IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18042

The Causal Impact of School-Meal Programmes on Children in Developed Economies: A Meta-Analysis

Sara Ayllón
Samuel Lado

JULY 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18042

The Causal Impact of School-Meal Programmes on Children in Developed Economies: A Meta-Analysis

Sara Ayllón
IZA and University of Girona

Samuel Lado
University of Girona

JULY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Causal Impact of School-Meal Programmes on Children in Developed Economies: A Meta-Analysis*

This paper is the first to meta-analyse the literature on the causal effects of school-meal programmes on children's behavioural, health and educational outcomes in developed countries, while addressing potential publication bias and heterogeneity between studies. We create a sample of 2,821 estimates from 42 studies and gather 59 aspects reflecting the context in which each estimate was obtained, including type of data, programme characteristics, student population, estimation method and publication quality, among others. We employ both linear and non-linear techniques to correct for publication bias, and we use Bayesian model averaging to study heterogeneous effects and address model uncertainty. The results are consistent with small publication bias — with the exception of studies devoted to analysing test scores, which appear more selective when reporting results. Once publication bias is accounted for, we find that school-meal programmes in high-income economies have minimal impact on students' behaviour, health and education. Our heterogeneity analysis documents the fact that means-tested programmes and breakfast initiatives yield the greatest benefits for children's outcomes.

JEL Classification: H42, H53, I38

Keywords: school-meal programmes, children, behaviour, health, education, meta-analysis, Bayesian model averaging, model uncertainty

Corresponding author:
Sara Ayllón
EQUALITAS
C/Universtat de Girona 10
17003
Girona
Spain
E-mail: sara.ayllon@udg.edu

* We are extremely grateful for the funding received from the Social Research Call 2022 of 'La Caixa' Foundation (project number LCF/PR/SR22/52570007) as well as the projects PID2022-137352NB-C41 (funded by MCIN/AEI/10.13039/501100011033) and 2021-SGR-570. Samuel Lado also acknowledges support from FPU21/00242 ('Ayudas para la Formaci´on de Profesorado Universitario', Ministerio de Universidades). Participants at the 2024 MAER-Net Colloquium (September 2024) are thanked for their useful comments. Any errors or misinterpretations are our own.

1 Introduction

School-meal programmes are often credited with offering benefits to children across several dimensions. They are supposed to level the playing field in education by improving attendance, engagement and achievement. Assuming that they are nutritionally balanced, school meals can also promote general health and tackle malnutrition, underweight and obesity. They are justified in terms of better behaviour and enhanced socialisation. Accordingly, recent years have seen a considerable shift towards increasing the coverage of school-meal programmes in developed economies, with some countries mandating that such programmes become universal. For example, in the US, the Community Eligibility Provision (CEP) programme now provides free meals to all students in schools where 25% of the pupil population receive income-based assistance, such as the Supplemental Nutrition Assistance Program (SNAP) (previously it was 40%). In the European context, since academic year 2014/2015, England and Scotland have provided universal free school meals for children aged 4—7. In that same year, in Estonia, where free meals were initially offered to pupils in grades 1 to 4, the programme was extended to cover all students up to grade 12. More recently, in 2020, Lithuania introduced free meals for pre-schoolers and first-grade pupils, while Wales began a phased rollout of universal free meals for all primary-school children in 2022. In London, as from academic year 2023/2024, the City Council has also introduced free meals for all pupils in primary education. And while the claims concerning the positive effects of school-meal programmes made by the general public, stakeholders and policy-makers are easy to understand and embrace, we do not know whether the bulk of empirical evidence is consistent with all these assertions; particularly in developed economies, where children's general level of health is high, the great majority of pupils are well fed, while truancy rates are very low in the majority of countries.

Multiple studies have examined the effects of school-meal programmes on a range of outcomes, such as programme participation (Ruffini, 2022; Corcoran et al., 2016); body mass index (BMI), nutrient intake, obesity, overweight and overall health (Abouk and Adams, 2022; Schanzenbach, 2009); academic performance and attendance (Gordanier et al., 2020; Imberman and Kugler, 2014); behaviour (Altindag et al., 2020; Norwood, 2020); income in adulthood (Lundborg et al., 2022; Bütikofer et al., 2018); and household finances (Marcus and Yewell, 2022). However, the findings from these studies are far from conclusive: while some indicate positive effects on the expected outcomes (Cuadros-Meñaca et al., 2023; Altindag et al., 2020; Belot and James, 2011; Hinrichs, 2010), others suggest limited, inconsistent or even negative impacts (Schanzenbach, 2009; McEwan, 2013). A synthesis of the evidence of this strand of literature is further complicated by methodological differences across studies, disparities in the institutional contexts, alternatives that the school-meal programmes crowd out, the characteristics of eligible children and diversity in meal quality, among other important features. Thus far, we do not know what the overall effect of school-meal programmes is on child outcomes in developed economies. Nor are we aware of the extent to which the existing results are largely context dependent. This paper answers these questions.

Estimating the causal effect of school-meal programmes in rich contexts has proved difficult because of three main challenges. First, there is limited variation in programme implementation across schools, regions, countries or over time (Ruffini, 2022; Schwartz and Rothbart, 2020; Bitler and Seifoddini, 2019). This has often stymied the possibility of using quasi-experimental designs from which to derive the causal effect of the programmes.

Second, the paucity of data linking participation in school-meal programmes with child outcomes has also limited the number of studies that can credibly estimate the causal impact of these schemes. Third, at the student level, there could be endogenous selection into the programme, as students who choose to participate differ along observable and non-observable characteristics, compared to their peers who do not participate, which further complicates the analysis. However, recent methodological advances, the increased availability of administrative data and a number of policy reforms have contributed to the growth of this body of research in recent years (Schwartz and Rothbart, 2020).

It is these studies that, for the first time in this strand of literature, are meta-analysed in this paper, with the objective of assessing the effectiveness of school-meal programmes in developed economies and of informing decision making. The potentially important consequences for children and the budget relevance of school-meal programmes completely justifies such a meta-analysis. For example, in the US, the National School Lunch Program (NSLP), which has annual expenditure of approximately $13.8 billion, is the second-largest food and nutrition assistance programme after SNAP (Marcus and Yewell, 2022). Moreover, we are the first to take into account potential publication bias and p-hacking in this strand of literature, with the aim of exploring the extent to which selective reporting may have shaped the existing body of evidence.

Publication bias is defined by Brodeur et al. (2020) as occurring when 'the statistical significance of a result determines the probability of publication', likely a reflection of the peer-review process. Instead 'p-hacking refers to a variety of practices that a researcher might (consciously or unconsciously) use to generate "better" p-values, perhaps (but not necessarily) in response to the difficulty of publishing statistically insignificant results' (Brodeur et al., 2020: 3634—3635). If left unaddressed, these issues can result in biased estimates and misleading confidence intervals, particularly in fields with strong underlying intuition, as in our case (Andrews and Kasy, 2019; Doucouliagos and Stanley, 2013).[1]

In view of these challenges, and following the guidelines set out by Irsova et al. (2024), we collect 2,821 estimates from 42 studies that credibly examine the causal impact of school-meal programmes on child outcomes in three main domains: behaviour, health and education. The number of studies available is not yet sufficient to be meta-analysed for other domains, such as the impact on earnings in adulthood, household finances, parental labour market outcomes or reliance on charitable services (Ruffini et al., 2025; Holford and Rabe, 2022; Lundborg et al., 2022; Marcus and Yewell, 2022; Bütikofer et al., 2018). We focus solely on developed economies, because the objectives pursued by school-meal programmes in high-income countries differ from those in developing economies.[2]

Importantly, we gather 59 aspects reflecting the context in which these estimates were obtained. Accounting for heterogeneity across contexts is important for three main reasons. First, variations in school-meal programmes may lead to differences in the benefits that children can potentially gain. Second, heterogeneity analysis can provide a certain

[1] According to Ioannidis et al. (2017), selective reporting in economics often exaggerates the typical reported estimates twofold.

[2] In contrast to our focus on developed countries, Wang et al. (2021) conduct a systematic review and meta-analysis that investigates the impacts of school-meal programmes on educational and health outcomes specifically in low- and middle-income countries. Their findings suggest that these programmes are associated with an increase in height, weight and school attendance among school-age children and adolescents. The authors find mixed evidence of publication bias, depending on the outcome under scrutiny. However, their analysis only includes estimates from randomised controlled trials and controlled before-after studies based on limited sample sizes. Moreover, publication bias is not explicitly tested, but is studied solely using funnel plot figures.

consensus and guidance in terms of the most effective design for school-meal programmes. And, third, context matters because policy-makers tend to dismiss research that does not apply to their country or region (Jackson and Mackevicius, 2024; Clarke, 2019). In this respect, we included data on programme and student characteristics (whether the school-meal programme covers lunch, breakfast or both; whether it is universal or means-tested; whether it is targeted at specific grades; whether it refers to a subgroup; and country), data characteristics (outcome of interest, type of treatment and whether the data is longitudinal or cross-sectional and administrative or from a survey), estimation characteristics (methodology and type of controls) and publication characteristics (type of publication, impact factor and number of citations). We investigated whether these aspects consistently influence the estimated values, using Bayesian model averaging (BMA), which allowed us to explore all possible models by computing every combination of explanatory variables, and thus addressing model uncertainty.

Our findings suggest that the expectations of school-meal programmes in developed economies are often too high: most of our tests indicate that, once we account for publication bias, school-meal programmes do not have a statistically significant effect on health, behaviour or educational outcomes. In those tests where a statistically significant effect is present, it is so small as to be not economically meaningful. Our results also indicate that publication bias is small in this strand of literature, which provides confidence that the evidence being built is not based on selective reporting — except for the studies devoted to research into standardised test scores, among which publication bias is rather clearer. We provide further evidence that even after controlling for a wide range of study characteristics, the bias-corrected impact of school-meal programmes on child development in rich economies is minimal. In other words, heterogeneity across studies does not drive our main findings. Nevertheless, means-tested programmes and breakfast initiatives are the two features most strongly linked with positive benefits. This means that the effectiveness of a school-meal programme depends significantly on where (the context) and how (the design) it is implemented, and for whom (the targeted population) it is intended.

This paper makes two main contributions to the literature. First, it offers the inaugural meta-analysis in this field. To the best of our knowledge, research on the impact of school-meal programmes on children in developed countries has only been studied in summaries and systematic reviews (see, for instance, Spill et al., 2024; Kurtz et al., 2022); however, no meta-analysis has been published to date. Therefore, this body of research still awaits (potential) correction for publication bias and p-hacking. We address both issues using the most recently developed meta-analytical techniques, including linear and non-linear funnel-based methods, as well as selection models. Second, apart from considering the extent of selective reporting, this is also the first study to examine how the various characteristics of school-meal programmes analysed in the literature influence the reported effects on children. This is of particular interest, as the literature on school meals has yet to reach a consensus on the most effective design for such programmes. As a result, school-meals provision and coverage vary greatly from country to country (Lundborg et al., 2022). In the context of tight public budgets, policy-makers may be particularly interested in identifying the most cost-effective approach. We document the finding that school-breakfast initiatives and means-tested programmes that target low-income children are the two approaches that are most beneficial to children's development.

This paper is organised as follows. After this introduction, Section 2 offers a description of the search strategy, the inclusion criteria and the data collected. Section 3 investigates publication bias, while in Section 4 we explore whether our results are influenced by heterogeneity between studies. Lastly, Section 5 offers some conclusions. In the Online Appendix, Section A provides additional information on the search and data collection process, Section B details the variables used to capture heterogeneity and Section C includes additional tables and figures.

2 Data

We conducted a thorough review of those econometric studies published since 2000 that empirically assess the causal effects of school-meal programmes on children's outcomes and, in particular, on behaviour, health and education. Our search was performed in Google Scholar using the search terms ('school meal*' OR 'school breakfast*' OR 'school lunch*') AND ('programme' OR 'program') AND ('impact') AND ('child*') AND ('outcome*') AND ('causal').[3] Following the recommendations of previous authors, we opted for Google Scholar because it inspects the full text of studies, rather than just the title, abstract and keywords (Opatrny et al., 2025; Irsova et al., 2024; Havranek et al., 2015).

The different stages for the inclusion of the primary studies proceeded as follows. We initially screened the abstracts of the first 500 studies that were identified through the Google Scholar query. We downloaded all the studies that could potentially contain estimates of the causal effects of school-meal programmes on child outcomes. This initial selection yielded a total of 112 studies. We subsequently extended our search by carefully examining the references within those studies, identifying an additional 12 relevant papers. Next, we skimmed the full text of the 124 prospective studies and discarded 70 that did not meet our inclusion criteria and 12 that were duplicates — see Figure A.1 in the Online Appendix A for a PRISMA flow diagram of our search strategy.[4]

The following are the six inclusion criteria that we imposed. First, the study must report an estimated relationship between school meals and child outcomes, including behaviour (e.g. school suspensions, misbehaviour), health (e.g. overweight, obesity) and education (e.g. attendance, test scores).[5] Other outcomes, such as programme participation, household food purchases or life-time earnings, were not examined in our analysis, because there are surprisingly few studies that include them (Lundborg et al., 2022; Marcus and Yewell, 2022; Bütikofer et al., 2018). Second, we restricted the analysis to studies that employ causal inference methods, namely difference-in-differences (DiD), regression discontinuity design (RDD) and instrumental variables (IV). Third, we only considered studies conducted in Organisation for Economic Co-operation and Development (OECD) countries, so that our results are relatively comparable across contexts with similar school systems. Fourth, the study must include either the standard error or an alternative measure that allows for its reconstruction. Fifth, the research should focus on the implementation of a school-meal programme, its expansion or the improvement of an existing one. Studies that examine the reduction or removal of school meals, or any regressive measures, were not considered (Maruyama and Nakamura, 2025; von Hinke Kessler Scholder, 2013). And sixth, we disregarded reports from NGOs or governments, as well as pilot

[3]Asterisks enable searches to include variations of a keyword. For instance, 'child*' will return results containing words such as 'child', 'children', 'childhood' and 'childcare'.

[4]Out of the final sample of 42 studies, 30 were identified through Google Scholar, while the remaining 12 were obtained via snowballing. Of those 30 articles, we had previously identified 22, or 73.3%, indicating that our search was effective in capturing the most important primary studies (Irsova et al., 2024).

[5]We consider student absenteeism to be an educational outcome, regardless of the reasons for school being missed.

studies (e.g. Brown et al., 2012). The final sample included the 42 studies listed in Table 1.

Table 1: Studies included in the meta-analysis

Abouk and Adams (2022)	Cuadros-Meñaca et al. (2023)	Imberman and Kugler (2014)
Altindag et al. (2020)	Davis (2019)	Kho (2018)
Anderson et al. (2018)	Davis and Musaddiq (2019)	Kim (2021)
Ayllón and Lado (2025)	Davis et al. (2024)	Kirksey and Gottfried (2021)
Bartfeld et al. (2020)	Domina et al. (2024)	Leos-Urbel et al. (2013)
Belot and James (2011)	Dotter (2013)	Lundborg et al. (2022)
Bethmann and Cho (2022)	Dunifon and Kowaleski-Jones (2003)	McEwan (2013)
Bhattacharya et al. (2006)	Frisvold (2015)	Norwood (2020)
Borbely et al. (2024)	Gordanier et al. (2020)	Ribar and Haldeman (2013)
Bütikofer et al. (2018)	Gordon and Ruffini (2021)	Rothbart et al. (2023)
Capogrossi (2012)	Gutierrez (2021)	Ruffini (2022)
Collante Zárate et al. (2024)	Hinrichs (2010)	Schanzenbach (2009)
Corcoran et al. (2016)	Holford and Rabe (2022)	Schanzenbach and Zaki (2014)
Cuadros-Meñaca et al. (2022)	Holford and Rabe (2024)	Schwartz and Rothbart (2020)

Note: The table lists all the primary studies included in the meta-analysis. See more details on the search strategy in Online Appendix A.

From the studies included, we gather 2,821 estimates that are used in the meta-analysis. For each estimate, we also collect 59 factors that reflect the context in which the estimate was obtained, the method of estimation, the student population under scrutiny, the type of data used, etc. See more detail in Section 4 and in Online Appendix B.[6]

Importantly, in the meta-analytical database that we produced, each estimate captures the impact of school-meal programmes on various child outcomes, and therefore different measures are used (for example, suspension rates, standardised mean test scores, number of days absent in a school year, percentage of obesity prevalence or BMI z-scores). As a result, these estimates are not directly comparable.[7] To address this concern, we use partial correlation coefficients (PCCs) to standardise effect sizes, as in previous meta-analyses, using the following formula:

$$r_{ij} = \frac{t_{ij}}{\sqrt{t_{ij}^2 + df_{ij}}} \tag{1}$$

where r_{ij} represents the partial correlation coefficient of the i-th estimate reported in study j, t_{ij} denotes the corresponding t-statistic, and df_{ij} indicates the sample size of a given estimate minus three (Stanley et al., 2023). However, as noted by Stanley et al. (2023, 2024), PCCs are inherently related to their standard errors, making meta-analysis of PPCs generally biased. To mitigate this bias, one approach is to transform the PCCs into Fisher's z, as follows (Borenstein et al., 2009):

$$z_{ij} = 0.5 \times \ln\left(\frac{1 + r_{ij}}{1 - r_{ij}}\right) \tag{2}$$

[6]These 160,000+ data points were gathered manually by the authors of this paper, who would read each paper individually and jointly discuss how data needed to be introduced in the main dataset.

[7]For several outcomes, we also adjust the signs of the reported estimates, so that they accurately reflect the direction of the effect — this allows us to compare, for instance, a reduction in misbehaviour with an increase in positive behaviour. In the case of average BMI, we adjust the coefficient, so that a decline in BMI is considered positive, given that the vast majority of children in the contexts analysed in this paper are not underweight or undernourished.

where z_{ij} represents the Fisher's z-transformation of a given estimate with its standard error being $SE(z_{ij}) = \sqrt{\frac{1}{n_{ij}-3}}$. We use this metric to assess publication bias and heterogeneity, then report the results in terms of Cohen's d, as it is a well-established metric in the social sciences for measuring effect sizes. To transform Fisher's z to Cohen's d, we follow two steps. First, we transform Fisher's z back to PCCs:

$$r_{ij} = \frac{e^{2z_{ij}} - 1}{e^{2z_{ij}} + 1} \tag{3}$$

Second, we convert PCCs to Cohen's d:

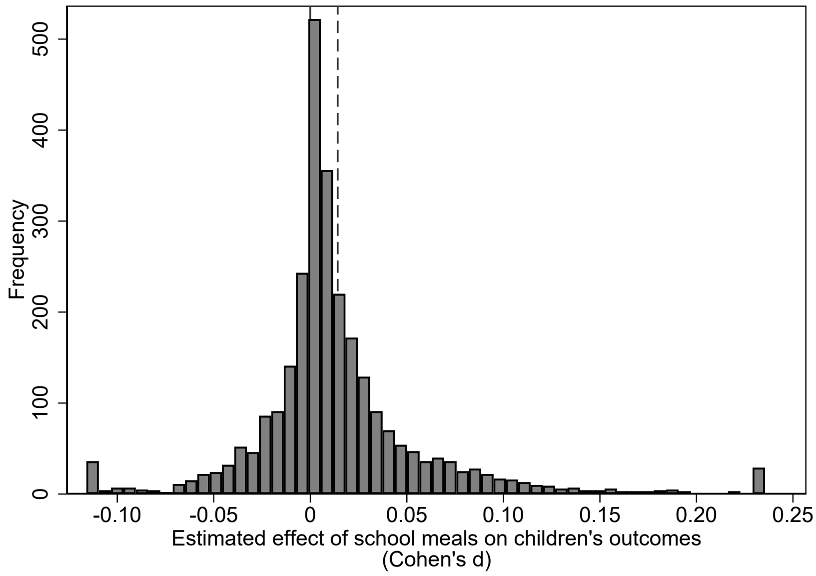
$$d_{ij} = \frac{2r_{ij}}{\sqrt{1 - r_{ij}^2}} \tag{4}$$

with $SE(d_{ij}) = \sqrt{\frac{4V_{r,ij}}{(1-r_{ij}^2)^3}}$. We interpret the results using the benchmark proposed by Kraft (2020), which defines an effect size below 0.05 as 'small', between 0.05 and 0.20 as 'medium' and 0.20 or greater as 'large'.

Figure 1 presents the histogram of effect sizes in terms of Cohen's d. We observe that estimates close to zero are prevalent in the literature, with the distribution showing relatively large tails, skewed slightly to the right. The dashed line represents the unconditional sample mean of 0.014 — providing a first indication that school-meal programmes in developed countries may have, on average, a small effect on child outcomes. In turn, Figure 2 shows the distribution of effect sizes reported in the individual studies, sorted by the period of time to which the data refers. More than half of the studies refer to the last decade, and no systematic differences are observed regarding the period of time under scrutiny. With few exceptions, most studies indicate either non-significant effects or small positive effects. Additionally, significant within-study variation is observed, which supports our strategy based on collecting all the estimates and not just one (or a few) per study.

Figure 3 shows the distribution of effect sizes, while accounting for potential sources of systematic heterogeneity. We thus consider: (a) whether the school-meal programme is means-tested or universal; (b) whether it covers breakfast, lunch or both; (c) the domain of the student outcome; and (d) the method employed. In Panel A, we observe that the distribution of the effects of both means-tested programmes and universal programmes follows a similar pattern. Although less frequent, the estimates of means-tested programmes are more skewed towards the right, possibly suggesting, on average, a more positive effect on children's outcomes than in the case of universal schemes. As for whether the programme includes breakfast, lunch or both (Panel B), estimates for breakfast and lunch appear predominantly positive, whereas those for programmes that offer both meals are, on average, closer to zero. In terms of student outcomes (Panel C), the distribution of estimates is highly uneven. Educational outcomes are studied more frequently and tend to lean towards the right. Results for behaviour are generally positive but less common, while health outcomes are skewed to the left, indicating the greater prevalence of negative results. With regard to the method employed (Panel D), the majority of studies use a DiD strategy, with estimates tending to be positive. Although less frequent, IV estimates also tend to be right-skewed. The RDD distribution is more heavily skewed to the right, suggesting that studies using this methodology report more positive effects.

Figure 1: Distribution of the effect sizes

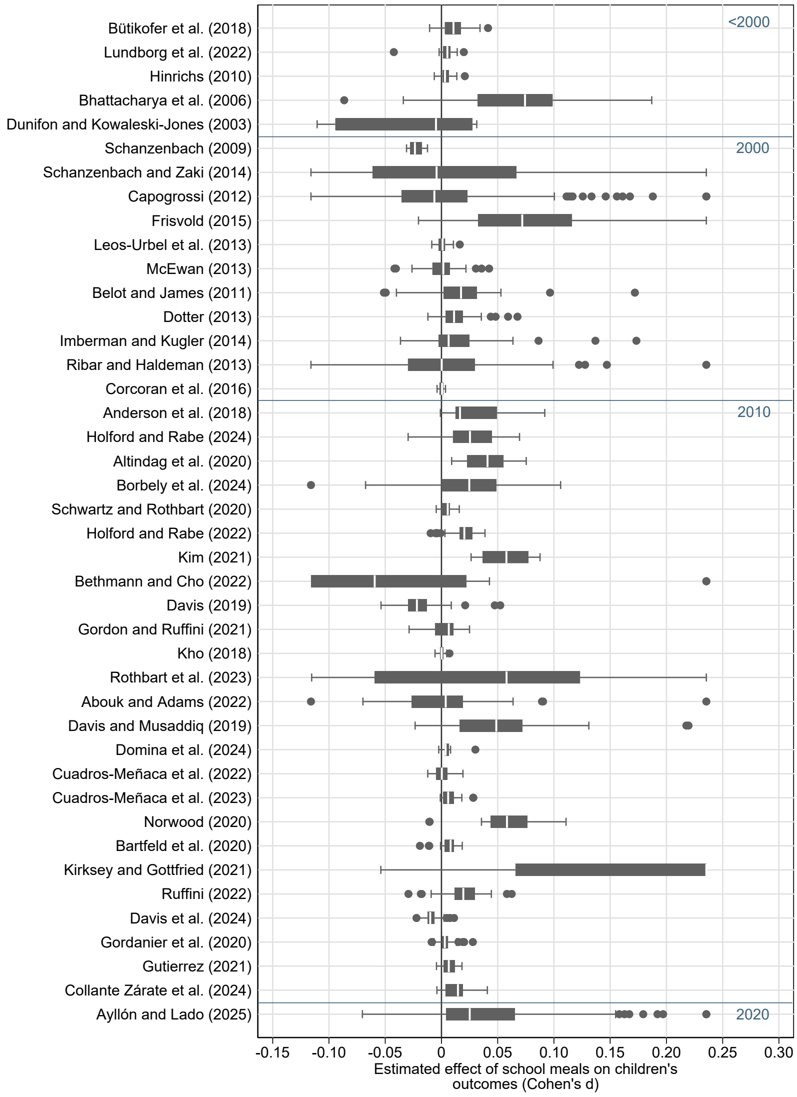


Note: The histogram displays the effect sizes from primary studies in terms of Cohen's d. The solid line is set at zero, while the dashed line represents the sample mean. Extreme outliers are winsorised at the 1st and 99th percentiles.

Table 2 presents summary statistics for all subsets of studies, allowing additional comparisons. Column (1) reports the number of observations, while Columns (2) and (3) show the unweighted mean Cohen's d for each subset of the literature and the 95% confidence interval. Columns (4) and (5) give the results weighted by the inverse number of estimates per study, so that each article has equal importance.[8] There are a few lessons that we can learn from the table. First and foremost, although the overall impact of school-meal programmes on children's outcomes is small, with an average of 0.014 (see the last row in the table), there is significant variability across subsets, which underlines the importance of considering heterogeneity — as we do in Section 4 of this paper. Second, school-meal programmes have, on average, a more positive effect on students' behaviour (0.017) and educational (0.017) outcomes, while their effects on health (0.008) are smaller. Third, breakfast (0.018) and means-tested (0.021) initiatives tend to show slightly stronger effects than programmes covering lunch (0.016) and those with universal coverage (0.010). Focusing on specific programmes, the US School Breakfast Program (SBP) and the Breakfast After the Bell (BAB) initiatives display more positive outcomes, while the National

[8] Weight selection is an important decision in meta-analysis (see, for instance, Opatrny et al., 2025). Three weighting schemes have commonly been used in meta-analytical research: i) equal weight to each estimate; ii) equal weight to each study; and iii) inverse-variance weighting. Our results from the first two approaches are similar, as shown in Table 2. Inverse-variance weighting, which assigns greater importance to studies with higher precision, is not presented in Table 2. When applied, the unconditional sample mean decreases by more than half (down to 0.006), indicating that if we were to consider only studies with the most precise estimates, the overall effect of school-meal programmes would be even closer to zero.

Figure 2: Effect sizes in primary studies

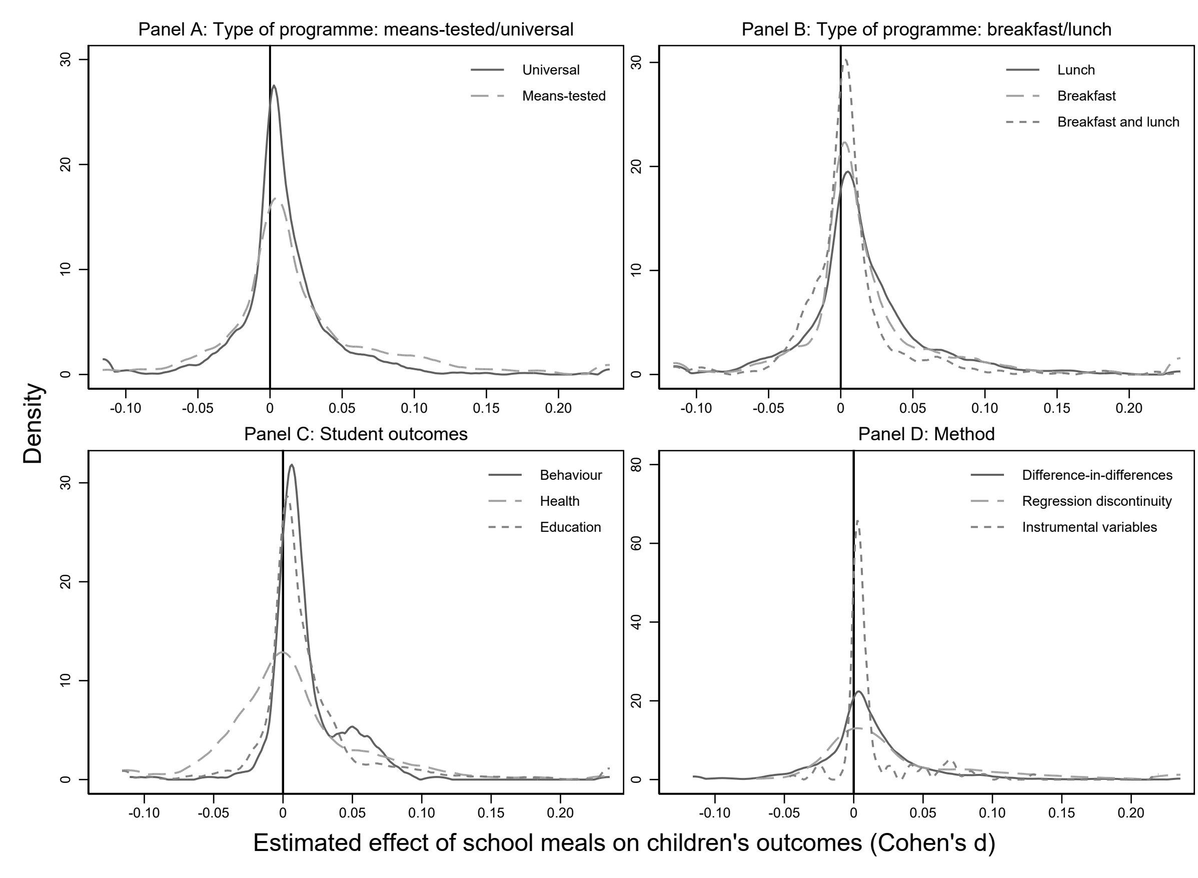


Note: The figure presents the effect sizes from primary studies in terms of Cohen's d, sorted by the time period to which the data analysed refers. The solid line is set at zero. Each box represents the interquartile range (P25-P75), with a line indicating the median. The whiskers extend to data points within 1.5 times the interquartile range. Extreme outliers are winsorised at the 1st and 99th percentiles.

School Lunch Program (NSLP) does not.[9] Interestingly, most of these effects are concentrated among high-school students. Fourth, in terms of student characteristics, we do not observe any significant differences, except in estimates that focus solely on male students (0.011), which are, on average, higher than those for females (0.001). Fifth, studies using cross-sectional data at the school or district level tend to yield more positive estimates.

[9] Table C.1 in Online Appendix C presents a description of the main school-meal programmes covered in this meta-analysis.

Figure 3: Distribution of effects sizes by potential sources of systematic heterogeneity



Note: The figure presents the distribution of effect sizes from primary studies in terms of Cohen's d, grouped by potential sources of systematic heterogeneity. The solid line is set at zero. Methodologies other than DiD, RDD and IV are excluded from Panel D, due to the limited number of observations. Extreme outliers are winsorised at the 1st and 99th percentiles.

Similarly, those employing an RDD (0.036) show such effects more often than studies using DiD (0.011) or IV (0.016) methods. Finally, studies published after 2019 (0.021) present higher coefficients than those published earlier (0.009). We do not observe any differences based on whether the article appears in a peer-reviewed journal or is a working paper.

3 Publication bias

In Section 2, we discussed the results derived from group means. However, these findings may be influenced by selective reporting (publication bias) and heterogeneity. In this and the following section we will address both issues.

Publication bias is often associated with the correlation between estimated values and standard errors, and there are at least two reasons for this relationship. First, researchers, editors or referees may have a preference for statistically significant results. In the face of

Table 2: Summary statistics for different subsets of the literature

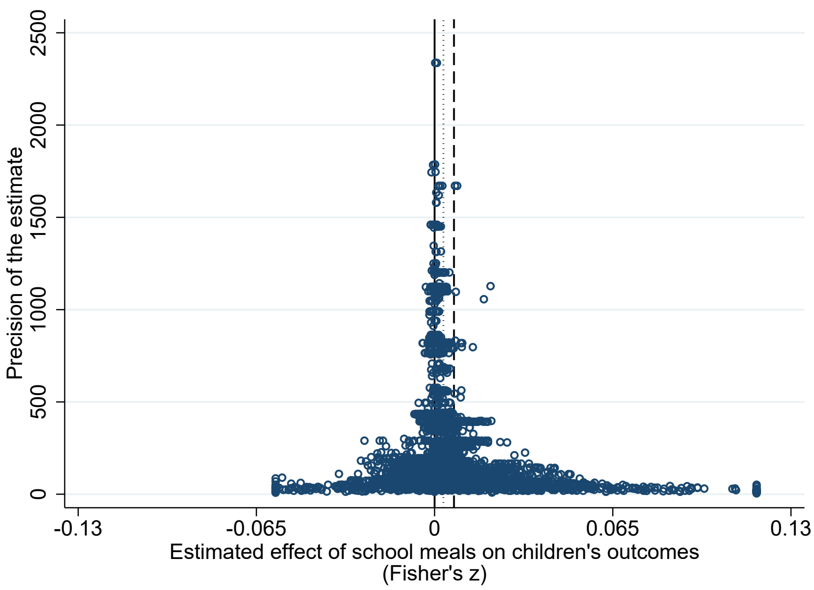
		Unweighted			Weighted		
	Observations	Mean	95% conf. int.		Mean	95% conf. int.	
	(1)	(2)	(3)		(4)	(5)	
Outcome analysed							
Outcome: Behaviour	258	0.017	0.013	0.020	0.013	0.008	0.017
Outcome: Health	943	0.008	0.004	0.011	0.011	0.008	0.015
Outcome: Education	1620	0.017	0.015	0.020	0.022	0.019	0.025
Programme variation							
Lunch	1265	0.016	0.013	0.018	0.012	0.009	0.014
Breakfast	925	0.018	0.015	0.022	0.030	0.026	0.034
Breakfast and lunch	631	0.005	0.002	0.008	0.010	0.007	0.013
Means-tested	1085	0.021	0.018	0.025	0.030	0.025	0.034
Universal	1736	0.010	0.008	0.011	0.012	0.010	0.014
Elementary school	2633	0.014	0.012	0.016	0.015	0.013	0.018
Middle school	1701	0.017	0.014	0.019	0.019	0.017	0.022
High school	611	0.027	0.022	0.031	0.033	0.027	0.038
Country: US	1982	0.011	0.008	0.013	0.017	0.014	0.019
Country: Other	839	0.023	0.020	0.025	0.018	0.015	0.021
Breakfast After the Bell	651	0.014	0.010	0.018	0.027	0.021	0.032
Community Eligibility Provision	555	0.007	0.004	0.010	0.010	0.007	0.014
National School Lunch Program	450	-0.000	-0.005	0.004	-0.012	-0.015	-0.008
School Breakfast Program	302	0.018	0.011	0.026	0.042	0.034	0.050
Student variation							
Female	160	0.001	-0.005	0.006	-0.009	-0.019	-0.000
Male	168	0.011	0.006	0.016	0.012	0.005	0.020
Not minority	90	0.019	0.013	0.026	0.019	0.013	0.026
Minority	133	0.014	0.005	0.023	0.011	0.004	0.017
Advantaged	242	0.009	0.003	0.014	0.007	0.002	0.013
Disadvantaged	308	0.012	0.006	0.019	0.025	0.017	0.032
Type of data							
Longitudinal data	2621	0.014	0.012	0.015	0.018	0.016	0.020
Cross-sectional data	200	0.023	0.017	0.028	0.014	0.006	0.022
Administrative data	2009	0.018	0.016	0.020	0.021	0.019	0.023
Survey data	812	0.005	0.000	0.009	0.005	0.000	0.010
Individual-level data	2267	0.009	0.007	0.011	0.006	0.004	0.008
School-level data	447	0.037	0.032	0.041	0.044	0.038	0.049
District-level data	103	0.027	0.016	0.038	0.035	0.020	0.049
Method and estimation characteristics							
Method: DiD	2051	0.011	0.009	0.012	0.016	0.014	0.017
Method: RDD	380	0.036	0.030	0.042	0.045	0.036	0.054
Method: IV	206	0.016	0.011	0.020	0.025	0.019	0.030
Method: Other	184	0.006	-0.005	0.017	-0.006	-0.015	0.003
Main result	2270	0.016	0.015	0.018	0.018	0.016	0.020
Robustness	551	0.005	0.001	0.009	0.014	0.009	0.020
Publication characteristics							
Published before 2019	1529	0.009	0.006	0.011	0.012	0.009	0.014
Published after 2019	1292	0.021	0.018	0.023	0.022	0.019	0.025
Published in journal	1709	0.015	0.013	0.017	0.016	0.014	0.019
Working paper	1112	0.013	0.010	0.016	0.020	0.017	0.022
Published in economics	1432	0.014	0.013	0.016	0.016	0.015	0.018
Not published in economics	1389	0.014	0.011	0.017	0.018	0.015	0.022
All estimates	2821	0.014	0.012	0.016	0.017	0.015	0.019

Note: The table presents the effect sizes from primary studies in terms of Cohen's d for each subset of the literature. In Columns (4) and (5), we weight each observation by the inverse number of estimates reported per study. Minority children include those from immigrant backgrounds, black, Hispanic or other ethnicities. Disadvantaged children are defined as those living in low-income households, alongside parents with limited educational attainment, experiencing disabilities, coming from disrupted families or having shown poor academic performance previously. Advantaged children are those who do not experience any of these circumstances. Extreme outliers are winsorised at the 1st and 99th percentiles. See Table B.1 in Online Appendix B for a definition of the variables, and Table C.1 in Online Appendix C for details on the main school-meal programmes covered by the primary studies.

some degree of imprecision in their data and methods, researchers might explore various combinations of control variables until they achieve an estimate large enough to counteract the standard error. Second, researchers, besides preferring statistically significant results, may favour estimates with intuitive signs, leading them to dismiss counterintuitive findings (Havranek et al., 2022).

In Figure 4, we visually examine the presence of publication bias in the literature on school meals using a funnel plot (Egger et al., 1997). This scatter plot illustrates the relationship between Fisher's z estimates (on the horizontal axis) and their precision (inverse of the standard error, on the vertical axis). The most precise estimates should cluster close to the true mean effect in the top portion of the graph, with variance increasing towards the bottom as precision decreases. In the absence of publication bias, the funnel plot should be symmetrical (Stanley, 2005). However, the plot in Figure 4 displays some slight asymmetry, skewed towards the right, potentially suggesting a small degree of selective reporting in the literature under scrutiny.

Figure 4: Funnel plot for the causal estimates of school-meal programmes



Note: The figure presents the relationship between effect sizes from primary studies in terms of Fisher's z and their precision. The solid line represents zero, the dotted line indicates the median and the dashed line represents the sample mean. Extreme outliers are winsorised at the 1st and 99th percentiles.

To evaluate the extent of publication bias, we first conduct a series of linear tests based on the regression of Fisher's z estimates against their respective standard errors, employing the following equation (Stanley, 2005; Egger et al., 1997; Card and Krueger, 1995):

$$z_{ij} = \alpha + \gamma SE(z_{ij}) + \epsilon_{ij} \tag{5}$$

where z_{ij} is the i-th Fisher's z of the effect of school-meal programmes on child outcomes from the j-th study, $SE(z_{ij})$ represents the corresponding standard error and ϵ_{ij} is the error term. Standard errors are clustered at the study level to account for within-study correlation (Irsova et al., 2024). We focus on two parameters: α and γ. The latter provides information on the existence, direction and magnitude of publication bias, while the former, reported in the tables in terms of Cohen's d to ease interpretation, captures the mean effect size corrected for publication bias. In the absence of selective reporting, the slope coefficient γ should be zero, indicating no relationship between the estimates and their standard errors. We follow the benchmark of Doucouliagos and Stanley (2013): if $|\widehat{\gamma}|$ is less than 1, it suggests 'little to modest' selectivity; if $|\widehat{\gamma}|$ falls between 1 and 2, it indicates 'substantial' selectivity; meanwhile values exceeding 2 point to 'severe' selectivity.

Table 3 presents the findings from various alternative ways of estimating Equation (5). Panel A details the results when using linear methods, while Panel B does the same for non-linear methods. In the first column of Panel A, we provide the standard ordinary least squares (OLS) results. However, if unobserved characteristics of the primary studies are correlated with the estimated effects, this approach may generate misleading outcomes. To overcome this limitation, in Column (2), we run a model with study fixed effects to account for unobserved heterogeneity at the study level. In Column (3), we use the meta-analysis instrumental variable estimator (MAIVE) (Irsova et al., 2023). This method uses the inverse of the square root of the sample size as an instrument for the reported standard error.[10] In the last two columns of Panel A, we estimate Equation (5) using two alternative weighting approaches. First, we weight each observation by the inverse number of estimates reported per study, giving equal weight to each study (Krueger, 2003). Second, we assign greater weight to more precise estimates by using the inverse of the standard error (Stanley, 2005). Results from Panel A confirm the intuitive interpretation of the funnel plot: there is a slight indication of publication bias in favour of positive estimates, which can be classified as 'little to modest' according to Doucouliagos and Stanley (2013). The corrected mean (effect beyond bias) shrinks from the unconditional sample mean of 0.014 to between -0.002 and 0.006 standard deviation units and is imprecisely estimated, suggesting that, after accounting for publication bias, school-meal programmes in developed countries have an impact on children's outcomes that is so small as not to be meaningful. These findings remain consistent across the alternative methods used.

So far, we have assumed that publication bias is a linear function of the standard error. However, this relationship may not hold universally (Andrews and Kasy, 2019; Bom and Rachinger, 2019). For instance, very precise estimates, concentrated at the top of the funnel plot, are less likely to suffer from publication bias when the true effect is non-zero (Bom and Rachinger, 2019; Stanley and Doucouliagos, 2014). Hence, in Columns (6) to (8) of Panel B, we present three tests that relax the linearity assumption. We use the weighted average of adequately powered (WAAP) approach by Ioannidis et al. (2017) in Column (6). This method estimates the true effect by considering studies with statistical power exceeding 80%. Below this threshold, statistically significant findings are more likely due to chance and bias.[11] Similarly, in Column (7), we employ the stem-based method (Furukawa, 2021), focusing solely on highly precise estimates. Unlike the

[10]Sample size is robust to selection, unaffected by measurement error, not influenced by changing methodology and not mechanically linked to effect sizes (Irsova et al., 2023). MAIVE also accounts for p-hacking (for example, selecting the error clustering that yields significant results), which could bias the estimates of the underlying mean if left unaddressed (Irsova et al., 2023).

[11]For a study to possess sufficient power, its standard error must be smaller than the absolute value of the underlying effect divided by 2.8 (Ioannidis et al., 2017).

Table 3: Publication bias tests

Panel A: Linear models					
	OLS	FE	MAIVE	wNOBS	WLS
	(1)	(2)	(3)	(4)	(5)
Publication bias	0.471**	0.669***	0.467**	0.448	0.473**
	(0.205)	(0.147)	(0.212)	(0.296)	(0.198)
Effect beyond bias	0.003	-0.002	0.003	0.006	0.002*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Observations	2821	2821	2821	2821	2821
Panel B: Non-linear models					
	WAAP	STEM	EK	AK	*p*-uniform*
	(6)	(7)	(8)	(9)	(10)
Publication bias	.	.	0.451***	.	L=32.404***
	(.)	(.)	(0.049)	(.)	(0.000)
Effect beyond bias	0.003***	0.016*	0.003***	0.005***	0.013***
	(0.000)	(0.004)	(0.000)	(0.000)	(0.001)
Observations	2821	2821	2821	2821	2821

Note: Panel A presents the results from the regression $z_{ij} = \alpha + \gamma SE(z_{ij}) + \epsilon_{ij}$, with z_{ij} as the i-th Fisher's z of the effect of school-meal programmes on children's outcomes from the j-th study and $SE(z_{ij})$ as its standard error. FE denotes study-level fixed effects, MAIVE refers to meta-analysis instrumental variable estimator (Irsova et al., 2023), wNOBS assigns weight to each observation based on the inverse number of estimates reported per study, WLS assigns weight to each observation based on the inverse of the standard error. In Panel B, WAAP stands for weighted average of adequately powered (Ioannidis et al., 2017), STEM refers to the stem-based technique (Furukawa, 2021), EK denotes the endogenous kink model (Bom and Rachinger, 2019); AK indicates the selection model by Andrews and Kasy (2019) and *p*-uniform* refers to the selection model by van Aert and van Assen (2023). The effect-beyond-bias estimates are reported in terms of Cohen's d. Extreme outliers are winsorised at the 1st and 99th percentiles. Standard errors are clustered at the study level. *** significant at 1%, ** at 5% and * at 10%.

WAAP approach, the number of estimates included in this method is determined by a bias-variance trade-off, balancing the reduction in publication bias against a potential increase in variance and vice versa. In Column (8), we resort to the endogenous kink (EK) meta-regression model (Bom and Rachinger, 2019), which determines a threshold for the standard error below which publication bias is unlikely to occur. Subsequently, it fits a piecewise linear regression of the estimates on their standard errors, incorporating a break at this threshold. When we relax the assumption that publication bias is a linear function of the standard error, the effects beyond bias become statistically significant, but remain of similar magnitude to those in Panel A, except for the STEM model. The EK model exhibits a similar level of publication bias to that reported in Panel A and still falls within the classification of 'little to modest' selectivity, as defined by Doucouliagos and Stanley (2013).

Thus far, the models we have employed to detect publication bias are categorised as funnel-based. They operate on the assumption that publication bias is influenced by the size of reported estimates, rather than the *p*-values (Irsova et al., 2023). In contrast, selection models suggest that estimates with varying levels of significance are subject to different probabilities of publication — see Figure C.1 in Online Appendix C for graphical

Table 4: Publication bias tests by outcome analysed

Outcome: Behaviour					
Panel A: Linear models	OLS	FE	MAIVE	wNOBS	WLS
Publication bias	0.294	0.390	0.307	-0.550	1.005
	(0.339)	(0.253)	(0.349)	(0.516)	(0.564)
Effect beyond bias	0.013*	0.012***	0.013*	0.022**	0.003
	(0.004)	(0.002)	(0.004)	(0.005)	(0.001)
Observations	258	258	258	258	258
Panel B: Non-linear models	WAAP	STEM	EK	AK	*p*-uniform*
Publication bias	.	.	1.550***	.	L=5.842**
	(.)	(.)	(0.130)	(.)	(0.016)
Effect beyond bias	0.002***	0.002	0.000	0.004***	0.018**
	(0.000)	(0.001)	(0.000)	(0.000)	(0.004)
Observations	258	258	258	258	258
Outcome: Health					
Panel A: Linear models	OLS	FE	MAIVE	wNOBS	WLS
Publication bias	0.370**	0.595***	0.361**	0.269	0.237
	(0.158)	(0.064)	(0.170)	(0.229)	(0.228)
Effect beyond bias	-0.004	-0.012***	-0.004	0.002	0.000
	(0.004)	(0.001)	(0.004)	(0.004)	(0.002)
Observations	943	943	943	943	943
Panel B: Non-linear models	WAAP	STEM	EK	AK	*p*-uniform*
Publication bias	.	.	0.092	.	L=5.129**
	(.)	(.)	(0.079)	(.)	(0.023)
Effect beyond bias	-0.000	-0.000	0.002***	0.001	0.012**
	(0.000)	(0.003)	(0.000)	(0.000)	(0.003)
Observations	943	943	943	943	943
Outcome: Education					
Panel A: Linear models	OLS	FE	MAIVE	wNOBS	WLS
Publication bias	0.674*	0.691**	0.666*	0.865**	0.707***
	(0.357)	(0.318)	(0.348)	(0.402)	(0.199)
Effect beyond bias	0.003	0.002	0.003	0.003	0.002
	(0.002)	(0.004)	(0.002)	(0.003)	(0.001)
Observations	1620	1620	1620	1620	1620
Panel B: Non-linear models	WAAP	STEM	EK	AK	*p*-uniform*
Publication bias	.	.	0.598***	.	L=21.879***
	(.)	(.)	(0.069)	(.)	(0.000)
Effect beyond bias	0.004***	0.016**	0.003***	0.006***	0.013***
	(0.000)	(0.004)	(0.000)	(0.001)	(0.001)
Observations	1620	1620	1620	1620	1620

Note: Panel A presents the results from the regression $z_{ij} = \alpha + \gamma SE(z_{ij}) + \epsilon_{ij}$, with z_{ij} as the i-th Fisher's z of the effect of school-meal programmes on children's outcomes from the j-th study and $SE(z_{ij})$ as its standard error. FE denotes study-level fixed effects, MAIVE refers to meta-analysis instrumental variable estimator (Irsova et al., 2023), wNOBS assigns weight to each observation based on the inverse number of estimates reported per study, WLS assigns weight to each observation based on the inverse of the standard error. In Panel B, WAAP stands for weighted average of adequately powered (Ioannidis et al., 2017), STEM refers to the stem-based technique (Furukawa, 2021), EK denotes the endogenous kink model (Bom and Rachinger, 2019); AK indicates the selection model by Andrews and Kasy (2019) and *p*-uniform* refers to the selection model by van Aert and van Assen (2023). The effect-beyond-bias estimates are reported in terms of Cohen's d. Extreme outliers are winsorised at the 1st and 99th percentiles. Standard errors are clustered at the study level. *** significant at 1%, ** at 5% and * at 10%.

evidence in our context. These models enable researchers to estimate the potential impact of missing estimates, had they been published in the absence of publication bias and therefore included in the meta-analysis (Irsova et al., 2024; Mathur and VanderWeele, 2020). In order to account for such a possibility, in the last two columns of Panel B, we apply two different selection models. First, we use the selection model developed by Andrews and Kasy (2019), which calculates the probability of publication based on how likely it is that an estimate falls within different intervals, determined by the critical values of t-statistics. Then, the model weights each estimate by the probability of its being published, assigning more weight to those estimates with a lower chance of being published. Second, we apply the p-uniform* model of van Aert and van Assen (2023). This model examines the distribution of p-values around the 5% threshold. If there is an over-representation of p-values below this threshold and an under-representation above it, that indicates publication bias. The model corrects for this bias by assigning different weights to estimates, based on their probability of publication. The results from both the Andrews and Kasy (2019) and the p-uniform* models are consistently in line with those from the non-linear, funnel-based models. In the case of the latter, the effect beyond bias is similar to the unconditional sample mean of 0.014 standard deviation units — indicating, once more, that even when estimates attain statistical significance, the overall impact of school-meal programmes in developed economies is not economically meaningful.

Table 4 explores whether publication bias differs across the three outcome dimensions considered in the literature: behaviour, health and education. Regarding behaviour, with the sole exception of the EK model, tests indicate that publication bias is not of concern. The same holds true for health, except in the OLS, the fixed effects and MAIVE models; in these cases, nonetheless, selective reporting is smaller than that computed for the full sample. The educational domain appears to be the primary driver of the overall publication bias found in Table 4, yet it can still be classified as 'little to modest' (Doucouliagos and Stanley, 2013). All in all, the effects adjusted for publication bias suggest that school-meal programmes in developed economies have minimal impact on student outcomes, regardless of the dimension analysed. Although the results for behaviour — and those from non-linear models in the educational domain — do attain statistical significance, the coefficients are so small as to represent no meaningful improvement in these areas.

In both this and the previous section, we have analysed the impact of school-meal programmes on child outcomes measured in different ways by computing a standardised artificial measure. First, we converted the effect sizes reported in the primary studies into PCCs. Next, we transformed the PCCs into Fisher's z, conducted publication bias tests, and presented the bias-corrected effect size estimates in terms of Cohen's d. However, a total of 811 estimates from 16 studies report results using the same metric, allowing analysis without the need to undertake such conversion. This applies to studies examining the impact of school-meal programmes on educational achievement measured in terms of test scores, where results are normally standardised at the academic year or the course-academic year level (see, for instance, Ayllón and Lado, 2025; Anderson et al., 2018; Corcoran et al., 2016). Figure C.2 in Online Appendix C presents the histogram for this subset of studies and shows that most of the estimates are between 0 and 0.10 standard deviations. The distribution of effect sizes has larger tails, skewed to the right, and fewer negative estimates than in the literature as a whole — shown in Figure 1. The unconditional sample mean is 0.052 standard deviations, suggesting that school-meal programmes have a positive medium effect on test scores (Kraft, 2020). Importantly, these findings may be shaped by publication bias. To consider such a possibility, we computed the funnel plot for these estimates — Figure C.3 in Online Appendix C — and performed all the same publication bias tests as before — detailed in Table 5. The results provide evidence of the presence of publication bias in this part of the literature, which is much more pronounced than for the full sample. After adjusting for publication bias, the main effect is reduced to zero, with eight out of the ten tests confirming the same result. According to the results of this meta-analysis, on average school-meal programmes do not improve children's test scores.

Taken together, our findings indicate that publication bias in the overall literature on the impact of school-meal programmes on children's outcomes is not of concern — with the exception of the group of studies analysing standardised test scores. Once we account for publication bias, the majority of tests indicate that the overall causal effect of school-meal programmes on behaviour, health and educational outcomes is null or so small as not to translate into any meaningful impact. This aligns with the mixed findings reported in the literature and the lack of consensus so far. It also underscores the importance of finding and understanding the features that may be key in the design of successful school-meal programmes. This is precisely what we do in the next section.

Table 5: Publication bias tests for standardised test scores

Panel A: Linear models					
	OLS	FE	MAIVE	wNOBS	WLS
	(1)	(2)	(3)	(4)	(5)
Publication bias	0.8585***	0.7839***	0.6453**	0.8539***	0.8657***
	(0.1238)	(0.1675)	(0.3038)	(0.1972)	(0.1289)
Effect beyond bias	0.0023	0.0066	0.0146	0.0049	0.0019
	(0.0048)	(0.0097)	(0.0194)	(0.0067)	(0.0040)
Observations	811	811	811	811	811
Panel B: Non-linear models					
	WAAP	STEM	EK	AK	p-uniform*
	(6)	(7)	(8)	(9)	(10)
Publication bias	.	.	0.977***	.	L=35.508***
	(.)	(.)	(0.082)	(.)	(0.000)
Effect beyond bias	-0.003	-0.002	-0.001	0.022***	0.035***
	(0.003)	(0.012)	(0.001)	(0.005)	(0.006)
Observations	811	811	811	811	811

Note: Panel A presents the results from the regression $e_{ij} = \alpha + \gamma SE(e_{ij}) + \epsilon_{ij}$, with e_{ij} as the i-th estimate of the effect of school-meal programmes on children's standardised test scores from the j-th study and $SE(e_{ij})$ as its standard error. FE denotes study-level fixed effects, MAIVE refers to meta-analysis instrumental variable estimator (Irsova et al., 2023), wNOBS assigns weight to each observation based on the inverse number of estimates reported per study, WLS assigns weight to each observation based on the inverse of the standard error. In Panel B, WAAP stands for weighted average of adequately powered (Ioannidis et al., 2017), STEM refers to the stem-based technique (Furukawa, 2021), EK denotes the endogenous kink model (Bom and Rachinger, 2019); AK indicates the selection model by Andrews and Kasy (2019) and p-uniform* refers to the selection model by van Aert and van Assen (2023). Extreme outliers are winsorised at the 1st and 99th percentiles. Standard errors are clustered at the study level. *** significant at 1%, ** at 5% and * at 10%.

4 Heterogeneity

In the previous section, we provided evidence that the literature on school-meal programmes suffers from small publication bias, and that the effect of such programmes on children's development is minimal in rich economies. However, as described in Section 2, the estimates in primary studies vary considerably, as they originate from research conducted in diverse settings, examine different outcomes, analyse various school-meal programmes, employ different datasets and methodologies, and focus on specific subgroups of children. Consequently, such variability across studies can lead to systematically different results.

To address this issue, we collect 59 variables that capture the context in which each estimate was obtained and perform multiple meta-regression analysis (MRA). The variables are selected on the basis of a thorough literature review (outlined in Online Appendix B) to cover the principal differences between studies. We categorise these variables into six groups. First, we gather data on the outcome analysed, including behaviour, health and education. Second, we include variables that reflect the characteristics of the programme: whether it provides only breakfast, only lunch or both; whether it is universal (available to all children regardless of family income) or means-tested; the educational stage of the children under examination (elementary, middle or high school); the country where the analysis was conducted; and whether the study covers the entire country or only a specific state or city. Third, we consider the student's socio-demographic characteristics, including gender, ethnicity and socio-economic status. Fourth, we assemble information on the characteristics of the data: whether it is longitudinal or cross-sectional, administrative or survey-based, and the level at which it is measured (individual, school or district), as well as the year mean. Fifth, we collect estimation characteristics, including the methodology used (DiD, RDD, IV or other methodologies), whether the authors treat the estimate as a main result or a robustness check, type of controls, and whether the study incorporates fixed effects and, if so, which type. Lastly, we gather the following information for each study: the publication year; whether it was published in an academic journal or as a working paper; the impact factor of the journal; whether or not the journal is in the field of economics; and the number of citations it has received according to Google Scholar.[12]

Table 6 summarises the variables capturing heterogeneity between studies.[13] More than half of the estimates relate to educational outcomes; health outcomes account for 33.4% of the total estimates; while behavioural outcomes comprise less than 10%. Regarding programme characteristics, a significant proportion of the programmes analysed cover lunch (44.8%) and are universal (61.5%). Most studies investigate the impact of school-meal programmes on the entire elementary school student population (93.3%), while 60.3% focus on middle schools and 21.7% on high schools. Notably, 70.3% of the estimates are from the US and 61.6% involve a specific state or city. Concerning student variation, 12% of the estimates focus on either boys or girls, 4.7% on those from a minority group and almost 11% on those from a disadvantaged socio-economic background. In contrast, 62.4% of estimates refer to the entire student body. Nearly all estimates rely on longitudinal data (92.9%), come from administrative sources (71.2%) and are at the

[12]This information was collected in January 2025.

[13]Certain variables (e.g. educational stage) do not total 100% because studies often analyse outcomes across multiple student groups, considering characteristics such as gender, ethnicity and socio-economic status, along with the inclusion of controls and fixed effects. For example, a study might conduct a subgroup analysis of Hispanic girls from socio-economically disadvantaged backgrounds in elementary and middle schools. These categories overlap rather than being mutually exclusive.

student level (80%). The data year mean is 2006. As for method and estimation characteristics, DiD analyses are the most prevalent (72.7%), with more than half of the estimates controlling for gender, minority and disadvantaged status. Regarding publication characteristics, 60.1% of the estimates come from a study published in a peer-reviewed journal with a mean impact factor of 1.5. On average, the studies have approximately 72 citations and the publication year mean is 2017.

To account for heterogeneity between studies, we incorporate these 59 variables into the following multiple MRA model:

$$z_{ij} = \gamma_0 + \gamma_1 SE(z_{ij}) + \gamma_2 X_{ij} + \epsilon_{ij} \tag{6}$$

where, as in Equation (5), z_{ij} is the i-th Fisher's z of the effect of school-meal programmes on children's outcomes from the j-th study, $SE(z_{ij})$ denotes the corresponding standard error and ϵ_{ij} is the error term. Importantly, X_{ij} represents the set of variables that capture heterogeneity between studies, which allows us to identify the most relevant characteristics in explaining differences among reported estimates, while assessing the robustness of the relationship between estimates and standard errors (Matousek et al., 2022).

Including all 59 variables in multiple MRA can significantly reduce the precision of the overall estimation, due to collinearity.[14] Moreover, prior to analysis there is often uncertainty about which explanatory variables truly belong in the underlying model. To account for these problems, we use BMA, which addresses model uncertainty by conducting numerous regressions that include all possible combinations of explanatory variables (Hoeting et al., 1999; Raftery et al., 1997). It then assigns weights to each model based on its posterior model probability, which reflects its goodness of fit. Using these weights, it calculates the posterior inclusion probability (PIP) for each covariate. This measure can be interpreted as the probability that a given variable is a useful predictor. When implementing BMA, we follow previous applied meta-analyses (see, for example, Opatrny et al., 2025; Yang et al., 2024; Kroupova et al., 2024) and employ the unit information prior (UIP) and the dilution model prior (George, 2010).[15] We also use the Markov chain Monte Carlo algorithm, which focuses solely on models with high posterior probabilities. This approach ensures computational feasibility, as it is impractical to estimate all possible model combinations (2^{59}).

In Figure 5 we illustrate the results from the baseline BMA model. The vertical axis lists the explanatory variables ranked by their PIP, with the highest at the top and the lowest at the bottom. The horizontal axis displays the cumulative posterior model probability, sorted from left to right, based on decreasing probability. Blue colour (darker in greyscale) denotes a positive impact of the variable on reported estimates, while red (lighter in greyscale) suggests a negative effect. Conversely, white indicates that this particular variable is not included in the model. As shown in the plot, publication bias ($SE(z)$), school-breakfast initiatives, means-tested programmes, a focus on high-school students, nationwide analyses, administrative data, units of observation at the school level, inclusion of school and individual fixed effects, controlling for minority status, publication year and number of citations are all positively associated with the reported

[14]The inclusion of variables with low variance can lead to increased collinearity and volatile results. However, in our case none of the variables show such low variance (Irsova et al., 2024).

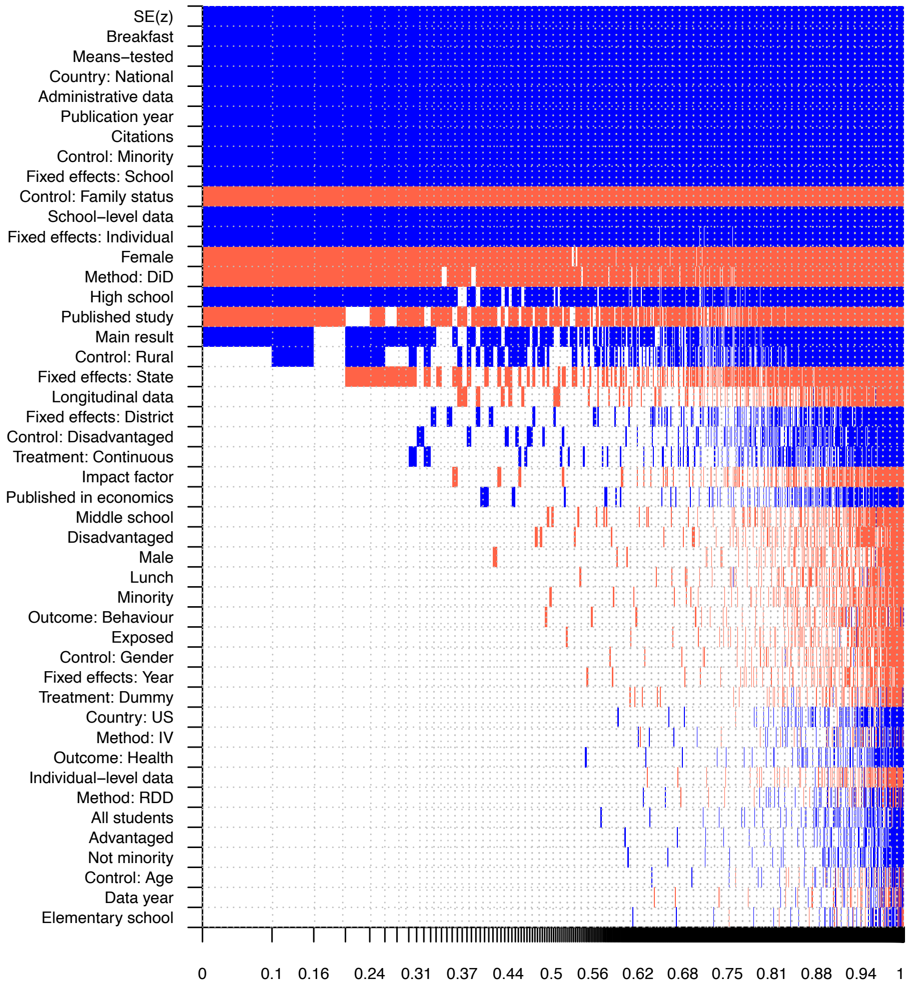
[15]The unit information prior provides the same amount of information as one observation of data (Eicher et al., 2011). The dilution model prior penalises models with significant collinearity, although in our case the correlation between individual variables is not substantial (George, 2010) — see Figure C.4 in Online Appendix C.

Table 6: Summary statistics of variables reflecting heterogeneity

	Mean (1)	Std. Dev (2)	Weighted mean (3)
Outcome analysed			
Outcome: Behaviour	0.091	0.288	0.166
Outcome: Health	0.334	0.472	0.299
Outcome: Education (ref.)	0.574	0.495	0.535
Programme variation			
Lunch	0.448	0.497	0.405
Breakfast	0.328	0.470	0.331
Breakfast and lunch (ref.)	0.224	0.417	0.264
Means-tested	0.385	0.487	0.287
Universal (ref.)	0.615	0.487	0.713
Elementary school	0.933	0.249	0.881
Middle school	0.603	0.489	0.593
High school	0.217	0.412	0.334
Country: US	0.703	0.457	0.713
Country: Other (ref.)	0.297	0.457	0.287
Country: National	0.384	0.487	0.359
Country: Regional (ref.)	0.616	0.487	0.641
Student variation			
Female	0.057	0.231	0.050
Male	0.060	0.237	0.055
Not minority	0.032	0.176	0.028
Minority	0.047	0.212	0.043
Advantaged	0.086	0.280	0.048
Disadvantaged	0.109	0.312	0.073
All students	0.624	0.484	0.727
Type of data			
Longitudinal data	0.929	0.257	0.928
Cross-sectional data (ref.)	0.071	0.257	0.072
Administrative data	0.712	0.453	0.782
Survey data (ref.)	0.288	0.453	0.218
Individual-level data	0.804	0.397	0.690
School-level data	0.158	0.365	0.260
District-level data (ref.)	0.037	0.188	0.048
Data year	2006.121	15.284	2006.217
Method and estimation characteristics			
Method: DiD	0.727	0.446	0.773
Method: RDD	0.135	0.341	0.097
Method: IV	0.073	0.260	0.053
Method: Other (ref.)	0.065	0.247	0.077
Main result	0.805	0.397	0.851
Robustness (ref.)	0.195	0.397	0.149
Exposed	0.093	0.290	0.169
Not exposed (ref.)	0.907	0.290	0.831
Treatment: Dummy	0.837	0.369	0.830
Treatment: Continuous	0.110	0.312	0.113
Treatment: Categorical (ref.)	0.055	0.227	0.065
Control: Gender	0.625	0.484	0.533
Control: Age	0.367	0.482	0.288
Control: Minority	0.722	0.448	0.649
Control: Disadvantaged	0.797	0.402	0.755
Control: Family status	0.138	0.345	0.086
Control: Rural	0.245	0.430	0.166
Fixed effects: Individual	0.159	0.366	0.159
Fixed effects: School	0.388	0.487	0.425
Fixed effects: District	0.107	0.309	0.162
Fixed effects: State	0.096	0.295	0.067
Fixed effects: Year	0.676	0.468	0.729
Publication characteristics			
Publication year	2017.572	5.157	2018.055
Impact factor	1.487	1.587	1.887
Citations	71.582	92.038	87.420
Published study	0.606	0.489	0.737
Working paper (ref.)	0.394	0.489	0.263
Published in economics	0.508	0.500	0.522
Not published in economics (ref.)	0.492	0.500	0.478

Note: The table presents the summary statistics of the 59 variables that capture the context in which each estimate was obtained. The total number of observations is 2,821. The notation '(ref.)' indicates the reference category for each variable included in the BMA analysis. In Column (3), we weight each observation by the inverse number of estimates reported per study. Extreme outliers are winsorised at the 1st and 99th percentiles. See Table B.1 in Online Appendix B for a detailed description of the variables.

Figure 5: Model inclusion in BMA



Note: The figure presents the results of the baseline BMA using the unit information prior and the dilution model prior (George, 2010). The vertical axis lists the explanatory variables ranked by their posterior inclusion probability, while the horizontal axis displays the cumulative posterior model probability, sorted from left to right, based on decreasing probability. Blue colour (darker in greyscale) denotes a positive impact of the variable on reported estimates, while red (lighter in greyscale) suggests a negative effect. Extreme outliers are winsorised at the 1st and 99th percentiles. See Table B.1 in Online Appendix B for a detailed description of the variables.

estimates. In contrast, conducting sub-analyses for girls, employing DiD, controlling for family status, and publication in academic journals are negatively associated with the reported estimates.

Table 7 presents the numerical results of the BMA analysis.[16] The first column displays the posterior means in terms of Cohen's d, which can be interpreted as the marginal

[16] In Table C.2 in Appendix C, we also report the BMA results for the subset of studies that examine the impact of school-meal programmes on standardised test scores.

Table 7: BMA results

	BMA			OLS		
	P. mean	P. Std. Dev	PIP	Coef.	SE	*p*-value
	(1)	(2)	(3)	(4)	(5)	(6)
SE(z)	1.3308	0.0358	1.0000	1.3390	0.0328	0.0000
Outcome analysed						
Outcome: Behaviour	-0.0001	0.0005	0.0335			
Outcome: Health	0.0001	0.0002	0.0219			
Programme variation						
Lunch	-0.0002	0.0006	0.0365			
Breakfast	0.0291	0.0013	1.0000	0.0292	0.0011	0.0000
Means-tested	0.0255	0.0017	1.0000	0.0251	0.0012	0.0000
Elementary school	0.0000	0.0002	0.0118			
Middle school	-0.0002	0.0006	0.0467			
High school	0.0084	0.0019	0.8876	0.0091	0.0011	0.0000
Country: US	0.0001	0.0005	0.0297			
Country: National	0.0214	0.0017	1.0000	0.0216	0.0012	0.0000
Student variation						
Female	-0.0140	0.0022	0.9681	-0.0160	0.0017	0.0000
Male	-0.0002	0.0006	0.0348			
Not minority	0.0000	0.0003	0.0150			
Minority	-0.0002	0.0006	0.0329			
Advantaged	0.0000	0.0002	0.0142			
Disadvantaged	-0.0002	0.0005	0.0426			
All students	0.0000	0.0002	0.0186			
Type of data						
Longitudinal data	-0.0025	0.0034	0.1402			
Administrative data	0.0154	0.0014	1.0000	0.0143	0.0012	0.0000
Individual-level data	-0.0002	0.0010	0.0280			
School-level data	0.0191	0.0018	0.9886	0.0201	0.0013	0.0000
Data year	-0.0000	0.0000	0.0164			
Method and estimation characteristics						
Method: DiD	-0.0108	0.0026	0.9121	-0.0093	0.0012	0.0001
Method: RDD	0.0001	0.0006	0.0248			
Method: IV	0.0001	0.0006	0.0244			
Main result	0.0059	0.0021	0.7283	0.0073	0.0011	0.0011
Exposed	-0.0002	0.0007	0.0350			
Treatment: Dummy	-0.0001	0.0003	0.0272			
Treatment: Continuous	0.0007	0.0012	0.1092			
Control: Gender	-0.0001	0.0004	0.0288			
Control: Age	0.0000	0.0002	0.0160			
Control: Minority	0.0167	0.0013	0.9997	0.0179	0.0010	0.0000
Control: Disadvantaged	0.0008	0.0012	0.1182			
Control: Family status	-0.0224	0.0024	0.9977	-0.0233	0.0018	0.0000
Control: Rural	0.0036	0.0021	0.4667			
Fixed effects: Individual	0.0143	0.0021	0.9828	0.0135	0.0014	0.0000
Fixed effects: School	0.0138	0.0016	0.9980	0.0129	0.0013	0.0000
Fixed effects: District	0.0012	0.0017	0.1387			
Fixed effects: State	-0.0055	0.0037	0.4000			
Fixed effects: Year	-0.0001	0.0004	0.0272			
Publication characteristics						
Publication year	0.0046	0.0002	1.0000	0.0045	0.0002	0.0000
Impact factor	-0.0001	0.0003	0.0767			
Citations	0.0002	0.0000	1.0000	0.0002	0.0000	0.0000
Published study	-0.0092	0.0029	0.7830	-0.0111	0.0014	0.0001
Published in economics	0.0007	0.0013	0.0836			

Note: Columns (1) to (3) provide the results of the baseline BMA using the unit information prior and the dilution model prior (George, 2010). Columns (4) to (6) report the results of an OLS model, focusing on variables with a posterior inclusion probability exceeding 0.5 in the BMA. 'P. mean' denotes posterior mean, 'P. Std. Dev' details the posterior standard deviation and 'PIP' refers to posterior inclusion probability. The posterior means and OLS coefficients are reported in terms of Cohen's *d*. Extreme outliers are winsorised at the 1st and 99th percentiles. See Table B.1 in Online Appendix B for a detailed description of the variables.

effects of these variables on the reported estimates. The second column presents the posterior standard deviations and the third column reports the PIP for each variable in the model. To identify the relevant variables, we follow the convention established by Jeffreys (1961). According to this classification, PIP values of 0.5—0.75 indicate a weak effect, 0.75—0.95 suggest a moderate effect, 0.95—0.99 represent a strong effect and values from 0.99 to 1 denote a decisive effect. Among all the variables considered, only 17 have PIP values above 0.5. The BMA results indicate that, relative to education, whether the study analyses behaviour or health outcomes is not systematically linked to higher estimates. In terms of programme variation, initiatives that provide only breakfast yield estimates that are more positive. This is consistent with the bulk of evidence indicating that school-breakfast programmes make students less likely to skip the meal (Bartfeld et al., 2009), which, in turn, improves their outcomes. Breakfast is often regarded as the 'most important meal of the day' (Martin et al., 2024). Additionally, means-tested programmes tend to have a greater positive effect than universal ones, which contradicts evidence highlighting the benefits of universal school-meal provision, such as reduced stigma, increased take-up and lower administrative costs (Altindag et al., 2020). Although the effect is relatively small, studies focusing on high-school students also report estimates that are more favourable. Regarding student characteristics, analyses restricted to female students generally show negative effects, while those for males do not report statistically significant results. These gender differences confirm that girls and boys respond differently to a school-meal intervention that involves changes in diet and social environment (Sørensen et al., 2016). Neither ethnic minority status nor socio-economic advantage appears to influence the overall effect. Studies that use administrative and school-level data are more likely to produce positive estimates. Regarding method and estimation characteristics, DiD analyses — commonly used in studies that analyse a transition from a means-tested to a universal free school-meal programme — are linked to poorer student outcomes. This is in line with the fact that means-tested programmes are more effective than universal ones. Estimates identified by the authors as 'main results' are positively associated with reported estimates, as opposed to 'robustness checks', which may account for part of the publication bias (Opatrny et al., 2025). However, this effect is small. Controlling for minority status and incorporating individual and school fixed effects also tends to produce more positive findings. Recent studies with more citations report positive estimates, while published studies in peer-reviewed journals do not, although the posterior means are practically zero. After accounting for heterogeneity between studies, we keep confirming small publication bias in this strand of the literature.

We conduct three robustness checks on the results that consider the whole sample. First, in Columns (4) to (6) of Table 7, we present the results from an OLS model that considers only variables with a PIP exceeding 0.5 in the BMA (Opatrny et al., 2025; Kroupova et al., 2024). Second, in Figure C.5 in Online Appendix C, we test the sensitivity of our baseline BMA to different priors. On the vertical axis we plot the PIP and on the horizontal axis we list the explanatory variables, sorted by their PIP. In addition to our baseline BMA model with the UIP prior and the dilution model prior (George, 2010), we present the results of a BMA using the UIP prior and the uniform model prior (Eicher et al., 2011), as well as the Bayesian risk information criterion (BRIC) prior and the random model prior (Ley and Steel, 2009; Fernández et al., 2001).[17] Third, in Table C.3

[17]The uniform model prior assigns equal probability to all models (Eicher et al., 2011). The BRIC prior penalises model complexity by applying a penalty to models with more explanatory variables. The random prior introduces uncertainty regarding the inclusion probability of each variable, allowing the

in Online Appendix C, we report findings using frequentist model averaging (FMA). Like BMA, FMA addresses model uncertainty, but does not require prior specification (Wang et al., 2009).[18] All three tests produce results that are consistent with our baseline BMA findings.

Overall, we observe little heterogeneity in the literature on the impact of school-meal programmes on children's outcomes. Most of the characteristics considered have an economically negligible effect on the reported estimates. Even after accounting for these factors, bias in favour of positive estimates is small.

5 Concluding remarks

This paper studies the extent to which school-meal programmes have a causal impact on children's outcomes in rich economies and, in particular, on education, health and behaviour. We use novel meta-analytical techniques to account for publication bias and model uncertainty in this strand of the literature. Drawing on 2,821 estimates from 42 studies and 150,000+ data points, we document the fact that publication bias is small and is mostly concentrated among studies that focus on the analysis of test scores. Once estimates are adjusted for publication bias, the overall causal effect of school-meal programmes on child development is minimal. However, our heterogeneity analysis indicates that students participating in means-tested programmes and those receiving breakfast tend to experience greater benefits than those in universal programmes and/or who are served only lunch.

Does this suggest that school-meal programmes are not worth the investment and should be withdrawn? First, our results do not imply that school meals should be removed, as we focus primarily on the expansion or improvement of a programme that was already in place. Thus, we lack evidence on what would happen if school meals were removed. The limited evidence available suggests a negative effect on mental health (Bethmann and Cho, 2022) and obesity (Maruyama and Nakamura, 2025). Second, the benefits of school-meal programmes may extend beyond immediate educational or health outcomes and operate on a medium- to long-term basis — a time frame that we could not consider in this meta-analysis, as there is an insufficient number of studies on this. Additionally, school-meal programmes may impact children through other channels, such as household finances, parental employment, work-family balance and enhanced social skills facilitated through socialisation. Third, the effectiveness of school-meal programmes may depend on their quality, the nutritional content of the food provided, the opportunities to socialise and communicate with peers, the inclusion of extracurricular activities and/or the effectiveness of mealtime supervision. These may be important features, but could not be considered, as such information is rarely collected. Lastly, it is important to note that our findings are specific to developed countries and may not be generalised to developing economies, where school-meal programmes serve other objectives, such as combating malnutrition and improving school attendance.

This strand of literature needs more research. Several issues should be addressed, such as the impact of school-meal programmes on other child outcomes, beyond those considered in this study, as well as their long-term effects. It will only be possible to reach

data to inform the model size (Ley and Steel, 2009).

[18]FMA estimators depend on the weights assigned to different models. To select these model weights, we minimise Mallow's criterion. This approach balances the trade-off between model complexity and goodness of fit (Hansen, 2007).

a conclusion on these other aspects when we have more studies in this area. Perhaps then we will be able to understand better the key contextual factors and features that ensure the success of a school-meal programme. This is essential if we are to provide evidence that can help policy-makers design policies that can improve children's well-being and reduce socio-economic inequalities.

References

Abouk, R., and S. Adams (2022): "Breakfast After the Bell: The effects of expanding access to school breakfasts on the weight and achievement of elementary school children," *Economics of Education Review*, 87, 102224.

Altindag, D. T., D. Baek, H. Lee, and J. Merkle (2020): "Free lunch for all? The impact of universal school lunch on student misbehavior," *Economics of Education Review*, 74, 101945.

Anderson, M. L., J. Gallagher, and E. Ramirez Ritchie (2018): "School meal quality and academic performance," *Journal of Public Economics*, 168, 81–93.

Andrews, I., and M. Kasy (2019): "Identification of and correction for publication bias," *American Economic Review*, 109(8), 2766–2794.

Ayllón, S., and S. Lado (2025): "More than just lunch: School-meal subsidies and language proficiency," IZA Discussion Papers No. 17631.

Bartfeld, J., M. Kim, J. H. Ryu, and H.-M. Ahn (2009): "The School Breakfast Program: Participation and impacts," Economic Research Service, US Department of Agriculture.

Bartfeld, J. S., L. Berger, and F. Men (2020): "Universal access to free school meals through the Community Eligibility Provision is associated with better attendance for low-income elementary school students in Wisconsin," *Journal of the Academy of Nutrition and Dietetics*, 120(2), 210–218.

Belot, M., and J. James (2011): "Healthy school meals and educational outcomes," *Journal of Health Economics*, 30(3), 489–504.

Bethmann, D., and J. I. Cho (2022): "The impacts of free school lunch policies on adolescent BMI and mental health: Evidence from a natural experiment in South Korea," *SSM — Population Health*, 18, 101072.

Bhattacharya, J., J. Currie, and S. J. Haider (2006): "Breakfast of champions? The School Breakfast Program and the nutrition of children and families," *Journal of Human Resources*, 41(3), 445–466.

Bitler, M. P., and A. Seifoddini (2019): "Health impacts of food assistance: Evidence from the United States," *Annual Review of Resource Economics*, 11, 261–287.

Bom, P. R. D., and H. Rachinger (2019): "A kinked meta-regression model for publication bias correction," *Research Synthesis Methods*, 10(4), 497–514.

Borbely, D., M. Gehrsitz, S. McIntyre, and G. Rossi (2024): "Does the provision of universal free school meals improve school attendance?," *Economics of Education Review*, 103, 102597.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein (2009): *Introduction to Meta-Analysis*. John Wiley & Sons.

Brodeur, A., N. Cook, and A. Heyes (2020): "Methods matter: p-hacking and publication bias in causal analysis in economics," *American Economic Review*, 110(11), 3634–3660.

Brown, V., C. Crawford, L. Dearden, E. Greaves, S. Kitchen, C. Payne, S. Purdon, and E. Tanner (2012): "Evaluation of the free school meals pilot: Impact Report," Department for Education Research Report DFE-RR227.

Bütikofer, A., E. Mølland, and K. G. Salvanes (2018): "Childhood nutrition and labor market outcomes: Evidence from a school breakfast program," *Journal of Public Economics*, 168, 62–80.

Capogrossi, K. L. (2012): "Childhood misnourishment, school meal programs and academic performance," Ph.D. thesis, Virginia Polytechnic Institute and State University.

Card, D., and A. B. Krueger (1995): "Time-series minimum-wage studies: A meta-analysis," *American Economic Review*, 85(2), 238–243.

Clarke, B. (2019): "Literature review: The evidence decision-makers want," Center for the Study of Social Policy.

Collante Zárate, S., C. Rodríguez Orgales, and F. Sánchez (2024): "The power of a meal. School feeding and its educational effects: Evidence from Colombia," Documento CEDE No. 24.

Corcoran, S. P., B. Elbel, and A. E. Schwartz (2016): "The effect of Breakfast In the Classroom on obesity and academic performance: Evidence from New York City," *Journal of Policy Analysis and Management*, 35, 509–532.

Cuadros-Meñaca, A., M. R. Thomsen, and R. M. Nayga Jr (2022): "The effect of breakfast after the bell on student academic achievement," *Economics of Education Review*, 86, 102223.

——— (2023): "School breakfast and student behavior," *American Journal of Agricultural Economics*, 105(1), 99–121.

Davis, W. (2019): "Should kids have their lunch and eat it too? Estimating the effect of universal free school meals on child weight," available at: `https://static1.squarespace.com/static/5ad910ec365f02f74f353357/t/5e1cf93e5ca3535f457303b9/1578957130509/jmp_Will_Davis_new.pdf`.

Davis, W., D. Kreisman, and T. Musaddiq (2024): "The effect of universal free school meals on child BMI," *Education Finance and Policy*, 19(3), 461–491.

Davis, W., and T. Musaddiq (2019): "Estimating the effects of universal free school meal enrollment on child health: Evidence from the Community Eligibility Provision in Georgia schools," SSRN Electronic Journal.

Domina, T., L. Clark, V. Radsky, and R. Bhaskar (2024): "There is such a thing as a free lunch: School meals, stigma, and student discipline," *American Educational Research Journal*, 61(2), 287–327.

DOTTER, D. D. (2013): "Breakfast at the desk: The impact of universal breakfast programs on academic performance," unpublished Working Paper.

DOUCOULIAGOS, C., AND T. STANLEY (2013): "Are all economic facts greatly exaggerated? Theory competition and selectivity," *Journal of Economic Surveys*, 27(2), 316–339.

DUNIFON, R., AND L. KOWALESKI-JONES (2003): "The influences of participation in the National School Lunch Program and food insecurity on child well-being," *Social Service Review*, 77(1), 72–92.

EGGER, M., G. D. SMITH, M. SCHNEIDER, AND C. MINDER (1997): "Bias in meta-analysis detected by a simple, graphical test," *British Medical Journal*, 315(7109), 629–634.

EICHER, T. S., C. PAPAGEORGIOU, AND A. E. RAFTERY (2011): "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants," *Journal of Applied Econometrics*, 26(1), 30–55.

FERNÁNDEZ, C., E. LEY, AND M. F. STEEL (2001): "Benchmark priors for Bayesian model averaging," *Journal of Econometrics*, 100(2), 381–427.

FIGLIO, D. N., AND J. WINICKI (2005): "Food for thought: The effects of school accountability plans on school nutrition," *Journal of Public Economics*, 89(2), 381–394.

FRISVOLD, D. E. (2015): "Nutrition and cognitive achievement: An evaluation of the School Breakfast Program," *Journal of Public Economics*, 124, 91–104.

FURUKAWA, C. (2021): "Publication bias under aggregation frictions: From communication model to new correction method," MIT Working Paper.

GEORGE, E. I. (2010): "Dilution priors: Compensating for model space redundancy," in *Borrowing Strength: Theory Powering Applications — A Festschrift for Lawrence D. Brown*, vol. 6, pp. 158–166. Institute of Mathematical Statistics.

GORDANIER, J., O. OZTURK, B. WILLIAMS, AND C. ZHAN (2020): "Free lunch for all! The effect of the Community Eligibility Provision on academic outcomes," *Economics of Education Review*, 77, 101999.

GORDON, N., AND K. RUFFINI (2021): "Schoolwide free meals and student discipline: Effects of the Community Eligibility Provision," *Education Finance and Policy*, 16(3), 418–442.

GUNDERSEN, C., B. KREIDER, AND J. PEPPER (2012): "The impact of the National School Lunch Program on child health: A nonparametric bounds analysis," *Journal of Econometrics*, 166(1), 79–91.

GUTIERREZ, E. (2021): "The effect of universal free meals on student perceptions of school climate: Evidence from New York City," EdWorkingPaper No. 21-430.

HANSEN, B. E. (2007): "Least squares model averaging," *Econometrica*, 75(4), 1175–1189.

Havranek, T., R. Horvath, Z. Irsova, and M. Rusnak (2015): "Cross-country heterogeneity in intertemporal substitution," *Journal of International Economics*, 96(1), 100–118.

Havranek, T., Z. Irsova, L. Laslopova, and O. Zeynalova (2022): "Publication and attenuation biases in measuring skill substitution," *Review of Economics and Statistics*, 106(5), 1187–1200.

Hinrichs, P. (2010): "The effects of the National School Lunch Program on education and health," *Journal of Policy Analysis and Management*, 29(3), 479–505.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian model averaging: A tutorial," *Statistical Science*, 14(4), 382–417.

Holford, A. (2015): "Take-up of free school meals: Price effects and peer effects," *Economica*, 82(328), 976–993.

Holford, A., and B. Rabe (2022): "Going universal. The impact of free school lunches on child body weight outcomes," *Journal of Public Economics Plus*, 3, 100016.

——— (2024): "Universal free school meals and children's bodyweight. Impacts by age and duration of exposure," *Journal of Health Economics*, 98, 102937.

Imberman, S. A., and A. D. Kugler (2014): "The effect of providing breakfast in class on student performance," *Journal of Policy Analysis and Management*, 33(3), 669–699.

Ioannidis, J. P. A., T. D. Stanley, and H. Doucouliagos (2017): "The power of bias in economics research," *Economic Journal*, 127(605), F236–F265.

Irsova, Z., P. R. D. Bom, T. Havranek, and H. Rachinger (2023): "Spurious precision in meta-analysis," IES Working Papers 5/2023.

Irsova, Z., H. Doucouliagos, T. Havranek, and T. D. Stanley (2024): "Meta-analysis of social science research: A practitioner's guide," *Journal of Economic Surveys*, 38(5), 1547–1566.

Jackson, C. K., and C. L. Mackevicius (2024): "What impacts can we expect from school spending policy? Evidence from evaluations in the United States," *American Economic Journal: Applied Economics*, 16(1), 412–446.

Jeffreys, H. (1961): *Theory of Probability.* Oxford Classic Texts in the Physical Sciences. Oxford University Press, third edition.

Kho, A. (2018): "Three essays on school reform," Ph.D. thesis, Vanderbilt University.

Kim, Y. (2021): "The effects of universal free lunch provision on student achievement: Evidence from South Korea," Working Paper, available at: `https://yoonjungkim.io/research/`.

Kirksey, J. J., and M. A. Gottfried (2021): "The effect of serving 'Breakfast After-the-Bell' meals on school absenteeism: Comparing results from Regression Discontinuity Designs," *Educational Evaluation and Policy Analysis*, 43(2), 305–328.

Kraft, M. A. (2020): "Interpreting effect sizes of education interventions," *Educational Researcher*, 49(4), 241–253.

Kroupova, K., T. Havranek, and Z. Irsova (2024): "Student employment and education: A meta-analysis," *Economics of Education Review*, 100, 102539.

Krueger, A. B. (2003): "Economic considerations and class size," *Economic Journal*, 113(485), F34–F63.

Kurtz, M. D., K. S. Conway, and R. D. Mohr (2022): "The academic effects of United States child food assistance programs — At home, school, and in-between," Oxford Research Encyclopedia of Economics and Finance, available at: `https://oxfordre.com/economics/view/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-865`.

Leos-Urbel, J., A. E. Schwartz, M. Weinstein, and S. Corcoran (2013): "Not just for poor kids: The impact of universal free school breakfast on meal participation and student outcomes," *Economics of Education Review*, 36, 88–107.

Ley, E., and M. F. Steel (2009): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regression," *Journal of Applied Econometrics*, 24(4), 651–674.

Lundborg, P., D.-O. Rooth, and J. Alex-Petersen (2022): "Long-term effects of childhood nutrition: Evidence from a school lunch reform," *Review of Economic Studies*, 89(2), 876–908.

Marcus, M., and K. G. Yewell (2022): "The effect of free school meals on household food purchases: Evidence from the Community Eligibility Provision," *Journal of Health Economics*, 84, 102646.

Martin, A. J., K. C. P. Bostwick, E. C. Burns, V. Munro-Smith, T. George, R. Kennett, and J. Pearson (2024): "A healthy breakfast each and every day is important for students' motivation and achievement," *Journal of School Psychology*, 104, 101298.

Maruyama, S., and S. Nakamura (2025): "Wholesome lunch to the whole classroom: Short- and longer-term effects on early teenagers' weight," *Health Economics*, forthcoming.

Mathur, M., and T. VanderWeele (2020): "Sensitivity analysis for publication bias in meta-analyses," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(5), 1091–1119.

Matousek, J., T. Havranek, and Z. Irsova (2022): "Individual discount rates: A meta-analysis of experimental evidence," *Experimental Economics*, 25(1), 318–358.

McEwan, P. J. (2013): "The impact of Chile's school feeding program on education outcomes," *Economics of Education Review*, 32, 122–139.

Millimet, D. L., R. Tchernis, and M. Husain (2010): "School nutrition programs and the incidence of childhood obesity," *Journal of Human Resources*, 45(3), 640–654.

Norwood, B. (2020): "Breakfast of champions: Universal free breakfast and student conflict and test scores in Texas schools," SSRN Electronic Journal.

Opatrny, M., T. Havranek, Z. Irsova, and M. Scasny (2025): "Publication bias and model uncertainty in measuring the effect of class size on achievement," *Journal of Labor Economics*, forthcoming.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997): "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, 92(437), 179–191.

Ribar, D. C., and L. A. Haldeman (2013): "Changes in meal participation, attendance, and test scores associated with the availability of universal free school breakfasts," *Social Service Review*, 87(2), 354–385.

Rothbart, M. W., A. E. Schwartz, and E. Gutierrez (2023): "Paying for free lunch: The impact of CEP universal free meals on revenues, spending, and student health," *Education Finance and Policy*, 18(4), 708–737.

Ruffini, K. (2022): "Universal access to free school meals and student achievement: Evidence from the Community Eligibility Provision," *Journal of Human Resources*, 57(3), 776–820.

Ruffini, K., O. Öztürk, and P. Pekgün (2025): "In-kind government assistance and crowd-out of charitable services: Evidence from free school meals," NBER Working Paper 33562.

Schanzenbach, D. W. (2009): "Do school lunches contribute to childhood obesity?," *Journal of Human Resources*, 44(3), 684–709.

Schanzenbach, D. W., and M. Zaki (2014): "Expanding the School Breakfast Program: Impacts on children's consumption, nutrition and health," NBER Working Paper 20308.

Schwartz, A. E., and M. W. Rothbart (2020): "Let them eat lunch: The impact of universal free meals on student performance," *Journal of Policy Analysis and Management*, 39(2), 376–410.

Spill, M. K., R. Trivedi, R. C. Thoerig, A. A. Balalian, M. B. Schwartz, C. Gundersen, A. Odoms-Young, E. F. Racine, M. J. Foster, J. S. Davis, and A. J. MacFarlane (2024): "Universal free school meals and school and student outcomes: A systematic review," *JAMA Network Open*, 7, e2424082.

Sørensen, L. B., C. T. Damsgaard, R. A. Petersen, S.-M. Dalskov, M. F. Hjorth, C. B. Dyssegaard, N. Egelund, I. Tetens, A. Astrup, L. Lauritzen, and K. F. Michaelsen (2016): "Differences in the effects of school meals on children's cognitive performance according to gender, household education and baseline reading skills," *European Journal of Clinical Nutrition*, 70(10), 1155–1161.

Stanley, T. D. (2005): "Beyond publication bias," *Journal of Economic Surveys*, 19(3), 309–345.

Stanley, T. D., C. Doucouliagos, and T. Havranek (2023): "Reducing the biases of the conventional meta-analysis of correlations," IES Working Paper No. 34/2023.

Stanley, T. D., and H. Doucouliagos (2014): "Meta-regression approximations to reduce publication selection bias," *Research Synthesis Methods*, 5, 60–78.

Stanley, T. D., H. Doucouliagos, M. Maier, and F. Bartoš (2024): "Correcting bias in the meta-analysis of correlations," *Psychological Methods*, forthcoming.

van Aert, R., and M. van Assen (2023): "Correcting for publication bias in a meta-analysis with the p-uniform* method," MetaArXiv Preprints.

von Hinke Kessler Scholder, S. (2013): "School meal crowd out in the 1980s," *Journal of Health Economics*, 32(3), 538–545.

Wang, D., S. Shinde, T. Young, and W. W. Fawzi (2021): "Impacts of school feeding on educational and health outcomes of school-age children and adolescents in low- and middle-income countries: A systematic review and meta-analysis," *Journal of Global Health*, 11, 04051.

Wang, H., X. Zhang, and G. Zou (2009): "Frequentist model averaging estimation: A review," *Journal of Systems Science and Complexity*, 22(4), 732–748.

Yang, F., T. Havranek, Z. Irsova, and J. Novak (2024): "Where have all the alphas gone? A meta-analysis of hedge fund performance," CEPR Discussion Paper 18979.

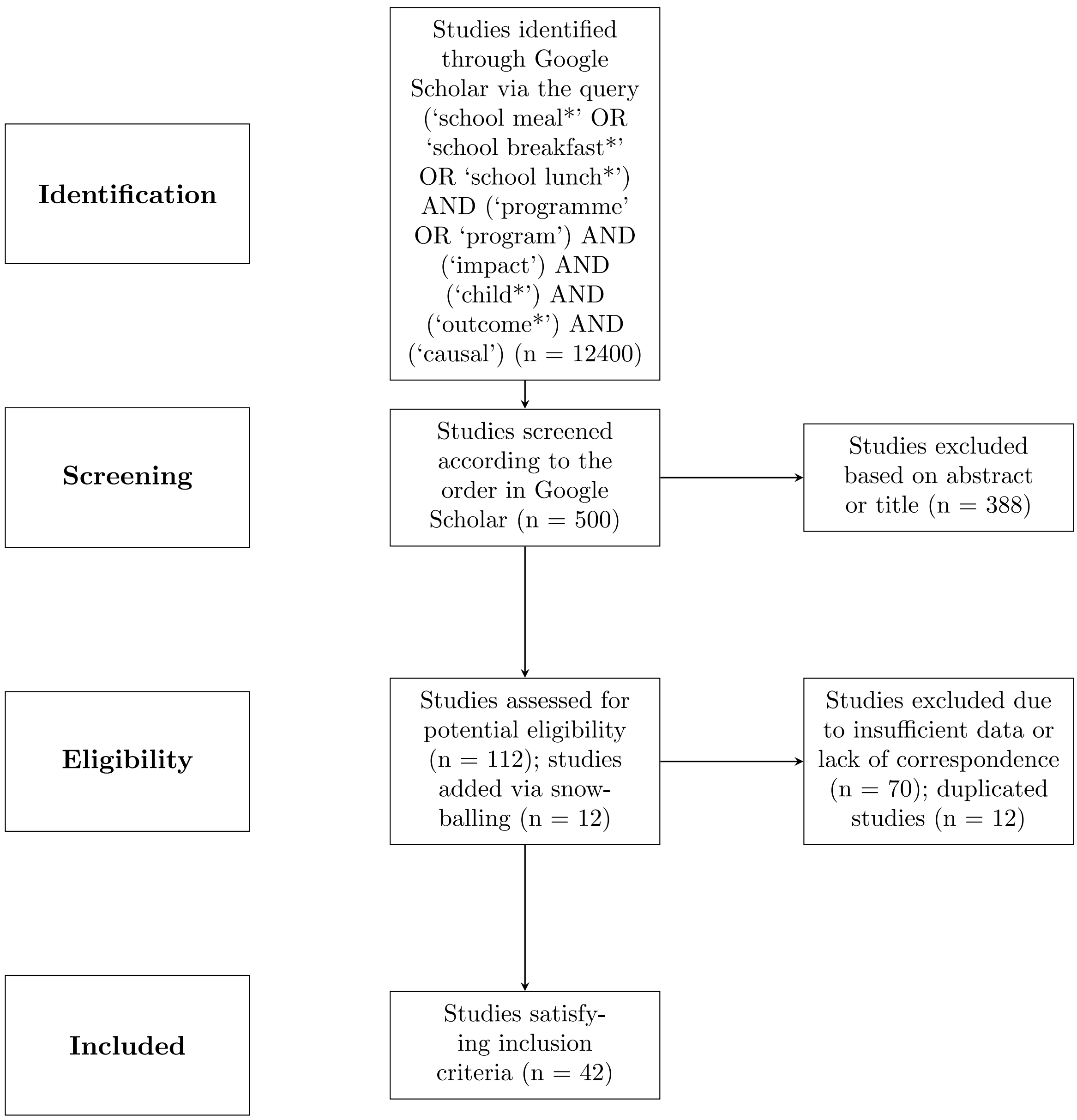
The causal impact of school-meal programmes on children in developed economies: A meta-analysis

Online Appendix

Sara Ayllón and Samuel Lado

A Search and data collection

Figure A.1: PRISMA flow diagram



Note: Asterisks allow searches to include variations of a keyword.

B Details on variables explaining heterogeneity

In this appendix, we describe the variables collected for the meta-analysis, which reflect the key differences between primary studies. These variables were selected following a review of the literature on the impact of school-meal programmes in developed economies. We begin by explaining the main outcomes of interest in the three domains under scrutiny — behaviour, health and education — as well as programme characteristics, such as whether the programme offers breakfast, lunch or both, and the type of provision: means-tested or universal. We also account for student-level variation. We then discuss the variables related to the type of data, methodology, estimation and publication characteristics, whose inclusion is now common practice in recent meta-analysis.

Outcome analysed. The first dimension on which the literature on the impact of school-meal programmes differs is the outcomes analysed. The majority of studies focus on three domains: behaviour, health and education. Regarding behaviour, research examines diverse aspects, including disciplinary infractions, suspension rates, bullying and fights. These studies use the implementation of free meals across schools as a means of identifying variation, and generally document positive impacts (Domina et al., 2024; Cuadros-Meñaca et al., 2023; Gordon and Ruffini, 2021; Gutierrez, 2021; Altindag et al., 2020).

In terms of health, body mass index (BMI) is the primary metric of interest. Nevertheless, changes in BMI are difficult to interpret, especially in school-level studies that compare children of different ages (Davis and Musaddiq, 2019). A change in the mean BMI does not provide information about where in the weight distribution the change is occurring. For instance, obese or underweight children losing weight can have the same impact on the mean BMI, though they have different health implications (Davis and Musaddiq, 2019). Hence, researchers also rely on alternative weight indicators, including underweight, overweight and obesity status. The findings from these studies range from positive (Holford and Rabe, 2022, 2024; Gundersen et al., 2012) to null or negative effects (Schwartz and Rothbart, 2020; Corcoran et al., 2016; Millimet et al., 2010; Schanzenbach, 2009). A minority of studies explore additional health outcomes, such as nutrient intake or health limitations (defined as factors that affect participation in childhood activities, school attendance or the performance of schoolwork).

As for educational outcomes, most of the analysis focuses on attendance and academic achievement. Regarding school attendance (or absenteeism), a few studies report positive impacts (Gordanier et al., 2020; Belot and James, 2011), while the majority indicate null effects (Cuadros-Meñaca et al., 2022; Corcoran et al., 2016; Imberman and Kugler, 2014; Leos-Urbel et al., 2013). In settings such as those analysed in these studies, where attendance rates are already high, school-meal programmes do not imply any improvement in attendance (Corcoran et al., 2016). Concerning academic achievement, nearly all the studies evaluate the impact of the provision of school meals on standardised test scores in Maths and Language (mainly English) and, to a lesser extent, in Science, and return mixed results. Part of the literature documents positive effects by which school-meal programmes enhance children's standardised test scores (Ayllón and Lado, 2025; Cuadros-Meñaca et al., 2022; Ruffini, 2022; Norwood, 2020; Gordanier et al., 2020; Schwartz and Rothbart, 2020; Frisvold, 2015; Imberman and Kugler, 2014; Belot and James, 2011).[19] Yet Abouk

[19]However, there is the possibility that these improvements may be short-lived due to the heightened caloric intake on exam days (Imberman and Kugler, 2014). In fact, Figlio and Winicki (2005) provide evidence that schools facing accountability sanctions increase the caloric content of their lunches on

and Adams (2022), Corcoran et al. (2016) and Leos-Urbel et al. (2013) find no evidence that universal free breakfasts enhance test scores. Other educational outcomes analysed include annual grades, grade retention and years of completed education (Bütikofer et al., 2018; McEwan, 2013; Hinrichs, 2010), again with mixed results.

Programme variation. The literature on the impact of school-meal programmes has not reached any consensus regarding the best way to design such programmes; this may be attributed to the relatively recent development of this field of study and differences in the contexts in which the programmes operate (Lundborg et al., 2022; Maruyama and Nakamura, 2025). Consequently, the studies included in the meta-analysis differ in terms of the characteristics of the school-meal programmes analysed. The first source of variation concerns the types of meal provided. Some programmes offer only breakfast, while others provide only lunch. Furthermore, there are initiatives that cover both breakfast and lunch, such as the Community Eligibility Provision (CEP) programme in the US.

The second source of variation between programmes is related to the eligibility criteria. Certain countries provide universal free meals to all students, regardless of their income level. Others restrict free meals to schools identified as disadvantaged. Additionally, many countries offer reduced-price or free meals based on income eligibility. The question of whether school-meal programmes should be universal or means-tested (whether at the individual or the school level) is a major topic of discussion in the literature. On the one hand, universal programmes may reduce administrative costs and the stigma associated with receiving free school meals, but they can be very costly (Altindag et al., 2020). On the other hand, means-tested programmes have the potential to reduce costs by efficiently targeting children in need; nonetheless, the bureaucracy and the potential for stigmatisation can increase non-take-up among needy families (Holford, 2015).[20]

Causal evidence of the impact of means-tested school-meal programmes is scarce and inconclusive.[21] Schanzenbach (2009) and Millimet et al. (2010) both identify negative effects of the NSLP on child weight outcomes. Conversely, Frisvold (2015), Bhattacharya et al. (2006) and Millimet et al. (2010) find positive effects of the SBP on child health and educational outcomes. Hinrichs (2010) shows similarly positive effects, albeit long term, on educational attainment for the NSLP. Research on universal free school meals is more extensive. Most of the literature exploits the staggered implementation of such programmes using a DiD approach. This evidence suggests that school-wide free meals increase programme participation, with effects on behaviour, health and education ranging from positive to null (Davis et al., 2024; Holford and Rabe, 2022, 2024; Gordon and Ruffini, 2021; Altindag et al., 2020; Leos-Urbel et al., 2013).

Primary studies also vary in terms of the educational level at which the school-meal programme is implemented and the geographical location. As for educational level, most analyses focus on elementary schools (e.g. Abouk and Adams, 2022; Frisvold, 2015), while a few examine the effects of school meals in middle or high schools (e.g. Gutierrez, 2021; Schwartz and Rothbart, 2020). Regarding geographical coverage, the majority of studies on developed economies have focused on the US, where data on participation is regularly

testing days to enhance short-term student performance.

[20] In the US, despite the 1970 amendments to the National School Lunch Act that prohibited the 'overt identification' of students receiving reduced-price or free school meals, at schools that allow both cash and non-cash transactions students may still be able to identify their peers who receive subsidised school meals (Gordon and Ruffini, 2021).

[21] Note that endogenous selection issues are more prevalent in means-tested school-meal programmes than in universal ones. Consequently, finding a convincing identification strategy becomes more challenging.

collected.[22] In contrast, evidence from outside the US is limited. Existing studies examine the effects of different programmes in Chile (McEwan, 2013), Colombia (Collante Zárate et al., 2024), Norway (Bütikofer et al., 2018), South Korea (Bethmann and Cho, 2022; Kim, 2021; Altindag et al., 2020), Spain (Ayllón and Lado, 2025), Sweden (Lundborg et al., 2022) and the UK (Borbely et al., 2024; Holford and Rabe, 2022, 2024; Belot and James, 2011). Notably, the vast majority of analyses are conducted at the regional level for two main reasons: i) the educational systems in the countries analysed are largely decentralised at the state or regional level; and ii) data unavailability at the national level implies that studies using administrative data are often conducted at the regional level, where access to data may be easier.

Student variation. Aside from programme attributes, the studies included in the meta-analysis differ in terms of the sub-analyses conducted, as the effectiveness of school-meal programmes may vary according to student characteristics, including gender, grade, age and socio-economic status. For example, gender disparities in health outcomes may arise from differences in development rates between boys and girls, or from their unique responses to potential stigma associated with means-tested school-meal programmes (Davis, 2019). Socio-economic background also influences the effectiveness of school-meal programmes. Children from economically disadvantaged families who receive free meals under means-tested schemes may experience feelings of exclusion or embarrassment, impacting their participation (Kho, 2018). The introduction of universal free school-meal programmes can help to reduce stigma. Additionally, in these programmes, non-low-income students may benefit from increased family resources, which could increase their participation in extracurricular activities. For example, Kim (2021) documents that the introduction of free school meals in South Korea caused more students to take part in after-school programmes.

Type of data. The literature on the causal effects of school-meal programmes on children's outcomes often relies on different sources of data. Most analyses use longitudinal administrative data, following children over several academic years (e.g. Cuadros-Meñaca et al., 2022). In contrast, studies that employ survey data are normally cross-sectional (e.g. Dunifon and Kowaleski-Jones, 2003), since longitudinal surveys that include information on school-meal participation are rarely available — with the exception of ECLS-K.[23] Another source of variation concerns the unit of observation at which data are collected, including individual-level (e.g. Abouk and Adams, 2022), school-level (e.g. Altindag et al., 2020) and district-level data (e.g. Ruffini, 2022). The time frame of the effect is also a key difference: the vast majority of studies focus on short-term outcomes, while Lundborg et al. (2022), Bütikofer et al. (2018) and Hinrichs (2010) are the only ones to examine the long-term implications of exposure to free school meals during childhood. This is also a piece of information that we consider in our meta-analysis.

Method and estimation characteristics. As explained in Section 2, we only included studies that provide credible causal estimates, employing techniques such as DiD, RDD and IV. Research relying on DiD usually analyses the effects on child outcomes of a transition from means-tested to universal free school meals. These studies identify the

[22] For example, the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K) includes school-level information on breakfast and lunch availability, as well as the percentage of students receiving school-meal subsidies.

[23] The OECD Programme for International Student Assessment (PISA) for Development collects cross-sectional, school-level data on meal provision and records parental involvement in meal preparation and distribution. This information is available only for low- and middle-income countries. PISA data for high-income countries does not contain these variables.

causal effect through two alternative identification strategies. Some compare regions or schools that transition with those that do not (e.g. Bethmann and Cho, 2022); others exploit variation in exposure times (e.g. Lundborg et al., 2022). In the latter case, some researchers narrow the sample to include only students exposed to universal free school meals at some point during the period of analysis (Rothbart et al., 2023), while others consider the entire population, comparing students treated at some point in time with those not yet treated or never treated (Cuadros-Meñaca et al., 2023). This is why we also collect information on exposure.

RDD and IV are normally used in studies that examine the effects of means-tested school-meal programmes that rely on eligibility criteria, usually defined by an income threshold (e.g. Schanzenbach, 2009). However, a number of studies using these techniques examine the CEP programme in the US, which provides universal free school meals (Davis and Musaddiq, 2019). Eligibility for this programme, as described in Table C.1 in Online Appendix C, depends on the identified student percentage (ISP), which represents the proportion of students who are eligible for means-tested free school meals or who participate in other federal means-tested public assistance programmes. Researchers often use this discontinuity in the ISP to examine the causal impact of universal free school meals.

Another key aspect is how treatment is defined and the inclusion of control variables in the regressions. Regarding treatment, most studies include a dummy variable, typically indicating whether or not the student receives free school meals. However, some researchers examine how outcomes vary with the duration of exposure. To do so, they employ both continuous (e.g. Hinrichs, 2010) and categorical (e.g. Holford and Rabe, 2022) treatment variables. We take all this information into account in our dataset. As for controls, we collect standard variables, including gender, age, whether the student is from a minority ethnic group or has a disadvantaged background, family status and whether the student lives in a rural area. Additionally, we gather information on individual, school, district, state and year fixed effects, which are commonly employed in DiD analyses using longitudinal administrative data.

Publication characteristics. As in previous meta-analyses, we assemble a range of variables that reflect the quality of each study. We begin by recording whether the study was published in a peer-reviewed journal or was a working paper. Excluding working papers, particularly older ones, might provide a skewed view of the literature, if the results are counterintuitive and face difficulties in being published in a peer-reviewed journal, thus increasing publication bias. We then consider whether the study has been influential, including the number of citations according to Google Scholar. Additionally, for published studies, we collect a range of variables to control for article quality, including the journal's impact factor and whether it was published in an economics journal. Finally, we also gather the publication year.

Table B.1: Description of variables reflecting heterogeneity

Variable	Description
Outcome analysed	
Outcome: Behaviour	= 1 if behavioural outcome (e.g. misbehaviour, discipline, school exclusions).
Outcome: Health	= 1 if health outcome (e.g. BMI, weight, height, nutrition).
Outcome: Education (ref.)	= 1 if educational outcome (e.g. test scores, annual grades, attendance).
Programme variation	
Lunch	= 1 if the programme only covers lunch.
Breakfast	= 1 if the programme only covers breakfast.
Breakfast and lunch (ref.)	= 1 if the programme covers breakfast and lunch.
Means-tested	= 1 if the programme is means-tested (i.e. programme eligibility is determined based on family income).
Universal (ref.)	= 1 if the programme is universal (i.e. available to all individuals without any means-testing or specific eligibility criteria).
Elementary school	= 1 if the effect is estimated for students in elementary school.
Middle school	= 1 if the effect is estimated for students in middle school.
High school	= 1 if the effect is estimated for students in high school.
Country: US	= 1 if the country of analysis is the US.
Country: Other (ref.)	= 1 if the country of analysis is not the US.
Country: National	= 1 if the analysis is at the national level.
Country: Regional (ref.)	= 1 if the analysis is at the regional level.
Student variation	
Female	= 1 if the effect is estimated for female students.
Male	= 1 if the effect is estimated for male students.
Not minority	= 1 if the effect is estimated for non-minority students.
Minority	= 1 if the effect is estimated for minority students (e.g. black, Hispanic).
Advantaged	= 1 if the effect is estimated for advantaged students (e.g. high-income, above-median income, highly educated parents, high academic achievement).
Disadvantaged	= 1 if the effect is estimated for disadvantaged students (e.g. low-income, below-median income, poorly educated parents, disabled, low academic achievement).
All students	= 1 if the effect is estimated for the whole population of students.
Type of data	
Longitudinal data	= 1 if longitudinal data is used.
Cross-sectional data (ref.)	= 1 if cross-sectional data is used.
Administrative data	= 1 if administrative data is used.
Survey data (ref.)	= 1 if survey data is used.
Individual-level data	= 1 if data is at the individual level.
School-level data	= 1 if data is at the school level.
District-level data (ref.)	= 1 if data is at the district level.
Data year	Year mean of the data used. In cases where the effects are long-term, 'Data year' refers to the time when individuals received the treatment.
Method and estimation characteristics	
Method: DiD	= 1 if the difference-in-differences approach is used for estimation.
Method: RDD	= 1 if the regression discontinuity design approach is used for estimation.
Method: IV	= 1 if the instrumental variables approach is used for estimation.
Method: Other (ref.)	= 1 if some other approach is used for estimation.
Main result	= 1 if the authors present the result as a main finding.
Robustness (ref.)	= 1 if the authors present the result as a robustness check.
Exposed	= 1 if the analysis excludes students who are never treated.

Continued on next page

Table B.1: Description of variables reflecting heterogeneity (continued)

Variable	Description
Not exposed (ref.)	= 1 if the analysis includes students who are never treated.
Treatment: Dummy	= 1 if the treatment is a dummy variable.
Treatment: Continuous	= 1 if the treatment is a continuous variable.
Treatment: Categorical (ref.)	= 1 if the treatment is a categorical variable.
Control: Gender	= 1 if estimation controls for students' gender.
Control: Age	= 1 if estimation controls for students' age.
Control: Minority	= 1 if estimation controls for students' minority status (e.g. black, Hispanic).
Control: Disadvantaged	= 1 if estimation controls for students' disadvantaged background (e.g. low-income, below-median income, poorly educated parents, disabled, low academic achievement).
Control: Family status	= 1 if estimation controls for students' family status.
Control: Rural	= 1 if estimation controls for students' rural status.
Fixed effects: Individual	= 1 if estimation includes individual FE.
Fixed effects: School	= 1 if estimation includes school FE.
Fixed effects: District	= 1 if estimation includes district FE.
Fixed effects: State	= 1 if estimation includes state FE.
Fixed effects: Year	= 1 if estimation includes year FE.
Publication characteristics	
Publication year	Publication year of the study.
Impact factor	The Journal Citation Reports (JCR) impact factor of the journal in which the primary study was published.
Citations	Number of Google Scholar citations.
Published study	= 1 if the study is published in a peer-reviewed journal.
Working paper (ref.)	= 1 if the study is a working paper.
Published in economics	= 1 if the study is published in a peer-reviewed economics journal.
Not published in economics (ref.)	= 1 if the study is not published in a peer-reviewed economics journal.

Note: The table details all the variables collected in the meta-analysis and used in the BMA analysis. The notation '(ref.)' indicates the reference category for each variable included in the BMA analysis.

C Additional tables and figures

Table C.1: School-meal programmes assessed in the primary studies

Programme	Country	Description
Ajuts individuals de menjador	Barcelona (Spain)	This subsidises 70% of the cost of the daily menu for pupils whose income is below a certain threshold, defined according to their family structure. Students whose income is below 60% of this threshold and who obtain 10 points in a so-called *family circumstances and social needs assessment* by the Institute of Social Services, receive free school meals.
Breakfast After the Bell (BAB)	US	The BAB programme provides free breakfasts after the school day begins. Its delivery model includes breakfast in the classroom (BIC), grab-and-go and additional breakfast periods.
Community Eligibility Provision (CEP)	US	This extends free school meals (breakfast and lunch) to all students within a school or district. Eligibility is determined by the identified student percentage (ISP), which represents the proportion of students within a school or district who qualify for free school meals based on their household's participation in federal means-tested public assistance programmes (e.g. Supplemental Nutrition Assistance Program or Temporary Assistance for Needy Families).
Eco-friendly Free School Meal Program	South Korea	This has provided free school meals to all students, regardless of family income, since 2011.
National School Lunch Program (NSLP)	US	This offers reduced-price school meals to pupils from families with incomes of between 130% and 185% of the federal poverty line (FPL) and free school lunches to those below 130% of the FPL.
'Oslo breakfast'	Norway	This provides free school breakfasts to all students in primary school. It was introduced during the 1920s and 1930s.
Programa de Alimentación Escolar (PAE)	Chile	This provides free school meals to pupils from 60% of the most vulnerable or socio-economically disadvantaged households. The caloric content of meals varies between schools, and is determined by a vulnerability index, which is a weighted average of socio-economic and anthropometric measures derived from survey data on first-year students.
Programa de Alimentación Escolar (PAE)	Colombia	This is a nationwide programme that provides free school meals. Full-day schools, rural schools and vulnerable urban schools are prioritised. Within each school, the School Feeding Committee determines which students will receive meals and the type of meal provided.
Swedish School Lunch Programme	Sweden	This provides free school lunches to all students in primary school. It was gradually rolled out across municipalities in the 1950s and 1960s.

Continued on next page

Table C.1: School-meal programmes assessed in the primary studies (continued)

Programme	Country	Description
School Breakfast Program (SBP)	US	This follows the same eligibility criteria as the NSLP. Families with incomes of between 130% and 185% of the FPL are eligible for reduced-price breakfasts, while those below 130% of the poverty line qualify for free school breakfasts.
Universal Infant Free School Meals (UIFSM)	England and Scotland	This provides free school lunches to pupils in reception, year 1 and year 2 in all government-funded schools. The means-tested system remains in place for children in year 3 and above.

Note: The table describes the most frequent school-meal programmes analysed in the literature.

Table C.2: BMA results for standardised test scores

	P. mean (1)	P. Std. Dev (2)	PIP (3)
SE(z)	0.7434	0.0705	1.0000
Programme variation			
Lunch	-0.1262	0.1050	0.7852
Breakfast	-0.0985	0.0468	0.8539
Means-tested	0.1742	0.1103	0.8580
Elementary school	0.0001	0.0029	0.0236
Middle school	-0.0001	0.0034	0.0255
High school	-0.0215	0.0313	0.3707
Country: National	0.1084	0.1041	0.6650
Student variation			
Female	-0.0000	0.0024	0.0202
Male	0.0009	0.0055	0.0437
Not minority	-0.0001	0.0023	0.0210
Minority	-0.0006	0.0042	0.0404
Advantaged	-0.0003	0.0026	0.0302
Disadvantaged	-0.0012	0.0050	0.0770
All students	0.0127	0.0108	0.6435
Type of data			
Administrative data	0.1481	0.0755	0.8474
Individual-level data	0.1101	0.0941	0.6383
School-level data	-0.0059	0.0274	0.1228
Data year	-0.0427	0.0088	1.0000
Method and estimation characteristics			
Method: DiD	0.0000	0.0042	0.0312
Method: RDD	0.2198	0.0840	0.9801
Method: IV	0.0003	0.0044	0.0208
Main result	-0.0010	0.0055	0.0489
Exposed	-0.0040	0.0137	0.1066
Treatment: Dummy	0.0004	0.0031	0.0319
Treatment: Continuous	-0.0007	0.0047	0.0386
Control: Gender	-0.0002	0.0035	0.0340
Control: Age	-0.0196	0.0276	0.3907
Control: Minority	0.0009	0.0073	0.0459
Control: Disadvantaged	-0.0018	0.0100	0.0550
Control: Rural	-0.0037	0.0123	0.1140
Fixed effects: Individual	0.0056	0.0124	0.2088
Fixed effects: School	-0.0019	0.0081	0.0782
Fixed effects: District	-0.0134	0.1411	0.2011
Fixed effects: Year	0.2212	0.0737	0.9900
Publication characteristics			
Publication year	0.0363	0.0097	1.0000
Impact factor	-0.0089	0.0194	0.2938
Citations	0.0002	0.0007	0.2647
Published study	0.0038	0.0168	0.0858
Published in economics	-0.0142	0.0476	0.1686

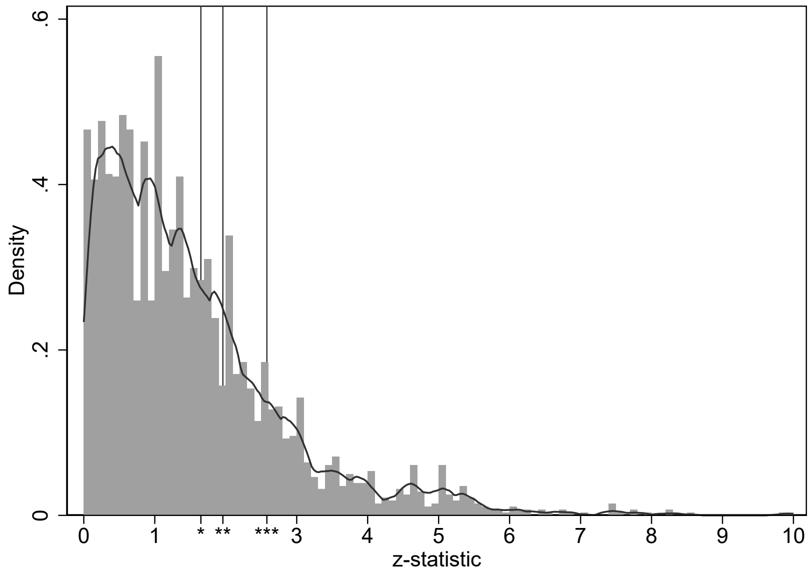
Note: The table presents the results of the baseline BMA using the unit information prior and the dilution model prior (George, 2010). The variables 'country: US', 'longitudinal data', 'fixed effects: state' and 'control: family status' are omitted due to low variance. 'P. mean' denotes posterior mean, 'P. Std. Dev' details the posterior standard deviation and 'PIP' refers to posterior inclusion probability. The posterior means are reported in terms of Cohen's *d*. Extreme outliers are winsorised at the 1st and 99th percentiles. See Table B.1 in Online Appendix B for a detailed description of the variables.

Table C.3: FMA results

	Coef. (1)	SE (2)	p-value (3)
SE(z)	1.4408	0.0407	0.0000
Outcome analysed			
Outcome: Behaviour	-0.0014	0.0020	0.7260
Outcome: Health	0.0018	0.0014	0.5200
Programme variation			
Lunch	0.0016	0.0025	0.7490
Breakfast	0.0242	0.0020	0.0000
Means-tested	0.0212	0.0025	0.0000
Elementary school	-0.0004	0.0020	0.9200
Middle school	-0.0068	0.0015	0.0230
High school	0.0100	0.0019	0.0080
Country: US	0.0106	0.0029	0.0680
Country: National	0.0284	0.0021	0.0000
Student variation			
Female	-0.0242	0.0039	0.0020
Male	-0.0132	0.0039	0.0910
Not minority	-0.0102	0.0044	0.2460
Minority	-0.0174	0.0042	0.0380
Advantaged	-0.0118	0.0042	0.1600
Disadvantaged	-0.0170	0.0042	0.0430
All students	-0.0138	0.0042	0.1000
Type of data			
Longitudinal data	-0.0120	0.0039	0.1240
Administrative data	0.0260	0.0021	0.0000
Individual-level data	0.0148	0.0041	0.0710
School-level data	0.0288	0.0038	0.0000
Data year	0.0002	0.0001	0.3170
Method and estimation characteristics			
Method: DiD	-0.0136	0.0024	0.0050
Method: RDD	0.0034	0.0032	0.5950
Method: IV	-0.0048	0.0035	0.4930
Main result	0.0064	0.0013	0.0140
Exposed	-0.0054	0.0027	0.3170
Treatment: Dummy	0.0062	0.0021	0.1400
Treatment: Continuous	0.0096	0.0025	0.0550
Control: Gender	-0.0034	0.0018	0.3450
Control: Age	0.0020	0.0016	0.5320
Control: Minority	0.0124	0.0019	0.0010
Control: Disadvantaged	0.0024	0.0020	0.5490
Control: Family status	-0.0242	0.0027	0.0000
Control: Rural	0.0054	0.0016	0.0920
Fixed effects: Individual	0.0160	0.0018	0.0000
Fixed effects: School	0.0160	0.0018	0.0000
Fixed effects: District	0.0180	0.0030	0.0030
Fixed effects: State	-0.0120	0.0029	0.0390
Fixed effects: Year	-0.0070	0.0019	0.0650
Publication characteristics			
Publication year	0.0048	0.0003	0.0000
Impact factor	-0.0020	0.0006	0.0960
Citations	0.0002	0.0000	0.0000
Published study	-0.0114	0.0025	0.0230
Published in economics	0.0124	0.0025	0.0130

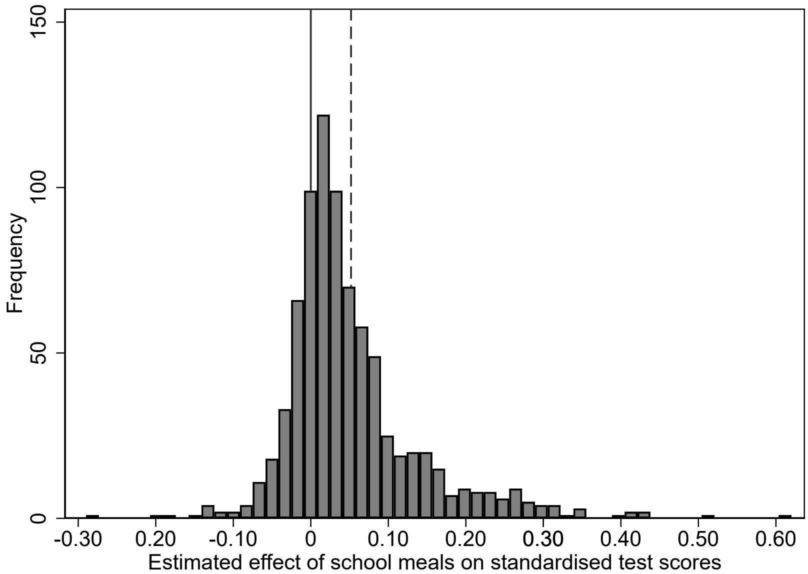
Note: The table presents the results of the FMA. The coefficients are reported in terms of Cohen's d. Extreme outliers are winsorised at the 1st and 99th percentiles. See Table B.1 in Online Appendix B for a detailed description of the variables.

Figure C.1: Distribution of the z-statistics



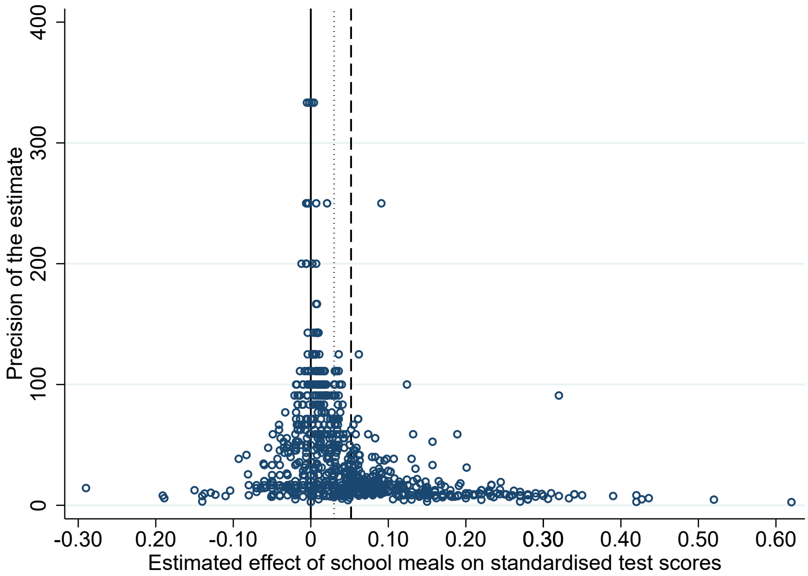
Note: The figure presents a histogram of z-statistics for each estimate from primary studies. Extreme outliers are winsorised at the 1st and 99th percentiles. *** denotes a z-statistic of 2.58, ** indicates a z-statistic of 1.96 and * represents a z-statistic of 1.65.

Figure C.2: Distribution of the effect sizes for standardised test scores



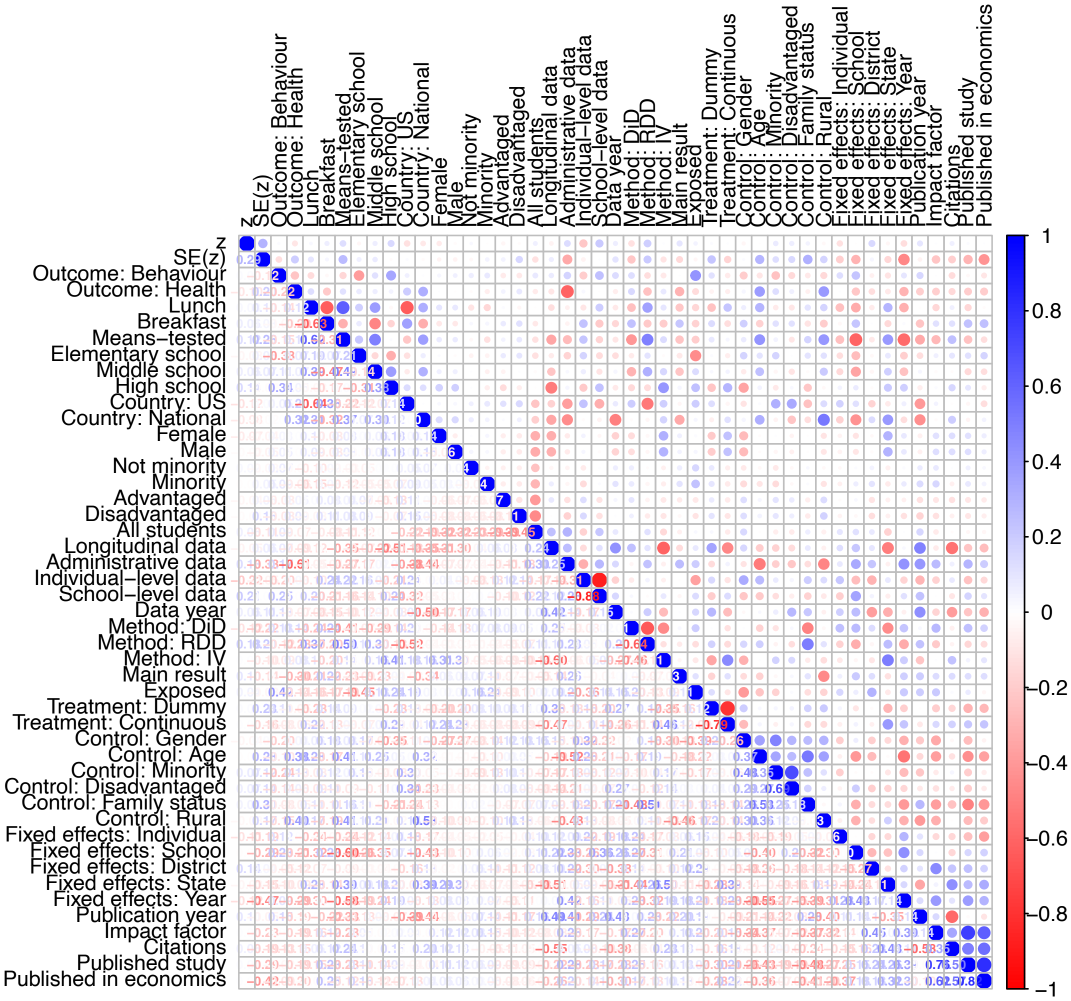
Note: The histogram displays the effect sizes from primary studies analysing the effects of school-meal programmes on standardised test scores. The solid line is set at zero, while the dashed line represents the sample mean. Extreme outliers are winsorised at the 1st and 99th percentiles.

Figure C.3: Funnel plot for standardised test scores



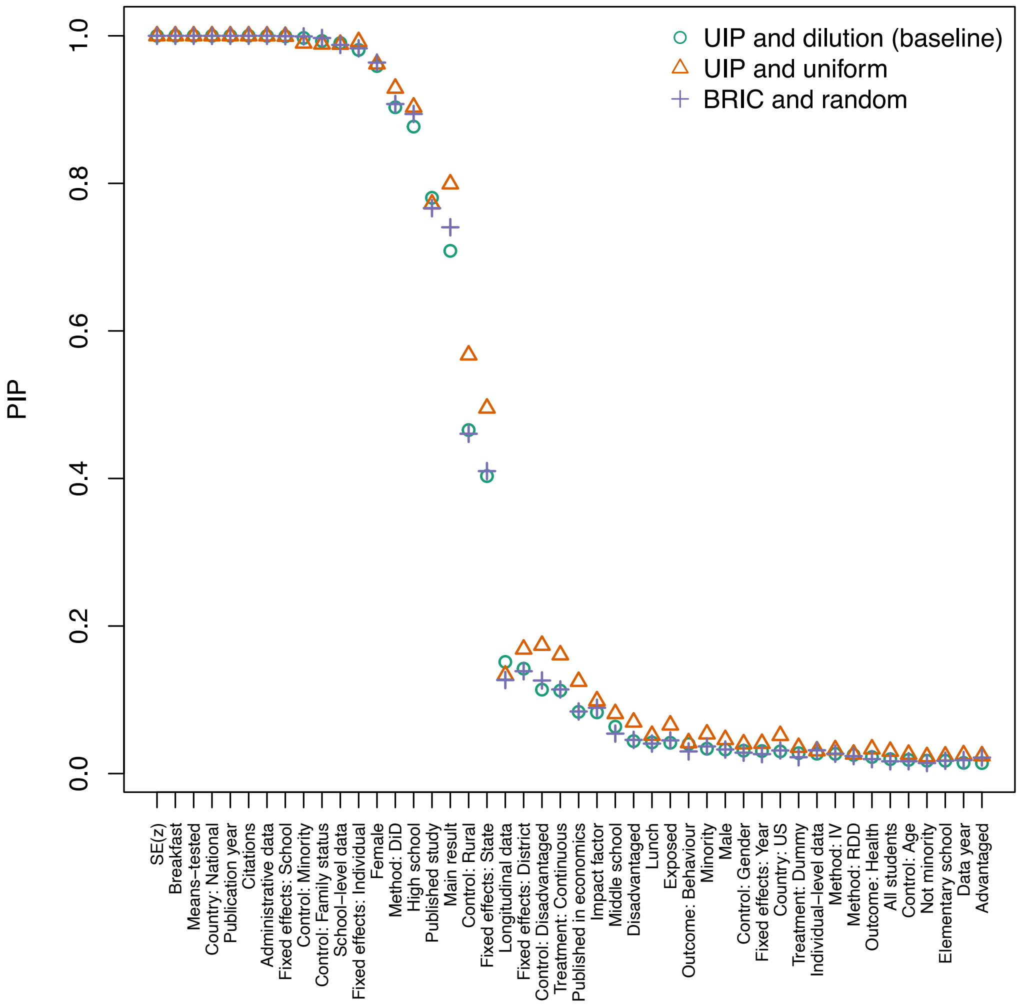
Note: The figure presents the relationship between effect sizes from primary studies analysing the effects of school-meal programmes on standardised test scores and their precision. The solid line represents zero, the dotted line indicates the median and the dashed line represents the sample mean. Extreme outliers are winsorised at the 1st and 99th percentiles.

Figure C.4: Correlation between multiple MRA variables



Note: This figure presents the correlation between all variables introduced in the multiple meta-regression. Extreme outliers are winsorised at the 1st and 99th percentiles. See Table B.1 in Online Appendix B for a detailed description of the variables.

Figure C.5: Sensitivity of BMA to different priors



Note: The figure presents the results of the BMA with different priors. On the vertical axis, we plot the PIP and on the horizontal axis, we list the explanatory variables, sorted by their PIP. Circles represent the BMA results using the unit information prior and dilution prior as proposed by George (2010). Triangles indicate results employing the UIP and uniform priors as suggested by Eicher et al. (2011). Crosses show the results using priors recommended by Fernández et al. (2001). Extreme outliers are winsorised at the 1st and 99th percentiles. See Table B.1 in Online Appendix B for a detailed description of the variables.