Rev Econ Household
https://doi.org/10.1007/s11150-026-09847-z

Early Childhood Education and Care and mother's labor supply: The role of publication and weak causal design biases

Pablo Brugarolas [1,2]

Received: 20 January 2025 / Accepted: 12 April 2026

Abstract
This paper conducts a meta-analysis of causal studies examining the impact of Early Childhood Education and Care (ECEC) programs on maternal employment in developed countries. Using 981 effect-size estimates from 35 studies, I harmonize employment outcomes into percentage-point changes and assemble 70 study-level characteristics describing program design, subgroups, contexts, and empirical strategies. I apply a battery of linear and non-linear publication-bias tests and use model-averaging methods to account jointly for publication bias and what I term weak design bias, defined by whether studies satisfy core identification diagnostics for RDD, IV, RCT, and DiD designs. Publication bias is modest but non-negligible: correcting only for selective reporting reduces the descriptive mean effect of around 5 percentage points (p.p.) to roughly 1 p.p., yielding insignificant implied intention-to-treat (ITT) effects and average treatment-on-the-treated (ATT) effects of about 4 p.p. Once I also reweight the literature toward designs that meet identification checks, the implied ITT effect rises to about 8 p.p., and the implied ATT effect stabilizes around 10 p.p. The corrected ATT effects are particularly pronounced for child care programs, delivered by public providers, and implemented in high-employment settings. Overall, the results suggest that modern ECEC expansions generate sizeable employment gains for already attached mothers facing binding care constraints. They also provide bias-corrected benchmarks for evaluating ECEC reforms and their contribution to mitigating child penalties and gender gaps in labour-market outcomes.

Keywords Preschool provision · maternal labor supply · meta-analysis · publication bias · ECEC · pre-k · child care

JEL Codes C83 · J13 · J18 · J22

✉ Pablo Brugarolas
pablo.brugarolas@udg.edu

[1] Department of Economics, University of Girona, Girona, Spain

[2] Social Policy Department, London School of Economics & Political Science, London, UK

Published online: 30 April 2026

1 Introduction

Despite spectacular increases in female employment since the 1970s, mothers' employment rates in most developed countries remain significantly lower than those of fathers (Albanesi et al. 2023; Cascio et al. 2015; Goldin 2006). Traditional gender roles assign child-rearing responsibilities disproportionately to women (Andrew et al. 2022; Sayer et al. 2004), creating barriers to labor market participation not typically faced by men. The lack of affordable child care remains a key obstacle for mothers seeking to work (more). Many OECD countries have responded by expanding Early Childhood Education and Care (ECEC) availability: today, about 25% of children aged 0-2 and 80% of children aged 3-5 are enrolled in early childhood education in these countries (OECD 2023). As policymakers consider further investments in ECEC, it is essential to assess the large body of causal research on whether expansions of access effectively supports mothers' labor supply, though this task is challenging due to (1) ambiguous theoretical predictions regarding maternal labor supply responses and (2) non-trivial variation in (quasi-)experimental methods, contexts, and focus among studies. How can we synthesize this evidence to assess whether ECEC provision effectively promotes maternal employment?

According to economic models of family decision-making, mothers balance employment and ECEC choices to maximize utility, considering consumption, leisure, and care quality within budget constraints (Becker 1993; Blau 2001). By reducing the effective price of formal care or expanding guaranteed access, ECEC policies expand mothers' choice sets, influencing the utility of work and care alternatives. However, as Blau (2003) notes, the substitution effect competes with an income effect: by reducing household child care expenses, subsidies may free up income, which some mothers could use to work fewer hours and prioritize leisure time. Whether substitution or income effects dominate ultimately depends on individual preferences, making the average ECEC impact on maternal labor supply theoretically ambiguous.

Synthesizing causal evidence is further complicated by methodological differences among studies. First, variations in the institutional context (e.g., age of children or baseline child care availability) or mothers' characteristics (e.g., socioeconomic backgrounds, age, or family structure) influence findings (Carta and Rizzica 2018; Nollenberger and Rodríguez-Planas 2015). Second, studies disagree on whether ECEC programs mainly affects the intensive or extensive margins of employment, as seen in comparisons between Carta and Rizzica (2018) and Bauernschuster and Schlotter (2015). Third, some studies note that ECEC can crowd out private alternatives, which may dampen labor supply effects (Nollenberger and Rodríguez-Planas 2015).

An underexplored factor challenging this body of evidence is publication bias, wherein statistically significant results are more likely to appear in print than null findings (Ashenfelter et al. 1999; Blanco-Perez and Brodeur 2020; Brodeur et al. 2023; De Long and Lang 1992; Doucouliagos and Stanley 2013; Elliott et al. 2022; Imai et al. 2021; Ioannidis 2005; Neisser 2021; Stanley et al. 2021; Ugur et al. 2020). Bartoš et al. (2024) find that publication bias is more severe in economics than in other empirical disciplines like medicine or psychology. Based on 90,000 effect size estimates in economics, they estimate that correcting for publication bias reduces the observe median probability of a genuine effect from 99.9% to 29.7%, and halves the effect sizes—see also Ioannidis et al. (2017).

Importantly, studies also vary in their rigor for the underpinning assumptions of their empirical strategies —a problem that I defined as "weak design bias." In RDD studies, credible inference depends on covariate balance at the cutoff (Lee and Lemieux 2010), avoiding high-order polynomials to prevent overfitting (Pei et al. 2022), and checking for score manipulation near the threshold (Cattaneo et al. 2020; McCrary 2008)–see Cattaneo and Titiunik (2022) for a survey. However, several studies in the meta-analysis lack these checks (e.g., Bauernschuster and Schlotter 2015; Fitzpatrick 2012). In DiD designs, weak design bias risks arises from untested pre-trends, differential timing, and composition instability in repeated cross-sectional data (de Chaisemartin and D'Haultfœuille 2023; Sant'Anna and Xu 2023). For example, Cascio (2009) and Akgündüz et al. (2021) do not confirm parallel trends; Schuss and Azaouagh (2021) and Andresen and Havnes (2019) face differential timing bias; and Baker et al. (2008) and Lefebvre and Merrigan (2008) do not verify compositional stability. Additionally, weak instruments, indicated by low first-stage F-statistics, can compromise causal estimates in IV studies (Andrews et al. 2019; Murray 2006), as observed in some studies in this strand of the literature (e.g., Goux and Maurin 2010; Sabol and Chase-Lansdale 2015).

I address these challenges by conducting a meta-analysis of empirical work examining the causal employment impact of ECEC programs in developed countries. The dataset includes 981 effect size estimates from 35 studies published over the past two decades, standardized as percentage-point changes in employment, part-time, and full-time work. The dataset also tracks 70 study characteristics, detailing program design, subgroups, publication traits, and empirical strategies. The study applies a battery of publication bias tests and model-averaging techniques to explore effect heterogeneity beyond publication and weak design biases, building upon recent methodological guidelines for meta-analysis in economics (Havranek et al. 2020; Irsova et al. 2023b).

The results reveal three main patterns. First, publication bias is modest but non-negligible: linear and non-linear tests in Section 3 show that correcting only for selective reporting reduces the descriptive average effect from about 5 p.p. to roughly 1 p.p., and renders the ITT effects insignificant, while ATT estimates shrink to around 4 p.p. Second, once both publication and weak design biases are taken into account, the ITT rises to roughly 8 p.p., while the ATT increases to about 10 p.p., reflecting sizeable gains among mothers who obtain access to formal care. Third, the meta-analysis uncovers pronounced heterogeneity in these beyond-bias returns. Employment gains are strongest among mothers who actually get access to the policy. Likewise, the design and delivery of ECEC programs–particularly child care programs, publicly provided, and implemented in high-employment settings–substantially shape the magnitude of these gains.[1]

[1] By high-employment settings, this paper refers to study contexts in which mothers' baseline employment rates are already relatively high before the policy intervention. Empirically, these are defined as cases falling in the top third of the distribution of pre-intervention employment levels in the sample. In practice, this corresponds to studies with pre-intervention employment above 60% for overall employment, above 40% for full-time employment, and above 50% for part-time employment.

This paper makes three main contributions to the literature. First, it provides the first meta-analysis of the causal impact of ECEC programs on mothers' employment–see qualitative reviews in Albanesi et al. (2023); Morrissey (2017); Olivetti and Petrongolo (2017). Because no prior quantitative synthesis exists, this body of research still awaits systematic correction for publication bias and p-hacking. I address both issues using the most recently developed meta-analytical techniques, including linear and non-linear funnel-based methods as well as selection models. Doing so yields bias-corrected estimates showing that ECEC programs meaningfully increase employment among mothers who actively enroll. This provides a credible benchmark for welfare analyses and for debates on how child care reforms can mitigate gender gaps after childbirth.

Second, beyond addressing selective reporting, this is also the first study to systematically examine how program characteristics, empirical strategies, and institutional contexts shape the effects reported in the literature. The meta-analysis reconciles findings that often appear inconsistent due to differences in population, design, or program features, and clarifies why the ECEC literature has yet to reach consensus on the most effective program design. A central insight concerns heterogeneity by baseline employment: prior qualitative reviews frequently argued that ECEC has limited impact in high-employment settings due to diminishing returns (Albanesi et al. 2023; Bousselin 2022; Cortés and Pan 2023). Given the secular rise in female labor force participation, this perspective left modern reforms expanding childcare provision in an ill-position. However, I show that after accounting for publication and design biases, substantial positive gains emerge precisely in settings with high pre-intervention employment, consistent with most recent studies (Andresen and Havnes 2019; Bousselin 2022; Carta and Rizzica 2018; Pihl 2022; Wikle and Wilson 2023). This result resonates with a growing literature on the role of preschool programs in mitigating child penalties, particularly in high-employment contexts where the effects appear more pronounced (Lim and Duletzki 2023). Such patterns may reflect the influence of conservative gender norms in previous cohorts. For example, Fervers and Kurowska (2022) find that childcare reforms in Poland had stronger impacts in less conservative areas—see also (Moriconi and Rodríguez-Planas 2021). This finding also connects to a literature on peer effects in mothers' labor market outcomes, highlighting the influence of local norms and networks on mothers' employment decisions (e.g., Boelmann et al. 2024; Nicoletti et al. 2018; Olivetti et al. 2020).

Finally, this paper contributes to a broadening body of work showing that publication bias is pervasive in empirical microeconomics research (e.g., DellaVigna and Linos 2022; Ferraro and Shukla 2020; Vivalt 2019). It also follows a strand of the applied economic literature that builts on recent meta-analytic guidelines in economics designed to address this challenge (de Batz and Kočenda 2024; Havranek et al. 2024; Matousek et al. 2022; Opatrny et al. 2024; Xue et al. 2021). My contribution to this field is to show that accounting for publication bias alone is insufficient: the credibility of identifying assumptions varies substantially across RDD, IV, and DiD designs, and weighting the literature by stronger causal designs materially counterbalances the effects of publication bias. Brodeur et al. (2020) provide initial evidence that publication bias differs across designs, with substantial issues for IV and, to a lesser degree, DiD. This paper goes further by formally evaluating the design-specific identification assumptions required for credible inference and incorporating these assessments into the synthesis. In doing so, the paper introduces a framework for integrating design quality into meta-analytic corrections, an area largely unexplored to date.

The remainder of the paper is organized as follows. Section 2 describes the data collection procedure. Section 3 analyzes whether publication bias contaminates the average impact of the policy on maternal employment. Section 4 estimates how the heterogeneous impacts of the policy beyond publication and weak design bias. Section 5 synthesizes the implications of the corrected ITT and ATT estimates, discusses why the largest gains appear in high-employment settings, evaluates the policy relevance of observed enrollment gaps, and identifies priorities for future research on ECEC and maternal labor supply. In the Online Appendix, Section A provides the search string used to carry out the systematic search. Section B includes supplementary tables and figures, while Section C reports the diagnostics and robustness checks for the Bayesian model-averaging analysis. The data and code required to reproduce the results in this paper are available at https://doi.org/10.17605/osf.io/k4f5p.

2 Data

2.1 ECEC programs and its features

In this review, early childhood education and care (ECEC) programs are defined as regulated arrangements that combine elements of supervision, education, or both for children below the age at which compulsory schooling begins—typically around age six. I distinguish three broad types of ECEC programs: child care, preschool (pre-k), and kindergarten. Table 1 summarizes how the distribution of these program types varies across Anglo-Saxon and European settings.

Child care programs are primarily designed to support parental employment by offering supervised, often full-day, care. Child care programs in the sample share several structural features across both Anglo-Saxon and European settings. In both regions they are overwhelmingly universal–close to 90% of Anglo-Saxon programs and 100% of European ones. Provision is predominantly public (78% and 75%, respectively). Daily schedules also look similar: full-day provision is the norm in both contexts, covering about 80% of Anglo-Saxon programs and 67% of European ones. Finally, child care is financed mainly through public subsidies on both regions. The key difference lies in the age of entry. In Anglo-Saxon countries, child care programs mostly serve children aged 3-6 (85%). In European countries, by contrast, child care programs overwhelmingly begin before age 3 (95%). Studies examining programs in this group include Baker et al. (2008), Herbst (2017), and Brewer et al. (2022) for the Anglo-Saxon region and Andresen and Havnes (2019), Bousselin (2022), and Müller and Wrohlich (2020) for European countries.

Preschool programs generally aim to support children's cognitive and socio-emotional development while easing their transition into formal schooling. In Anglo-Saxon systems, all preschool programs in the sample operate for children aged 3-6 and are overwhelmingly half-day (93%), targeted (90%), and publicly provided

Table 1 ECEC program characteristics by region and program type

	Anglo-Saxon countries			European countries		Total
	Child care	Preschool	Kindergarten	Child care	Preschool	
N	59 (6.0%)	267 (27.2%)	198 (20.2%)	260 (26.5%)	197 (20.1%)	981 (100.0%)
Age of entry						
0 to 2	9 (15.3%)	0 (0.0%)	0 (0.0%)	247 (95.0%)	45 (22.8%)	301 (30.7%)
3 to 6	50 (84.7%)	267 (100.0%)	198 (100.0%)	13 (5.0%)	152 (77.2%)	680 (69.3%)
Scope						
Targeted	6 (10.2%)	239 (89.5%)	0 (0.0%)	0 (0.0%)	33 (16.8%)	278 (28.3%)
Universal	53 (89.8%)	28 (10.5%)	198 (100.0%)	260 (100.0%)	164 (83.2%)	703 (71.7%)
Provision						
Private/mixed	13 (22.0%)	0 (0.0%)	0 (0.0%)	65 (25.0%)	22 (11.2%)	100 (10.2%)
Public	46 (78.0%)	267 (100.0%)	198 (100.0%)	195 (75.0%)	175 (88.8%)	881 (89.8%)
Intensity						
Half-day	12 (20.3%)	249 (93.3%)	166 (83.8%)	87 (33.5%)	110 (55.8%)	624 (63.6%)
Full-day	47 (79.7%)	18 (6.7%)	32 (16.2%)	173 (66.5%)	87 (44.2%)	357 (36.4%)
Financing mode						
Publicly subsidized	57 (96.6%)	26 (9.7%)	0 (0.0%)	260 (100.0%)	39 (19.8%)	382 (38.9%)
Publicly provided	2 (3.4%)	241 (90.3%)	198 (100.0%)	0 (0.0%)	158 (80.2%)	599 (61.1%)

Columns split the sample into Anglo-Saxon vs European countries and, within each region, program type (child care, preschool, kindergarten); the last column reports totals. Cells show counts with column percentages in parentheses. "Age of entry" groups programs starting before age 3 vs ages 3–6. "Scope" denotes eligibility (Targeted vs Universal). "Provision" refers to provider ownership (Public vs Private/mixed). "Intensity" is the typical daily schedule (Half-day vs Full-day). "Financing mode" distinguishes demand-side support (Publicly subsidized) from direct public provision (Publicly provided). Kindergarten appears only in Anglo-Saxon systems.

(90%) –see (Chor et al. 2016; Fitzpatrick 2010; Pihl 2022; Wikle and Wilson 2023). By contrast, European preschool programs are almost entirely universal (83%), serve children both aged 0-2 and 3-6, and more frequently run full-day schedules (44%). They are publicly provided in nearly 90% of cases (e.g., Bauernschuster and Schlotter 2015; Carta and Rizzica 2018; Goux and Maurin 2010; Havnes and Mogstad 2011; Nollenberger and Rodríguez-Planas 2015).

Lastly, kindergarten refers to the year of education immediately preceding compulsory schooling, and it appears in the sample only in Anglo-Saxon settings. All kindergarten programs are universal and publicly provided, with 84% offering half-day schedules — see (Cascio 2009; Fitzpatrick 2012; Gelbach 2002; Soldani 2021).

2.2 Search strategy, screening, and inclusion criteria

This section describes the steps taken to gather data for the meta-analysis, adhering to the methodologies recommended by Havranek et al. (2020) and Irsova et al. (2023b).

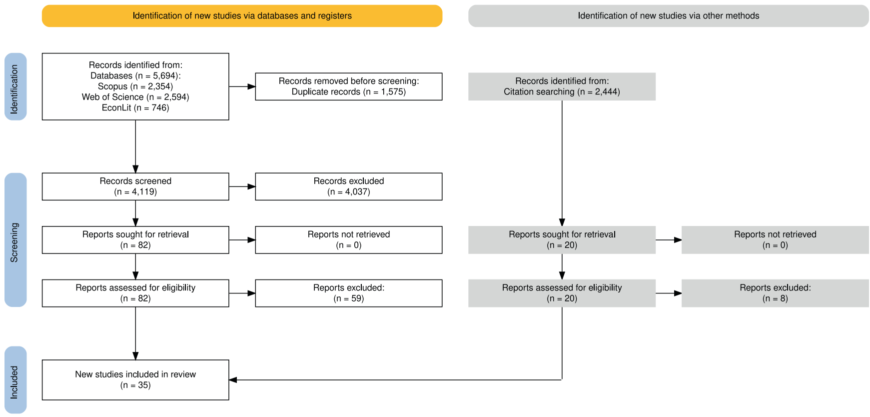

Fig. 1 PRISMA flow diagram. *Note*: The PRISMA flow diagram provides a visual summary of the data collection process. The initial search across Scopus, Web of Science, and EconLit identifies 5694 articles (left part). After duplicate removal and preliminary screening, 4119 articles remain. A title and abstract review narrows this down to 82 articles for full-text assessment, of which 59 are excluded, leaving 23 eligible for the meta-analysis. An additional 12 studies are included through snowballing using PURE-suggest (right part), bringing the final dataset to 35 studies. The diagram was generated with the Shiny app developed by Haddaway et al. (2022)

A visual overview of the data collection process is presented in Fig. 1 through a PRISMA 2020 flow diagram (Page et al. 2021).

To identify studies examining the impact of ECEC programs on maternal employment, I begin by compiling an initial set of references drawn from three previous literature reviews (Albanesi et al. 2023; Morrissey 2017; Olivetti and Petrongolo 2017). Building on this foundation, I develop a search string designed to capture the various ways these studies describe the topic in their titles, keywords, and abstracts— the search string is provided in Section A of the Appendix. Using this search string, I extract peer-reviewed articles from Scopus, Web of Science, and EconLit, covering the period between January 2000 and December 2023, resulting in 2354, 2594, and 746 records, respectively.

From the studies identified through the search string, only those meeting the following inclusion criteria are incorporated into the meta-analysis. First, the analysis focuses on the policy's extensive margin, selecting studies that define treatment as participation or an increased likelihood of enrollment in an ECEC program, while excluding those that explore the intensive margin, such as levels of participation. Second, eligible studies must examine employment outcomes, both at the extensive and intensive margins, including the probability of being employed, working part-time or full-time, hours worked, and weeks worked. Third, the dataset is restricted to studies employing credible causal inference methods based on exogenous variation, such as experiments, instrumental variables (IV), difference-in-differences (DiD), and regression discontinuity designs (RDD). Fourth, the meta-analysis includes studies that evaluate the effects of preschool programs across diverse maternal populations, capturing variations in policy impacts by socioeconomic status, family structure, or age. Fifth, only studies assessing preschool reforms in regions with comparable ECEC systems–specifically Europe, the United States, Canada, Australia, and New Zealand–are included. Finally, estimates must be accompanied by a measure of statistical significance, such as a t-statistic, standard error (SE), or *P*-value, to qualify for inclusion.

The screening process is conducted using Rayyan, a software tool designed for systematic reviews (Ouzzani et al. 2016). In Rayyan, I import all records retrieved from Scopus, Web of Science, and EconLit, resulting in 4,119 unique articles after duplicates are removed. Through title and abstract screening, 82 articles are shortlisted for full-text review based on their potential to meet the inclusion criteria. Of these, 59 are excluded for failing to satisfy the criteria, leaving 23 studies eligible for inclusion in the meta-analysis database. To further enhance the comprehensiveness of the dataset, I apply a snowballing technique using PURE-suggest, a citation network tool (Beck and Krause 2022). This approach systematically screens all papers citing or cited by those identified through the initial search string, yielding an additional 12 primary studies that meet the selection criteria following full-text review.

2.3 Outcome harmonization and standardization

The meta-analytical database captures estimates reflecting various dimensions of mothers' employment. Approximately 68% of these estimates are expressed as percentage point changes in labor market outcomes, such as the probability of being employed, working part-time, or working full-time. The remaining 32% pertain to hours worked per week or weeks worked per year. To ensure consistency across different employment outcomes, I standardize the effect sizes by converting measures of hours and weeks worked into probabilities of full-time or part-time employment.

First, I classify the effects of hours worked and weeks worked as indicative of either full-time or part-time employment. This process calculates the post-intervention mean by adding the coefficient to the pre-intervention mean. Estimates with a post-intervention mean of 35 or higher classify as full-time employment, while those below 35 classify as part-time employment. For weeks worked, I use a threshold of 48 weeks to define full-time employment, assuming that employees typically receive 3-4 weeks off per year. Accordingly, post-intervention means of 48 or more weeks classify as full-time, while those below 48 classify as part-time.

I then calculate the standardized probability effect size, $P(\text{employed})_{f,p}$, by dividing the original coefficient by the corresponding unit, u, which is set to 40 for hours worked and 52 for weeks worked:

$$P(employed)_{f,p} = \frac{coeff}{u} \tag{1}$$

The standard error is scaled as $SE_{f,p} = \frac{se}{u}$. Estimates of hours (or weeks) worked with post-intervention means above the 35 (or 48) threshold classify as a full-time probability, while those below classify as a part-time probability. Observations missing post-intervention mean values in these categories are removed (80) because harmonization is infeasible.

Table 2 The studies used in the meta-analysis

Akgündüz et al. (2021)	Gambaro et al. (2021)	Sabol and Chase-Lansdale (2015)
Andresen and Havnes (2019)	Gelbach (2002)	Sall (2014)
Baker et al. (2008)	Goux and Maurin (2010)	Schiman (2022)
Bauernschuster and Schlotter (2015)	Haeck et al. (2015)	Schochet (2022)
Bousselin (2022)	Havnes and Mogstad (2011)	Schuss and Azaouagh (2021)
Brewer et al. (2022)	Herbst (2017)	Soldani (2021)
Carta and Rizzica (2018)	Lefebvre and Merrigan (2008)	Pihl (2022)
Cascio (2009)	Lefebvre et al. (2009)	Washbrook et al. (2011)
Chor et al. (2016)	Lovász and Szabó-Morvai (2019)	Wikle and Wilson (2023)
Fervers and Kurowska (2022)	Müller and Wrohlich (2020)	Yeh and Wodtke (2023)
Fitzpatrick (2010)	Nollenberger and Rodríguez-Planas (2015)	Zimmert (2023)
Fitzpatrick (2012)	Ravazzini (2018)	

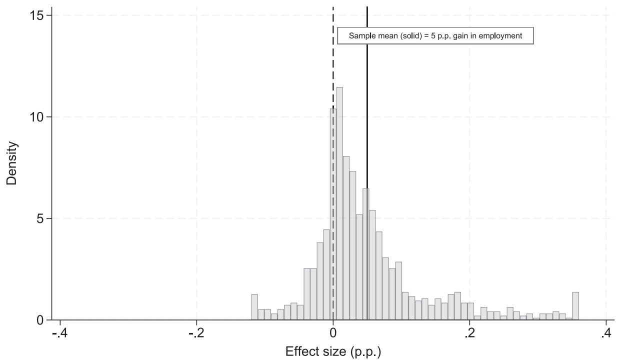

Fig. 2 Distribution of effect sizes. *Note*: The histogram in the figure displays the effect sizes from individual studies (p.p.), with solid and dashed lines indicating the sample mean and zero, respectively

After standardization, the dataset captures gains in three outcomes–probability of being employed, probability of being employed full-time, and probability of being employed part-time–all measured in the same unit (percentage points). An effect size of 0.100 represents a 10 p.p. increase in employment outcomes. To reduce the influence of outliers, I follow the convention of winsorizing estimates at the 1% level. The final sample includes 981 estimates drawn from 35 studies, as detailed in Table 2.

Figure 2 presents a histogram of the standardized estimates in the sample, with the solid line indicating the unconditional sample mean. On average, ECEC enrollment has a small positive impact on mothers' employment, estimated at 5 percentage points. The distribution appears approximately normal, with many estimates clustered near the mean. However, the histogram also highlights relatively large tails,

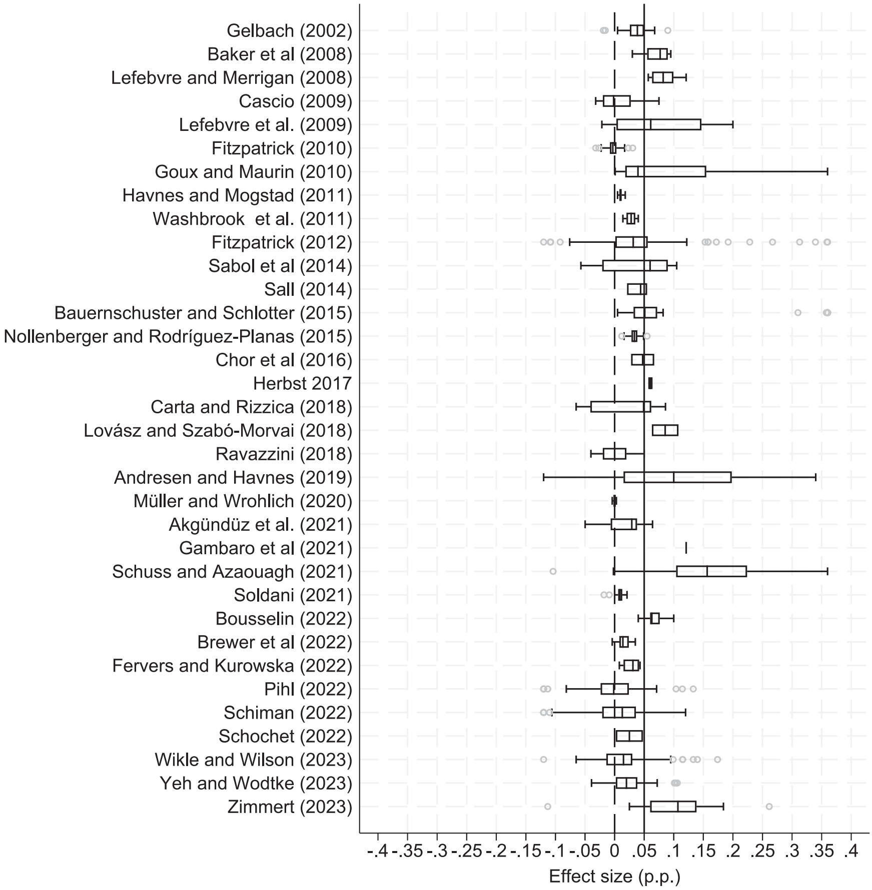

Fig. 3 Effect sizes in primary studies (p.p.) *Note*: This figure presents the effect sizes from primary studies (arranged by publication date). Each box illustrates the interquartile range (P25-P75) divided by a line at the median. Whiskers extend to data points within 1.5 times the interquartile range, with solid and dashed lines indicating the sample mean and zero, respectively

particularly on the right. While the smallest negative effect reported in the primary studies is around -12.0 p.p., the largest positive effect reaches approximately 36.0 p.p.

Figure 3 presents the distribution of alpha estimates from primary studies, organized by publication date. Three notable observations arise. First, there is no clear time trend in the reported estimates; results from more recent studies do not appear systematically different from older ones. Second, the data show significant variation both across and within studies. Many interquartile ranges do not overlap with the unconditional sample mean (solid line), and several are entirely above or below zero (dashed line). This highlights the need to consider the full range of reported estimates. As emphasized by Irsova et al. (2023b), relying on a single representative estimate per study risks overlooking the broader variability. Finally, the considerable heterogeneity across studies underscores the importance of a meta-analysis that synthesizes these diverse findings and adjusts for potential publication bias.

3 Publication bias

This section explores how publication bias may influence the reported effects of ECEC programs on maternal employment. Publication bias occurs when the selection process for studies—shaped by editors, referees, or authors—systematically favors certain types of results, whether consciously or unconsciously. Typically, findings that align with policy expectations or achieve statistical significance are more likely to be published, while studies reporting null or unexpected effects may be overlooked (Chopra et al. 2024). For example, evidence suggesting that ECEC programs have no significant impact on maternal employment might be disregarded if there is a prevailing belief in their positive effects. In the absence of such bias, one would expect no systematic relationship between the precision of an estimate and its magnitude, with both "expected" and "unexpected" findings appearing uniformly across levels of precision.

To evaluate publication bias in this literature, I employ two complementary sets of formal tests. The first set consists of linear meta-regression techniques.The first linear test asks whether imprecise estimates tend to be systematically larger or smaller than precise ones. This is implemented using an ordinary least squares (OLS) regression of effect sizes on their standard errors (Stanley 2005). However, if unobserved characteristics of the primary studies are correlated with the estimated effects, this approach may be misleading. To overcome this limitation, next test runs a model with study fixed effects to account for unobserved heterogeneity at the study level (FE). The third test gives each study equal influence–regardless of how many estimates it reports–by weighting observations by the inverse of the number of estimates per study (wNOBS) (Stanley 2008). The last linear test is the weighted least squares model (WLS), where more precise estimates receive more weight (Stanley and Doucouliagos 2017). In these models, a non-zero slope provides insights into the presence, direction, and magnitude of publication bias. Assuming that publication bias varies linearly with the standard error and that the effect sizes are not heterogeneous—assumptions later relaxed—the constant term in these regressions represents the "effect beyond bias," that is, the estimated average effect of ECEC programs on maternal employment once the selective publication of certain estimates is accounted for.

In addition to linear methods, I apply several nonlinear approaches that relax the linearity assumption to better capture complex patterns of publication bias. The first approach relaxes the linearity assumption by focusing only on estimates that are precise enough to be informative. The weighted average of adequately powered estimates (WAAP) only retains observations with at least 80% statistical power (Ioannidis et al. 2017). This test is based on the intuition that when a study lacks power, statistically significant results are more likely to arise from noise or selective reporting. The second approach (STEM) also concentrates on precise estimates, but it chooses how many to include in a data-driven way. Instead of imposing an 80% power threshold, the stem-based method identifies the subset of highly precise estimates that jointly minimizes bias and variance (Furukawa 2021). The third approach relax the linearity assumption by allowing the relationship between effect sizes and their standard errors to change across levels of precision. The endogenous kink (EK) model first estimates the point at which publication bias is likely to stop influencing results–that is, the threshold below which estimates become sufficiently precise that selection pressures fade (Bom and Rachinger 2019). Once this cutoff is identified, the method fits a piecewise linear regression with a break at that endogenous point. Additionally, the Andrews-Kasy selection model (AK) (Andrews and Kasy 2019) estimates publication probabilities at varying significance levels, reweighting estimates to adjust for selectivity. Finally, the p-uniform* (P-UNI*) model evaluates whether the distribution of *p*-values deviates from uniform expectations, with significant skewness serving as evidence of publication bias (van Aert and van Assen 2023).

These linear and nonlinear tests commonly assume that estimates and their standard errors are unrelated in the absence of bias. However, this assumption can break down if both are shaped by shared observable or unobservable factors, such as characteristics of the studies themselves. For instance, studies reporting intention-to-treat effects are generally more precise and small in magnitude than those focusing on average treatment effects on the treated. I address this issue by implementing the meta-analysis instrumental variable estimator (MAIVE) (Irsova et al. 2023a). This test instruments the standard errors using the inverse square root of the sample size, which is highly correlated with precision while less susceptible to selection bias. Section 4 will further reinforce the robustness of the independence assumption by including detailed study-level variables to evaluate whether systematic publication bias persists once these observable factors are accounted for.

Table 3 provides evidence of existing yet mild publication bias in studies analyzing the effects of ECEC programs on mothers' employment. Panel A shows that all linear tests produce publication bias slopes below one, which Doucouliagos and Stanley (2013) classify as indicative of 'modest' bias. For comparison, Havranek et al. (2024) report substantially higher bias coefficients, ranging from 1 to 3, in natural experiment studies estimating the elasticity of substitution between skilled and unskilled labor, reflecting stronger selective pressures.[2] The nonlinear tests in Panel A also reveal that publication bias is positive yet modest, suggesting a tendency for existing studies to overestimate the effect of ECEC programs on mothers' employment by modestly favoring imprecise positive results over equally imprecise negative findings. Taken together, these tests paint a consistent picture: publication bias exists, but it is mild, and the literature does not appear to selectively inflate the employment benefits of ECEC programs in a substantial way.

[2] Based on their meta-meta-analysis of 87 economics literatures, Doucouliagos and Stanley (2013) propose three bands for interpreting selective reporting: a publication bias slope below one indicates "little to modest" selectivity; slopes between one and two indicate "substantial" selectivity; and slopes greater than two indicate "severe" selectivity. Intuitively, a slope around two corresponds to a literature in which most reported *t*-statistics lie just above conventional significance thresholds (roughly what one would observe if published results were selected on $t \approx 2$). Across the 87 literatures they survey, the mean slope is 1.6–classified as "substantial"–and about 40% fall into the "little to modest" range. To give readers additional yardsticks, I contrast this with recent meta-analyses. Elminejad et al. (2023), studying the Frisch elasticity of labor supply (the elasticity governing how much individuals adjust desired hours of work in response to wage changes while holding their marginal utility of wealth constant), find evidence of "severe" selectivity. By contrast, recent meta-analyses of education policies–such as class size (Opatrny et al. 2024), school-meal programs (Ayllón and Lado 2025), and quasi-experimental school-spending reforms (Jackson and Mackevicius 2024)–generally report publication bias slopes below one and characterize publication bias as "little to modest."

Table 3 Tests that detect and correct for publication bias

Panel A: Linear models

	(1) OLS	(2) FE	(3) MAIVE	(4) WLS	(5) wNOBS
Publication bias	0.807***	0.466	− 0.034	1.008***	0.971***
	(0.240)	(0.422)	(0.784)	(0.163)	(0.216)
		[−5.509, 1.535]	[0.553, 1.337]	[0.323, 1.490]	
			{−5.545, 0.819}		
Effect beyond bias	0.016	0.030	0.052	0.007	0.007
	(0.015)	(0.018)	(0.036)	(0.005)	(0.007)
	[−0.008, 0.051]			[−0.000, 0.022]	[−0.006, 0.033]

Panel B: Nonlinear models

	(6) WAAP	(7) STEM	(8) KINK	(9) AK	(10) P-UNI*
Effect beyond bias	Underpowered	0.002	− 0.001	0.006***	0.037***
	Underpowered	(0.002)	(0.000)	(0.002)	(0.008)

N=981; Studies=34; mean reported by primary studies = 0.05. In Panel A, the tests linearly regress effect sizes on standard errors. The FE specification incorporates study-level fixed effects, while MAIVE instruments the standard error using the inverse square root of the sample size. WLS applies weights based on the inverse of the standard error, and wNOBS assigns weights inversely proportional to the number of estimates per study. Panel B presents nonlinear methods. WAAP computes the weighted average of adequately powered estimates, STEM applies the stem-based technique for optimal power thresholds, KINK uses the endogenous kink model to allow for nonlinearities in the relationship between estimates and standard errors, AK represents the selection model from Andrews and Kasy (2019), and P-UNI* tests the uniformity of p-values in the distribution. Standard errors are clustered at the study level and reported in parentheses. For MAIVE, the first-stage robust F-statistic is 7.046. Confidence intervals are reported in square brackets for wild bootstrap estimates (Roodman et al. 2019) and in curly brackets for two-step weak-instrument-robust intervals (Andrews 2016, 2018; Andrews et al. 2019). * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$

As a final assessment of selective reporting, I implement the randomization tests proposed by Brodeur et al. (2020). The basic idea is to test whether estimates are unnaturally clustered just above conventional significance thresholds, which would be consistent with p-hacking or selective reporting. The intuition is the following. Take a conventional two-sided cutoff for statistical significance, such as $|z| \geq 1.65$ (the 10% level, often marked with one star), $|z| \geq 1.96$ (the 5% level), or $|z| \geq 2.58$ (the 1% level). Now look only at estimates whose test statistics lie in a very narrow symmetric window around that cutoff. For example, with a 0.5 window around the one-star cutoff, we keep all estimates with $|z| \in [1.65 - 0.5, 1.65 + 0.5] = [1.15, 2.15]$. In the absence of p-hacking, the share of "significant" estimates (those above the cutoff) should be close to 50% of all estimates in that band. The randomization test formalizes this idea. For each conventional cutoff $z \in \{1.65, 1.96, 2.58\}$ and each bandwidth $h \in \{0.5, 0.4, \ldots, 0.05\}$, I collect all absolute z-statistics with $|z| \in [z - h, z + h]$. I then compute: (i) the share of these estimates with $|z| \geq z$, and (ii) a one-sided binomial p-value for the null that this share equals 0.5. A small p-value would

Table 4 Randomization tests around conventional significance cutoffs

	Bandwidth h: half-width of symmetric window around cutoff						
	0.5 (1)	0.4 (2)	0.3 (3)	0.2 (4)	0.1 (5)	0.075 (6)	0.05 (7)
Panel A: 10% significance cutoff ($\|z\| \geq 1.65$)							
Share with $\|z\| \geq 1.65$ within window	0.543	0.554	0.542	0.543	0.511	0.529	0.524
One-sided binomial p-value[a]	0.112	0.087	0.191	0.233	0.500	0.432	0.500
Number of estimates in window	219	175	131	92	47	34	21
Panel B: 5% significance cutoff ($\|z\| \geq 1.96$)							
Share with $\|z\| \geq 1.96$ within window	0.470	0.466	0.470	0.511	0.565	0.611	0.760
One-sided binomial p-value[a]	0.830	0.835	0.783	0.459	0.231	0.121	0.007
Number of estimates in window	215	178	132	92	46	36	25
Panel C: 1% significance cutoff ($\|z\| \geq 2.58$)							
Share with $\|z\| \geq 2.58$ within window	0.354	0.345	0.370	0.410	0.406	0.524	0.286
One-sided binomial p-value[a]	1.000	1.000	0.996	0.938	0.892	0.500	0.971
Number of estimates in window	144	110	92	61	32	21	14

Notes: For each cutoff $z \in \{1.65, 1.96, 2.58\}$ and bandwidth h, the table reports: (i) the share of estimates with $|z| \geq z$ in the symmetric window $[z - h, z + h]$; (ii) the associated sample size; and (iii) a one-sided binomial test of equal mass above and below the cutoff

[a]Tests the null that $\Pr(|z| \geq z) = 0.5$ in the selected window

indicate bunching above the threshold and thus raise concerns about p-hacking or selective reporting.

Table 4 reports the results. In Panel A (10% significance threshold, $|z| \geq 1.65$), using a 0.5 window ($h = 0.5$) around the cutoff, 54.3% of the 219 estimates in that band are significant, and the one-sided p-value for testing equality of mass above and below the cutoff is 0.112. This means that the slight excess of "just significant" results is well within what we would expect by chance. Narrower windows lead to similar conclusions. Panels B and C repeat the exercise at the 5% ($|z| \geq 1.96$) and 1% ($|z| \geq 2.58$) thresholds. Across all thresholds and bandwidths, I generally fail to reject the null of equal mass above and below the cutoff. Overall, these results indicate little evidence of strategic bunching or p-hacking around conventional significance cutoffs in this literature. Importantly, this approach does not rely on functional form assumptions and does not require independence across estimates, so it serves as a robustness check for the main publication bias results in Table 3.

The relatively modest publication bias in this literature may stem from the theoretical ambiguity in the economic model of family decision-making, which reduces selective reporting pressures. Unlike fields with dominant theoretical models, where studies aligning with expected results are more likely to be published (Doucouliagos and Stanley 2013), the impacts of ECEC subsidies on mothers' labor supply lack a clear theoretical direction. Depending on the relative strength of substitution and income effects, labor supply responses can plausibly vary, leaving no strong expectation for positive or negative impacts. This absence of a dominant theory appears to have attenuated the bias toward publishing specific results.

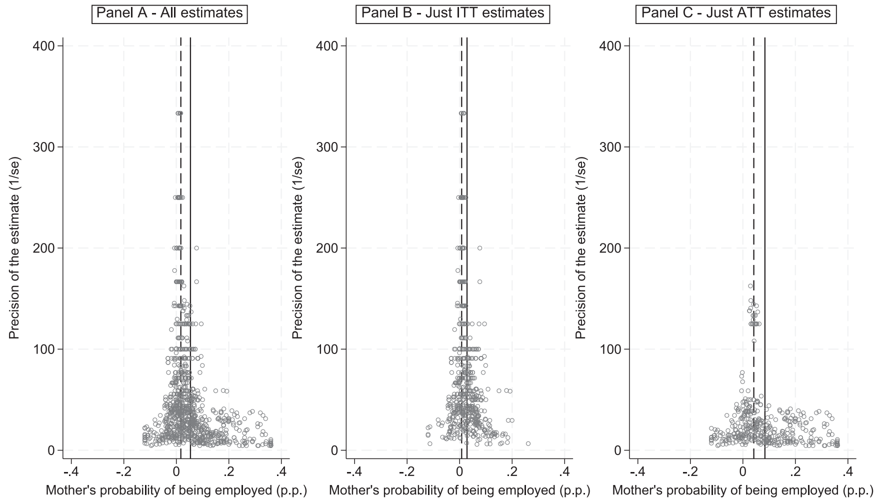

Fig. 4 Funnel plots examining publication bias. Entire sample vs. ITT and ATT estimates. *Note*: The plot illustrates the relationship between effect size estimates and their precision, where an absence of publication bias would produce a symmetric inverted funnel. In Panel **A**, the mean reported effect size (solid line) is visibly skewed above the most precise estimates (dashed line at 0), signaling the presence of publication bias. Panel **B** and **C** show that part of this bias is attibuted to other characteristics beyond publication bias (whether studies estimate an ITT or ATT in this case)

Despite this modest bias, the average causal effect of ECEC programs on mothers' employment is highly sensitive to the publication bias corrections in the tests provided in Table 3. applying these corrections reduces the average effect from the 5.1 p.p. descriptive gain reported in primary studies to a small and largely insignificant 1 p.p., suggesting that most of the observed impact reflects noise concentrated among the least precise estimates. Tests based on statistical power–WAAP and STEM–likewise indicate effects beyond bias that are either underpowered or extremely small. Altogether, this reduction underscores that, even with modest publication bias, initial estimates of effect size may be inflated. The effect beyond bias should be however interpreted with caution until next section, as additional study characteristics may violate the independence assumption relied upon in these tests.

To illustrate this point, I use a funnel plot—a visual tool that maps estimates (y-axis) against their precision (x-axis), defined as the inverse of each estimate's standard error (Egger et al. 1997). In a bias-free context, the plot would resemble a symmetric inverted funnel, where precise estimates cluster around the mean and less precise estimates spread outward. Figure 4 presents funnel plots for the entire sample (Panel A) as well as for ITT (Panel B) and ATT estimates (Panel C). Panel A shows a clear asymmetry: the average reported effect (solid line) is misaligned with the most precise estimates (dashed line), suggesting modest publication bias that inflates the estimated average effect of ECEC programs on mothers' employment. The pronounced tail at higher effect sizes hints at possible left censoring of negative effects. Skewness within each estimand is, however, much less severe. Panel B reveals that ITT estimates are nearly symmetric, while Panel C shows that larger estimates in

Panel A predominantly originate from ATT estimates, as impacts tend to be stronger for recipients of the policy. In the next section, I delve into how heterogeneity in study designs and characteristics may impact the estimated true effects of ECEC programs on maternal employment.

4 Heterogeneity

The relationship between estimates and standard errors observed in the previous section, though suggestive of mild publication bias, may also reflect underlying heterogeneity in the literature on ECEC programs and maternal employment. The linear publication-bias tests relate effect sizes only to their standard errors; in that setting, any systematic differences in study context or design that are correlated with both magnitude and precision can show up as a spurious publication-bias slope. In this sense, the gap between the mean reported effect and the effect-beyond-bias effects in Table 3 may partly reflect omitted moderators in the meta-regression, rather than selective reporting alone. Funnel plots in Fig. 4 and the distributions in Figs. 2 and 3 highlight substantial variability in reported estimates, suggesting that employment outcomes are shaped by specific study contexts. The heterogeneity analysis in this section therefore extends the publication-bias framework by explicitly conditioning on observed study characteristics, thus separating genuine contextual heterogeneity from selectivity.

This analysis addresses two levels of model uncertainty: at the study level and within the meta-analytic framework. At the study level, research on the employment impacts of ECEC programs exhibits substantial variation in design and context. To account for this, I systematically link differences in study outcomes to variations in methodologies and datasets across studies. At the meta-analytic level, uncertainty revolves around identifying the characteristics that best explain this heterogeneity. To tackle this issue, I employ model averaging techniques, particularly Bayesian Model Averaging (BMA), to comprehensively explore effect heterogeneity. Through BMA, I estimate "implied effects," which quantify the impact of ECEC programs on maternal employment while correcting for publication bias and weak design biases. These effects capture variations across different contexts, such as specific subsamples of mothers or varying baseline employment levels. Following this approach helps to systematically examine how specific policy design features influence the effectiveness of ECEC programs in improving employment outcomes.

Table 5 presents the 70 characteristics included in the model, reflecting the diverse contexts from which estimates of maternal employment are drawn. To enhance clarity, these variables are organized into six categories. The first category covers outcome measures, such as distinctions between part-time and full-time employment. The second includes program characteristics, like the age of entry, intensity, or geographic scope of the ECEC program. The third focuses on subgroups analyzed in primary studies, such as socioeconomic status (SES), family structure, or the mother's age. The fourth addresses causal design elements, including the use of methods like instrumental variables (IV), difference-in-differences (DiD), or regression discontinuity designs (RDD), as well as robustness checks such as pre-trend analysis in DiD studies. The fifth category accounts for intervention

Table 5 70 variables reflecting the estimation context

Category	Characteristics
Outcome	Employed, employed part-time, employed full-time
Program	Program type (child care, preschool, kindergarten)
	Age of entry (0 to 2, 3 to 6); Effect timing (0 to 6, 7 to 15)
	Scope (universal, targeted); Evidence of crowd-out effects (no, yes)
	Counterfactual (targeted childcare, informal care, family care)
	Provision (private, public); Financing mode (publicly subsidized, publicly provided)
	Period (pre-2000s, post-2000s); Intensity (half-day, full-day)
	Geographic location (Anglo-Saxon countries; European countries)
Subgroup	SES (low, high); Family type (single mothers, mothers in couples);
	Mother's age (< 30 y.o., > 30 y.o.)
Causal Design	IV; RDD; RCT; DiD; Weak first-stage (yes, no);
	DiD similar pre-trends (yes, no);
	DiD no compositional change (yes, no);
	DiD accounts for differential timing bias (yes, no);
	RD bandwidth size (small, medium, large);
	RD polynomial degree (linear, non-linear)
	RD manipulation of the score (yes, no); RD covariate balance (yes, no);
	Estimand (ATT, ITT)
Intervention	Outcome pre-intervention mean (low, medium, high);
	Bite absolute change (< 10 p.p., 10-20 p.p., > 20 p.p.)
	Bite relative change (< 50%, 50-200%, > 200%);
	Bite pre-intervention mean (low, medium, high)
Publication	Citations per year; Impact factor; Result reported in the Appendix

characteristics, such as the pre-intervention mean of the outcome variable. The pre-intervention mean of the outcome is taken directly from each primary study and refers to the reported pre-reform level of the corresponding labour market outcome for mothers (e.g., the employment rate of mothers prior to the ECEC expansion). For heterogeneity analysis, I classify these values into three roughly equally sized categories (low, medium, high) separately for each outcome. For instance, for overall employment the "high" group captures pre-intervention rates above 60% (see Online Appendix Table B2). Finally, the sixth category considers publication characteristics, including annual citation rates or the impact factor of the publishing journal, which serve as proxies for study quality beyond the directly observed characteristics. For detailed definitions, means, and standard deviations of these variables, see Table B2 in the Online Appendix.

Determining which variables explain the heterogeneity in estimates reported in the literature presents a significant challenge. One approach is to include all 70 explanatory variables in a single regression. However, this poses a problem: we cannot determine in advance which variables truly belong to the underlying model. While all 70 variables may seem relevant for explaining heterogeneity, it is likely that only a subset has meaningful explanatory power. Including all variables in a single regression would reduce the precision of the estimation and make it difficult to draw clear inferences, even for the most critical predictors. An alternative approach is to select variables subjectively or through ad hoc methods, but this risks increasing both model uncertainty and the potential for bias in the meta-analytic results.

To address model uncertainty, I instead apply model-averaging techniques as recommended by Irsova et al. (2023b), using both Bayesian and frequentist frameworks as outlined in Steel (2020). This approach is commonly applied in other meta-analyses, such as Xue et al. (2021), Matousek et al. (2022), Opatrny et al. (2024), Havranek et al. (2024), and de Batz and Kočenda (2024). Bayesian Model Averaging (BMA) evaluates all possible combinations of explanatory variables, assigning a weight to each model based on how well it fits the data. These weights are then used to compute a weighted average across models. Unlike Bayesian methods, frequentist model averaging does not rely on prior beliefs, making it a useful robustness check. However, it is computationally more intensive. Consistent with standard practice, I primarily focus on the Bayesian framework and later confirm that the frequentist results align, reinforcing the robustness of findings under different modeling assumptions.

Bayesian Model Averaging (BMA) addresses model uncertainty by evaluating all possible combinations of explanatory variables in regression models (Raftery 1995). This involves estimating a large number of regressions, each incorporating a different subset of covariates, and then calculating a weighted average across these models. The weights are based on posterior model probabilities, which indicate how well each model fits the data compared to the alternatives (Eicher et al. 2011; Raftery 1995).[3] Using these posterior model probabilities, BMA computes weighted averages for each covariate's posterior mean, standard deviation, and posterior inclusion probability (PIP). These metrics correspond to the estimates, standard errors, and significance levels in traditional regressions.

Figure 5 presents the results of the BMA analysis for maternal employment outcomes. Variables are displayed on the vertical axis, ranked by their Posterior Inclusion Probability (PIP) in descending order. Those at the top have the strongest associations with changes in employment outcomes. Each column represents a regression model, ordered from left to right by decreasing posterior model probability, with the width of each column reflecting its probability. Darker shading highlights variables frequently included in high-probability models, indicating a

[3] Posterior model probabilities are calculated as the product of two factors: the integrated likelihood, which measures the probability of the observed data given the model, and the prior model probability, which reflects researchers' assumptions about the model before analyzing the data. To avoid bias, a uniform prior is used, assigning equal weight to all models (Eicher et al. 2011). Additionally, I apply a dilution g-prior to address potential collinearity among covariates (George 2010).

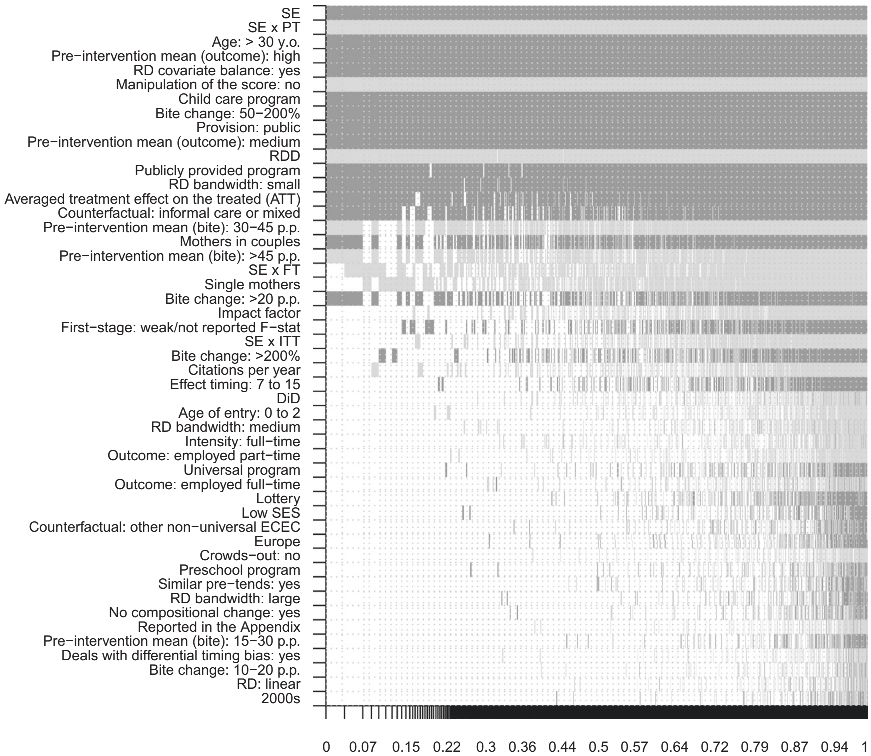

Fig. 5 BMA model: variable inclusion and effect directions. *Note*: This figure presents the baseline results of the Bayesian Model Averaging (BMA) analysis. Variables are arranged along the y-axis in descending order of their Posterior Inclusion Probability (PIP), with those most strongly linked to maternal employment appearing at the top. The x-axis shows the cumulative posterior model probability, displaying models from left to right in decreasing order of likelihood. The shading of the cells conveys the direction of the association: darker cells represent variables positively associated with maternal employment, lighter cells indicate negative associations, and white cells denote variables excluded from the model

stronger influence on employment outcomes. As in the publication-bias analysis, the standard error is the dominant predictor, and its interaction with the part-time margin also receives a PIP close to one, pointing to systematic variation in selective reporting across employment margins. Beyond publication bias, several design features and program characteristics appear prominently. RDD studies with covariate balance and small bandwidths rank among the strongest predictors of larger reported effects, while studies documenting no manipulation of the score tend to yield smaller impacts. Some program characteristics also feature near the top of the distribution: child care programs and publicly provided or publicly operated settings appear frequently in high-probability models, indicating that institutional design and provider structure account for a substantial share of cross-study heterogeneity. Overall, roughly half of the characteristics display non-negligible inclusion probabilities, underscoring that differences in program type, causal design, and intervention context jointly shape the distribution of reported effects, even after accounting for selective reporting.

Table 6 Explaining heterogeneity after accounting for model uncertainty

	BMA			Frequentist check (OLS)			FMA		
Response: reported estimate	P.M	P.SD	PIP	Coef.	SE	p-val.	Coef.	SE	p-val.
Intercept	−0.14		1.00	−0.15	0.01	0.00	−0.17	0.05	0.00
Standard error (SE)	0.83	0.10	1.00	0.82	0.07	0.00	0.87	0.10	0.00
SE * ITT	−0.08	0.18	0.20				−0.38	0.18	0.03
SE * FT	−0.15	0.16	0.53	−0.30	0.10	0.00	−0.17	0.16	0.26
SE * PT	−0.79	0.09	1.00	−0.85	0.08	0.00	−0.67	0.11	0.00
Outcome variation									
Outcome: employed full-time	−0.00	0.00	0.04				−0.01	0.01	0.57
Outcome: employed part-time	−0.00	0.00	0.05				−0.01	0.01	0.21
Program variation									
Age of entry: 0 to 2	−0.00	0.00	0.06				−0.03	0.01	0.01
Effect timing: 7 to 15	0.00	0.01	0.10				0.02	0.01	0.10
Universal program	0.00	0.00	0.04				−0.02	0.02	0.38
Counterfactual: targ. childcare	0.00	0.00	0.02				0.03	0.04	0.42
Counterfactual: informal care	0.01	0.01	0.71	0.02	0.01	0.00	0.02	0.01	0.03
Provision: public	0.04	0.01	1.00	0.05	0.01	0.00	0.08	0.01	0.00
Publicly provided program	0.05	0.02	0.97	0.05	0.01	0.00	0.05	0.01	0.00
Intensity: full-day	−0.00	0.00	0.05				−0.00	0.01	0.67
Crowds-out: no	−0.00	0.00	0.02				−0.02	0.01	0.08
Program type: child care	0.08	0.02	1.00	0.08	0.01	0.00	0.14	0.03	0.00
Program type: Preschool	0.00	0.00	0.02				0.02	0.03	0.58
Region: Europe	0.00	0.00	0.02				0.03	0.02	0.16
Period: 2000s	0.00	0.00	0.01				0.00	0.01	0.63
Subgroup variation									
Low SES	0.00	0.00	0.02				0.01	0.01	0.08
Mothers in couples	0.02	0.01	0.60	0.02	0.01	0.01	0.02	0.01	0.04
Single mothers	−0.01	0.02	0.53	−0.02	0.01	0.00	−0.00	0.01	0.63
Age: > 30 y.o.	0.05	0.01	1.00	0.04	0.01	0.00	0.08	0.01	0.00
Casual design variation									
Method: DiD	−0.00	0.01	0.06				−0.06	0.05	0.22
Method: RDD	−0.05	0.01	0.99	−0.05	0.01	0.00	−0.10	0.05	0.04
Method:Lottery	0.00	0.01	0.04				−0.02	0.05	0.75
DiD: no differential timing	−0.00	0.00	0.01				−0.01	0.01	0.60
DiD: similar pre-tends	0.00	0.00	0.02				0.02	0.02	0.21
DiD: no compositional change	0.00	0.00	0.02				0.03	0.01	0.04
RD bandwidth: small	0.04	0.01	0.95	0.04	0.01	0.00	0.10	0.05	0.05
RD bandwidth: medium	−0.00	0.01	0.05				0.06	0.05	0.26
RD bandwidth: large	0.00	0.00	0.02				0.08	0.05	0.11
RD: linear	−0.00	0.00	0.01				−0.01	0.01	0.46
RD: no score manipulation	−0.12	0.02	1.00	−0.13	0.02	0.00	−0.10	0.03	0.00

Table 6 continued

Response: reported estimate	BMA P.M	BMA P.SD	BMA PIP	Frequentist check (OLS) Coef.	Frequentist check (OLS) SE	Frequentist check (OLS) p-val.	FMA Coef.	FMA SE	FMA p-val.
RD: covariate balanced	0.16	0.02	1.00	0.17	0.02	0.00	0.10	0.05	0.05
First-stage: weak	0.01	0.01	0.21				0.02	0.01	0.13
ATT	0.02	0.01	0.83	0.03	0.01	0.00	0.01	0.01	0.56
Intervention variation									
Outcome at baseline: medium	0.03	0.01	1.00	0.03	0.01	0.00	0.03	0.01	0.00
Pre-mean (outcome): high	0.05	0.01	1.00	0.05	0.01	0.00	0.05	0.01	0.00
Bite change: 10-20 p.p.	−0.00	0.00	0.01				0.02	0.01	0.13
Bite change: > 20 p.p.	0.01	0.01	0.42				0.04	0.01	0.01
Bite change: 50-200%	0.03	0.01	1.00	0.03	0.01	0.00	0.02	0.01	0.16
Bite change: > 200%	0.00	0.01	0.19				−0.02	0.01	0.22
Pre-mean (bite): 15-30 p.p.	0.00	0.00	0.01				0.00	0.01	0.66
Pre-mean (bite): 30-45 p.p.	−0.03	0.02	0.67	0.01	0.01	0.15	−0.05	0.02	0.00
Pre-mean (bite): > 45 p.p.	−0.01	0.01	0.54	−0.02	0.01	0.00	−0.01	0.01	0.47
Publication characteristics									
Citations per year	−0.00	0.00	0.18				−0.00	0.00	1.00
Impact factor	−0.00	0.01	0.22				−0.02	0.01	0.01
Reported in the Appendix	−0.00	0.00	0.01				−0.00	0.01	0.61

Response = estimate of the causal effect of enrolling in an childcare program on mother's employment outcomes. *PM* posterior mean, *PSD* posterior standard deviation, *PIP* posterior inclusion probability. Section C in the Online Appendix describes the variables.

Table 6 highlights the characteristics that consistently influence reported impacts on mothers' employment, comparing results from three approaches: Bayesian Model Averaging (BMA), ordinary least squares (OLS) regression limited to variables with a PIP above 0.5, and Frequentist Model Averaging (FMA). The first robustness check combines Bayesian variable selection with frequentist estimation techniques, while the second relies entirely on frequentist methods. Following Havranek et al. (2024), a characteristic is considered a consistent driver of employment outcomes if it meets three criteria: a PIP greater than 0.5 in BMA and statistical significance at the 10% level in both the frequentist robustness check and FMA.

As shown in the table, the standard error remains strongly and positively associated with reported employment gains across all approaches, reinforcing the presence of selective reporting discussed in Section 3. The publication-bias slope in the BMA model is 0.83, consistent with the "modest" bias documented in Table 3. Moreover, the interaction between the standard error and the ITT indicator shows no evidence that selective reporting differs by estimand type: ITT and ATT estimates appear similarly sensitive to statistical imprecision.

A different pattern emerges, however, when distinguishing between extensive- and intensive-margin outcomes. The interactions between the standard error and the indicators for part-time and full-time employment are significant and negative, implying that the association between imprecision and reported effect size is strongest for extensive-margin outcomes and substantially weaker for intensive-margin outcomes. In other words, funnel asymmetry is most pronounced for employment entry, while publication bias is noticeably attenuated for full-time work and nearly absent for part-time work. This margin-specific pattern aligns with the theoretical discussion in the Introduction. The canonical labor-supply model implies a clear substitution response at the extensive margin, whereas responses at the intensive margin depend on the balance of substitution and income effects and are therefore theoretically ambiguous. If publication incentives are stronger when theoretical predictions are clearer, one would expect selective reporting pressures to be larger for extensive-margin outcomes and weaker for intensive-margin outcomes. The results in Table 6 fit this logic: empirical estimates of employment entry exhibit the clearest funnel asymmetry, while intensive-margin estimates show comparatively little selective reporting.

Several program characteristics emerge as consistent predictors of larger effects, as evidenced by their high PIP values and substantial posterior mean estimates in the BMA. Child care programs have among the largest posterior means, and both publicly provided programs and publicly operated services pass all three robustness checks. Studies using informal or mixed care as the counterfactual also tend to report larger effects. On the subgroup dimension, estimates for mothers in couples frequently appear in high-probability models and exhibit positive coefficients across methods. Older mothers also show consistently larger employment gains. Single mothers, by contrast, display negative and statistically significant coefficients, indicating weaker employment responses once publication bias is accounted for. Lastly, the policy environment matters. Both medium and high baseline employment rates pass the selection criteria and display large posterior means, suggesting that child care expansions have particularly strong effects in settings where mothers are already closely attached to the labour market.

Causal-design characteristics play an important role as well. RDD studies with covariate balance and small bandwidths show significant and positive associations with employment gains, whereas RDD studies reporting no score manipulation tend to yield smaller impacts. Overall, whether weaker causal designs lead to upward or downward adjustments in implied effect sizes depends on the relative prevalence of methods like IV, lottery-based designs, RDD, and DiD within the sample. Later in this section, Fig. 7 examines these dynamics in more detail, illustrating how weak design characteristics influence the estimated effect beyond bias.

The final step of this meta-analysis involves calculating "implied effects," which are derived by constructing a synthetic study that integrates all estimates while assigning greater weight to those based on stronger identification strategies. This method enables me to explicitly define best practices and estimate effect sizes using the parameters from the BMA. Recognizing the subjectivity in defining best practices, I follow two established strategies. First, I use a transparent definition that just corrects for publication bias. Second, I extend this definition to also account for weak causal designs. As illustrated in Table 6, both factors significantly contribute to variation in effect sizes across the model-averaging approaches. Estimates that fail to consider these influences risk conflating employment gains for mothers with selective

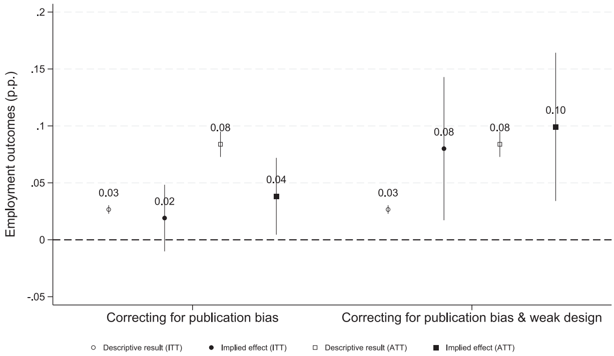

Fig. 6 Average implied effects under the two definitions of best practice. *Note*: The figure displays the implied effect sizes calculated using Bayesian Model Averaging, based on a transparent definition of best practice. It illustrates the expected employment gains if the literature were adjusted to account for publication bias and weak identification strategies. The 90% credible intervals, shown in parentheses, provide a measure of uncertainty around these estimates

reporting or weak identification strategies. All remaining characteristics from the BMA are set to their sample means.

Figure 6 illustrates the average implied effects for ITT and ATT under two scenarios. On the right, the first set of estimates shows the effect of correcting for publication bias. The ITT–which averages over all eligible mothers, including those who do not enroll in ECEC–declines from 3 percentage points (p.p.) to a non-significant 2 p.p. In contrast, the ATT–which captures the effect on mothers who actually take up a place–falls from 8 p.p. to 4 p.p. As expected, the ATT effects are larger, reflecting that treatment-on-the-treated effects measure the impact among those who directly receive the program. These corrections align closely with the effects observed among the most precise estimates in Fig. 4. On the left, the figure presents results after additionally adjusting for weak design bias. Under this "best-practice" definition, the implied ITT increases to 8 p.p., while the implied ATT rises to 10 p.p., consistent with more sizable gains among mothers who obtain access to formal care. Taken together, these results underscore the importance of accounting for the degree to which studies satisfy the identification assumptions of their quasi-experimental designs: had the meta-analysis corrected only for publication bias, implied effects would have appeared smaller, whereas incorporating both corrections reveals substantially larger underlying impacts.

Figure 7 shows that the employment effects of ECEC programs vary meaningfully with the institutional features of the interventions included in the meta-analysis. Panel A compares different program types and reveals a clear gradient: policies oriented toward child care for the youngest children produce the largest gains, while preschool and kindergarten programs generate more modest, though still positive, effects. This pattern is consistent with the different aims and operating hours of these services. child care programs are typically designed to support parental employment and offer full-day coverage, whereas preschool and kindergarten often operate for shorter hours and are organised with educational rather than employment goals in mind. Even so, the adjusted ATT estimates show that access to any structured and

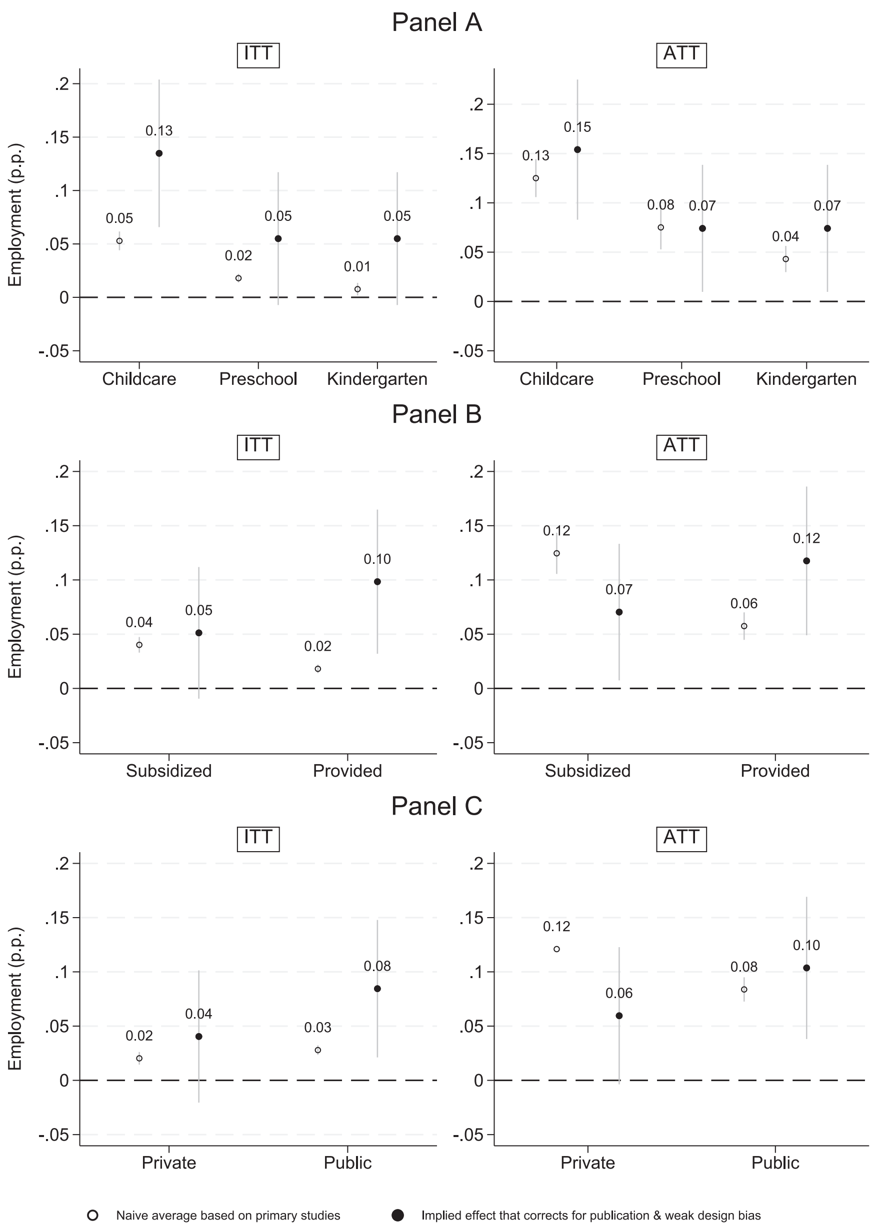

Fig. 7 Implied effects by institutional design of ECEC programs. *Note*: The figure displays the implied effect sizes calculated using Bayesian Model Averaging, based on a transparent definition of best practice. It illustrates the expected employment gains if the literature were adjusted to account for publication bias and weak identification strategies. Panel A compares program types (childcare, preschool, and kindergarten). Panel B compares financing modes (subsidized vs. publicly provided). Panel C compares provider type (private vs. public). The 90% credible intervals, shown in parentheses, provide a measure of uncertainty around these estimates

predictable care arrangement can facilitate mothers' labour-market participation. Panel B highlights differences across financing modes. After correcting for publication and weak-design biases, publicly provided programs yield markedly larger employment effects than programs that rely on subsidies to private providers. This difference is consistent with the idea that provision expands mothers' feasible care options not only by reducing costs but also by guaranteeing availability and stable hours–features that subsidies alone cannot ensure. The estimates therefore highlight the importance of the form through which support is delivered, not merely its monetary value. A similar pattern appears in Panel C, which compares provision by public versus private or mixed providers. Public provision again produces larger implied effects, suggesting that the reliability and standardisation associated with publicly operated services may be particularly salient for mothers considering a return to work. Taken together, the three panels indicate that the design and delivery of ECEC programs–who they serve, how they are financed, and who provides them–play an important role in shaping their labour-supply impact.

Figure 8 turns to differences in employment responses across outcomes, subgroups, and baseline conditions. Panel A shows that, once publication and design biases are accounted for, employment gains are relatively similar across overall, part-time, and full-time work. The implied ITT estimates point to a sizeable and statistically meaningful increase of about 8 p.p., while the implied ATT effects cluster around an even larger 10 p.p. This pattern indicates that all mothers eligible for ECEC experience meaningful employment gains, but that effects are strongest among those who actually take up a place. Panel B examines heterogeneity across groups traditionally considered more or less attached to the labour market. Here, too, a clear pattern emerges: vulnerable mothers exhibit the largest implied gains, but substantial effects also appear for mothers in couples and for single mothers. While the naïve averages suggest moderate differences across groups, the adjusted estimates paint a more unified picture–once study quality and selection issues are addressed, all three groups display meaningful improvements in labour-market participation. Panel C introduces baseline employment levels as a source of heterogeneity. The ITT estimates show little systematic variation, but the implied ATT effects rise steadily with pre-intervention employment rates. High-employment contexts exhibit the largest gains, counter to the common presumption that such settings leave little room for additional labour-supply responses. In Section 5, I examine why this pattern might emerge by drawing on insights from the primary studies and related strands of the literature that point to likely mechanisms.

5 Discussion

This meta-analysis synthesizes 981 effect-size estimates from 35 studies on the employment effects of preschool and child care programs, focusing on two central sources of bias: publication bias, which inflates estimates toward statistically significant findings, and weak design bias, which reflects differences in causal identification across studies. The implied-effect analysis shows that correcting only for publication bias reduces the average ITT effect from about 3 percentage points (p.p.) to a small and statistically insignificant, while the ATT effect halves. When

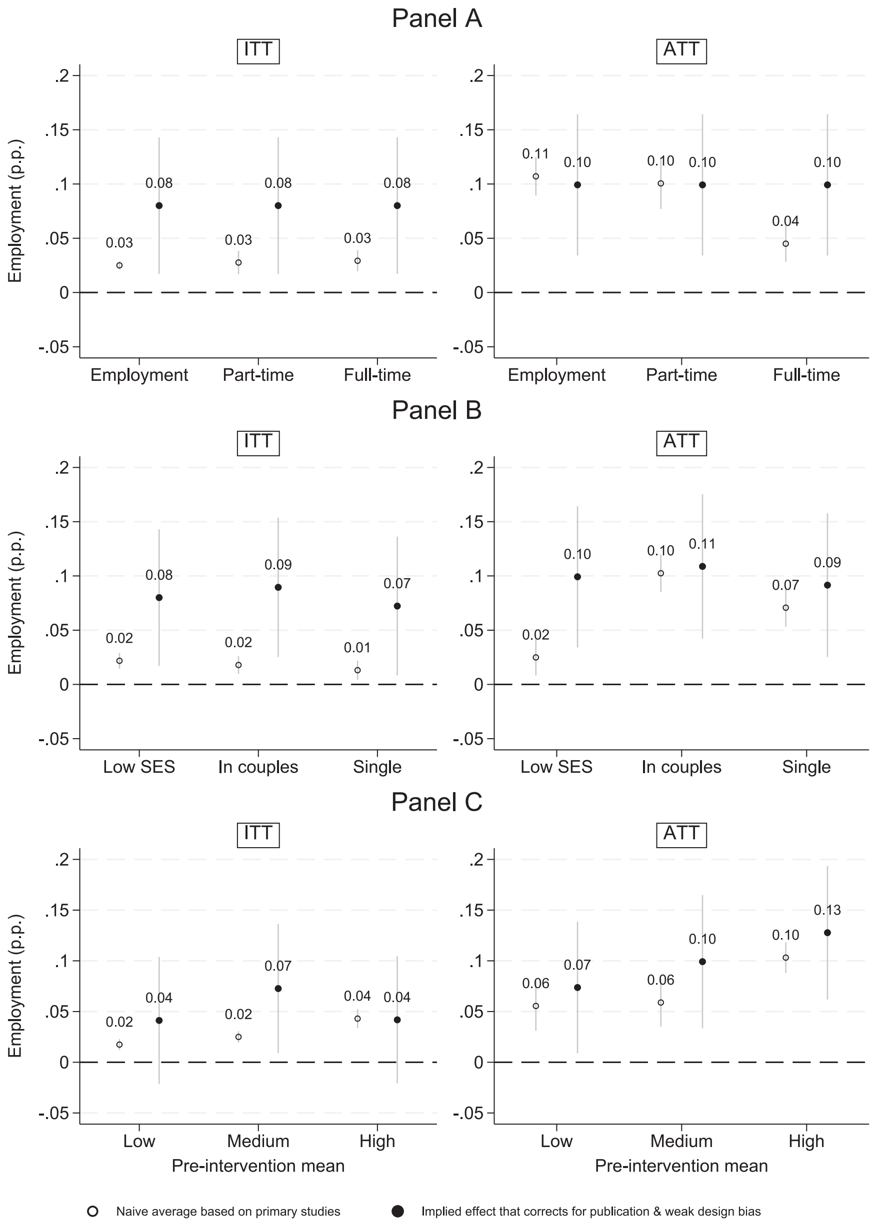

Fig. 8 Implied effects by outcome, subgroup, and baseline employment. *Note*: The figure displays the implied effect sizes calculated using Bayesian Model Averaging, based on a transparent definition of best practice. It illustrates the expected employment gains if the literature were adjusted to account for publication bias and weak identification strategies. Panel A compares employment outcomes (overall employment, part-time employment, and full-time employment). Panel B compares subgroups (low-SES mothers, mothers in couples, and single mothers). Panel C compares contexts by pre-intervention mean employment (low, medium, and high). The 90% credible intervals, shown in parentheses, provide a measure of uncertainty around these estimates

additionally accounting for weak identification strategies, the pattern changes: the implied ITT increases to about 8 p.p., and the implied ATT rises to around 10 p.p., consistent with larger gains among mothers who take up a slot in ECEC programs. Taken together, these results highlight that adjusting only for publication bias would understate the true impact of ECEC programs, whereas incorporating both sources of selectivity–publication patterns and adherence to identification assumptions–reveals substantially larger underlying effects. Corrected effects reported at the end of the result section reveal that meaningful labour-supply responses are concentrated among treated mothers, and that the design and delivery of ECEC programs–particularly those oriented toward younger children, publicly provided, offering reliable hours, and implemented in high-employment settings–substantially shape the magnitude of these gains. The remainder of this section first discusses why implied ATT effects are largest in high pre-intervention employment settings; second, outlines the policy lessons that follow from the presence of sizable enrollment gaps; and third, highlights priorities for future research on ECEC and maternal labor supply.

5.1 Interpreting large ATT effects in high-employment contexts

I begin by examining why the largest implied ATT effects arise in high pre-intervention employment settings and what this pattern reveals about which mothers remain constrained at the margin of child care use. One might expect child care reforms to have limited scope in high-employment settings, since the pool of mothers not working is smaller and often assumed to consist primarily of those with low labour-market attachment. The primary studies in this meta-analysis show that this assumption systematically mischaracterises who the remaining non-users are. Non-users in these settings are not "traditional stay-at-home" families but mothers who are already attached to the labour market or have strong incentives to work, yet remain out of formal care because the child care system leaves them without a feasible option. Across primary studies in high pre-intervention employment contexts, the non-users are overwhelmingly mothers who would use child care if it were available, affordable, or appropriate for their child's age. In Norway, formal care for toddlers was severely rationed, and cohabiting mothers of two-year-olds who wished to work were routinely placed on waiting lists while receiving cash-for-care that compensated only non-use; once slots expanded, take-up was nearly universal and employment surged (Andresen and Havnes 2019). In France and Italy, the non-users are mothers of two-year-olds stuck in the "gap year" before universal preschool begins at age three; they intend to return to work or sustain employment, but the missing toddler-year service obliges them to delay labour-market re-entry or reduce hours until formal kindergarten becomes available (Carta and Rizzica 2018; Goux and Maurin 2010). In Quebec, non-users of formal care are mothers priced out of the regulated system, particularly low-educated mothers for whom the market price of child care exceeds the return to work, forcing them into informal, fragile arrangements or inactivity (Lefebvre and Merrigan 2008). In Luxembourg, non-users are mothers who cannot cover schedule gaps for younger children or for after-school hours, rather than mothers with low attachment (Bousselin 2022). In the US studies, the non-users are low-income single mothers working in unstable jobs who rely on unreliable informal care and live in areas with insufficient Head Start funding (Pihl 2022; Wikle and Wilson 2023). In all these cases, non-use reflects a binding child care constraint rather than a preference for home-based care.

These mothers also face especially strong constraints that amplify the employment impact when access improves. They are often mothers of the youngest children, for whom informal or family-based care is a weak substitute because toddlers require intensive supervision that grandparents or neighbours cannot reliably provide in high-employment societies where female relatives are also working. In Norway, the authors explicitly show that informal care for toddlers is extremely limited, making mothers themselves the default caregivers when slots are unavailable (Andresen and Havnes 2019). In France, early access affects precisely the group of mothers whose employment is interrupted at age two because no public option exists until age three (Goux and Maurin 2010). In Italy, the high price and shortage of slots combined with rigid labour markets mean that mothers who would otherwise continue working are forced into inactivity during the toddler year (Carta and Rizzica 2018). In Quebec, the constraint is primarily price; without the $5/day fee, many less-educated mothers face high effective tax rates and cannot make work pay (Lefebvre and Merrigan 2008). In Luxembourg, the constraint takes the form of time–particularly gaps in hours and after-school coverage–which prevents mothers from sustaining full-time work even when overall participation is high (Bousselin 2022). In the US expansions, the constraint is instability: mothers are pushed out of work when informal arrangements fail, and stable Head Start access directly reduces the risk of job loss (Pihl 2022; Wikle and Wilson 2023). These constraints operate most sharply in high-employment settings because alternative arrangements are scarce, costly, or unreliable precisely when most other women are already working.

Expanding access therefore taps into a substantial pool of "latent" labour supply: mothers who are already close to the margin of employment but held back by specific institutional barriers. Once those barriers are relaxed–through expanded slots, or reduced fees–large adjustments follow. In Norway, cohabiting mothers increase employment by more than 30 percentage points relative to a baseline above 60 percent, consistent with rapid movement from constrained inactivity to full-time work (Andresen and Havnes 2019). In France, granting access at age two prevents the year-long employment interruption that would otherwise accompany a lack of child care, leading to sizeable effects on mothers' labour supply trajectories (Goux and Maurin 2010). In Italy, early kindergarten lowers the fixed cost of re-entry at a critical age and reduces the risk of long spells out of work, producing meaningful increases in participation. In Quebec, lowering child care prices dramatically increases employment among precisely those mothers whose labour supply was suppressed by high fees (Carta and Rizzica 2018). In Luxembourg, the reform brings previously non-employed mothers into work by solving the problem of incompatible schedules rather than by shifting preferences (Bousselin 2022). And in the US, the availability of reliable preschool stabilises employment for single mothers who were already working but vulnerable to child care breakdown (Pihl 2022; Wikle and Wilson 2023). Across these contexts, once the constraint that binds for already-attached mothers is relaxed, labour supply responds strongly, even when aggregate participation seems close to a ceiling.

Although the meta-analysis does not identify mechanisms driving this result directly, this pattern is consistent with three strands of recent evidence. First, an emerging literature shows that ECEC policies are particularly effective at reducing child penalties in environments where female labour force participation is already high, but mothers still face strong constraints on their labour supply after childbirth. Lim and Duletzki (2023) show that the massive expansion of public ECEC programs for children under three in West Germany explains around 40-56 percent of the observed decline in mothers' earnings penalties, with effects persisting up to seven years after birth. Their evidence points to two channels: in the short run, ECEC programs raises maternal labour supply at the intensive margin (more days worked, more full-time and regular employment), while in the longer run it enables mothers to sort into higher-paying firms and occupations by reducing career interruptions. Crucially, these gains are not uniform across space. Within Germany, the effects of ECEC expansion are much stronger in regions with high initial child penalties and limited access to alternative child care. Second, related work shows that ECEC reforms tend to be more effective in areas with less conservative gender norms, where institutional changes more easily translate into mothers' labour-supply responses (Fervers and Kurowska 2022; Moriconi and Rodríguez-Planas 2021). Third, evidence on peer effects in mothers' employment decisions suggests that the behaviour of other mothers in the local environment shapes labour-supply choices (e.g., Boelmann et al. 2024; Nicoletti et al. 2018; Olivetti et al. 2020).

5.2 Implications for policy

Given that the largest effects materialize for mothers who enroll their children, understanding and addressing the barriers that prevent eligible families from doing so becomes a central policy priority. A body of evidence shows that enrolment gaps in ECEC programs are substantial and structured along familiar socioeconomic lines. Even in systems that provide nominally universal places, copayments and residual fees typically depress use of early education and care, and sizeable SES gradients in attendance persist (Magnuson and Waldfogel 2016; Stahl and Schober 2018). In Germany, for example, children of parents without college entrance qualifications are about 14 percentage points (roughly 37%) less likely to be enrolled in early child care than children of parents with such qualifications, and reducing fees narrows–but does not eliminate–this gap (Jessen et al. 2020). Beyond prices, behavioral and informational frictions loom large. Lower-SES families are more likely to lack clear information about the benefits of education, the details of application and verification procedures, or their own eligibility (Bettinger et al. 2012; Currie and Gahvari 2008; Hoxby and Turner 2015; Jensen 2010). In New Orleans, publicly funded ECEC requires families to "pick-apply-verify-enroll," and many drop out at each step: most eligible 0-3 year-olds never apply at all, and only about two thirds of applicants verify by the main deadline, pointing to real learning and compliance costs (Weixler et al. 2020, 2024). Similar patterns appear in Germany, where fragmented, non-transparent local procedures and rationed slots create navigation hurdles that fall more heavily on lower-SES parents, even when fees are low (Hermes et al. 2025). On top of these hurdles, recent audit evidence suggests that discrimination in the "contact" stage can reduce take-up before any formal application is filed: in an audit study, Hermes et al. (2023) show that email inquiries from migrant-sounding parents are less likely to receive an offer than otherwise identical inquiries from natives.

Closing enrollment gaps thus requires a mix of financial, informational, and institutional reforms. Evidence on co-payment reductions suggests that modest fee cuts typically reshuffle children across care types, whereas large reductions that make a full school-day of care free generate both higher take-up and measurable employment gains. In England, extending entitlements from part-time to full-time free care increased formal-care use by roughly 11 percentage points and raised employment among mothers who gained access to fully subsidized full-time places by up to 3.5 points, while earlier expansions of part-time free hours mainly crowded out informal care without increasing total hours (Brewer et al. 2022). In Germany, abolishing fees in the final preschool year boosted full-day attendance by about 3.2 points and increased mothers' full-time work by around 2.3 points (Huebener et al. 2020). At a broader scale, staggered synthetic difference-in-differences estimates for 80 countries suggest that child care laws raise women's labor-force participation and that provisions improving affordability deliver the largest and most persistent gains (Anukriti et al. 2025). Price reforms, however, are not enough if families cannot navigate the system. Information and support interventions that target specific points in the application sequence have been shown to increase completion at each stage. Timely prompts can raise initial application rates, but without continued support many families stall at verification (Weixler et al. 2024). At verification, simple weekly reminders focused on needed documents–plus staff answering questions by text–clear up confusion about rules and logistics and help families finish the paperwork (Weixler et al. 2020). Complementary efforts in Germany show that structured information and hands-on help with center choice, timing, and visits enable lower-SES families to navigate the system more like their higher-SES peers, raising their acceptance chances (Hermes et al. 2025). Finally, institutional design can curb discrimination and reduce the returns to insider knowledge. Centralized assignment mechanisms with transparent, welfare-relevant priorities–such as sibling status, parental employment, and travel-time tie-breakers–can increase disadvantaged families' participation by neutralizing late-application and provider-selection disadvantages that persist in decentralized systems (De Groote and Rho 2024).

5.3 Future directions for research

This review highlights several important avenues for future work. First, there is substantial scope for experimental research on ECEC and maternal labor supply. In this meta-analysis, only five studies rely on randomized identification, and all of them evaluate U.S. programs (Sabol and Chase-Lansdale 2015; Schiman 2022; Schochet 2022; Wikle and Wilson 2023; Yeh and Wodtke 2023). In this regard, a particularly promising avenue is to exploit lottery-based assignment in oversubscribed preschool programs (Weiland et al. 2024). Lottery designs offer clean identification while allowing researchers to study program quality and curriculum, heterogeneous effects by setting and subgroup, and longer-run outcomes, often with larger samples and richer administrative linkages than traditional trials. Embedding maternal labor-market outcomes into such lottery-based evaluations would generate a new generation of high-quality evidence on how different ECEC designs shape parents' employment trajectories.

A second priority is to better understand how ECEC programs affect employment and careers for high-income mothers. Only 50 estimates in this meta-analysis speak to this group. Some expansions find clear positive effects on the extensive margin among higher-income mothers (Akgündüz et al. 2021; Baker et al. 2008; Havnes and Mogstad 2011; Lefebvre and Merrigan 2008; Nollenberger and Rodríguez-Planas 2015; Schuss and Azaouagh 2021; Zimmert 2023), whereas other estimates are statistically indistinguishable from zero (Carta and Rizzica 2018; Goux and Maurin 2010; Lefebvre et al. 2009; Müller and Wrohlich 2020). This margin is central for understanding gender inequality: recent work shows that child penalties in earnings have become the dominant source of gender gaps in advanced economies, arising precisely when highly educated, high-earning women adjust hours, participation, occupation, and promotion trajectories around childbirth (Kleven et al. 2019a, b). Related career-cost and allocation-of-talent literatures argue that motherhood-induced reductions in labor supply among high-ability women generate sizable losses in human capital accumulation and aggregate productivity (Adda et al. 2017; Hsieh et al. 2019). Future research could therefore examine how ECEC reforms affect hours, job quality, occupation, wage growth, and promotion at the top of the income distribution, ideally by linking policy variation to rich administrative earnings and firm data.

A third critical gap concerns the impacts of subsidized early childhood education on maternal labor supply over the life cycle. Only six studies in this review follow mothers long enough to observe outcomes when their children are between 7 and 15 years old (Nollenberger and Rodríguez-Planas 2015; Pihl 2022; Sabol and Chase-Lansdale 2015; Schiman 2022; Soldani 2021; Wikle and Wilson 2023). Recent reviews of parental leave policies indicate that while moderate, job-protected leave helps preserve women's formal attachment to the labor market, it has done little to eliminate the large motherhood earnings penalty, which continues to arise through reductions in hours, moves into more flexible but lower-paying occupations, and slower promotion in environments shaped by policy design, job protection rules, and workplace and gender norms (Albanesi et al. 2023; Cortés and Pan 2023). From this perspective, early child care subsidies may yield large dynamic returns if they shift mothers onto higher-experience, higher-wage career paths. Longitudinal designs that track mothers over a decade or more after an ECEC expansion–and explicitly connect short-run employment responses to cumulative earnings and occupation trajectories–are essential to assess whether early child care policies durably reshape maternal careers.

Lastly, at a meta-analytic level, future work could make important progress on two related questions that this review is not able to address systematically: the intensive margin of child care use and the interaction between preschool provision and the availability of toddler care. This study focuses on the extensive margin–ECEC enrollment–because it provides the cleanest counterfactual for assessing how introducing or expanding access to ECEC affects mothers' employment across countries with very different institutional architectures. A future meta-analysis could extend this work by reviewing causal studies that examine how reforms affect the intensity of ECEC use, once a common taxonomy of dosage becomes feasible. A similar limitation arises for interactions between preschool provision and the availability of care for younger children. Conceptually, the employment effects of preschool for 3-5-year-olds should depend on whether younger siblings already have access to subsidized or publicly provided care, yet primary studies seldom report this information in a harmonized way. As a result, the current evidence base does not support a sufficiently granular classification of "full ECEC systems" versus "preschool-only systems." Future research could therefore contribute by developing richer, cross-nationally comparable measures of ECEC intensity and by documenting infant/toddler care availability at the time of each reform, allowing analysts to study how dosage and system-level complementarities jointly shape maternal labor supply.

Data Availability Data and code is provided as a supplement for full replication.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s11150-026-09847-z.

Acknowledgements This paper has benefited from helpful feedback and suggestions from Sara Ayllón, Mary-Alice Doyle, Tomas Havranek, Chiara Orsini, Ismael Sanz, Tom Stanley, and Bob Reed. I thank participants at the LSE Social Policy Quantitative Reading Group (SPQRG) and the the 2024 Meta-Analysis in Economics Research Network Colloquium for their useful comments. This work was supported by the project 2021-SGR-570 (University of Girona). I thank the IFUdG2022 PhD and Amics del País grants for financial support. I am particularly grateful to two anonymous referees and the editor-in-chief and co-editor of the journal for their constructive comments and suggestions. Any errors or misinterpretations are my own.

Author Contributions The author was responsible for all aspects of the manuscript, including research, analysis, writing, and preparation of figures and tables. The author reviewed and approved the final manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Conflict of interest The authors declare no competing interests.

References

Adda, J., Dustmann, C., & Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, *125*, 293–337.

Akgündüz, Y. E., Van Huizen, T., & Plantenga, J. (2021). Who'll take the chair? Maternal employment effects of a Polish (pre) school reform. *Empirical Economics*, *61*, 1097–1133.

Albanesi, S., Olivetti, C., and Petrongolo, B. (2023) Families, labor markets, and policy, in *Handbook of the Economics of the Family* Volume 1: Elsevier, 255–326.

Andresen, M. E., & Havnes, T. (2019). Child care, parental labor supply and tax revenue. *Labour Economics*, *61*, 101762.

Andrew, A., Cattan, S., Costa Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., & Sevilla, A. (2022). The gendered division of paid and domestic work under lockdown. *Fiscal Studies*, *43*, 325–340.

Andrews, I. (2016). Conditional linear combination tests for weakly identified models. *Econometrica*, *84*, 2155–2182.

Andrews, I. (2018). Valid two-step identification-robust confidence sets for GMM. *Review of Economics and Statistics*, *100*, 337–348.

Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, *109*, 2766–2794.

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, *11*, 727–753.

Anukriti, S., Dinarte-Diaz, L., Montoya-Aguirre, M., and Sakhonchik, A. (2025) Filling the Gaps: Childcare Laws for Women's Economic Empowerment, in *AEA Papers and Proceedings* Volume 115, 689–694, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Ashenfelter, O., Harmon, C., & Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics*, *6*, 453–470.

Ayllón, S., and Lado, S. (2025) The Causal Impact of School-Meal Programmes on Children in Developed Economies: A Meta-Analysis, Working Paper DP No. 18042, IZA Institute of Labor Economics.

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

Bartoš, F., Maier, M., & Wagenmakers, E.-J. (2024). Footprint of publication selection bias on meta-analyses in medicine, environmental sciences, psychology, and economics. *Research Synthesis Methods*, *15*, 500–511.

Bauernschuster, S., & Schlotter, M. (2015). Public child care and mothers' labor supply–Evidence from two quasi-experiments. *Journal of Public Economics*, *123*, 1–16.

Beck, F., and Krause, C. (2022) Visually Explaining Publication Ranks in Citation-based Literature Search with PURE Suggest, in *EuroVis 2022 - Posters* ed. by Krone, M., Lenti, S., and Schmidt, J.: The Eurographics Association.

Becker, G. S. (1993): *A treatise on the family: Enlarged edition*, Harvard: Harvard University Press, 1st edition.

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics*, *127*, 1205–1242.

Blanco-Perez, C., & Brodeur, A. (2020). Publication bias and editorial statement on negative findings. *The Economic Journal*, *130*, 1226–1247.

Blau, D. M. (2001): *The child care problem: An economic analysis*, New York: Russell Sage Foundation.

Blau, D. M. (2003) Child care subsidy programs, in *Means-tested transfer programs in the United States*: University of Chicago Press.

Boelmann, B., Raute, A., and Schonberg, U. (2024) Wind of change? Cultural determinants of maternal labor supply, *American Economic Journal: Applied Economics*, Forthcoming.

Bom, P. R., & Rachinger, H. (2019). A kinked meta-regression model for publication bias correction. *Research Synthesis Methods*, *10*, 497–514.

Bousselin, A. (2022). Access to universal childcare and its effect on maternal employment. *Review of Economics of the Household*, *20*, 497–532.

Brewer, M., Cattan, S., Crawford, C., & Rabe, B. (2022). Does more free childcare help parents work more?. *Labour Economics*, *74*, 102100.

Brodeur, A., Carrell, S., Figlio, D., & Lusher, L. (2023). Unpacking p-hacking and publication bias. *American Economic Review*, *113*, 2974–3002.

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, *110*, 3634–3660.

Carta, F., & Rizzica, L. (2018). Early kindergarten, maternal labor supply and children's outcomes: Evidence from Italy. *Journal of Public Economics*, *158*, 79–102.

Cascio, E. U. (2009). Maternal labor supply and the introduction of kindergartens into American public schools. *Journal of Human Resources*, *44*, 140–170.

Cascio, E. U., Haider, S. J., & Nielsen, H. S. (2015) The effectiveness of policies that promote labor force participation of women with children: A collection of national studies.

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, *115*, 1449–1455.

Cattaneo, M. D., & Titiunik, R. (2022). Regression discontinuity designs. *Annual Review of Economics*, *14*, 821–851.

Chopra, F., Haaland, I., Roth, C., & Stegmann, A. (2024). The null result penalty. *The Economic Journal*, *134*, 193–219.

Chor, E., Andresen, M. E., & Kalil, A. (2016). The impact of universal prekindergarten on family behavior and child outcomes. *Economics of Education Review*, *55*, 168–181.

Cortés, P., & Pan, J. (2023). Children and the remaining gender gaps in the labor market. *Journal of Economic Literature*, *61*, 1359–1409.

Currie, J., & Gahvari, F. (2008). Transfers in cash and in-kind: Theory meets the data. *Journal of Economic Literature*, *46*, 333–383.

de Batz, L., & Kočenda, E. (2024). Financial crime and punishment: A meta-analysis. *Journal of Economic Surveys*, *38*, 1338–1398.

De Groote, O., and Rho, M. (2024) Matching Disadvantaged Children to Day Care: Evidence from a Centralized Platform, Working Paper DP No. 17547, IZA Institute of Labor Economics.

De Long, J. B., & Lang, K. (1992). Are all economic hypotheses false?. *Journal of Political Economy*, *100*, 1257–1272.

DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, *90*, 81–116.

de Chaisemartin, C., & D'Haultfœuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, *26*, C1–C30.

Doucouliagos, C., & Stanley, T. D. (2013). Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys*, *27*, 316–339.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634.

Eicher, T. S., Papageorgiou, C., & Raftery, A. E. (2011). Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, *26*, 30–55.

Elliott, G., Kudrin, N., & Wüthrich, K. (2022). Detecting p-hacking. *Econometrica*, *90*, 887–906.

Elminejad, A., Havranek, T., Horvath, R., & Irsova, Z. (2023). Intertemporal substitution in labor supply: A meta-analysis. *Review of Economic Dynamics*, *51*, 1095–1113.

Ferraro, P. J., & Shukla, P. (2020). Feature–is a replicability crisis on the horizon for environmental and resource economics?. *Review of Environmental Economics and Policy*, *14*, 339–351.

Fervers, L., & Kurowska, A. (2022). Local cultural context as a moderator of the impact of childcare on maternal employment: Evidence from a natural experiment. *Journal of European Social Policy*, *32*, 239–253.

Fitzpatrick, M. D. (2010). Preschoolers enrolled and mothers at work? The effects of universal prekindergarten. *Journal of Labor Economics*, *28*, 51–85.

Fitzpatrick, M. D. (2012). Revising our thinking about the relationship between maternal labor supply and preschool. *Journal of Human Resources*, *47*, 583–612.

Furukawa, C. (2021) Publication bias under aggregation frictions: From communication model to new correction method, working paper, MIT working paper.

Gambaro, L., Neidhöfer, G., & Spieß, C. K. (2021). The effect of early childhood education and care services on the integration of refugee families. *Labour Economics*, *72*, 102053.

Gelbach, J. B. (2002). Public schooling for young children and maternal labor supply. *American Economic Review*, *92*, 307–322.

George, E. I. (2010) Dilution priors: Compensating for model space redundancy, in *IMS Collections Borrowing Strength: Theory Powering Applications - A Festschrift for Lawrence D. Brown* Volume 6: Institute of Mathematical Statistics, 158–165.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American economic review*, *96*, 1–21.

Goux, D., & Maurin, E. (2010). Public school availability for two-year olds and mothers' labour supply. *Labour Economics*, *17*, 951–962.

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*, *18*, e1230.

Haeck, C., Lefebvre, P., & Merrigan, P. (2015). Canadian evidence on ten years of universal preschool policies: The good and the bad. *Labour Economics*, *36*, 137–157.

Havnes, T., & Mogstad, M. (2011). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics*, *95*, 1455–1465.

Havranek, T., Irsova, Z., Laslopova, L., & Zeynalova, O. (2024) Publication and attenuation biases in measuring skill substitution, *Review of Economics and Statistics*, 1–37.

Havranek, T., Stanley, T. D., & Doucouliagos, H. (2020). Reporting guidelines for meta-analysis in economics. *Journal of Economic Surveys*, *34*, 469–475.

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*, 519–564.

Hermes, H., Lergetporer, P., Mierisch, F., Peter, F., & Wiederhold, S. (2023) Discrimination in Universal Social Programs? A Nationwide Field Experiment on Access to Child Care, *CESifo Working Paper*.

Hermes, H., Lergetporer, P., Peter, F., & Wiederhold, S. (2025). Application barriers and the socio-economic gap in child care enrollment. *Journal of the European Economic Association*, *23*, 1133–1172.

Hoxby, C. M., & Turner, S. (2015). What high-achieving low-income students know about college. *American Economic Review*, *105*, 514–517.

Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, *87*, 1439–1474.

Huebener, M., Pape, A., & Spiess, C. K. (2020). Parental labour supply responses to the abolition of day care fees. *Journal of Economic Behavior & Organization*, *180*, 510–543.

Imai, T., Rutter, T. A., & Camerer, C. F. (2021). Meta-analysis of present-bias estimation using convex time budgets. *The Economic Journal*, *131*, 1788–1814.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*, e124.

Ioannidis, J. P., Stanley, T., & Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, *605*, F236–F265.

Irsova, Z., Bom, P. R., Havranek, T., & Rachinger, H. (2023a) Spurious precision in meta-analysis, Working Paper No. DP17927, Centre for Economic Policy Research.

Irsova, Z., Doucouliagos, H., Havranek, T., & Stanley, T. (2023b). Meta-analysis of social science research: A practitioner's guide. *Journal of Economic Surveys*, *00*, 1–20.

Jackson, C. K., & Mackevicius, C. L. (2024). What impacts can we expect from school spending policy? Evidence from evaluations in the United States. *American Economic Journal: Applied Economics*, *16*, 412–446.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, *125*, 515–548.

Jessen, J., Schmitz, S., & Waights, S. (2020). Understanding day care enrolment gaps. *Journal of Public Economics*, *190*, 104252.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019a) Child penalties across countries: Evidence and explanations, In *AEA Papers and Proceedings* Volume 109, 122–126, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Kleven, H., Landais, C., & Søgaard, J. E. (2019b). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, *11*, 181–209.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*, 281–355.

Lefebvre, P., & Merrigan, P. (2008). Child-care policy and the labor supply of mothers with young children: A natural experiment from Canada. *Journal of Labor Economics*, *26*, 519–548.

Lefebvre, P., Merrigan, P., & Verstraete, M. (2009). Dynamic labour supply effects of childcare subsidies: Evidence from a Canadian natural experiment on low-fee universal child care. *Labour Economics*, *16*, 490–502.

Lim, N., and Duletzki, L.-M. (2023) The Effects of Public Childcare Expansion on Child Penalties: Evidence from West Germany, working paper, Boston University.

Lovász, A., & Szabó-Morvai, Á (2019). Childcare availability and maternal labor supply in a setting of high potential impact. *Empirical Economics*, *56*, 2127–2165.

Magnuson, K., & Waldfogel, J. (2016). Trends in income-related gaps in enrollment in early childhood education: 1968 to 2013. *AERA Open*, *2*, 2332858416648933.

Matousek, J., Havranek, T., & Irsova, Z. (2022). Individual discount rates: A meta-analysis of experimental evidence. *Experimental Economics*, *25*, 318–358.
McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, *142*, 698–714.
Moriconi, S., & Rodríguez-Planas, N. (2021) Gender norms and the motherhood employment gap, Working Paper No. 14898, IZA Institute of Labor Economics.
Morrissey, T. W. (2017). Child care and parent labor force participation: A review of the research literature. *Review of Economics of the Household*, *15*, 1–24.
Müller, K.-U., & Wrohlich, K. (2020). Does subsidized care for toddlers increase maternal labor supply? Evidence from a large-scale expansion of early childcare. *Labour Economics*, *62*, 101776.
Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, *20*, 111–132.
Neisser, C. (2021). The elasticity of taxable income: A meta-regression analysis. *The Economic Journal*, *131*, 3365–3391.
Nicoletti, C., Salvanes, K. G., & Tominey, E. (2018). The family peer effect on mothers' labor supply. *American Economic Journal: Applied Economics*, *10*, 206–234.
Nollenberger, N., & Rodríguez-Planas, N. (2015). Full-time universal childcare in a context of low maternal employment: Quasi-experimental evidence from Spain. *Labour Economics*, *36*, 124–136.
OECD (2023) *Education at a Glance 2023: OECD Indicators*, Paris: OECD Publishing.
Olivetti, C., Patacchini, E., & Zenou, Y. (2020). Mothers, peers, and gender-role identity. *Journal of the European Economic Association*, *18*, 266–301.
Olivetti, C., & Petrongolo, B. (2017). The economic consequences of family policies: Lessons from a century of legislation in high-income countries. *Journal of Economic Perspectives*, *31*, 205–230.
Opatrny, M., Havranek, T., Irsova, Z., & Scasny, M. (2024) Publication bias and model uncertainty in measuring the effect of class size on achievement, *Journal of Labor Economics*, Forthcoming.
Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan–a web and mobile app for systematic reviews. *Systematic reviews*, *5*, 1–10.
Page, M. J., McKenzie, J. E., Bossuyt, P. M. et al. (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, *BMJ*, 372.
Pei, Z., Lee, D. S., Card, D., & Weber, A. (2022). Local polynomial order in regression discontinuity designs. *Journal of Business & Economic Statistics*, *40*, 1259–1267.
Pihl, A. M. (2022). Head Start and mothers' work: Free child care or something more?. *Labour Economics*, *75*, 102138.
Raftery, A. E. (1995) Bayesian model selection in social research, *Sociological Methodology*, 111–163.
Ravazzini, L. (2018). Childcare and maternal part-time employment: A natural experiment using Swiss cantons. *Swiss Journal of Economics and Statistics*, *154*, 1–16.
Roodman, D., Nielsen, M. Ø, MacKinnon, J. G., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *The Stata Journal*, *19*, 4–60.
Sabol, T. J., & Chase-Lansdale, P. L. (2015). The influence of low-income children's participation in Head Start on their parents' education and employment. *Journal of Policy Analysis and Management*, *34*, 136–161.
Sall, S. P. (2014). Maternal labor supply and the availability of public pre-k: Evidence from the introduction of prekindergarten into American public schools. *Economic Inquiry*, *52*, 17–34.
Sant'Anna, P. H. C., & Xu, Q. (2023) Difference-in-Differences with Compositional Changes, https://arxiv.org/abs/2304.13925.
Sayer, L. C., Bianchi, S. M., & Robinson, J. P. (2004). Are parents investing less in children? Trends in mothers' and fathers' time with children. *American Journal of Sociology*, *110*, 1–43.
Schiman, C. (2022). Experimental evidence of the effect of head start on mothers' labor supply and human capital investments. *Review of Economics of the Household*, *20*, 199–241.
Schochet, O. N. (2022). Case management or child care: Which has the greater impact on parental human capital and self-sufficiency in two-generation programs?. *Social Service Review*, *96*, 703–743.
Schuss, E., & Azaouagh, M. (2021). Combining parenthood and work: Transmission channels and heterogeneous returns to early public childcare. *Review of Economics of the Household*, *19*, 641–676.
Soldani, E. (2021). Public kindergarten, maternal labor supply, and earnings in the longer run: Too little too late?. *Labour*, *35*, 214–263.
Stahl, J. F., & Schober, P. S. (2018). Convergence or divergence? Educational discrepancies in work-care arrangements of mothers with young children in Germany. *Work, Employment and Society*, *32*, 629–649.
Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys*, *19*, 309–345.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, *70*, 103–127.

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research Synthesis Methods*, *8*, 19–42.

Stanley, T. D., Doucouliagos, H., Ioannidis, J. P., & Carter, E. C. (2021). Detecting publication selection bias through excess statistical significance. *Research Synthesis Methods*, *12*, 776–795.

Steel, M. F. (2020). Model averaging and its use in economics. *Journal of Economic Literature*, *58*, 644–719.

Ugur, M., Churchill, S. A., & Luong, H. M. (2020). What do we know about R&D spillovers and productivity? Meta-analysis evidence on heterogeneity and statistical power. *Research Policy*, *49*, 103866.

van Aert, R. C., & van Assen, M. (2023) Correcting for publication bias in a meta-analysis with the p-uniform* method, Technical report.

Vivalt, E. (2019). Specification searching and significance inflation across time, methods and disciplines. *Oxford Bulletin of Economics and Statistics*, *81*, 797–816.

Washbrook, E., Ruhm, C. J., Waldfogel, J., & Han, W.-J. (2011) Public policies, women's employment after childbearing, and child well-being, *The BE Journal of Economic analysis & Policy*, 11.

Weiland, C., Unterman, R., & Dynarski, S. (2024). Lottery-Based Evaluations of Early Education Programs: Opportunities and Challenges for Building the Next Generation of Evidence. *AERA Open*, *10*, 23328584241231933.

Weixler, L., Valant, J., Bassok, D., Doromal, J. B., & Gerry, A. (2020). Helping parents navigate the early childhood education enrollment process: Experimental evidence from New Orleans. *Educational Evaluation and Policy Analysis*, *42*, 307–330.

Weixler, L., Valant, J., Doromal, J. B., & Gerry, A. (2024). Increasing access in the ECE enrollment process: Evidence from an information intervention in New Orleans. *Early Childhood Research Quarterly*, *68*, 54–64.

Wikle, J., & Wilson, R. (2023). Access to Head Start and maternal labor supply: Experimental and quasi-experimental evidence. *Journal of Labor Economics*, *41*, 1081–1127.

Xue, X., Cheng, M., & Zhang, W. (2021). Does education really improve health? A meta-analysis. *Journal of Economic Surveys*, *35*, 71–105.

Yeh, C. T., & Wodtke, G. T. (2023). The Effects of Head Start on Low-Income Mothers. *Socius*, *9*, 23780231231192392.

Zimmert, F. (2023). Early child care and the employment potential of mothers: Evidence from semi-parametric difference-in-differences estimation. *Journal for Labour Market Research*, *57*, 19.