschool places per child on math and Spanish test scores (uptake close to 1 though)	Math: **0.24** Spanish: **0.23**	*Share living in poverty by municipality* At the 75th percentile, effects are 0.08 (Math) and 0.16 (Spanish) larger than at the median (statistical significance not reported for this result)	*Girls* Math: **0.26** Spanish: **0.27** (Estimates are not significantly different from boys' test scores, which are not reported separately).
Bietenbeck et al. (2017)	13-16 years	TOT effect of attending preschool on a composite score from a standardized literacy test and a numeracy test	*Kenya*: **0.12** *Tanzania*: **0.080**	*Household wealth above or below median* No consistent differences (results shown in figure only).	Not reported
Bladen et al. (2016)	11 years	ITT effect of availability of free preschool places in an area of residence on standardized tests of reading and math	Math: -0.002 Reading: **0.006**	*Free school meals* *Eligible* Reading: 0.008 Math: 0.003 *Non-eligible* Reading: **0.006** Math: -0.002	*Girls* Reading: **0.007** Math: -0.002 *Boys* Reading: 0.005 Math: -0.001 (Estimates are not significantly different for boys and girls).
Cascio and Schanzenbach (2013)	8th grade	ITT effect of being more exposed on math and reading test scores	Math: 0.9 points (not standardized and average effect for reading is not reported).	*Free/reduced-price lunches* *Eligible* Math: 2.2 points Reading: 0.82 points *Not eligible* Math: -1.3 points	Not reported

(1) Study	(2) Age/grade	(3) Type of effect	(4) Average effect	(5) SES	(6) Gender
				Reading: -0.81 points	
Felfe & Lalive (2010)	9-10 years old	LATE of having spent some time in preschool during 0-3 years of age on school grades	Significant beneficial effect on grades but the scale of the effect is unclear (raw standard deviation is missing)	Not reported	Not reported
Felfe et al. (2015)	15 years	ITT effect of being more exposed on PISA scores in math and reading	<u>Math</u>: 0.049 <u>Reading</u>: **0.15**	*Parents with-out/with a second-ary school degree* *Without* <u>Math</u>: 0.041 <u>Reading</u>: **0.17** *With* <u>Math</u>: 0.025 <u>Reading</u>: 0.11	*Girls* <u>Math</u>: 0.11 <u>Reading</u>: **0.19** *Boys* <u>Math</u>: -0.011 <u>Reading</u>: **0.12**
Fort et al. (2018)	8-14 years (mean 10.7)	LATE of 1 extra month of pre-school on IQ test score	**-0.045**	*Lower/higher in-come* Lower: -0.02 Higher: **-0.13**	Girls: **-0.07** Boys: -0.04
Havnes & Mogstad (2015)	18-20 years (males only)	ITT effect on being more ex-posed to pre-school on cogni-tive skills	Not available	Not available	Quantile effects for males reported in figure only. Small and insignificant effects in all quantiles.

Note: Effect sizes measured in standard deviation units. Statistically significant effects ($p < 0.05$), as reported by the studies, are shown in bold. Type of SES heterogeneity is shown in italics in column 5. The calculation of effect sizes for low and high SES, and for boys and girls use the same standard deviation as the average effect size.

3.5   Primary and Secondary School Progression

Effects of universal preschool programs on outcomes related to primary and secondary school progression were measured by an indicator of making age-adequate progress (e.g., being on-grade and probability of not being retained), by an indicator of having been retained (one excep-tion, Dumas & Lefranc, 2012, use the number of retentions), or by an indicator for having graduat-ed/being enrolled, or for having dropped out. To make the estimates as comparable as possible, we transformed measures of making age-adequate progress into measures of grade retention and dropout measures into measures of graduation or being enrolled. Progress and retention rates mirror each other in the sense that the probability of adequate progress equals the probability of never being retained. Graduation and dropout could differ, if there were students who did not graduate on time but had not yet dropped out. However, high school dropout was always meas-ured several years after appropriate graduation in the included studies, making such problems unlikely. The range for child age at measurement was 9-39 years.

Most effect estimates were reported as percentage point changes. In Table 3.4, we converted all estimates to percentage points and also reported the relative effects in percent. The negative estimates represent beneficial effects regarding grade retention, while positive estimates represent beneficial effects regarding graduation and being enrolled.

The included studies indicated that universal preschool programs have beneficial effects on measures related to primary and secondary school progression. All estimates of the average ef-fects for the general population indicated beneficial effects of either attending a preschool program (TOT/LATE estimates) or growing up in a more exposed area (ITT estimates), and 7 out of 12

estimates were statistically significant on a five percent level. The effects were larger for children from low SES families in all but two cases in the six studies reporting heterogeneous effects. Two studies reported harmful effects for high SES children, but none of the estimates were significant. Effects were less beneficial for girls in three studies and more beneficial in two.

Table 3.4   Effects on school enrollment, grade retention, being on-grade, and dropout

(1) Study	(2) Age/grade	(3) Type of effect	(3) Average effect	(4) SES	(5) Gender
Bastos et al. (2017)	12 years	ITT estimate of having access to a preschool on the probability of primary school enrollment and being retained	Enrollment 3.0 (3.5%) Retention **-2.4** (-2.7%)	*Share of adults with no education in community* *Low* Enrollment: **5.1** (5.9%) Retention: **-3.6** (-4.1%) *High* Enrollment: 0.51 (0.58%) Retention: -1.1 (-1.2%)	*Girls* Enrollment: 2.1 (2.4%) Retention: **-2.5** (-2.8%) *Boys* Enrollment: 4.3 (5.0%) Retention: **-2.5** (-2.8%)
Berlinski et al. (2008)	15 years	TOT estimate of attending pre-school 1-3 years (Mean = 1.75) compared to 0-1 years on the probability of being enrolled	**27** (30%)	*Mother's education* *Low*: **27** (30%) *High*: 8.4 (9.2%)	*Girls*: **24** (27%) *Boys*: **36** (40%)
Bietenbeck et al. (2017)	13-16 years	TOT effect of attending pre-school on probability of being enrolled	*Kenya*: **2.0** (2.1%) *Tanzania*: **9.0** (10.1%)	Not reported.	Not reported.
Bingley et al. (2018)	Age 35	ITT of daycare availability on the probability of obtaining a high school/vocational degree	**0.9** (1.2%)	Not reported	Not reported
Borraz & Cid (2013)	15 years	LATE estimate of attending pre-school on the probability of being retained	-4.4 (-15%)	*Mother's education* Less educated: 7.5 (25%) (No separate results for more educated)	*Girls*: -2.5 (8.3%) *Boys*: 16 (54%)
Dumas & Lefranc (2012)	16 years (number of grade repetitions) or after high school.	LATE of attending one more year of preschool on the number of grade repetitions and probability of high school graduation	No. of grade repetitions: -0.076 (-9.4%) High school graduation: **15** (20%)	Not reported (for IV specification).	Not reported (for IV specification).
Felfe et al. (2015)	Secondary school (not further specified)	ITT effect of being more exposed on probability of being retained	-3.2 (-10.9%)	*Parents without/with a secondary school degree* *Without*: -3.7 (-12.6%) *With*: -1.9 (-6.5%)	*Girls*: -4.5 (-15%) *Boys*: -1.9 (-6.5%)
Fitzpatrick (2008)	4th grade (typically 9	ITT effect of being more exposed on	-0.7 (-4.5%)	*Free/reduced-price lunch eligibility*	No differential effects by gender (results

(1) Study	(2) Age/grade	(3) Type of effect	(3) Average effect	(4) SES	(5) Gender
	years old)	probability of being retained		*White* *Eligible*: **-2.0** (-13%) *Not eligible*: 0.1 (0.6%) *Black* *Eligible*: **-2.5** (-16%) *Not eligible*: **-6.0** (-38%)	only mentioned in text)
Havnes & Mogstad (2011)	30-39 years	ITT effect on being more exposed on probability of high school graduation	**1.0** (1.4%)	*Mother's high school degree* *No degree*: **1.3** (1.7%) *Degree*: 0.21 (0.29%)	*Girls*: **0.81** (1.1%) *Boys*: **1.2** (1.7%)
Herbst (2017)	24-39 years	ITT effect of $100 more in spending on probability of high school graduation	**2.1** (2.7%)	Not reported	Not reported

Note: Absolute effects are reported in percentage points, relative effects in percent and in parentheses. Positive estimates represent beneficial effects of preschool for enrollment and graduation. Negative estimates represent beneficial effects for the number of grade repetitions and probability of being retained. Significant estimates (p < 0.05), as reported by the studies, are shown in bold. The effect is the increase or decrease in percent, calculated by dividing the effect estimate by the mean in the estimation sample. Type of SES heterogeneity is shown in italics in column 5.

3.6   Years of Schooling and Highest Grade Completed

Table 3.5 shows the estimates from four studies that reported effects on years of schooling and two studies that reported effects on the highest grade attained. As the highest grade was typically measured in years, the two outcomes are fairly comparable and are reported in years in Table 3.5. Furthermore, the table includes two estimates of the probability of obtaining a college and bachelor's degree. Age at measurement ranged from 13 to 66.

All six studies that included an estimate of the average effects of attending or being more exposed to universal preschool, indicated significant increases in the years of schooling and highest grade completed. The estimates of years of schooling range from 0.07 (Havnes & Mogstad, 2011) to 3.3 years (Bietenbeck et al. 2017).

Six studies examined how the effect differed over SES. Four studies found that the effect was largest for low SES children, one study found that children with highly educated mothers gained the most, and one study found no consistent differences over SES. Four studies reported heterogeneous effects across gender. All found beneficial effects for both genders, and the differences were mostly small. One study found larger effects for boys.

Table 3.5 Effects on years of schooling and highest grade completed

(1) Study	(2) Age	(3) Type of effect	(4) Average effect	(5) SES	(6) Gender
Berlinski et al. (2008)	Age 15	TOT effect of attending preschool on years of schooling	**0.79**	*Mother's education* Low: **0.74** High: 0.25	Girls: **0.88** Boys: **0.89**
Bietenbeck et al. (2017)	Age 13-16	TOT effect of attending preschool on highest grade completed	*Kenya*: **0.12** *Tanzania*: **0.11**	*Household wealth above/below median* (Results shown in figure only). *Kenya*: Children below median have insignificantly higher effects. *Tanzania*: Children below median have significantly lower effects.	Not reported
Bingley et al. (2018)	Age 35	ITT of daycare availability on years of schooling and the probability of obtaining a college degree	Years of schooling **0.092** College degree Percentage points (%) **0.017** (5.4%)	*Maternal schooling level* Years of schooling *Basic*: 0.021 *High school/vocational training*: **0.064** *College/university*: **0.077** Not reported for college degree	Years of schooling *Girls*: **0.049** *Boys*: **0.13** Not reported for college degree
Bingley & Westergård-Nielsen (2012)	Age 23-30	ITT of being more exposed to preschool on years of schooling	Not reported	(Results reported for more finegrained categories than in other studies) Children of less educated mothers have significantly more years of schooling. Preschool does not have any significant effect on education for children with higher educated mothers or fathers.	No general pattern found
Havnes & Mogstad (2011, 2015)	Age 30-39 & age 33-42	ITT effect on being more exposed to preschool on years of schooling and the probability of attending college	Years of schooling **0.074** Attending college Percentage points (%) **1.2** (3.3%)	*Mother high school (HS) education or not and family income* Years of schooling *Low*: **0.24** *Mid*: **0.081** *High*: 0.018 Attending college Percentage points (%) *No HS*: **1.4** (3.7%) *HS*: 0.33 (0.88%)	Years of schooling *Girls*: **0.066** *Boys*: **0.084** Attending college Percentage points (%) *Girls*: **1.2** (3.3%) *Boys*: **1.2** (3.2%)
Haimovich Paz (2015)	Age 30-66	ITT effect of exposure to kindergarten on maximum grade attainment	**0.18**	*Mother tongue* *Non-English*: **0.29** *English*: **0.14**	The sample consists only of boys
Herbst (2017)	Age 24-39	ITT effect of $100 more in spending on	Percentage points (%)	Not reported	Not reported

(1) Study	(2) Age	(3) Type of effect	(4) Average effect	(5) SES	(6) Gender
		probability of obtaining a bachelor's degree	**1.9** (27%)		

3.7   Employment and Earnings

Table 3.6 shows the estimates from studies that examined the effect of universal preschool programs on measures related to earnings, employment, and welfare, measured at ages from 23 to 59 years. The estimates for earnings were reported in percent and in percentage points for the probability of being employed and being on welfare. Positive estimates in Table 3.6 indicate beneficial effects on earnings and employment, while negative estimates indicate beneficial effects on the probability of being on welfare or receiving public assistance.

All estimates for the three outcomes indicated beneficial effects for the general population of children. Most estimates were statistically significant. However, the average effect contains hid substantial heterogeneity in some cases: all but two studies found larger effects for low SES children, and some estimates indicated significantly harmful effects for high SES children. In contrast, no consistent pattern was found in the studies that examined heterogeneity across gender.

Table 3.6   Effect on earnings (percent), employment, and being on welfare

(1) Study	(2) Age	(3) Type of effect	(4) Average effect	(5) SES	(6) Gender
Bingley et al. (2018)	Age 35	ITT effect of living in a neighborhood with a preschool when 4 years-old on (log) earnings and probability of having no earnings.	Earnings: **1.2%** No earnings: -0.2 (-1.6%)	*Maternal education* Earnings: *Basic*: 0.00% *High school*: 1.1% *College*: 1.5% No earnings: Not reported	Earnings: *Girls*: 0.1% *Boys*: **2.2%** No earnings: *Girls*: **0.005** (4.1%) *Boys*: **-0.008** (-6.6%)
Bingley & Westergård-Nielsen (2012)	Age 23-30	ITT of being more exposed to daycare on (log) earnings	Not reported	Interaction terms between dummies for parental earnings quartile and pre-school density are largely negative for low earnings and positive for high.	No general pattern found
Havnes & Mogstad (2011, 2015)	Age 30-39 & age 33-42	ITT effect on being more exposed on earnings and the probability of being on welfare	Earnings: 0.092% Being on welfare: **-0.91** (-5.6)	*Mother high school (HS) education or not and family income* Earnings: *Low*: **2.9%** *Mid*: -0.50% *High*: -2.0% The quantile treatment effects indicate	Earnings: *Girls*: 0.22% *Boys*: -0.22% Being on welfare: Girls: **-1.2** (-7.2%) Boys: **-0.63** (-3.9%)

(1) Study	(2) Age	(3) Type of effect	(4) Average effect	(5) SES	(6) Gender
				larger effects in the lower earnings quantiles. The effects start to turn negative around the 80th quantile and are substantial and significant at the top. <u>Being on welfare:</u> *No HS*: **-0.84** (-5.1%) *HS*: **-1.2** (-7.6%)	
Haimovich Paz (2015)	25-45 years	ITT effect of being more exposed on earnings for white males	**1.5%**	*Mother tongue* *Non-English*: **4**% English: 1%	The sample consists of boys only
Herbst (2017)	44-59 years	ITT effect of $100 more in spending on ln(earnings), being employed, or receiving public assistance.	<u>Earnings:</u> **2.5**% <u>Employed last year:</u> **0.5** (0.61%) <u>Public assistance:</u> -**0.2** (-7.1%)	<u>Earnings:</u> The quantile treatment effects (reported in a figure) are positive for all quintiles, but the magnitudes are larger for low SES. Not reported for employment and public assistance.	Not reported

Note: Effects on earnings are given in percent and are either calculated by dividing the effect estimate by mean earnings, or from beta-coefficients where the outcome variable is transformed to ln(earnings). Effects on the probability of being employed, being a recipient of welfare benefits, or having no earnings are given in percentage points. Statistically significant effects ($p < 0.05$), as reported in the studies, are shown in bold. Type of SES heterogeneity is shown in italics in column 5.

3.8   Benefit-Cost Analyses

Universal preschool programs involve a substantial amount of public spending, and one of the most important question for policy makers is whether the total benefits outweigh the cost of implementation. Three studies in our sample included a Benefit-Cost Analysis (BCA): Berlinski, Galiani, and Manacorda (BGM, 2008) examined a program in Uruguay; Cascio and Schanzenbach (CS, 2013) examined programs in the US states of Georgia and Oklahoma; and van Huizen, Plantenga, and Dumhs (vHPD, 2017) used estimates from Felfe, Nollenberger, and Rodríquez-Planas (2015) to analyze a Spanish program.

In all three studies, the estimated benefit-to-cost ratio was clearly above one, meaning that for every dollar the government invested in the universal preschool program society received more than one dollar in return. The ratios therefore indicated that the universal preschool programs were a worthwhile investment (Akinyemi, 2013). However, the three BCAs build on several assumptions and estimates. We discuss the main assumptions and compare the estimates used to other estimates from the included studies below.

The three studies extrapolated child earnings from the effect of universal preschool programs on either test scores (vHPD, CS) or years of schooling (BGM, 2008) when children were around 15 years of age and assumed that that the relationship between test score/years of schooling and earnings was constant over a child's career. BGM used a TOT estimate, while vHPD and CS transformed ITT estimates to TOT estimates by dividing the ITT with the differential take-up rate between treatment and control groups. The increases in lifetime earnings were estimated to be 1.3

(CS), 6.0 (vHPD), and 7.9 percent (BGM). To compare them to the other estimates of earnings reported in Table 3.6, we converted the TOT estimates in CS and vHPD back to ITT estimates. The test scores amount to around 0.03 standard deviations (CS) and 0.15 (vHPD), and the earnings estimate to around 0.3% (CS) and 1.5% (vHPD). The CS estimates are not particularly large compared to our other estimates, while vHPD is among the largest. We have fewer estimates to which we can compare BGM's estimates, but both of them seem to be larger than most of our other estimates. The larger effects might by fully reasonable though, given their developing country context.

BGM and CS did not include effects on maternal employment or tax revenues. The program studied by vHPD increased maternal employment, and vHPD included increased earnings for mothers, extrapolated from the employment estimate, as an additional benefit. Furthermore, vHPD included increased tax revenues from the increased child and maternal income, as well as benefits to tax payers from improving graduation and retention rates. The main share of benefits came from improved child earnings though; tax revenues and maternal earnings made up less than 35 percent of total benefits.

All three studies assumed that the estimates extended to all treated children and that any spill-over or general equilibrium effects of the intervention were ignorable. The studies did not include effects on for example welfare dependency, crime, health, and well-being, and there were no estimates of intergenerational effects included in the analyses (Rossin-Slater & Wüst, 2017, found beneficial intergenerational effects on educational attainment from a targeted Danish program). These omissions seem likely to have understated the total benefits of the programs, as the omitted outcomes tend to be positively associated with test scores and years of schooling.

All studies included the direct cost of the program for tax payers and parents (net of any decreased costs due to, for instance, out-of-pocket spending on other programs for parents). Only BGM included a cost of children staying in school for more years and a cost of obtaining revenue to finance the program (in their case the projected interest on a loan). Raising tax revenue to pay for operating costs, or the interest on a loan, may, depending on how the tax is designed, be costly due to deadweight losses. It is not obvious how large such losses would be, but some government guidelines use 10-20 percent of the costs funded by general taxation (e.g., Finansministeriet, 2017; Treasury of New Zealand, 2015).

The discount rate typically has a great impact on the results of BCAs. BGM and vHPD used a 3 percent discount rate for their baseline scenarios, and CS used the 30-year return on US Treasury bills, which was 3.4 percent at the time. The benefit-to-cost ratios in the baseline scenarios were 3.2 in CS, 4.3 in vHPD, and 19 in BGM. All ratios were above one, also when substantially higher discount rates were used.

Summing up, both benefits and costs appeared to be underestimated in the three studies. The omitted posts on the benefit side were, in our view, potentially more substantial than the omitted posts on the cost side. Increasing program costs by 20 percent to account for deadweight losses of taxation would, for example, not drive the ratios below one in any of the studies. Regarding the extrapolation of test scores and years of schooling to earnings, CS did not stand out in comparison to our other estimates, but still produced a ratio quite far above one. Other universal preschool programs showing beneficial effects may therefore also have benefit-to-cost ratios comfortably over one. In turn, this is a sign that the magnitude of the included effects was often substantial.

3.9   Comparison of Preschool Types

Datta Gupta and Simonsen (DGS, 2012, 2016) compared outcomes for children who attended public center-based care to children who attended family day care, exploiting the variation in the

composition of the type of child care municipalities provided. Children in family day care were cared for in the homes of child-minders employed and regulated by the Danish municipalities.

The two care arrangements investigated in DGS, center-based and family day care, differed mostly with regard to the preschool teachers' training. While core staff in the center-based pre-school setting had to hold a pedagogical degree, the caregivers in the family day care setting generally had less education, and no specific training was required (DGS, 2016). The two settings have approximately the same staff-to-child ratios (cf. Appendix Table 1). As most Danish children eventually enroll in center-based care, DGS estimated how an additional 1.5 years of early center-based care affected the child.

The comparison in DGS is not completely different from the comparison in some of the other studies, as several studies compared a formal universal preschool program to a more informal kind of preschool (c.f. Appendix Table 1). However, the family day care arrangement in DGS setting was highly regulated in contrast to informal out-of-home care in other included studies. The difference between formal preschool and informal care may be more pronounced in settings where the informal care setting is less regulated (DGS, 2016).

DGS (2016) found that enrollment in center-based care at age two increased enrollment in the academic track in high school at age 17 by 11 percentage points (17%), and the average grade in Danish (Math) by 0.23 (0.10) standard deviations at age 15. The effects on high school enrollment and average grade in Danish were significant on a five percent level. Significant effects for boys were found on all outcomes, while only the increase in the average grade in Danish was significant for girls. In addition, DGS found that the effects were larger for children of mothers with no more than high school education compared to children of mothers with some higher education. DGS (2012) found that children who attended center-based preschool at age three liked school significantly more than children who attended family day care, at age 11, but found no significant differences on a number of other outcomes, such as the Strengths and Difficulties Questionnaire, language and cognitive skills, delayed school entry, smoking, alcohol, and petty theft and vandalism.

Heckman et al. (2017) compared the Reggio Emilia approach, originating from the Italian city of the same name, with preschool approaches given to children in the nearby cities of Padova and Parma. The Reggio Emilia approach is based on a perception of the child as an individual with rights and potentials. The approach is notable for its investments in staffing, early inclusion of children with disabilities, and high rates of provision of early childhood services. However, non-Reggio Emilia approaches have historically shared many features with the Reggio Approach, and the similarities have increased over time. For that reason, the differences between the two preschool settings may be small and possibly negligible (Heckman et al. 2017).

Heckman et al. (2017) estimated the effect of preschool (age 0-6) for different cohorts at different ages; from a child cohort (aged seven) up until an age 50 cohort. Heckman et al. (2017) found that, compared to other types of preschool, Reggio Emilia approach preschool increased the child's depression score at age 19 (Parma), increased the probability of ever having voted in municipal and Regional elections at age 30 (Padova). There were no significant effects on a number of measures, including IQ, educational attainment, and health. The authors concluded that the differences in quality between the Reggio Emilia approach and the alternative programs were not sufficiently large to show substantial differences in outcomes for the adult population (Heckman et al. 2017, p. 60).

4   Discussion

We found 25 studies examining the effects of universal preschool programs on child outcomes from third grade to adulthood. The studies examined a broad range of programs, implemented in different countries and time periods. Below we discuss our most important findings, first regarding the average effects for the general population of children, and second regarding the heterogeneity in terms of SES and gender. We then discuss the limitations of the review, and lastly offer some concluding remarks and suggestions for further research.

4.1   Effects for the General Population of Children

We have two main findings regarding the average effects: Firstly, the effects on test scores and school grades, and on measures related to health, well-being, and behavior varied across (and sometimes within) studies. The magnitudes also varied, and the majority of estimates were not statistically significant (as reported in the studies). Secondly, all estimates for outcome measures related to adequate primary and secondary school progression, years of schooling and highest degree completed, and earnings and employment indicated beneficial average effects. The magnitudes of these estimates were often substantial, as well as statistically significant. Furthermore, the three included BCAs indicated benefits-to-costs ratios clearly above one. While the majority of studies and estimates thus indicated that universal preschool programs have beneficial long-term effects, the differences between outcome types are important to understand, and we discuss potential explanations below.

A simple explanation for the differences between outcome types could be that some programs were of a low enough quality to be harmful on average. No harmful effects were detected in adulthood, but the few studies that included estimates in primary and secondary school as well as in adulthood showed consistent beneficial average effects over time. If harmful effects are similarly persistent, this pattern indicates a crucial point: universal preschool programs need to be of better quality than children's counterfactual mode of care to produce beneficial effects.

Another interpretation could be that the full effects of universal preschool are better captured by the longer-term measures. Measures like graduation, earnings, and employment are arguably influenced by a broader set of skills than some of the measures for which studies found harmful effects. For example, improved personality skills seem to be the best explanation for the patterns in the Perry Preschool program of on the one hand enduring beneficial effects on crime, health, and earnings, and on the other hand short-term but quickly fading effects on cognitive skills (Heckman et al., 2013). However, as the included studies found some harmful effects on crime, health, and behavioral measures, lasting effects on personality skills cannot explain all the differences found between outcome types.

An additional reason related to the timing of measurement is that harmful effects may wane, either because other interventions are later given to children who fall behind, or naturally as children get older. For example, there could be short-term harmful effects on health and socialization from being around other children, but such effects may pass or even turn beneficial over time (e.g., Strachan, 1989; Baker et al., 2008). Although there was some evidence of fadeout of initial harmful effects (see e.g., Lebihan et al., 2017), most of them seem too long-term for waning effects to be a major explanation of the differences between outcomes.

Some outcome measures may be more variable and therefore more likely to produce both harmful and beneficial estimates by chance, despite the true effect being beneficial or harmful. Test score and grades are typically measured on one occasion, while school progression, graduation, employment, and earnings are the result of more continuous processes. They are therefore

less prone to chance results. Included cognitive skills tests were often not high stakes for students, and incentives and motivation to perform well matter for test results (e.g., Kautz et al., 2014). If children do not put in a lot of effort, the chance component of test scores may be substantial. However, some harmful effects were found on outcomes that were not measured at one test occasion and that are high stakes. The increased crime rates found in Baker et al. (2015) is perhaps the best example.

In our view, the differences between outcome types are not due to upward bias in studies showing beneficial effects. Individual studies may of course be biased, but the risk of upward bias in studies showing beneficial effects did not seem to be higher than the risk of downward bias in studies showing harmful effects. Studies were more likely to systematically overstate statistical significance, due to for example multiple hypothesis testing and problems with properly adjusting for clustering. These problems also pertained to studies showing harmful effects and would not change the direction of the effects.

Publication bias may mean that we should temper our conclusions regarding both beneficial and harmful effects, as null effects might be missing from the literature. While this may be the case, the treatment group in the included studies was often just more exposed than the control group. Given an effect, such study designs underestimate both beneficial and harmful effects. The distribution of true effects may therefore contain both larger harmful and beneficial effects, as well as more null and small effects. In any case, unless harmful effects were less likely to be published or written up – of which we have no evidence – publication bias cannot explain the differences between outcome types.

We cannot rule out a combination of the other explanations, but the simplest explanation of the differences between outcome types is that they were caused by different universal preschool programs having different effects. Indeed, given the variation in factors related to quality in the studied programs it would have been surprising if we had not found some differences. It was perhaps more surprising, also in relation to the message from prior reviews, that the results were not more mixed. We return to the causes of quality differences below, where we discuss heterogeneity in terms of SES and gender.

4.2   Heterogeneity over Socioeconomic Status and Gender

Having beneficial effects on average does not imply that universal preschool is equally good for all children. Previous reviews emphasized that beneficial effects of universal preschool programs were stronger for disadvantaged, or low SES, children. Our synthesis showed that this tendency was present for many outcomes, although there were examples of opposite effects. The relatively large beneficial effects found in studies from developing countries, where more or most children are low SES in comparison with developed countries, were also consistent with the pattern of more beneficial effects for low SES children. In line with previous reviews, we did not find a consistent pattern of gender differences in the effects of universal preschool programs. Our discussion in the introduction implied that gender differences may be quite subtly dependent on the features of both the preschool program and the counterfactual mode of care. The information needed to tease out these conditions was rarely present in the included studies, and we therefore focus the discussion below on the differences between children with high and low SES.

The few studies that showed more beneficial effects for high SES children can be rationalized by examining the quality differences between preschool programs and the counterfactual mode of care. The larger beneficial effects for high SES children in for example Bingley, Jensen, and Sander (2018) were found in a context where high SES children's counterfactual mode of care was low quality informal or private care (and not parental care like for most low SES children).

In most contexts, low SES children seemed more likely to improve their quality of care more by attending universal preschool programs than high SES children, indicating that universal preschool reduces socioeconomic inequalities. This is good news for governments looking for ways to provide equal opportunities to all children, but it matters greatly whether the reduction in inequality is caused by relative improvements of low SES children's skills or by an absolute worsening of high SES children's skills. Havnes and Mogstad (2015) suggested that differences in the effects over SES could either be explained by differences between high and low SES children in the take up of preschool programs, or in the effects of preschool programs. Distinguishing between the two explanations requires data for attendance, which most studies in our sample lack. We will therefore not be able to settle this issue. However, as the case for targeted programs become stronger if universal preschool provides lower quality than the counterfactual mode of care for high SES children (i.e., the harmful effects are caused by an absolute worsening of skills), we think it is still worth discussing the second explanation further in terms of relative and absolute effects of universal preschool on skills. Because nearly all examples of significant harmful effects were found for children with relatively high SES (Fort et al., 2018; Havnes & Mogstad, 2015; Lebihan et al., 2017) or, in the case of Baker et al. (2015), was most likely driven by this group, we center the discussion on these studies.

For outcomes where there is some form of rivalry (like getting into popular college programs), one group doing better than the other could be explained in terms either relative or absolute gains. For non-rival outcomes, where one child's attainment does not decrease other children's attainment, harmful effects cannot be explained by relative gains. Havnes & Mogstad (2011, 2015) found significant harmful effects on total earnings and the probability of being a high and top earner for high SES children, while the effects on standardized tests, years of schooling, the probability of attending college, dropping out of high school, being a low earner, and being on welfare indicated beneficial effects (although many effects were small and insignificant). As one would expect more rivalry on local labor markets over (high) earnings than rivalry over the other measures, this pattern of results is consistent with harmful effects being due to a relatively greater improvement of skills for low SES children. The pattern is also consistent with preschool lowering the absolute level of some skills for high SES children that are important for earnings, but not (as important) for the other measures.

By the same reasoning, the significant harmful effects found by Fort et al. (2018) in Bologna, Italy, on a cognitive skills test, and by Baker et al. (2015) and Lebihan et al. (2017) in Quebec, Canada, on measures of anxiety, quality of life, and crime are more likely due to an absolute lowering of skills for high SES children, as these outcome measures have low or no degree of rivalry. The program in Quebec received low quality assessments, especially at its inception (e.g., Almond et al., 2017; Cascio, 2015; Lebihan et al., 2017; van Huizen & Plantenga, 2015) and had a low staff-to-child ratio. The program in Bologna is often considered to be of a high quality, although it has a relatively low staff-to-child ratio for the 0-2 age group studied by Fort et al. (2018). The samples in Baker et al. (2015) and Lebihan et al. (2017) also included younger children (their age range was 0-4 years), and, as discussed in the introduction, preschool may be more likely to have harmful effects for very young children. Relatively few other included studies examined 0-2-year-olds, and most of those that did also included older children. It is therefore difficult to tell from our sample whether these harmful effects were program or age specific, or a combination of the two.

The discussion of heterogeneous effects again highlights the importance of quality of care. However, as many universal preschool programs lower the cost of child care for families, beneficial effects may partly be explained by increased incomes. We cannot rule out such effects, but pure income effects have received relatively little support in related literatures (e.g., Heckman & Mosso, 2014, but see Black, Devereux, Løken, & Salvanes, 2014, for a counterexample in a preschool context). Some of our included studies examined programs with clearly positive income

effects, but still found significant harmful effects (e.g., Baker et al., 2015; Lebihan et al., 2017), which further suggests that the quality of universal preschool programs is of first-order importance.

4.3   Limitations

We limited the scope of the search to the most relevant databases and by requiring that studies contained keywords matching the three dimensions of population, intervention type, and type of comparison. A broader search would have included all studies that matched any of these dimensions, but was infeasible given the extremely large number of hits (see the section *Search Strings* in the appendix). It is possible that we missed relevant studies because of this. However, as we searched the most relevant databases and performed an extensive citation tracking effort, it is unlikely that we missed essential parts of the literature.

We included studies comparing different types of preschools against each other, but found very few of this kind, and no conclusions about the long-term effects can be drawn. Although we found relatively many studies comparing universal preschool programs to informal or parental care, few studies examined the same outcomes and measured those outcomes in the same way. Furthermore, the content and contexts of the included programs differed widely. There were thus few fully comparable outcomes, and we refrained from meta-analysis primarily for this reason. Therefore, we were unable to report effect sizes averaged across studies and to formally examine the consistency of effects across studies.

All reviews are limited by the topics examined by the included studies. No study in our sample examined whether there were long-term spillover effects from expanding universal preschool programs. For example, if universal preschool has beneficial (harmful) effects on skills, we could, given the strong connection between human capital and growth (e.g., Hanushek & Woessmann, 2008), expect increased (decreased) growth rates. Such effects would not just affect treated areas and children, but whole regions and countries.

An inherent limitation in a review of long-term outcomes is that it is unclear how the universal preschool programs examined in the included studies relate to present day programs. That is, extrapolation of the results to the universal preschool programs of today should be done with caution.

5   Conclusion

Our review indicated that universal preschool programs can yield important long-term benefits in many contexts and for many children. Beneficial effects may occur when the programs are of higher quality than children's counterfactual mode of care. This was more often the case for children from low SES families, whereas we found no consistent gender heterogeneity. Furthermore, developing country programs showed relatively large beneficial effects. Increasing access to universal preschool programs for low SES children may therefore reduce inequality, both within and between countries.

We want to stress, however, that it matters greatly how inequality is reduced. The harmful effects we found for primarily high SES children, although they constitute a minority of all effects, should be taken seriously. More research about the causes of these effects and how they can be avoided would be important for policy. More studies with attendance data and including young children would be especially valuable additions to the literature.

Choosing among preschools, rather than choosing whether to put their child in preschool or not, is the choice facing many parents in large parts of the world. The quality of preschools may be difficult to observe for parents (Mocan, 2007), especially as there is doubt about whether available quality indicators accurately reflect preschool quality (e.g., Sabol et al., 2013). Preschool types, based on a clearly advertised approach, for example, would be easier to observe. To validate information about preschool types, studies comparing them over long-term outcomes are needed.

6   References

Studies included in the synthesis are marked with *.

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *Review of Economic Studies*, *72*(1), 1–19. doi:10.1111/0034-6527.00321

Akinyemi, S. (2013). *The economics of education*. Strategic Book Publishing & Rights Agency, LLC.

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education* (Vol. 4, pp. 1–181). Amsterdam: Elsevier. doi:10.3386/w16822

Almond, D., Currie, J., & Duque, V. (2017). *Childhood circumstances and adult outcomes: Act II* (NBER Working Paper No. w23017). Retrieved from http://www.nber.org/papers/w2301.

Apps, P., Mendolia, S., & Walker, I. (2013). The impact of pre-school on adolescents' outcomes: Evidence from a recent English cohort. *Economics of Education Review*, *37*, 183–199. doi:10.1016/j.econedurev.2013.09.006.

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

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

*Baker, M., Gruber, J., & Milligan, K. (2015). *Non-cognitive deficits and young adult outcomes: The long-run impacts of a universal child care program* (NBER Working Paper No. 21571). Retrieved from http://www.nber.org/papers/w21571.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, *333*(975), doi: 10.1126/science.1204534.

*Bastos, P., Bottan, N. L., & Cristia, J. (2017). Access to Preprimary Education and Progression in Primary School: Evidence from Rural Guatemala. *Economic Development and Cultural Change*, *65*(3), 521-547. doi:10.1086/691090.

Bauchmüller, R., Gørtz, M., & Rasmussen, A. W. (2014). Long-run benefits from universal high-quality preschooling. *Early Childhood Research Quarterly*, *29*(4), 457–470. doi:10.1016/j.ecresq.2014.05.009

Belsky, J. (2001): Emanuel Miller Lecture: Developmental risks (still) associated with early child care. *Journal of Child Psychology and Psychiatry*, *23*, 396-404. doi:10.1111/1469-7610.00782

*Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education of primary school performance. *Journal of Public Economics, 93*, 219-234. doi:10.1016/j.jpubeco.2008.09.002.

*Berlinski, S., Galiani, S., & Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics, 92*, 1416-1440. doi:10.1016/j.jpubeco.2007.10.007.

Bertrand, M., & Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*, *5*(1), 32-64. doi:10.1257/app.5.1.32

Björklund, A., & Salvanes, K. G. (2011). Education and family background: Mechanisms and policies. In *Handbook of the Economics of Education* (Vol. 3, pp. 201-247). Amsterdam: Elsevier. doi:10.1016/b978-0-444-53429-3.0000

*Bietenbeck, J., Ericsson, S., & Wamalva, F. (2017). *Preschool attendance, school progression, and cognitive skills in East Africa* (IZA Discussion Paper Series, DP No. 11212). Retrieved from http://ftp.iza.org/dp11212.pdf.

*Bingley, P., Jensen, V. M., & Sander, S. (2018). One size fits all? Effects of universal daycare on long-run child and mother outcomes. Forthcoming in Sander, S. (2018). *Essays on investments in universal early childhood education*. Dissertation, University of Copenhagen.

*Bingley, P., & Westergaard-Nielsen, N. (2012). Intergenerational transmission and day care. In: Ermisch, J., M. Jäntti, and T. Smeeding (eds). *From Parents to Children: The Intergenerational Transmission of Advantage* (p. 190–204). New York: Russell Sage Foundation.

Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., ... & Devercelli, A. E. (2017). Early childhood development coming of age: Science through the life course. *Lancet*, *389*(10064), 77–90. doi:10.1016/s0140-6736(16)31389-7

Black, S. E., Devereux, P. J., Løken, K. V., & Salvanes, K. G. (2014). Care or cash? The effect of child care subsidies on student performance. *Review of Economics and Statistics*, *96*(5), 824–837. doi:10.1162/rest_a_00439

Black, S. E., Grönqvist, E., & Öckert, B. (2017). Born to lead? The effect of birth order on non-cognitive abilities. *Review of Economics and Statistics*, forthcoming. doi:10.1162/rest_a_00690

*Blanden, J., Del Bono, E., McNally, S., & Rabe, B. (2016). Universal pre-school education: The case of public funding with private provision. *Economic Journal, 126*, 682-723. doi:10.1111/ecoj.12374.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Hoboken, NJ: John Wiley & Sons, Ltd.

Borghans, L., Golsteyn, B. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, *113*(47), 13354–13359. doi:10.1073/pnas.1601135113

*Borraz, F., & Cid, A. (2013). Preschool attendance and school-age profiles: A revision. *Children and Youth Services Review*, *35*, 816–825. doi:10.1016/j.childyouth.2013.01.023.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual review of Psychology*, *53*(1), 371–399.

Bradley, R. H., & Vandell, D. L. (2007). Child care and the well-being of children. *Archives of Pediatrics & Adolescent Medicine*, *161*(7), 669–676. doi:10.1001/archpedi.161.7.669

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, *90*(3), 414–427. doi:10.1162/rest.90.3.414.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372. doi:10.3368/jhr.50.2.317

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science*, *343*(6178), 1478–1485. doi:10.1126/science.1248429

Carneiro, P., & Ginja, R. (2014). Long-term impacts of compensatory preschool on health and behavior: Evidence from Head Start. *American Economic Journal: Economic Policy*, *6*(4), 135–173. doi:10.1257/pol.6.4.135

Cascio, E. U. (2009). *Do investments in universal early education pay off? Long-term effects of introducing kindergartens into public schools* (NBER Working Paper No. 1495). Retrieved from http://www.nber.org/papers/w14951.

Cascio, E. U. (2015). The promises and pitfalls of universal early education. *IZA World of Labor*, *116*, 1–16. doi:10.15185/izawol.116

Cascio, E. U. (2017). *Does Universal Preschool Hit the Target? Program Access and Preschool Impacts* (No. w23215). National Bureau of Economic Research. doi:10.3386/w23215

*Cascio, E. U. & Schanzenbach, D. W. (2013). The impacts of expanding access to high-quality preschool education. *Brookings Papers on Economic Activity*, 127–178. doi:10.1353/eca.2013.0012.

Clarke-Stewart, A., Gruber, C. P., & Fitzgerald, L. M. (1994). *Children at home and in day care*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Costa, J. C., da Silva, I. C. M., & Victora, C. G. (2017). Gender bias in under-five mortality in low/middle-income countries. *BMJ Global Health*, 2(2), e000350. doi:10.1136/bmjgh-2017-000350

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

Currie, J., & Thomas, D. (1995). Does Head Start make a difference?. *American Economic Review*, *85*(3), 341–364. doi:10.3386/w4406

*Datta Gupta, N. & Simonsen, M. (2012). The effects of non-parental child care on pre-teen skill and risky behavior. *Economics Letters, 116*, 622-625. doi:10.1016/j.econlet.2012.06.020.

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

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

Devaux-Spatarakis, A. (2014). L'experimentation "telle qu'elle se fait": Lecons de trois experimentations par assignation aléatoire. *Formation Emploi: Revue Francaise de Sciences Sociales*, (126), 17–38.

Development Partners Group Tanzania (2007). *The Teacher Education Development and Management Strategy – 2007/2008 to 2010/2011*. Dar es Salaam: The Ministry of Education and Vocational Training. Retrieved from:
http://www.tzdpg.or.tz/fileadmin/documents/dpg_internal/dpg_working_groups_clusters/cluster
_2/education/3-Core_Documents/2.02-
Education_Sector/Teacher_Education_Development_and_Management_Strategy_2007-
08_2010-11.pdf.

Dodge, K. A., Bai, Y., Ladd, H. F., & Muschkin, C. G. (2017). Impact of North Carolina's early childhood programs and policies on educational outcomes in elementary school. *Child Development*, *88*(3), 996–1014. doi:10.1111/cdev.12645

*Dumas, C. & Lefranc, A. (2012). Early schooling and later outcomes: Evidence from pre-school extension in France. In: Ermisch, J., M. Jäntti, and T. Smeeding (eds). *From Parents to Children: The Intergenerational Transmission of Advantage* (p. 164–189). New York: Russell Sage Foundation, 2012.

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2015). *Early childhood education* (NBER Working Paper No. 21766). Retrieved from http://www.nber.org/papers/w21766.

*Felfe, C., & Lalive, R. (2010). *How does early child care affect child development? Learning from the children of German unification*. Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Economics of Child Care and Child Development, No. B11-V2.

*Felfe, C., Nollenberger, N., & Rodríquez-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. doi: 10.1007/s00148-014-0532-x

Finansministeriet (2017). *Vejledning i samfundsøkonomiske konsekvensvurderinger*. Copenhagen: Finansministeriet. Retrieved from https://www.fm.dk/publikationer/2017/vejledning-i-samfundsoekonomiske-konsekvensvurderinger.

*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 and Policy, 8*(1), Article 46. doi:10.2202/1935-1682.1897

Flaherty, S. C., & Sadler, L. S. (2011). A review of attachment theory in the context of adolescent parenting. *Journal of Pediatric Health Care*, *25*(2), 114–121. doi:10.1016/j.pedhc.2010.02.005

*Fort, M., Ichino, A. & Zanella, G. (2018). *The cognitive cost of daycare 0-2 for children in advantaged families.* Unpublished manuscript. Retrieved from
http://www.andreaichino.it/research_progress/fort_ichino_zanella_rev1.pdf.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. doi:10.1126/science.1255484.

García, J. L., Heckman, J. J., Leaf, D. E., & Prados, M. J. (2016). *The life-cycle benefits of an influential early childhood program* (NBER Working Paper No. 22993). Retrieved from http://www.nber.org/papers/w22993.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., ... & Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science*, *344*(6187), 998–1001. doi:10.1126/science.1251178

Goossens, F. A., & Ijzendoorn , M. H. (1990). Quality of infants' attachments to professional caregivers: Relation to infant-parent attachment and day-care characteristics. *Child Development*, *61*(3), 832–837. doi:10.2307/1130967

*Haimovich Paz, F. (2015). The long-term return to early childhood education: Evidence from the first US kindergartens. In Haimovich Paz, F., *Three Essays on the Economics of Education and Early Childhood* (Ch. 1), Doctoral dissertation, University of California. Retrieved from https://escholarship.org/uc/item/4vd7d3x9.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, *46*(3), 607–68. doi:10.1257/jel.46.3.607

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4–9.

*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. doi: 10.1257/pol.3.2.97.

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

Heckman, J. J., Biroli, P., Del Boca, D., Heckman, L. P., Koh, Y. K., Kuperman, S., Moktan, S., Pronzato, C. D. & Ziff, A. L. (2017).* Evaluation of the Reggio Approach to early education*. Research in Economics,* forthcoming*.*

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, *94*(1-2), 114–128. doi:10.1016/j.jpubeco.2009.11.001.

Heckman, J. J., & Mosso, S. (2014). The economics of human development and social mobility. Annual Review of Economics, *6*(1), 689–733. doi:10.3386/w19925

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), 2052–2086. doi:10.1257/aer.103.6.2052

Henry, G. T., & Rickman, D. K. (2007). Do peers influence children's skill development in preschool?. *Economics of Education Review*, *26*(1), 100–112. doi:10.1016/j.econedurev.2005.09.006

*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. doi:10.1086/689478.

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration. Retrieved from http://handbook.cochrane.org.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62*(2), 467–475. doi:10.2307/2951620

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success* (OECD Education Working Papers, No. 110). OECD Publishing, Paris. Retrieved from http://dx.doi.org/10.1787/5jxsr7vr78f7-en.

Kline, P., & Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of Head Start. *Quarterly Journal of Economics, 131*(4), 1795–1848. doi:10.1093/qje/qjw027

*Lebihan, L., Haeck, C. & Merrigan, P. (2017). Universal childcare and long-term effects on child well-being: Evidence from Canada. *Journal of Human Capital*, forthcoming.

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics, 122*(1), 159–208. doi:10.1093/qje/qjw027

New Zealand Treasury (2015). *Guide to social cost benefit analysis*. Wellington: New Zealand Treasury. Retrieved from http://purl.oclc.org/nzt/g-cba

Magnuson, K. A., Kelchen, R., Duncan, G. J., Schindler, H. S., Shager, H., & Yoshikawa, H. (2016). Do the effects of early childhood education programs differ by gender? A meta-analysis. *Early Childhood Research Quarterly, 36*, 521–536. doi:10.1016/j.ecresq.2015.12.021

MacKinnon, J. G., & Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2), 233–254. doi: 10.1002/jae.2508

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., ... & Shonkoff, J. P. (2017). Impacts of early childhood education on medium- and long-term educational outcomes. *Educational Researcher, 46*(8), 474–487. doi:10.3102/0013189x17737739

Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., ... & Leseman, P. (2015). *A review of research on the effects of early childhood education and care (ECEC) upon child development*. CARE project; Curriculum quality analysis and impact review of European early childhood education and care (ECEC). Retrieved from http://ecec-care.org/fileadmin/careproject/Publications/reports/new_version_CARE_WP4_D4_1_Review_on_the_effects_of_ECEC.pdf.

Mocan, N. (2007). Can consumers detect lemons? An empirical analysis of information asymmetry in the market for child care. *Journal of Population Economics, 20*, 743–780. doi:10.1007/s00148-006-0087-6

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine 6*(7), e1000097. doi:10.1371/journal.pmed1000097

NICHD Early Child Care Research Network. (2002). Child-care structure→ process→ outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science*, *13*(3), 199–206.

OECD (2016). *Society at a Glance 2016: OECD Social Indicators*. OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264261488-en.

OECD (2017). *Public spending on childcare and early education*. OECD family database. Retrieved from http://www.oecd.org/els/family/database.htm.

Phillips, D., Lipsey, M., Dodge, K., Haskins, R., Bassok, D., Burchinal, M., & Weiland, C. (2017). *The current state of scientific knowledge on pre-kindergarten effects*. Brookings Institution and the Duke Center for Child and Family Policy. Retrieved from https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf.

Reynolds, A. J., & Ou, S.-R. (2011). Paths of effects from preschool to adult well-being: A confirmatory analysis of the Child–Parent Center Program. *Child Development*, *82*(2), 555–582. doi:10.1111/j.1467-8624.2010.01562.x

Reynolds, A. J., & Temple, J. A. (2008). Cost-effective early childhood development programs from preschool to third grade. *Annual Review of Clinical Psychology*, *4*, 109–139. doi:10.1146/annurev.clinpsy.3.022806.091411

Rossin-Slater, M., & Wüst, M. (2017). *What is the added value of preschool? Long-term impacts and interactions with an infant health intervention* (NBER Working Paper No. 22700). Retrieved from https://pdfs.semanticscholar.org/dfee/70be7360060910e2615110a69da8b42f1a59.pdf.

Ruhm, C., & Waldfogel, J. (2012). Long-term effects of early childhood care and education. *Nordic Economic Policy Review*, *1*(1), 23–51. doi:10.6027/9789289329873-2-en

Sabol, T. J., Hong, S. S., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-K programs predict children's learning?. *Science*, 341(6148), 845–846. doi:10.1126/science.1233517

*Smith, A. (2015). *The long-run effects of universal pre-K on criminal activity*. Unpublished manuscript. Retrieved from https://ssrn.com/abstract=2685507.

Strachan, D. P. (1989). Hay fever, hygiene, and household size. *BMJ: British Medical Journal*, *299*(6710), 1259–1260. doi:10.1136/bmj.299.6710.1259

UNESCO (2005). *Policy review report: Early childhood care and education in Kenya.* Early Childhood and Family Policy Series. Retrieved from: http://unesdoc.unesco.org/images/0013/001390/139026e.pdf.

UNESCO (2006a). *Argentina – Early childhood care and education (ECCE) programmes.* UNESCO International Bureau of Education (IBE). Retrieved from: http://unesdoc.unesco.org/images/0014/001480/148089e.pdf.

UNESCO (2006b). *Guatemala – Early childhood care and education (ECCE) programmes.* UNESCO International Bureau of Education (IBE). Retrieved from: http://unesdoc.unesco.org/images/0014/001481/148101e.pdf.

UNESCO (2006c). *Uruguay – Early childhood care and education (ECCE) programmes.* UNESCO International Bureau of Education (IBE). Retrieved from: http://unesdoc.unesco.org/images/0014/001481/148111e.pdf.

UNESCO (2018). Gross enrolment ratio, pre-primary, both sexes (%). UNESCO Institute for Statistics. Retrieved from: http://data.uis.unesco.org/index.aspx?queryid=142#

van Huizen, T., & Plantenga, J. (2015). *Universal child care and children's outcomes – A meta-analysis of evidence from natural experiments* (U.S.E Discussion Paper Series nr: 15-13). Retrieved from https://dspace.library.uu.nl/handle/1874/324613.

*van Huizen, T., Plantenga, J. & Dumhs, L. (2017). The costs and benefits of investing in universal preschool: Evidence from a Spanish reform. *Child Development,* forthcoming. doi:10.1111/cdev.12993.

Waldfogel, J. (2015). The role of preschool in reducing inequality. *IZA World of Labor*, 219. doi:10.15185/izawol.219

World Bank (2012). *Tanzania – Early childhood development.* Retrieved from: http://documents.worldbank.org/curated/en/986161468134082667/pdf/799320WP0SABER0box0379795B00PUBLIC0.pdf.

World Bank (2017). *Gross enrolment ratio, pre-primary, both sexes (%).* UNESCO Institute for Statistics. Retrieved from: https://data.worldbank.org/indicator/SE.PRE.ENRR

Young, A. (2017). *Consistency without inference: Instrumental variables in practical application.* Unpublished manuscript, London: London School of Economics and Political Science. Retrieved from: http://personal.lse.ac.uk/YoungA/.

Appendix

The contents of this appendix are as follows: The section Information about Included Studies describes the included studies and examined preschool programs in more detail. The section Examples of Excluded Studies provides examples of studies that we screened in full text but did not include. The section Additional Results from the Search and Screening Process contains additional details about the results of the search and screening process. The section Included Estimates describes the motivation for choosing one estimate over another in the cases where the choice did not obviously follow the principles described in the section Analysis in the main text. The section Search Strings contains the full search strings for the electronic databases.

Information about Included Studies

In Appendix Table 1 below, the included studies are described in terms of what country and region were studied, the preschool program and control condition, staff-to-child ratios, group sizes and staff education, and the natural experiment and estimation strategy used. Studies are listed in alphabetical order, except that studies of the same preschool programs are grouped together. When information about certain details about a preschool program was not included in a study, we included related information (e.g., group sizes instead of staff-to-child ratio, or, if possible, information used from other sources. These are referenced in the table. All other information was taken from the included studies. We used a few acronyms: difference-in-differences (DID), intention-to-treat (ITT), local average treatment effect (LATE), and regression discontinuity (RD). N denotes the number of areas included in the estimations, and n the number of child observations. Both numbers refers to the sample sizes used in the estimations of the mean effects. Ranges refer to the minimum and maximum N and n used in a study.

Appendix Table 1   Characteristics of included studies

Included study	Country/region, period, and sample	Preschool program(s) & control condition	Staff-to-child ratio & staff education in preschool program(s)	Identification and estimation
Baker, Gruber & Milligan (2015) Lebihan, Haeck & Merrigan (2015)	*Country/Region:* Canada *Period*: 1997-2001 Preschool program introduced in 1997 and phased in over a period of four years to 2001. *Sample*: 0-4 years old. The program was open for four-year-olds in 1997 and became available for 0-1 years in 2000-2001. *N* = 10 *n* = 10,857-45,242 (not reported per specification in Baker et al., 2015)	*Preschool program:* Quebec introduced a subsidy on universal preschool for children aged 0-4 in 1997, making preschool available for everyone for 5 dollars a day. The program was introduced step-wise by age. Preschool under the program was provided in two venues: preschool centers (centres de la petite enfance, CPE) and home-based care. *Control condition*: Children in Quebec shift from informal care into center-based care. The proportion of 0-4-year-olds in care rose by 14 percentage points, or roughly one-third of the baseline rate. There are no substantial changes in the number of children that were cared for in their own home (Baker et al., 2008, Table 2, p. 724), indicating that the introduction of publicly available preschool crowds out informal care arrangements/privately provided child care.	*Staff-to-child ratio*: 0-3-year-olds: 1:8 4-5-year-olds: 1:10 (Baker et al., 2008, p. 717). *Staff education*: Two-thirds of staff must have a college diploma or university degree in early childhood education (Baker et al., 2008, p. 717).	*Identification:* Exploits the introduction of the subsidy on universal preschool for children aged 0-4 in Quebec. They use the rest of Canada as a control group. *Estimation*: Use a DID strategy to estimate an ITT effect of being more exposed to a universal program, as the sample comprises all children and not only those that attend a preschool program.
Bastos, Bottan & Cristia (2016)	*Country/Region:* Guatemala, Rural communities *Period*: 1992-2000 *Sample*: 4-6-year-olds. *N* = 960 *n* = 8,543	*Preschool program:* Guatemala expanded their provision of public pre-primary schools from 5,300 to 11,500 during the period 1998-2005. The beneficiary communities were selected by the central government with no strict guidelines. *Control condition:* Mainly parental care, as 0.8-1.2% of the communities had a private preprimary in 2005. Little or no crowding out of informal or private alternatives.	*Staff-to-child ratio*: Not reported. *Staff education:* Teachers must have a pre-primary education qualification; this is obtained in teacher-training colleges (UNECO, 2006b). However, this information is from 2006, a few years after the period examined in this article. Staff requirements may have been different in the period examined in the paper.	*Identification:* Exploits the large expansion of pre-primary schools and the variation over time and between communities. *Estimation:* Use a DID strategy with trimming and propensity score reweighting to estimate an ITT effect. The authors also estimate a TOT effect, but without any data on actual attendance.
Berlinski, Galiani & Manacorda (2008) Borraz & Cid (2013)	*Country/Region:* Uruguay *Period*: 1995-2004 Berlinski et al. (2008): *Sample*: The preschool program comprises 3-5-year-olds.	*Preschool program:* Following a reform in the mid-1990s, the Government of Uruguay made pre-primary education universally available. Enrollment in public preschool rose by 76% over nine years. The expansion attracted mostly children from disadvantaged backgrounds. *Control condition:* Private provision/Informal care (not explicitly described). Private fee-based education was common. In Montevideo, around one third of the children	*Staff-to-child ratio*: 3-year-olds: groups of 20 4-year-olds: groups of 25 5-year-olds: groups of 35 *Staff education:* Early education teachers study in teacher training colleges to earn a qualification at the	*Identification:* Exploits the expansion in the provision of public pre-primary education. Comparing siblings that had different access to pre-primary education. Main specification contrasts having attended preschool for 1-3 years (treatment) with the group that have 0-1 year (control).

Included study	Country/region, period, and sample	Preschool program(s) & control condition	Staff-to-child ratio & staff education in preschool program(s)	Identification and estimation
	N=55 *n*= 23,042 Borraz & Cid (2013) *Sample:* 4-5-year-olds *N* = not reported *n* = 19,732	in primary education attended a private institution.	non-university tertiary level (UNESCO, 2006c. However, this information is from 2006, a few years after the period examined in this article. Staff requirements and ratios may have been different in the period examined in the paper.)	*Estimation:* Berlinski et al. (2008) use a sibling fixed-effects strategy to estimate a TOT effect. Borraz & Cid (2013) instrument preschool attendance with the mean preschool attendance by child age in each locality.
Berlinski, Galiani & Gertler (2009)	*Country/Region:* Argentina *Period*: 1994-1999 *Sample*: 3-5-year-olds *N* = 407-417 municipalities; 2750-3024 schools *n* =117,515-145,292	*Preschool program:* Argentina increased the number of preschool classrooms during the period 1993-1999. The increase in pre-primary enrollment varies between provinces. All provinces increased enrollment by at least 10 percentage points. *Control condition:* family care (not explicitly described).	*Staff-to-child ratio*: Not reported, but class size is 25. *Staff education:* Preschool teachers must be trained on teacher training colleges or at universities (UNESCO, 2006a). However, this information is from 2006, a few years after the period examined in this article. Staff requirements may have been different in the period examined in the paper.	*Identification:* Exploits the variation of treatment intensity across regions and cohorts following the expansion of pre-primary school facilities. They are unable to separate one, two or three years of exposure. *Estimation:* Use a DID strategy to estimate an ITT effect. The authors write that they cannot reject that the take-up rate was one, which would result in the estimates being close to a TOT estimate.
Bietenbeck, Ericsson & Wamalva (2017)	*Country/Region:* Kenya and Tanzania *Period*: Kenya: 2000-2013 Tanzania: 2000-2012 *Sample*: Kenya: 3-6-year-olds *N* = not reported *n* = 223,339 Tanzania: 5-6-year-olds *N* = 120 *n* = 293,757	*Preschool program:* There are three types of preschool in Kenya: public preschool, private preschool, and information neighborhood schools. In Tanzania, the vast majority of preschools are public. During the period 1997-2004, preschool enrollment increased from 79-84% in Kenya and from 61-69% in Tanzania. *Control condition:* Family care (not explicitly described).	*Staff-to-child ratio*: Kenya: 1:25-27 Tanzania: increased from 1:45 in 2007 to 1:100 in 2011 in state schools. *Staff education:* Kenya: Primary or secondary education. 41.4% Trained teachers Tanzania: Teachers must have completed lower-secondary school. Information for Kenya: (UNESCO, 2005) Information for Tanzania: (World Bank, 2012).	*Identification:* Compare differences between siblings. The authors argue that differences between siblings are due to changes in the local availability of preschool because of an expansion of the pre-primary sector during the studied period. *Estimation:* Use a sibling fixed-effect strategy to estimate a TOT effect.

Included study	Country/region, period, and sample	Preschool program(s) & control condition	Staff-to-child ratio & staff education in preschool program(s)	Identification and estimation
Bingley, Jensen & Sander (2018) Bingley & Westergård-Nielsen (2012)	*Country/Region:* Denmark Bingley et al. (2018): *Period:* 1967-1979 *Sample:* 3-6-year-olds *N* = 1,098 *n* = 403,241 Bingley & Westergård-Nielsen (2012): *Period:* 1976-1989 *Sample:* 0-6-year-olds *N* = 275 *n* = 531,733	*Preschool program:* A reform from 1964 increased the number of preschool slots. From 1966 to 1979, the number of institutions tripled (2018). From 1976-1989, preschool coverage tripled for the youngest children (age 1-2) and doubled for the oldest (age 3-6). *Control condition:* parental care/informal/private non-maternal care (mainly for high SES mothers with work). Private alternatives to public institutions existed.	*Staff-to-child ratio:* Not reported *Staff education:* The duration of the education was three years. Each institution was to have a certain minimum proportion of educated personnel but could also employ untrained helpers.	*Identification:* Exploit the step-wise roll-out of reforms increasing universal preschool provision in Denmark. Use variation over time and between municipalities. *Estimation:* Use a DID strategy to estimate an ITT effect.
Bladen, Bono McNally & Rabe (2016)	*Country/Region:* England *Period:* 2002-2007 *Sample:* 3-4-year-olds *N* = 888 *n* = 2,900,000	*Preschool program:* England implemented universal part-time preschool for three-year-olds in the early 2000s. The government funded private and voluntary institutions to provide free early education places. The expansion happened entirely in the private sector. *Control condition:* private or parental care. The expansion in preschool mainly crowds out private provision of preschool, as 82% of 3-years-olds already attend some type of preschool education before the reform. The expansion increases the enrollment of three-year-olds by 14.4 percentage points.	*Staff-to-child ratio:* Public sector: 1:13. Private sector: 1:8 if no qualified teacher, 1:13 if qualified teacher. *Staff education:* Public sector: Almost all employed staff hold a degree. Private sector: 10-20% hold a degree.	*Identification:* Exploits the staggered implementation of universal part-time preschool education for 3-year-olds across Local Education Authorities (LEAs) in England. Compare low and high intensity areas. *Estimation:* Use a DID strategy to estimate an ITT effect.
Cascio & Schanzenback (2013) Fitzpatrick (2008)	*Country/Region:* Cascio & Schanzenback (2013): US, Georgia & Oklahoma Fitzpatrick (2008): US, Georgia *Period:*	*Preschool program:* Georgia and Oklahoma introduced universal preschool for 4-year-olds in the 1990s. The program in Georgia and Oklahoma increased the likelihood of enrollment in preschool at age four by 19-20 percentage points for low SES children and 11-14 percentage points for high SES children. The enrollment in pre-kindergarten in Georgia increased from 13.9% in 1995 to 53.0% in 1999. *Control condition:* informal/formal care, different for different subgroups of children. High SES children move from	*Staff-to-child ratio:* 1:10. *Staff education:* In both states, classroom lead teacher must hold a bachelor degree and participate in annual training.	*Identification:* Cascio & Schanzenback (2013): Compare changes in preschool enrollment in the two states that introduced universal preschool initiatives with the rest of the country over the same period. Fitzpatrick (2008): The article compares children in Georgia that were

Included study	Country/region, period, and sample	Preschool program(s) & control condition	Staff-to-child ratio & staff education in preschool program(s)	Identification and estimation
	Cascio & Schanzenback (2013): Georgia: 1995-2005 Oklahola: 1998-2005 Fitzpatrick (2008): 1995-1999 *Sample*: 4-year-olds *N* = 50 Cascio & Schanzenback (2013): *n* = 295-334 state-years Fitzpatrick (2008): *n* = 537,112-1,241,994	private to public preschool.		offered the public pre-kindergarten to children in other states and children before the program was introduced. *Estimation:* Use a DID strategy to estimate an ITT effect. Cascio & Schanzenback (2013) also perform a benefit-cost analysis.
Datta Gupta & Simonsen (2012, 2016)	*Country/Region:* Denmark *Period*: 1996-1997 2012: *Sample*: 3-year-olds *N*= not reported *N* = 2,571-3,784 2016*: Sample:* 2-year-olds *N* = 253 *N* = 60,907	*Type comparison:* Compare center-based preschool to non-center-based but municipally-regulated family day care. *Preschool program:* Most children enrolled in family day care eventually enroll in center-based care. The interpretation of the result is an additional 1.5 years of early center-based care. They have data on actual attendance. At age 2 (3), 25% (33%) of enrolled children attend center-based child care arrangements, and 75% (67%) attend family day care.	*Staff-to-child ratio*: Center-based: 1:3.5 Family day care: 1:5 or less *Staff education:* Center-based: Most of core center staff hold a pedagogical degree Family day care: No formal education, but are offered vocational courses.	*Identification:* 19-30% of the municipalities offer guaranteed access to center-based preschool. In the municipalities that offer the guaranteed access, children have a higher probably of getting access to center-based preschool. *Estimation:* Use an IV strategy to estimate LATEs of center-based care relative to family day care for the group of children whose parents choose center-based care when access is guaranteed, but not otherwise.
Dumas & Lefranc (2012)	*Country/Region:* France *Period*: 1952-1983	*Preschool program:* During the 1960s and 1970s the enrollment in preschool for 3-year-olds rose from 35% to 90%. The increase varied between regions.	*Staff-to-child ratio*: Class size: 25 children. *Staff education:* Preschool teachers	*Identification:* In a robustness check, the authors exploit regional variation in access to preschool.

Included study	Country/region, period, and sample	Preschool program(s) & control condition	Staff-to-child ratio & staff education in preschool program(s)	Identification and estimation
	Sample: 2-5-year-olds *N* = 95 *n* = 6,799-21,710	*Control condition:* parental care. The contrast is described as getting one more year of preschool.	have a bachelor's degree.	*Estimation:* Use an IV strategy to estimate a LATE.
Felfe & Lalive (2010)	*Country/Region:* Germany, every state expect Berlin *Period:* 1996-2000 *Sample:* 0-3-year-olds *N* = Not reported *n* = 850	*Preschool program:* A substantial difference exists in child care offer rates across Germany, due to historical differences in the separated East and West Germany. Child care coverage rates are of the order of 40% in the former East Germany and below 10% in the former West Germany. There is also variation within regions. *Control condition:* Informal/parental care, although informal care arrangements were rarely used.	*Staff-to-child ratio*: East: 1:6.8 West: 1:5.1 *Staff education:* Staff have to undergo special training before being allowed to work in the sector. *Staff with a degree in child care:* East: 90% West: 84%	*Identification:* The authors use the difference in child care offer rates across Germany induced by the former East/West division as an instrument for attending preschool. *Estimation:* Use an IV strategy to estimate the effects of formal care for children of mothers who use formal care because of an increase in the child care offer rate.
Felfe, Nollenberger & Rodrígues-Planas (2015) Van Huizen, Duhms & Plantenga (2017)	*Country/Region:* Spain *Period:* 1991-1996 *Sample:* 3-year-olds. *N* = 15 (treatment: 8, control: 7) *n* = 20,458-40,340	*Preschool program:* Spain expanded their subsidized full-time, high quality universal child care supply in the early 1990s. The enrollment of 3-year-olds in public child care increased from 8.5 to 67.1% from 1990/1991 to 2002/2002. *Control condition:* mainly parental care, but part of the control group might have been in preschool.	*Staff-to-child ratio*: Maximum number of children per class is 20. *Staff education:* Preschool teachers are required to have a college degree in pedagogy.	*Identification:* Exploits the variation in the speed of expansion across states. Divide 15 states into treatment and control based on their increase in public child care enrollment of 3-year-olds. *Estimation:* Use a DID strategy to estimate an ITT effect. Estimates the effect of having a greater opportunity of one year of preschool when the child is three, no data over actual preschool attendance. Van Huizen et al. (2017) perform a benefit-cost analysis.
Fort, Ichino & Zanellax (2018)	*Country/Region:* Italy, Bologna *Period*: 2001-2005 *Sample*: 0-2-year-olds *N* = 1 *n* = 444	*Preschool program:* Parents in Bologna apply for a preferred child care program. Acceptance into a preferred child care program depends on the Family Affluence Index. Less affluent families get offered a spot first. This creates a threshold. On average, children that get offered the preferred spot will be in child care for a longer time, compared to children that are not offered their preferred spot. *Control condition:* informal/parental care. Private day care	*Staff-to-child ratio*: 0-year-olds: 1:4 1-2-year-olds: 1:6 *Staff education:* Not reported	*Identification:* Use the threshold in the admission system that determines whether children are offered a preschool slot as an instrument for attendance. *Estimation:* Use a fuzzy RD strategy to estimate a LATE.

Included study	Country/region, period, and sample	Preschool program(s) & control condition	Staff-to-child ratio & staff education in preschool program(s)	Identification and estimation
		is almost absent; extended family services are the most relevant substitution for day care.		
Haimovich Paz (2015)	*Country/Region:* USA *Period:* 1890-1910 *Sample:* White males, 4-6-year-olds *N* = 220 *n* = 20,263-239,390	*Preschool program:* The kindergarten movement provided preschool for children aged 4-6. The increase in enrollment in the years following the incorporation of public kindergartens was rapid in many cities, ranging from 20 to 80 percentage points. *Control condition:* The mothers were most likely the care providers before the kindergarten movement. Some crowding out of private alternatives.	*Staff-to-child ratio*: Not reported *Staff education:* Most kindergarten teachers were high school graduates with two years of specific training that included child psychology, music, and children's literature.	*Identification:* Exploit geographical variation and variation over time in the number of public kindergartens in cities following the kindergarten movement. *Estimation:* Use a DID strategy to estimate ITT effects.
Havnes & Mogstad (2011, 2015)	*Country/Region:* Norway *Period:* 1976-1979 *Sample:* 3-6-year-olds. *N* = 414 *n* = 499,026 (2011) *n* = 341,170 (2015)	*Preschool program:* A reform from 1975 increased the federal subsidy for child care. The local government was responsible for offering child care. The reform created large variation in the access to child care across municipalities and over time. *Control condition:* The analysis suggests that the new subsidized child care crowded out informal child care arrangements with almost no net increase in total use or maternal labor supply.	*Staff-to-child ratio*: 1:8 with at least one educated preschool teacher per 18 children. *Staff education:* Every formal child care institution had to be run by an educated preschool teacher responsible for day-to-day management. Preschool education is a college degree with supervised practice in a formal preschool institution included.	*Identification:* Compare municipalities with high coverage to municipalities with low coverage (above or below median percentage point increase in preschool coverage rates) *Estimation:* Use a DID strategy to estimate an ITT effect.
Heckman et al. (2017)	*Country/Region:* Italy, Reggio Emilia, Parma & Padova *Period*: 1954-2000 *Sample*: 0-6-year-olds *N*=3 *n:* Adolescents = 836 Adults 30s = 782 Adults 40s = 791 Adults 50s = 449	*Type comparison:* Reggio Emilia approach is compared to the approaches in the nearby cities of Padova and Parma. *Preschool program:* The Reggio Emilia approach is notable for its investment in staffing, early inclusion of children with disabilities, and high rates of provision of early childhood services.	*Staff-to-child ratio*: 3-year-olds: 1:12-13 *Staff education:* On a biweekly basis, a pedagogista with at least a bachelor's degree in psychology or pedagogy supports the professional development for the educational staff of approximately 4-5 municipal preschools.	*Identification:* Compare children from Reggio Emilia with children from Parma and Radova, who received different kinds of child care approaches. *Estimation:* Use a DID with matching strategy to estimate an ITT effect.
Herbst (2017)	*Country/Region:* USA, all states except New Mexico *Period:* 1943-1946	*Preschool program:* During World War 2, The Lanham Act established center-based preschool for children aged 0-5 and after-school services for children aged 6-12. The intensity differed between states.	*Staff-to-child ratio*: 1:10. *Staff education:* Program employed certified school teachers and con-	*Identification:* The article exploits the variation between states with low/high spending on the preschool program for children in states with

Included study	Country/region, period, and sample	Preschool program(s) & control condition	Staff-to-child ratio & staff education in preschool program(s)	Identification and estimation
	Sample: 0-12-year-olds *N* = 47 *n:* age 24-39 = 456,070 age 34-49 = 2,500,553 age 44-59 = 2,481,049	*Control condition:* Parental care (not explicitly described).	tracted with universities to establish formal training programs.	high spending. *Estimation:* Use a DID strategy to estimate an ITT effect.
Smith (2015)	*Country/Region:* USA, Oklahoma *Period:* 1998-1999 *Sample:* 4-year-olds *N* = 1 *n* = 365	*Preschool program:* Oklahoma introduced universal pre-kindergarten in the 1998-1999 school year. To attend kindergarten, the child had to be five by 1 September. This created a birthday cut-off at the year of the implementation, where children born on or before 1 September were assigned to kindergarten, while children born after 1 September were assigned to pre-kindergarten. Around 60% of students offered pre-kindergarten attended. *Control condition:* Formal/private/parental care. The prior conditions were a mix of Head Start, private preschool and no preschool (approximately 20%, 25% and 50%)	*Staff-to-child ratio:* Maximum 1:10 *Staff education:* Pre-kindergarten teachers are required to be certified in early childhood education.	*Identification:* The author uses the birthday cut-off at the year of the implementation of pre-kindergarten. *Estimation:* Use a RD strategy to estimate an ITT effect.

Note: Included studies in alphabetical order, except that studies of the same preschool programs are grouped together. When information about, for instance, staff education was not available in an included study, we used, if possible, information from other sources. These are referenced in the table. All other information is taken from the included studies. Acronyms: difference-in-differences (DID), intention-to-treat (ITT), local average treatment effect (LATE), regression discontinuity (RD). N denotes the number of areas included in the estimations, n the number of child observations. Both numbers refer to the sample sizes used in the estimations of the mean effects.

Examples of Excluded Studies

To illustrate how we applied the inclusion criteria, we give examples of excluded studies for each criterion below. Note that studies could have been excluded by several criteria, but we only mention one below.

Primary empirical research: Bradley and Vandell (2007) was excluded because it did not contain primary empirical research, but a review of child care studies on the impact of age at entry and amount, quality, and type of care on children's adaptive functioning.

Preschool programs: Cascio (2009) studied the long-run effect of introducing kindergarten programs as a part of (public) primary school. As these programs were an integrated part of primary school, they did not count as preschool according to our definition, and we excluded the study from the analysis.

Universal programs: Dodge, Bai, Ladd, and Muschkin (2017) studied the long-term effects of North Carolina's Smart Start and More at Four early childhood programs. These programs primarily targeted disadvantaged children and high-risk children, and the study was therefore excluded.

Long-term child outcomes: Baker, Gruber, and Milligan (2008) studied the same introduction of highly subsidized preschool in Quebec as Baker et al. (2015) and Lebihan et al. (2017), but reported outcomes for younger children (primarily 0-4 years) and therefore the study was excluded.

Types of comparisons: Similar to the above-mentioned studies from Canada, Black et al. (2014) used a subsidy scheme to study long-term child outcomes. However, as there were no effects on preschool utilization from a sharp discontinuity in the subsidy scheme the study did not examine any effects of different types of care and was therefore excluded.

Country, period, publication status, and language: We did not restrict inclusion by country, time period, or publication status of the study, but included only studies written in a language that at least two members of the research team understand (Danish, English, German, Norwegian, and Swedish). Devaux-Spatarakis (2014) was only in available in French and was therefore excluded.

Methods: Apps, Mendolia, and Walker (2013) used an elaborate matching procedure to control for a very rich set of child and family characteristics and to estimate the impact of preschool on adolescent outcomes. However, they did not use any natural experiment or randomized experiment in the identification and estimation of the effects, and we excluded the study for this reason.

Additional Results from the Search and Screening Process

The search of the electronic databases yielded 1,516 unique records (1,861 before duplicates were removed). Appendix Table 2 shows the distribution of records in databases. We identified an additional 86 records from other sources and screened a total of 145 studies in full text. Of these, 25 were included. The full search and screening process is illustrated in Appendix Figure 1 below (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009).

Appendix Table 2   The distribution of records per database.

Database	Hits
Academic Search Premier	434
ECONLIT	238
ERIC	381
PsycINFO	694
SocIndex	100
Teacher Reference Center	14
Total	**1,861**

Appendix Figure 1   Flowchart of the search and screening process
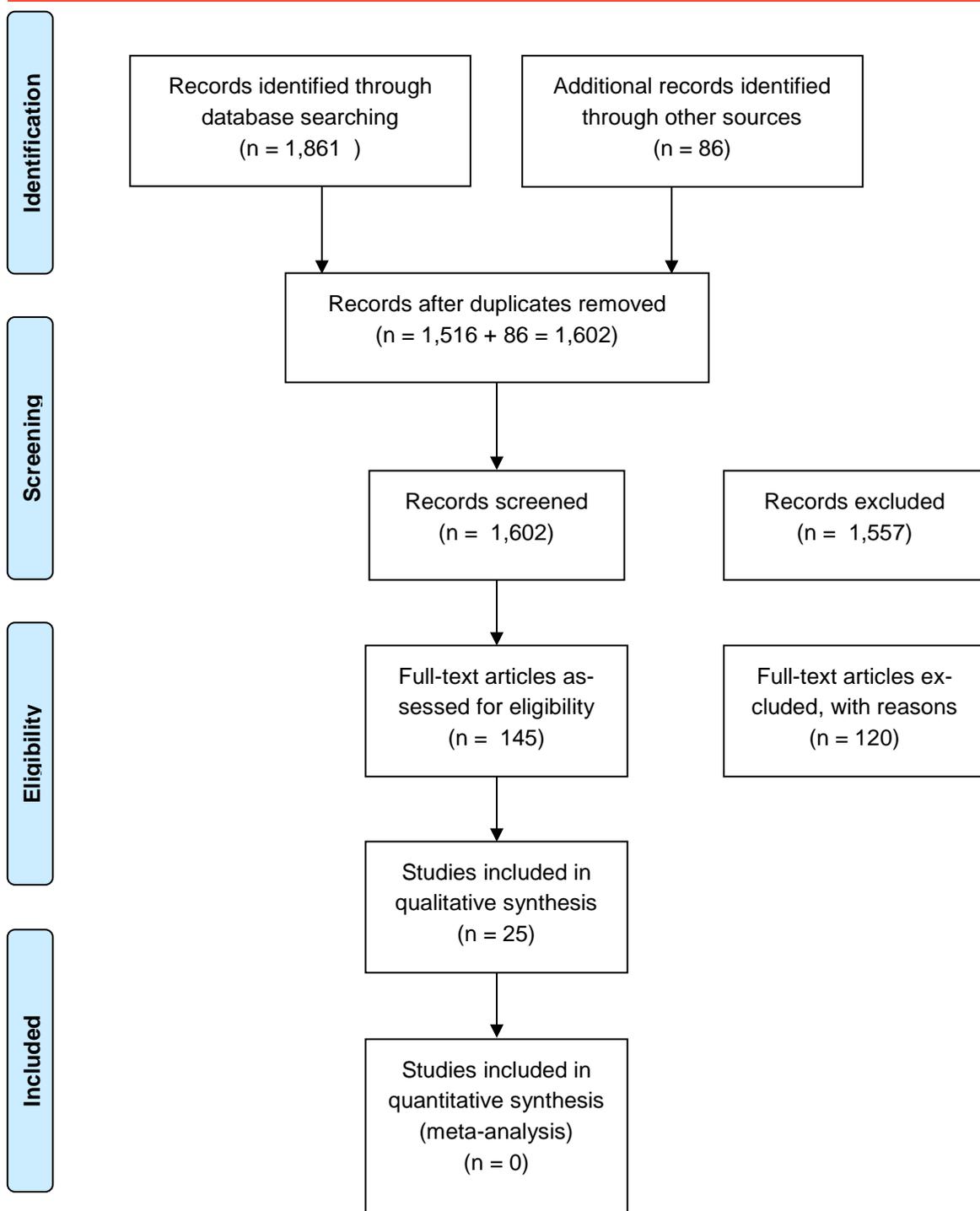
Included Estimates

This section provides a motivation of our choice of included estimates in the cases where there were overlapping samples between two studies, or where the choice was not obvious from the principles laid out in the section Analysis.

Health, Well-Being, and Behavior

Baker et al. (2015) and Lebihan et al. (2017) examined the effects of a preschool reform in Quebec, Canada, used similar estimation methods, and reported outcomes from partially overlapping samples. We included Lebihan et al.'s estimates in the analysis of problem behavior, as they provided separate estimates for children aged 8-9, and in the analysis of health, healthy behaviors, and well-being, as they had access to a further survey wave. Except for life satisfaction/quality of life, where Lebihan et al.'s estimates indicate insignificant beneficial effects and Baker et al. significant harmful effects, the signs of the estimates were always the same. Baker et al. (2015) included estimates on both the probability of being accused and convicted of a crime. The accused in Baker et al. are those charged, plus those dealt with through the use of extrajudicial measures. The latter seemed closer to the measures used by Smith (2015), and we thus included them. The direction of the results in Baker et al. was similar for both measures.

Test Scores and School Grades

Fitzpatrick (2008) reported results from the same preschool program and 4th grade tests as Cascio and Shanzenbach (2013). We included the results from Cascio and Shanzenbach (2013), as both Georgia and Oklahoma were in the treatment group in their study, while Fitzpatrick (2008) only included Georgia. Baker et al. (2015) reported two estimates from the PISA tests, one where the 2009 cohort was considered treated and one where this cohort was in the control group, because not all students in this cohort were exposed to treatment. As most estimates in Table 3 were based on contrasts between children who live in areas that were more or less exposed to universal preschool programs, we reported the former estimates. Using the latter estimates yielded effect sizes of smaller absolute magnitude.

Primary and Secondary School Progression

Borraz and Cid (2013) study the same expansion of universal preschool in Uruguay as Berlinski et al. (2008) but used data from only one survey wave. We therefore used the latter for all estimates of overlapping outcomes.

Years of Schooling and Highest Grade Completed

Bingley and Westergård-Nielsen (2012) and Bingley et al. (2018) examined the effect of universal preschool programs in Denmark on years of schooling. The two studies exploited a similar type of expansion/reform, but used non-overlapping samples (cohorts), which was why both studies were included in the analysis. Havnes and Mogstad (2011) and Havnes and Mogstad (2015) studied the same reform and used an overlapping sample. We used estimates from the latter regarding years of schooling, as they had access to a longer sample. Havnes and Mogstad (2015) did not include estimates of the probability of attending college. Consequently, those estimates are taken from their 2011 article.

Employment and Earnings

Havnes and Mogstad (2011) and Havnes and Mogstad (2015) studied the same reform and used an overlapping sample. We focused on the latter study in the analysis of earnings, as it used more

years of earnings in the estimations and because Havnes & Mogstad (2011) estimated the effects on the probability of being a top, high, average, and low earner, not on (log) earnings as the other included studies. As mentioned in the previous section, both Bingley and Westergård-Nielsen (2012) and Bingley et al. (2018) used data from Denmark, but their samples did not overlap. Bingley et al. (2018) provided heterogeneity estimates both across maternal education and earnings quartiles. We reported the former, as they were closer to the definition of SES used in most other articles.

Comparison of Preschool Types

Heckman et al. (2017) reported results from a within-Reggio Emilia comparison between children attending and children not attending preschool, which did not use a natural experiment and was therefore not included in the analysis. In the type comparison between the Reggio Emilia approach and alternative approaches, they compared the Reggio Emilia approach separately for Parma and Padova, using different methods; DID, matching and DID with matching. In the analysis we reported the results from the DID method with matching, and the statistical significance was based on p-values adjusted for multiple hypothesis testing.

Search Strings

We searched the following electronic databases for relevant studies: Academic Search Premier, EconLit, ERIC, PsycINFO, SocIndex, and Teacher Reference Center. All searches were performed in EBSCO-host and limited to 1980-2018 in November 2017. The search strings for the six databases follow below.

Academic Search Premier

Search	Search Terms	No. of hits
NS13	S4 AND S8 AND S12	434
S12	S9 OR S10 OR S11	703,897
S11	SU ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	60,875
S10	AB ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	656,011
S9	TI ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	41,423
S8	S5 OR S6 OR S7	1,821,837
S7	SU (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	108,124
S6	AB (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	1,707,851
S5	TI (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	221,871
S4	S1 OR S2 OR S3	77,686
S3	SU (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	40,738
S2	AB (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	45,548
S1	TI (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	28,484

ECONLIT

Search	Search Terms	No. of hits
S13	S4 AND S8 AND S12	238
S12	S9 OR S10 OR S11	26,243
S11	SU ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR random-ized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differ-ences")	334
S10	AB ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR random-ized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differ-ences")	25,651
S9	TI ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR random-ized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differ-ences")	1,117
S8	S5 OR S6 OR S7	364,749
S7	SU (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	317,119
S6	AB (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	111,888
S5	TI (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	17,833
S4	S1 OR S2 OR S3	21,408
S3	SU (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "child-hood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "child-hood initiative*")	20,451
S2	AB (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "child-hood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "child-hood initiative*")	2,163
S1	TI (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "child-hood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "child-hood initiative*")	1,129

ERIC

Search	Search Terms	No. of hits
S13	S4 AND S8 AND S12	381
S12	S9 OR S10 OR S11	29,271
S11	SU ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	729

S10	AB ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	34,012
S9	TI ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	2,018
S8	S5 OR S6 OR S7	164,085
S7	SU (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	17,024
S6	AB (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	199,972
S5	TI (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	21,740
S4	S1 OR S2 OR S3	65,921
S3	SU (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	71,441
S2	AB (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	43,281
S1	TI (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	19,393

PsycINFO

Search	Search Terms	No. of hits
S13	S4 AND S8 AND S12	694
S12	S9 OR S10 OR S11	144,671
S11	SU ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	12,166
S10	AB ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	138,237
S9	TI ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	12,278
S8	S5 OR S6 OR S7	455,654
S7	SU (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR	60,192

Search	Search Terms	No. of hits
	community-wide OR statewide)	
S6	AB (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	494,562
S5	TI (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	42,555
S4	S1 OR S2 OR S3	102,863
S3	SU (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	91,700
S2	AB (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	47,848
S1	TI (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	23,856

SocINDEX

Search	Search Terms	No. of hits
S13	S4 AND S8 AND S12	100
S12	S9 OR S10 OR S11	25,035
S11	SU ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	1,776
S10	AB ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	26,399
S9	TI ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	1,890
S8	S5 OR S6 OR S7	164,819
S7	SU (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	6,353
S6	AB (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	199,604
S5	TI (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	19,668
S4	S1 OR S2 OR S3	20,801
S3	SU (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	15,230
S2	AB (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood	14,822

	program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	
S1	TI (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	5,599

Teacher Reference Center

Search	Search Terms	No. of hits
S13	S4 AND S8 AND S12	14
S12	S9 OR S10 OR S11	5,325
S11	SU ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	364
S10	AB ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	4,920
S9	TI ("treatment-control" OR "treatment-comparison" OR "random* control* trial*" OR randomized field" OR "experiment*" OR "quasi-experiment*" OR "quasi-random* control* trial*" OR "Sibling sample design*" OR "sibling fixed effect*" OR "family fixed effect*" OR "instrumental variable*" OR "random-assignment design" OR "program effect*" OR "intervention* effect*" OR "instrument*" OR "IV" OR "exogenous variation" OR "evaluate" OR "discontinuity" OR "difference-in-difference*" OR "with-in household difference" OR "within household differences")	400
S8	S5 OR S6 OR S7	42,118
S7	SU (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	5,959
S6	AB (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	38,212
S5	TI (universal OR general OR comprehensive OR expan* OR nationwide OR large-scale OR community-wide OR statewide)	5,505
S4	S1 OR S2 OR S3	19,419
S3	SU (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	6,588
S2	AB (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	6,266
S1	TI (preschool* OR "childhood program" OR "child* develop* program*" OR pre-kindergarten OR childcare OR daycare OR "early childhood care" OR "pre-primary education" OR "childhood program*" OR "early education" OR prekindergarten OR "early childhood education" OR Pre-K OR "childhood care" OR "center based day care" OR "family day care" OR "childhood initiative*")	12,046