Population Research and Policy Review (2025) 44:21
https://doi.org/10.1007/s11113-025-09935-5

ORIGINAL RESEARCH

Does Cutting Child Benefits Reduce Fertility in Larger Families? Evidence from the UK's Two-Child Limit

Mary Reader[1,2] · Jonathan Portes[3] · Ruth Patrick[4]

Received: 4 November 2023 / Accepted: 6 January 2025

Abstract
We study the fertility effects of restricting child-related social assistance to the first two children in the family. As of 2017, all third and subsequent children born on or after 6 April 2017 in the UK were made ineligible for approximately 3000 GBP of means-tested child benefits per year. Using a triple difference and regression discontinuity design, we leverage administrative births microdata to identify the impact of the two-child limit on higher-order births. We find little to no decline in higher-order fertility among low-income families, with our estimates indicating at most small elasticities relative to the literature.

Keywords Fertility · Family size · Social assistance · Welfare reform

JEL J13 · J18 · H31 · H53

Introduction

How do financial incentives and constraints affect household fertility decisions? The basic economic model of fertility, as developed by Becker (1960), treats children as analogous to consumer durables, with associated costs and benefits. Government policies, in particular the treatment of children in the tax and benefit system, change these costs and benefits and therefore affect fertility decisions. However, even in this simple model, the magnitude of these effects will depend both on individual and household preferences and on other variables affecting family incomes; they are therefore likely to depend on the specific economic and social context. The

✉ Jonathan Portes
jonathan.portes@kcl.ac.uk

1 Department of Economics, Stanford University, Stanford, CA 94305, USA

2 London School of Economics, STICERD, Houghton Street, London WC2A 2AE, UK

3 Department of Political Economy, King's College London, Strand, London WC2R 2LS, UK

4 Department of Social Policy, University of York, Heslington, York YO10 5DD, UK

Published online: 04 March 2025

complexity of the factors involved, and the relative rarity of substantial exogenous changes to financial incentives that can be used to identify causal effects, mean empirical estimates remain limited.

In this paper we examine the impact of a major and internationally significant change to the treatment of children in the UK benefit system: the introduction of the "two-child limit" for cash benefits to low-income families. While a handful of countries restrict means-tested child benefits to the third or fourth child, this policy is the first attempt internationally to cap means-tested child benefits at the second child (Stewart, Patrick & Reeves 2023b).

The impact of this change on families' fertility decisions is of considerable interest, both from a policy and a research perspective. In contrast to other changes to the UK benefit system for low-income families, it was justified, at least in part, by the argument that benefit payments to low-income families with children incentivized higher levels of fertility, and that removing them would have the opposite effect (Treasury & DWP 2015). Implicitly, the objective of the policy was not only to reduce expenditure but also to reduce fertility among low-income households.

By contrast, opponents of the change made both normative arguments—that it was inappropriate for the state to attempt to reduce fertility among low-income households by cutting benefits, and that if a certain level of support per child was appropriate, it should apply to all children—and positive ones, that the policy change was unlikely to have a large impact on fertility decisions. Opponents predicted that the main impact of the change would be an increase in poverty among larger families. Establishing to what extent these narratives are consistent with the empirical evidence is a vital prerequisite to evaluate the policy effectively and inform wider debate on the future of the welfare system. It is also relevant to international debates about trends in fertility in advanced economies: the change came as overall fertility rates in the UK were falling steadily, with the UK's total fertility rate falling from 1.9 in 2011 to 1.65 in 2019 (Office for National Statistics, 2022).

The nature of the two-child limit makes it well suited to causal identification of the impact of financial incentives in the welfare system on fertility. It was a large change, introduced at a single point in time, and it affected some groups, while leaving others (those with fewer than two children, or not receiving benefits) entirely unaffected. In 2022, 450,000 households (including 1.4 million children) were affected by the policy (Department for Work and Pensions and Her Majesty's Revenue and Customs, 2024). The policy only applies to third or subsequent children who were born on or after 6 April 2017, creating a plausibly exogenous source of variation in treatment status by date of birth. In this paper, we leverage variation over date of birth, birth order and socio-economic status to isolate the causal impact of the two-child limit on fertility, using a difference-in-difference-in-differences (or triple difference) and a regression discontinuity design.

Our results suggest that the impact of the policy change on fertility, if any, was relatively small. The period before and after the policy change saw steady falls in the overall number of births in the UK. However, we do not find evidence of a substantial reduction in the relative number of births among those affected. Our triple difference estimates are either positive in sign or close to zero. Estimates from the regression discontinuity design meanwhile suggest a small negative effect, but our preferred specification is not statistically significant, and the magnitude is extremely small relative to the literature.

These results are novel and surprising, and out of line with the findings of most previous research both internationally and in the UK, which usually, by looking at changes which increased the generosity of benefits, suggest relatively large elasticities. The contribution of our paper is that it is the first, and we think only, attempt to evaluate the impact of what amounts to an internationally significant policy on fertility: restricting benefit payments to two children in a family. The evidence generated is important for all those interested in the state's role in influencing fertility decision making, and, beyond this, in how social security systems are structured to support or constrain child bearing. One interpretation of our results is that this response may not be symmetric: benefit cuts may not have equal and opposite effects to benefit expansions. Our results also undermine the implicit policy rationale for the reform: rather than causing a major reduction in the number of children born into low-income families, the main impact has been to increase the depth and incidence of child poverty (Chzhen & Bradshaw, 2024). Drawing on parallel qualitative research into the two-child limit, we speculate that imperfect information about the policy, the "stickiness" of fertility attitudes (particularly among larger families), and an erroneous understanding of how "choice" is exercised in fertility decision-making may all play a part (Patrick & Andersen, 2022).

These results—and the possible drivers behind them—have significant economic implications. At a time of declining and below replacement fertility in almost all advanced economies, the impact of financial incentives is of increasing interest and importance from both a research and policy perspective; our results highlight the importance of the interaction of such policies with the wider socio-economic environment and imply that simply focusing on financial incentives alone is unlikely to deliver the desired results from a policy perspective.

The paper is structured as follows. We briefly describe the increase, and subsequent reduction, in the generosity of the welfare system in the UK over the last two decades, and the context for the introduction of the two-child limit. We then review the evidence on the impact of the structure and generosity of the benefit system on fertility. Our principal contribution is our empirical analysis, which leverages comprehensive administrative data on births in England and Wales both before and after the policy change. To conclude, we discuss possible interpretations and implications of our analysis, from both a research and policy perspective.

Background

Since the establishment of the modern welfare state after the Second World War, the UK has used the tax-benefit system to support families with children through a combination of universal (Family Allowance and Child Benefit) and means-tested support. The latter was initially for families with essentially no income, but was extended to low-income working families with the introduction of Family Income Supplement in the 1970s. This provision was made significantly more generous in the 2000s with the introduction of Working Families Tax Credit, which was modelled on the US's Earned Income Tax Credit.[1]

By the early 2010s, most families with children were receiving some form of support from the tax credit system (in addition to Child Benefit). Entitlement extended well above the median of the income distribution, although support was concentrated on those on low incomes, whether in or out of work. The objective of this expansion of benefits was both to reduce poverty among low-income families and to increase employment. Though there is no evidence to suggest that increased fertility was an objective of these policies, Brewer et al. (2012) found that those most likely to have been affected saw a differential increase in fertility, with an estimated rise of 1.2 to 3.6 percentage points in the annual probability of those most affected having a child, equating to an increase of approximately 10,000 to 35,000 in annual births. This was a period in which overall fertility rates in the UK were rising, with an increase in annual births of over 100,000 a year in the 2000s. ONS (2009)'s descriptive analysis of these trends attributes about a third of the rise to the greater affordability of children for low-income households (including improved childcare provision as well as higher benefit levels), though this was not a causal estimate.

After the global financial crisis, a Conservative-led Coalition government enacted significant cuts to welfare benefits as part of its wider program of fiscal consolidation. The initial program of cuts included changes to benefits for sick and disabled people, capping the amount of benefits that could be received by individual households, reductions in the uprating of benefits, and the so-called "bedroom tax" that restricted housing benefit payments for those with "extra" bedrooms. This program did not focus explicitly on larger families, though larger families were disproportionately affected due to their higher household needs and lower work intensity (Stewart, Patrick & Reeves 2023a, Gaffney, 2015, Reed & Portes, 2015).

While presented primarily as an economic necessity—the prevailing levels of benefits were described as unaffordable, given the impact of the recession on government borrowing—a parallel narrative also suggested that the benefit system had undesirable moral hazard effects. The Prime Minister, David Cameron, argued in 2011:

"The benefit system has created a benefit culture. It doesn't just allow people to act irresponsibly, but often actively encourages them to do so." (UK Government, 2021).

In conjunction with some sections of the press, the government sought to justify benefit cuts by arguing that very significant payments were being made to people who were either making fraudulent claims or who were (legally) "milking the system" (Gaffney & Portes, 2013). Frequently, the latter referred to larger families or

[1] In the 2010s, Working Families Tax Credit was subsumed within a new benefit, Universal Credit; which includes a Child Allowance. With both tax credits and Universal Credit, the benefits are gradually withdrawn with earnings, and this also applies to the child-related support (meaning that households will receive less than the full child element, depending on their earned income).

so-called "benefits broods"; such a case, and the wider background, is described at length in Jensen and Tyler (2015).

The Two-Child Limit

After winning the 2015 General Election, in July a majority Conservative government announced a further program of benefit cuts, including the introduction of a "two-child limit". Previously, the system of means-tested tax credits (which was in the process of being replaced with a new, integrated benefit, Universal Credit) paid an equal amount for each child. After the reform, payments would only be made in respect of the first two children for each household. Previously, the entitlement would have increased by approximately 3000 GBP per year when an additional child was born. After the two-child limit, for children born on or after 6 April 2017, no child element would be paid if the household already had 2 + children. The policy reform therefore reduced the maximum entitlement to child-related benefits for a family with three children at the bottom of the income distribution from approximately 12,000 GBP to approximately 9,000 GBP. For families higher up the income distribution, the loss would be less, due to the operation of the means test or "taper", which reduced entitlement by 48 pence for each additional pound of net income. Those families whose income was too high to claim tax credits or Universal Credit were unaffected by the policy change.

At 2023–24 benefit rates, the two-child limit amounts to a loss of up to 3,235 GBP a year in benefits income (Department for Work & Pensions, 2022).[2] Households are exempt from the two-child limit if their third child is a multiple birth, an adopted child, in a non-parental caring arrangement, or if they are the result of non-consensual conception (see Treasury & DWP (2015) for details of the policy's scope and impact).

The government described the objective of the two-child limit as being "to ensure that families in receipt of benefits faced the same financial choices about having children as those supporting themselves solely in work" (Treasury, 2015a, 2015b). Implicit in this was the assumption that incentives within the benefits system influenced fertility. This view was made explicit in published Impact Assessments, which (based in part on Brewer et al., 2012) suggested that the policy change would result in reduced fertility, although no quantitative predictions were made:

> "In practice people may respond to the incentives that this policy provides and may have fewer children...Given that families are aware of the policy they may make the choice not to have (further) children." (Treasury & DWP 2015)

The extent to which reducing fertility was an objective of the two-child limit is unclear. Indeed, in response to a Work and Pensions Select Committee report on the two-child limit, the government denied that curbing fertility was a key policy aim:

[2] For certain families in paid employment, the loss is less than 3,235 GBP per child because the child element is steadily withdrawn as earnings rise above a certain income threshold.

> "This policy does not attempt to limit the number of children people have. Claimants are able to have as many children as they choose, in the knowledge of the support available." (Work & Pensions Select Committee, 2019).

There was therefore ambiguity over how far fertility changes were themselves an explicit policy objective, perhaps reflecting political sensitivities over welfare policy and the role of the state in influencing fertility decision making. Nevertheless, irrespective of whether the reduction of fertility was intended, in shifting fertility incentives it is plausible that it may have achieved this in practice, particularly in light of existing evidence, to which we now turn.

Existing Evidence on Welfare and Fertility

A review of the literature on the impact of financial incentives on fertility (Stone, 2020) found that almost all studies on this question identified positive and significant results, with an increase in the present value of child benefits of 10 percent of a household's annual income producing between 0.5 and 4.1 percent higher birth rates. A number of other empirical studies in developed countries suggest positive fertility effects from the expansion of benefits (González & Trommlerová, 2021). Evidence of reductions in fertility due to benefit cuts or withdrawals are more mixed (Cohen, Deejia & Romanov 2013, González & Trommlerová, 2021, Sandner & Wiynck, 2023).

In general, evidence suggests that there is heterogeneity in the magnitude of fertility responses to welfare by family size, income, and institutional context. Laroque and Salanié (2014) take a structural approach and, using data from France, find a substantial response of fertility to financial incentives in the French tax and benefit system. The response differs according to birth order, with first and (especially) third births particularly responsive. Simulating the impacts of an unconditional child credit of 150 Euros a month per child, they estimate that this would raise birth rates among the most affected group (i.e., those with two children) by 4.8 percentage points.

Fertility responses are also heterogeneous across the income distribution. Milligan (2005), using a triple difference approach to examine the introduction and subsequent withdrawal of a universal C$8000 cash payment to families with newborn children in Quebec, found large fertility effects, with an extra C$1000 leading to an increase of about 17 percent in fertility. Effects were considerably larger among better-off families. While this may seem counterintuitive, it is in fact consistent with the canonical Becker model. Where families can choose to spend on both the quantity and the "quality" of children (where "quality" is influenced by per-child spending), the impact of a simple per-child benefit is ambiguous. While the substitution effect (having children becomes less costly relative to working) unambiguously increases fertility, the income effect might either increase or decrease fertility, depending on whether the income elasticity of fertility is negative or positive (just as a cut in income taxes can either increase or decrease hours worked, depending on the income elasticity of demand for leisure). So while intuitively one would expect a cut in child benefits to reduce fertility, this need not be the case for all families

(Cohen et al., 2013; Milligan, 2005; Riphahn & Wiynck, 2017). In particular, if low-income families face a particularly sharp "quality-quantity" trade off, they may respond to increases in welfare by reducing fertility. That is, the increased income resulting from a per-child benefit for poorer families with one or two children may mean that they choose not to have a subsequent child.

However, for the policy change we examine—the removal of benefits for third and subsequent children—the income effect is less relevant. There is no income effect for the first two children, so the choice to have a third child should be driven solely by the substitution effect, and the impact of the policy change should be unequivocally negative.[3]

Finally, fertility elasticities appear to be sensitive to institutional context. Research in the US is less suggestive of large impacts. Following the introduction of Temporary Assistance for Needy Families (TANF) program in the 1990s, a number of states introduced measures to deny additional social assistance to any children born after a benefit claim was made. Research on these family caps has broadly found no effect on fertility (Camasso & Jagannathan, 2016; Dyer & Fairlie, 2004; Joyce et al., 2004; Kearney, 2004). However, even leaving aside the very different social and economic context of the US, it is not clear that the two-child limit will have the same fertility effects as family caps. While the policies are similar in their aims, they are different in their application: for family caps, the determinant of whether you are capped is the timing of pregnancy relative to your benefits claim; for the two-child limit, it is the size of your family. The interaction with other aspects of the welfare system (notably work requirements and childcare) is also very different.

The most directly relevant study for our purposes is Brewer et al. (2012), which examined the impact of increases in the generosity of the UK benefit system in 1999 (part of the expansion of the tax credit system described in Sect. 2 above). They used a difference-in-differences approach, exploiting the fact that the reforms were targeted at low-income households, and that the changes were likely to differentially affect the fertility of women in couples and single women due to the opportunity cost effects of the welfare-to-work element. Using data from the Family Resources Survey, they found no increase in births among single women, but an increase in births among coupled women on relatively low incomes. Their main estimates are an increase in the probability of having a child of approximately 2 percentage points per year in the affected groups (an increase in fertility rates of about 15 percent). However, given the relatively small number of births observed in the survey data, there is considerable uncertainty around this estimate.

A more recent, closely related study is Sandner and Wiynck (2023), who look at a German welfare policy change which, like the two-child limit, resulted in substantial cuts to household income for families with children. They find a fertility reduction of 6.8 percent for a large (18 percent) reduction in household income, implying a much lower elasticity of 0.38. They conclude that welfare recipients' fertility reacts

[3] In principle, the removal of benefits for the third child could, via the aforementioned income effect, lead to some families choosing to have a fourth child. However, this is highly unlikely to be relevant over the time period we examine.

less strongly to financial incentives than the fertility of overall populations. This idea is corroborated by a qualitative study of fertility choices by those subject to the two-child limit by Patrick and Andersen (2022), which found that few families took the policy into account when deciding how many children to have. The study found that low awareness about the policy and a mismatch between the policy presentation and everyday lives on a low-income may have contributed.

Given existing theoretical and empirical work, our prior is that the two-child limit is likely to have reduced fertility among affected households, although there is considerable uncertainty given the various dimensions of heterogeneity discussed above. If the effects were of similar magnitudes to those found in most of the papers discussed above, we would expect to see large changes among affected households. The financial impact of the two-child limit is almost twice as large in PPP terms as that of the benefit increase modelled by Laroque and Salanié (2014) (OECD, 2021), thereby implying an impact of more than 7 percentage points. The scale of the policy change is also larger than the reforms considered by Brewer et al. (2012) and Milligan (2005). However, Sandner and Wiynck (2023), examining a large cut to benefits rather than an increase in their generosity, find much smaller impacts. With this in mind, we turn to our empirical analysis.

Method

We use two identification strategies to test the effect of the two-child limit on fertility: a triple difference design and a regression discontinuity design.

Triple Difference Design

Our objective is to identify the impact of the two-child limit on fertility of those "treated" by the policy—that is, women (and/or couples) who, if they were to have a child on or after April 2017, were no longer eligible for child-related social assistance for that child. Since the policy applied only to those in receipt of benefits with 2 + existing children and was rolled out over time, a difference-in-differences setup is a natural choice for an identification strategy.

The treatment group here is low-income women (single parents or members of a couple household) with 2 + children. In a simple difference-in-differences, there would be two main candidates for a control group: low-income women with 0–1 children, and high-income women with 2 + children. Neither group offers a compelling counterfactual: there are numerous factors other than the policy change (other changes to the benefits system, as well as wider social and demographic trends) which are likely to change living standards and the desired number of children over the period in question, and which would introduce bias into a simple difference-in-differences strategy of this kind.

Our preferred approach is therefore a difference-in-difference-in-differences (or triple difference) approach, which creates four groups, only one of which is treated.

We therefore estimate the impact of the policy change by looking at the change in fertility between low-income women with 2 + children and high-income women with 2 + children, at the change in fertility between low-income women with 0–1 children and high-income women with 0–1 children, and then comparing these two changes.[4] This set-up mirrors that of Gruber (1994), which introduced the triple-difference estimator as a method for estimating the causal impact of policy changes: coincidentally, Gruber was also looking at the impact of a policy change (the introduction of mandatory maternity pay) which affected some (but not all) mothers in some (but not all) US states.

In principle, this should abstract both from differential changes in fertility between those on a low income and those not, and differential changes in fertility between larger and smaller families. This triple difference approach should be more robust than a simple difference-in-differences strategy (Milligan, 2005).

Algebraically, we present the probability of having a child for a woman in group g as follows:

$$Y_g = \beta_0 + \beta_1 \mathbf{1}_{[B_g=1]} + \beta_2 \mathbf{1}_{[L_g=1]} + \beta_3 \mathbf{1}_{[B_g=1]} \times \mathbf{1}_{[L_g=1]} + \varepsilon_{it} \tag{1}$$

where Y_{gt} is the birth rate for group g, B_g is a dummy variable equal to one for women in low income occupation families, L_g is a dummy variable equal to one for women in families with 2 + children, and ϵ_g is the error term. Our coefficient of interest is β_3: it represents the marginal impact of being on a low income and having 2 + children on the probability of having a child.

In a standard difference-in-differences, the identification assumption is often tested by the parallel trends assumption: while β_0, the constant/base probability, can vary (the "trend"), the subgroup specific parameters β_1, β_2 and β_3, remain constant (so the trends are parallel for different groups) except when affected by the policy change.

Our triple difference set-up allows us to partially relax this assumption. In particular, β_0, β_1 and β_2—the base probability, but also the marginal effect of either being on a low-income or having 2 + children—can vary over time. As Olden and Moen (2022) state, "even though the triple difference is the difference between two difference-in-differences, it does not need two parallel trend assumptions. It requires the relative outcome of group B and group A in the treatment state to trend in the same way as the relative outcome of group B and group A in the control state in the absence of treatment." This means that all we need is that β_3—the additional probability of having a child that results from the *interaction* of being on benefits and of already having had 2 + children—remains constant except when affected by the policy change; this is our key identifying assumption.

In comparing the size of β_3 before and after the policy change, our preferred triple difference specification is the following:

[4] It does not matter which difference is taken first; that is, we could equally compare fertility rates for low-income women with no or one child with those with 2 + children; do the same for high-income women; and then compare the two differences.

$$
\begin{aligned}
Y_{gt} =&\beta_0 + \beta_1 \mathbf{1}_{[B_g=1]} + \beta_2 \mathbf{1}_{[L_g=1]} + \beta_3 \mathbf{1}_{[B_g=1]} \times \mathbf{1}\left[L_g = 1\right] + \beta_4 \mathbf{1}_{[Post_t=1]} \\
&+ \beta_5 \mathbf{1}_{[B_g=1]} \times \mathbf{1}_{[Post_t=1]} + \beta_6 \mathbf{1}_{[L_g=1]} \times \mathbf{1}_{[Post_t=1]} \\
&+ \beta_7 \mathbf{1}_{[B_g=1]} \times \mathbf{1}_{[L_g=1]} \times \mathbf{1}_{[Post_t=1]} + \varepsilon_{gt}
\end{aligned} \tag{2}
$$

where Y_{gt} is the birth rate for group g in quarter t, B_g is a dummy variable equal to one for women in low income occupation families, L_g is a dummy variable equal to one for women in families with 2 + children, $Post_t$ is a dummy variable equal to one if the quarter is after April 2017 and zero otherwise, and ϵ_{gt} is the error term. Our coefficient of interest is β_7: it represents the change in the probability of having a child after the two-child limit for low-income women with 2 + existing children.

Regression Discontinuity Design

Our second identification strategy is a regression discontinuity (RD) design. This enables us to non-parametrically identify the effect of the two-child limit by exploiting the plausible exogeneity of the date of birth criterion (Lee & Lemieux, 2010). The downside of this approach is that RD estimates are by definition local, and do not account for the possibility of gradual fertility effects.

We restrict our RD analysis to low-income women with 2 + children (the treatment group from the triple difference) and estimate various regression discontinuity specifications at the date of birth level. Our preferred specification is a non-parametric local linear donut RD, which allows the slope to vary either side of the cut-off and excludes observations within 3 days of the cut-off[5]:

$$
\begin{aligned}
&Y_t = \beta_0 + \beta_1 \mathbf{1}[t \geq 0] + \beta_2 t + \beta_3 t \times \mathbf{1}[t \geq 0] + \varepsilon_t, \\
&\forall t \in \{\{-b \leq t \leq -3\}, \{3 \leq t \leq b\}\}
\end{aligned} \tag{3}
$$

Y_t is the probability of having a child with a date of birth of t, t is the date in days relative to the introduction cut-off $\overline{x}$ (6 April 2017), b is the bandwidth, ϵ_t is the error term. β_1 is the coefficient of interest and represents the intention-to-treat estimate of the effect of the two-child limit on fertility.

As recommended by the latest RD literature, we select the bandwidth b in a data-driven way, by optimizing the coverage error rate (CER) and use robust bias-corrected standard errors (Calonico, Cattaneo & Farrell 2021, Cattaneo, Idrobo & Titiunik 2019, Hyytinen et al., 2018).

[5] We use a donut RD specification because, as shown in Supplementary Figure S11, in the raw data there is a high outlier observation three days prior to the introduction of the policy. While this could suggest bunching before the cut-off, we are skeptical that individuals' control of birth timing is sufficiently precise to explain this result; instead the most plausible explanation is noise. To avoid bias, we exclude small numbers of observations local to the cut-off and test for sensitivity to this. We select the 3-day radius donut as a representative example and document other donut sizes, including no donut, in detail in Table 3.

Data

Our relevant outcome of interest is the probability of a woman having a child in each time period. For each group, we combine estimates of the number of births from a random 10 percent sample of administrative birth registry data with estimates of the number of women from the Annual Population Survey (APS). For each of the four groups in the triple difference, we calculate the birth probabilities as follows: the denominator is the number of women who could potentially have children, from the APS; the numerator is the number of births, from the administrative births data.[6]

Administrative Birth Registry Data, 2013–2019

Our main data source on the number of births to mothers at different income levels is administrative birth registrations microdata from 2013–2019 in England and Wales.[7] This provides individual-level data for all births (both live and stillbirths) in England and Wales over this period, with variables on the date of birth of the child, the number of previous live births born to the mother, and postcode of residence of the mother. For a random 10 percent sample of live births, the data includes socio-economic status (SES), as defined by the National Statistics Socio-economic Classification (NS-SEC)'s measure of occupation. NS-SEC is a standard and long-standing measure of SES in the UK and has been validated as a predictor of a wide range of medium and long-term social and economic outcomes. It is not a perfect measure of income or benefit receipt, but it is strongly correlated with both (Office for National Statistics, 2016; Social Mobility Commission, 2022). It is likely to be more predictive over long periods of time (since occupational status abstracts from short-term fluctuations in income) and at the household or family level (since it abstracts from individual idiosyncrasies in how an occupation reflects an individual's income). In our case, we use the highest NS-SEC occupation of the mother and father, which enables us to consider the resources of the whole household.

We do not use data from 2020 and beyond given the likely impact of the pandemic on fertility, which could introduce systematic bias into our results by disproportionately affecting low-income and larger families (Reader & Andersen, 2022).

Figure 1 charts the number of total births (left axis) alongside the number of higher-order (3+) births (right axis) in England and Wales by year from 2013 to 2019. Higher-order births started to fall in 2015, as did total births in 2016. The share of higher-order births fell over the period. However, the aggregate data do not suggest a large impact of the two-child limit in 2017.

[6] A random 10 percent sample of the administrative births data are coded with NS-SEC occupation status, while the APS is a household survey which does not sample the full population. To account for the consequent sampling uncertainty, we calculate standard errors for the number of births and the number of women by quarter by extracting the standard error from the proportion of women/births in each category and scaling this up to a population estimate.

[7] This work was undertaken in the Office for National Statistics Secure Research Service using data from ONS and other owners and does not imply the endorsement of the ONS or other data owners.

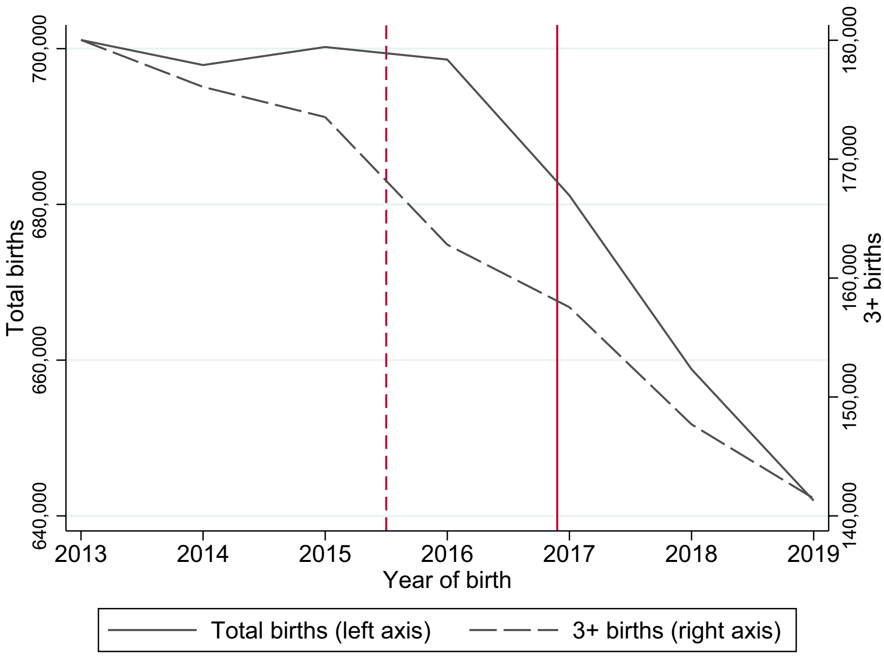

Fig. 1 Number of total and 3 + births in England and Wales by year, 2013–2019. The figure plots the total number of births in England and Wales on the primary axis (solid line) and the total number of third or subsequent births (dashed line) on the secondary axis by year. Data are presented annually, since there is seasonal variation in births by month and quarter of birth. The red solid line indicates the introduction of the two-child limit in April 2017 (here it is displayed just before 2017, as the data is annual, and the policy affected the majority of 2017). The red dashed line indicates the announcement of the two-child limit in July 2015 (for the same reasons, here it is displayed just before 2016). Data from ONS administrative birth registry microdata for England and Wales

The birth registry data do not include data on benefits receipt; to proxy for this, we use a measure of combined occupational status by selecting the highest-ranked occupation of the mother or father, where available. We do this on grounds that family occupation is strongly correlated with benefits receipt. Based on the observed probabilities in Supplementary Material Figure S1 and Figure S2, we group observations into those with a family occupation with a higher probability of benefits receipt (NS-SEC 3–8) and those with a lower probability of benefits receipt (those in NS-SEC 1 and 2). We test for sensitivity to alternative definitions in Sect. 5.1.

Table 1 shows summary statistics for the universe of births in England and Wales from 2013–2019. Of the 10 percent sample with coded occupation, approximately 50.4 percent of families are in a 'high-income occupation' (NS-SEC 1–2), while 49.6 percent are in a 'low-income occupation' (NS-SEC 3–8).

Table 1 Summary Statistics: Births in England and Wales, 2013–2019

	N	Mean	SD
Multiple birth	5,303,199	0.030	0.171
Stillbirth	5,303,199	0.004	0.066
Previous live births	5,281,659	0.992	1.162
Maternal age (years)	5,303,082	30.357	5.706
Paternal age (years)	5,015,288	33.273	6.771
Income deprivation score	5,303,199	14.992	9.994
Low-income occupation (NS-SEC 3–8)	514,799	0.496	0.500
High-income occupation (NS-SEC 1–2)	514,799	0.504	0.500
Single parent	5,303,198	0.158	0.365

The table presents summary statistics for the administrative birth registry data from 2013–2019. The number of observations is lower for low-income and high-income occupation because occupation is only coded for a 10 percent random sample of the data. Data from ONS administrative birth registry micro-data for England and Wales

The Annual Population Survey, 2013–2019

While the administrative data tells us the number of births, it does not tell us how many women are in the relevant groups, and so does not give us fertility rates. To construct fertility probabilities by group, we need data on the number of women in each group, to provide the denominator for the birth data above.[8] For this, we use the Annual Population Survey (APS), the largest household survey in the UK, with its annual sample including approximately 320,000 individual respondents in 122,000 households. The APS contains data at the individual level on the sex, age and NS-SEC occupation of the respondent, which benefits they receive (if any), the number of dependent children under 16 in the family unit, and the type of family unit (e.g. single parent or couple family). The family unit is the relevant unit of observation for benefits receipt in the APS. We match respondents by family unit and construct the same combined occupation measure as the births data, selecting the highest-ranked occupation of members of the family. We restrict our sample to women of approximate childbearing age (16–45) in England and Wales.

In Supplementary Figure S3, we chart the number of women by family size (2 + vs 0–1 existing children) and family occupation (low-income vs high-income) status (high-income vs low-income occupation) from 2012 to 2019. The population

[8] In principle, we could have used the APS alone, since it has all the variables we need (benefits receipt, number and age of children). Unfortunately, however, this is not feasible or desirable in practice. First, sample sizes, while large for most purposes, would not give a large enough sample of families with 2 + children born close to the policy change. Second, there are well-documented concerns about survey non-response, item non-response and self-reporting errors in household surveys, particularly among vulnerable and disadvantaged groups (Gorman et al., 2014). Meanwhile, registering births in England and Wales is a statutory responsibility, so the administrative births data offer comprehensive data on the universe of births over the period. This means the administrative births data are more likely to pick up changes in fertility among vulnerable groups, such as larger families in receipt of benefits.

of women is extremely stable across this period with the exception of 2015, which exhibits a discontinuity in the number of women in smaller families by low-income occupation status. Since the discontinuity reverses in 2016, we do not consider it plausible that this is a genuine reflection of a change in the population of women. We therefore take the mean number of women in each group from 2012–2019 as the relevant denominator. In practice, this means that we are estimating fertility effects assuming that the underlying population of women is relatively stable: Figure S3 suggests that this assumption is justified.

Results

Triple Difference Design

Figure 2 combines the number of births (from the births data) and the number of women (from the APS) to chart four-quarter rolling average fertility probabilities for

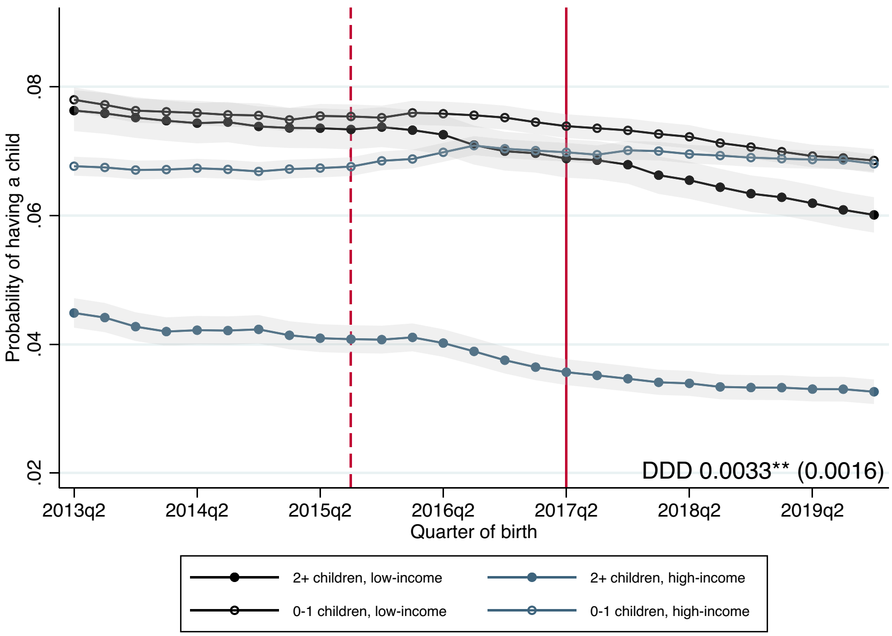

Fig. 2 Fertility trends by family size and family occupation. The figure charts the four-quarter rolling average probability of a woman aged 16–45 having an additional child by quarter of birth, by family size (2 + vs 0–1 existing children) and family occupation (low-income vs high-income) status. A birth is defined as low-income if the highest-ranking occupation of the family is NS-SEC 3–8 and high-income if it is NS-SEC 1 or 2. The number of births by quarter (and hence the probabilities) are annualized. Shaded areas represent 95 percent confidence intervals. The two red lines show the data the policy it was announced and the date it was implemented. The figure also shows the triple difference estimate β_7 from estimating Eq. 2, with standard errors in parentheses. ***p < 0.01, ** p < 0.05, * p < 0.1. Data from ONS administrative birth registry microdata and Annual Population Survey for England and Wales

the four groups from the triple difference design. The underlying numbers of births by quarter are shown in Supplementary Figure S4. Trends among the four groups are broadly parallel prior to the introduction of the policy. This is confirmed by Supplementary Figure S5, which shows there are no significant pre-trends in β_3. As a robustness check, Supplementary Figure S6 shows results when de-seasonalizing the data (residualizing on season fixed effects), instead of the four-quarter moving average; results are almost identical.

Figure 2 seems to indicate a downward trend in births for families with 2+children after Q3 of 2016. Given this, we start in Table S1 by implementing a simple (naïve) Difference-in-Differences (DD) in which we compare fertility trends among individuals in low-income families with 0–1 vs 2+children. We find a small negative effect of 0.49 percentage points, statistically significant at the 1 percent level. Meanwhile, Table S2 repeats the DD exercise but by comparing families with 2+children with low- vs high-income. Here, we find a negative effect of 0.24 percentage points, significant at the 10 percent level. In other words, within the set of low-income families, the fertility of larger families has declined the most, and within the set of larger families, the fertility of low-income families has declined the most.

At first glance, this suggests that there may have been effects of the two-child limit. But this does not follow. In particular, if—as seems plausible—there are wider trends, unrelated to the policy, that affect the fertility of lower income women, and/or reducing the number of higher order births, then this will affect the results of these simple DD estimates. It is clear from visual inspection of Fig. 2 that such wider trends are indeed impacting the results. Inspection of the tables also shows that fertility fell among low-income families with 0–1 children and with high-income families with 2+children. Neither of these falls can be directly attributed to the policy, since neither group was affected by the policy. In order to construct an estimate that is causally attributable to the two-child limit, rather than being driven by broader trends affecting all low-income families, or all families with 2+children, we need to implement a triple difference specification.

In Table 2, we therefore show the results of our main specification, the triple differences (Eq. 2). While the treatment group experience the largest reduction in fertility over the period, the triple difference estimate of the impact of the policy is actually positive (statistically significant at the 5 percent level), albeit small. This is because, as shown in Fig. 2, while births to women with 2+children fell more than births to women with 0–1 children among low-income women, the relative fall was even greater for high-income women, who were unaffected by the policy.

Put another way, while the treatment group did experience the largest fall in fertility of any of the four groups, this fall is more than explained by the sum of the relative fall in higher-order births (regardless of income) and the relative fall in low-income births (regardless of birth order), and neither of these relative falls can be directly attributed to the policy change, which only affected low-income women with 2+children. This appears inconsistent with the hypothesis that the policy has a measurable negative impact on fertility.

Note that the triple difference estimate would have been identical in magnitude if we had taken differences in the reverse order (this follows directly from Eq. 2). That is, if we had first compared low income women with 0–1 children with low income

Table 2 Triple difference design

	2 + children		0–1 children	
	Low-income	High-income	Low-income	High-income
Pre	0.0729	0.0405	0.0753	0.0681
	(0.0006)	(0.0006)	(0.0004)	(0.0004)
Post	0.0634	0.0334	0.0708	0.0693
	(0.0007)	(0.0007)	(0.0005)	(0.0005)
First difference	−0.0095***	−0.0070***	−0.0046***	0.0012**
	(0.0009)	(0.0009)	(0.0007)	(0.0006)
Second difference	−0.0024*		−0.0058***	
	(0.0013)		(0.0009)	
Third difference	0.0033**			
	(0.0016)			

The table shows triple difference estimates of estimating Eq. 2. A birth is defined as low-income if the highest-ranking occupation of the family is NS-SEC 3–8 and high-income if it is NS-SEC 1 or 2. Standard errors are shown in parentheses. ***$p<0.01$, ** $p<0.05$, * $p<0.1$. Data for the numerator (number of births) is from ONS administrative birth registry microdata (N = 514,799); data for the denominator (number of women) is from Annual Population Survey (N = 456,612) for England and Wales

women with 2 + children, and similarly for high-income women, as in Table S1, we would find that while higher-order births to low income women fell more than higher-order births to high-income women, the relative fall was even greater for lower-order births, which were unaffected by the policy.

Close examination of Fig. 2 suggests that it is possible fertility rates among low-income families with 2 + children started to decline after the announcement of the policy in July 2015.[9] If families learned of the two-child limit in July 2015, the earliest effect on fertility would show up 9 months later. To test for this possibility, we run the same triple differences specification as in Eq. 2, but instead of defining the post period based on the introduction of the policy, we define it from 9, 12, 15, or 18 months after July 2015. Figure S7 plots the triple difference estimates from this exercise. The effect sizes are positive in magnitude and are very similar to our main results in Table 2. This suggests that there were no significant anticipation effects of the policy.

These triple difference results are robust to various sensitivity tests. First, we test for robustness of to the inclusion of time trends. As noted above, our triple difference

[9] The policy was announced in the July 2015 budget for implementation in 2017. Some media coverage, particularly in the broadsheet press, stated that the policy would be introduced in 2017 and that "families who have a third child after April 2017 could be caught" (Grice 2015; Slack 2015). However, other media sources did not specify this date, meaning that some families may not have known about the birth cut-off (Whitehouse 2015). On the (strong) assumption that parents had perfect information about the policy announcement, this would give families almost two years to respond. Parents who were considering having a third child might decide to do so quickly in advance of the birth cut-off to ensure they received benefits for that child. Alternatively, parents may have interpreted the policy as a signal of a lack of support for larger families and decided not to have a third child. The direction of these anticipation effects is theoretically unclear.

specification allows for pre-existing trends in fertility rates, provided there is no significant trend in β_3 prior to the policy change. As shown in Supplementary Figure S5, we find no statistically significant pre-trends in β_3. However, visual inspection of Fig. 2 suggests the possibility of a small divergence prior to the policy. Given this, we re-estimate our triple difference specification, allowing for each of our four groups to have a (different) linear trend in the run up to the policy (see, for example Dobkin et al., 2018). As shown in Supplementary Figure S8, our triple difference estimate remains positive, but now becomes statistically insignificant. Again, this points in the direction of a null result.

Second, we test for sensitivity to different categorizations of family occupation. As Figure S1 highlights, within the group we define as “low-income occupation” (NS-SEC 3–8), there is variation in the probability of receiving benefits, with benefits receipt increasing in NS-SEC category. Given this, in Figure S9 we test for sensitivity to the choice of NS-SEC group by excluding NS-SEC 3–5 from the treatment group. In doing so, we are comparing a more “extreme” treatment group of NS-SEC 6–8 to the control group NS-SEC 1–2. Results are very similar to the main results, with the triple difference estimate remaining positive in sign (more positive in fact). Since our results do not change when we focus on groups with higher probabilities of receiving benefits, this suggests that the null result is not driven by measurement error in the likelihood of benefits receipt.

Finally, we test for robustness to excluding childless women from the control group. Intuitively, this exercise makes the treatment group and control group more similar, thereby strengthening the hypothesis that the control group are a valid counterfactual for the treatment group. Figure S10 shows that this exercise has no impact on our triple difference estimate, which remains positive in sign.

Regression Discontinuity Design

Our second identification strategy makes use of the plausibly exogenous variation by date of birth for identification. Our strategy here is to zoom in on the treatment group (low-income women with 2 + children) and estimate a regression discontinuity design in the date of birth of the child. The unit of observation here is the daily probability of having a child.

Table 3 summarizes RD estimates of the effect of the two-child limit on the probability of low-income women with 2 + existing children having an additional child. Column 1 runs a local linear RD for the bandwidth that optimizes the coverage error rate (CER), as recommended by Calonico et al. (2021). This model indicates a decline in the probability of having a child among this group of 1.1 percentage points (16 percent relative to the control mean), which is statistically significant at the 10 percent level.

In Figure S11 of the Supplementary Material, we plot results by date of birth for the specification from Column 1 of Table 3. Prior to the introduction of the policy, at day −3 there is a high outlier, which has potential to generate downward bias in our estimation of the RD effect, thereby exaggerating any decline in fertility. We therefore test for robustness to a series of “donut holes”, by progressively

Table 3 Regression discontinuity design

	Regression discontinuity						Control mean
	(1)	(2)	(3)	(4)	(5)	(6)	
	− 0.011* (0.006)	− 0.011 (0.007)	− 0.011 (0.007)	− 0.007 (0.006)	− 0.014* (0.008)	0.005 (0.012)	0.070
Bandwidth (days)	55.96	56.91	56.91	63.93	58.32	64.38	
N	111	110	110	120	102	86	
CER-optimal	X	X	X	X	X	X	
Donut-hole radius (days)	0	1	2	3	7	21	

The table reports the coefficient β_1 from Eq. 3 for various regression discontinuity models. The regression discontinuity restricts the sample to the treatment group (women with 2 + children in low-income occupations) and exploits the 6 April 2017 date of birth cut-off. The running variable is date of birth relative to 6 April 2017. Column 1 is a local linear RD for a CER-optimal bandwidth, using all available data. Column 2 is a local linear RD for a CER-optimal bandwidth, excluding observations within 1 day of the cut-off (i.e. for a 1-day donut radius); Column 3 does the same for a 2-day radius; Column 4 for a 3-day radius; Column 5 for a 7-day radius; and Column 6 for a 21-day radius. All standard errors are robust bias-corrected and are listed in parentheses. ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Data from ONS administrative birth registry microdata and Annual Population Survey for England and Wales

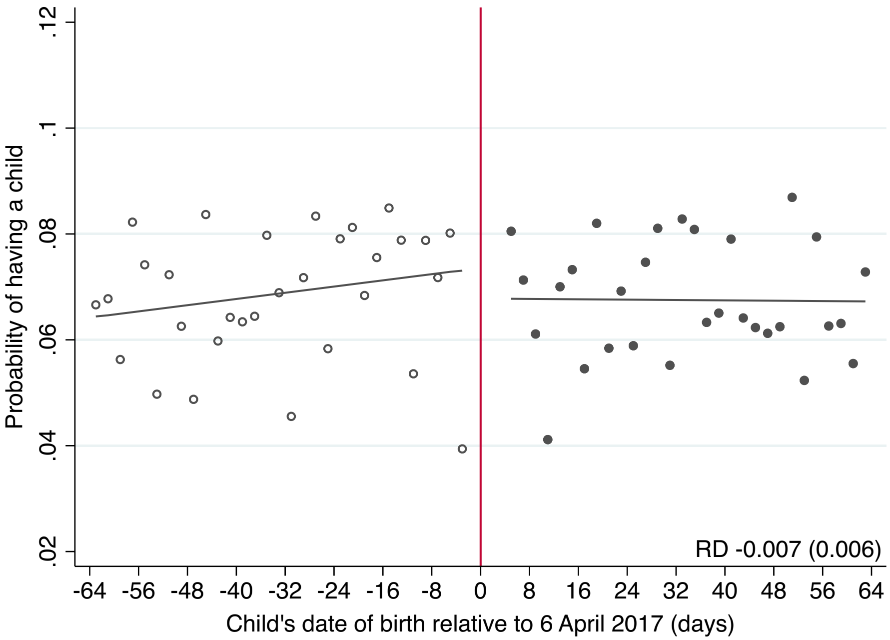

Fig. 3 Regression discontinuity design: low-income women with 2+children around 6 April 2017. The figure charts the annualized probability of low-income women with 2+existing children having an additional child by 3-day DOB bin, for the specification from Column 4 of Table 3 (a donut RD with a 3-day radius). A linear spline is fitted either side of the cut-off. The red vertical line indicates the first date of birth (6 April 2017) when the two-child limit became effective. A birth is defined as low-income if the highest-ranking occupation of the family is NS-SEC 3–8. The shaded areas represent 95 percent confidence intervals. The figure also reports the relevant RD coefficient from Table 3, with associated robust-bias-corrected standard errors in parentheses. ***$p<0.01$, ** $p<0.05$, * $p<0.1$. Data from ONS administrative birth registry microdata and Annual Population Survey for England and Wales

excluding small bands of observations local to the cut-off. Column 2 shows results when excluding a 1-day donut radius; the magnitude of the effect stays the same, but significance is lost. Column 3 excludes a 2-day donut radius, with results unchanged. Column 4 excludes a 3-day radius, which omits the outlier at day −3 and reduces the size of the RD effect to 0.7 percentage points. After this point, larger donuts lead to noisier RD estimates in sign and magnitude: the 7-day donut radius in Column 5 indicates a negative effect of 1.4 percentage points, significant at the 10 percent level, while a 21-day donut in Column 6 suggests an insignificant positive effect of 0.5 percentage points. Taken together, these results suggest a small negative effect of the policy on fertility for births very close to the 6 April cut-off. We select Column 3 as our preferred model on grounds that it is the fairest summary of these results. This indicates a negative, but statistically insignificant, effect of 0.7 percentage points (10 percent relative to the control mean).

Figure 3 illustrates this preferred RD specification graphically, by charting the probability of having a child for the relevant CER-optimal bandwidth (64 days) with a 3-day donut radius (Cattaneo, Idrobo & Titiunik 2019, Hyytinen et al., 2018). Consistent with the above results, a noticeable but small discontinuity is visible.

In Table S3 of the Supplementary Material, we test the robustness of the regression discontinuity to different bandwidths and types of smoothing kernel. Estimates are very similar to our preferred specification across these models, suggesting a negative effect in the region of 0.6–0.9 percentage points. RD estimates are also very similar across different functional forms, as shown for quadratic specifications in Table S4 and for cubic specifications in Table S5.

To test the validity of the regression discontinuity design further, in Figure S12 of the Supplementary Material, we compare the discontinuity experienced by the treatment group with that of the three other groups in the triple difference. Intuitively, credible fertility effects from the two-child limit should present as a negative discontinuity for the treatment group, but not for the other groups. This is precisely what we find: the only group to see a negative effect according to the RD is the treatment group (low-income women with 2 + children); the RD estimate is positive and insignificant for all other groups. This gives some reassurance that our identification strategy is valid.

Finally, we implement placebo cut-off tests at 6 April in four years in which the two-child limit was not implemented: 2013, 2014, 2015 and 2018. This tests whether there is discontinuous seasonal variation around the April cut-off, which could spuriously drive our RD results. As shown in Supplementary Table S6, there are no significant negative effects at any of the placebo cut-offs. This suggests that the small negative RD effect documented in Table 3 is likely to be attributable to the policy.

Discussion

In our triple difference design, our DDD estimates are consistently positive in sign or close to zero, which goes against the hypothesis that low-income women with 2 + children reduced their fertility due to the policy. Meanwhile, our main regression discontinuity estimate suggests a non-significant decline in fertility of 0.7 percentage points (10 percent in relative terms). Taking our results from the triple differences and regression discontinuity design together, we interpret that the introduction of the two-child limit had at most a small negative impact on fertility (it seems highly implausible on theoretical or common-sense grounds that it could have had a positive impact).

It is nevertheless of interest to evaluate the quantitative significance of the potential impacts suggested by our estimates. We therefore take our RD estimate as an upper bound and compute an elasticity of fertility with respect to income by comparing it to the level of benefit cap (the level at which total benefits are capped in the UK, which is 20,000 GBP a year for most of the country)—a useful albeit approximate estimate of income levels among families likely to be affected. This indicates that the loss of income (compared to the previous system) for parents choosing to have a third child is approximately 16 percent.[10] The implied elasticity of fertility with respect to the change in income resulting from having an additional child is therefore 0.6. Given that our triple difference estimate is actually positive (implying, implausibly, a positive fertility effect and hence a negative elasticity) we regard this as very much an upper bound.

By contrast, the elasticity implied by Brewer et al. (2012)—which found an increase in fertility among those affected of more than 15 percent in response to an income increase of about 10 percent—is closer to 1.5. Other papers—notably Milligan (2005) whose triple difference identification strategy is closest to our own—find even larger elasticities, with considerably larger fertility responses in response to much smaller changes in family income. A review of the literature by Stone (2020) found that almost all studies on this question identified positive and significant results, with an increase in the present value of child benefits of 10 percent of a household's annual income producing between 0.5 and 4.1 percent higher birth rates, with Brewer et al. (2012) towards the lower end of the range. Extrapolating Stone (2020)'s methodology, the impact of the two-child limit could easily be in excess of 100 percent of annual household incomes over the life of the child (although it could be less if parental earnings increase sufficiently to reduce benefit entitlement). Overall, our results look very much at the extreme low end of the range found by the literature.

On the face of it, this is surprising. As discussed above, the most closely related research suggests that increases in child-related benefits led to immediate, significant, and large increases in fertility among affected groups, even though (in the case of the changes examined by Brewer et al., 2012) the increases were quite widely spread. Our results, by contrast, imply that large cuts, with large and unambiguous impacts on the financial incentives to have children for the affected group, had at most small and gradual impacts. What could explain this asymmetry?

One possible explanation is that Brewer et al. (2012), in common with the vast majority of studies reviewed by Stone (2020), are estimating the impacts of more generous welfare policies. By contrast, Sandner and Wiynck (2023) look at a German welfare policy change which, as in the UK, resulted in substantial cuts to household income for families with children. They find a fertility reduction of 6.8 percent for a large (18 percent) reduction in household income, implying an elasticity of 0.38, which is not dissimilar from our results. They conclude that welfare recipients' fertility reacts less strongly to financial incentives than the fertility of overall populations.

A further potential explanation is imperfect information about the policy. In the absence of information prior to conception, families may not find out about the policy until they notify the authorities of the new child and their change of circumstances, thus removing the possibility of a fertility response. Recent qualitative research found limited awareness of the policy: approximately half of the participants affected by the two-child limit did not know about the policy when they

[10] In reality, average household income among those affected by the two-child limit is likely to be higher than this, so this may be a slight underestimate.

conceived their third or subsequent child (Patrick & Andersen, 2022). The nature of the qualitative sample—which was restricted to larger families affected either by the two-child limit or the benefit cap—means that there is likely to have been some selection bias (since women who had been deterred from having another child by the policy will have been excluded). Nevertheless, these findings suggest that imperfect information may have contributed towards the limited fertility response. This is consistent with broader research suggesting that levels of information among welfare recipients about the structure of the welfare system are relatively low, and that this may attenuate the impact of financial incentives (Card, 2020).

In some cases, families may find out about the policy during their pregnancy if they have contact with welfare advisers or civil servants (Patrick & Andersen, 2022). In such cases, abortion is the only available fertility response. There are likely to be low numbers of people who become aware of the policy during their pregnancy and are willing to have an abortion directly as a result. A survey of women who had abortions during the pandemic suggested that 57 percent who were likely to be affected by the two-child limit said it was a relevant factor in their decision (British Pregnancy Advisory Service, 2020). However, this was a small survey with a highly selective sample. Abortion microdata is very difficult to access in the UK and published data do not include a breakdown by both the number of previous live births and socio-economic status. Nevertheless, we examined published data on abortions by the number of previous live births in England and Wales (Department of Health & Social Care, 2021). There is no evidence in the data of a substantial abortion response among those with 2+children; existing trends (towards more abortions, and a shift in the age distribution of abortions towards older women) do not appear to have shifted substantially at the time of the policy change, as shown in Figure S13 of the Supplementary Material. While awareness of the policy is likely to grow over time, our data do not indicate an acceleration of fertility effects within the first two years of the policy.

Moreover, qualitative work by Patrick and Andersen (2022) found a disconnect between the policy presentation—which presents people making active choices about a conception—and the everyday reality, which is often more complex and far from frictionless. For instance, Patrick and Andersen (2022) find examples of people becoming pregnant due to contraception failures and in the context of coercive and abusive relationships, where they did not feel able to make use of exemption clauses. Indeed, the exemption clause for domestic abuse requires the claimant to have left the abusive relationship.

Other possible explanations are more speculative. It may be that, in contrast to Laroque and Salanié (2014)'s findings in France, in the UK women considering having a third or subsequent child are *less* responsive to economic incentives in fertility decisions. Religious and cultural factors may be relevant here: the two-child limit disproportionately affects orthodox Jewish and Muslim families (Work & Pensions Select Committee, 2019). Several of the affected families in a qualitative study of the policy had religious beliefs which meant that they reported the two-child limit as having no impact on their conception decisions; adherence to their religious faith to them was more important than the financial incentives at play (Patrick & Andersen, 2022). Indeed, Cohen et al. (2013) found that the ultra-Orthodox Jewish population in Israel were less responsive to a benefit cut affecting larger families.

Conclusion

Over the last thirty years, welfare states in the developed world have become concerned with the potential for moral hazard within the context of welfare and fertility: the assumption that expansions in benefits lead to expansions in fertility, and vice versa. In the 1990s, this motivated several US states to introduce "family caps" under Clinton's program of welfare reform; today it can be seen in the handful of countries that cap means-tested child benefits at the third or fourth child (Stewart & Reeves 2023b). However, until 2017 no country has attempted to cap child benefits at the second child (Stewart, Patrick & Reeves 2023b). Identifying the impact of the UK's two-child limit on fertility therefore offers a unique opportunity to identify the causal effects of capping child benefits by family size.

In this paper we show that capping child benefits at the second child leads to much smaller fertility effects than one might expect. This is surprising given the literature to date, which through the examination of benefits expansions has suggested relatively large elasticities between benefits and fertility (Brewer et al., 2012; Laroque & Salanié, 2014). We are not able to identify the reasons behind this asymmetry with our quantitative data, but qualitative research conducted in parallel with our research (Patrick & Anderson, 2022) suggests that imperfect information, the relative "stickiness" of attitudes towards abortion (and fertility more generally among larger families), and the disconnect between the policy narrative and the everyday realities of fertility decision-making may have been important. This links to a wider evidence base indicating that welfare policy often fails to account for the everyday lived experiences of life on a low income, assuming and anticipating behavioral responses which are not realized (e.g. see Patrick, 2017; Millar & Bennett, 2017; Dwyer, et al, 2022).

What are the policy implications of this? If capping child benefits does not have large impacts on fertility, then the fiscal savings of the policy result almost exclusively from lower payments to poorer families, with only a marginal additional impact from reduced fertility. In having greater household needs and lower work intensity on average, larger families already faced a disproportionate risk of poverty prior to the policy, and child poverty in the UK among this group has increased sharply since 2013–14 (Bradshaw, 2020, Stewart, Patrick & Reeves 2023a). Since the two-child limit does not appear to have significantly reduced fertility in larger families, it appears inevitable that it will increase child poverty further among larger families. Existing estimates suggest that this indeed happening (Child Poverty Action Group, 2022; Chzhen & Bradshaw, 2024). Evidence from qualitative longitudinal research also indicates that the two-child limit is having negative material, emotional and social effects on children in affected households (Andersen, Patrick & Reeves 2023). This has significant implications for inequalities in children's outcomes and development (Cooper & Stewart, 2017).

Finally, in potentially implying an asymmetry between the effects of benefits expansions and benefits reductions, our results have wider policy implications in showing that it cannot simply be assumed that doing the opposite of a policy will lead to equal and opposite effects. The UK government understood and framed the two-child limit as a policy that would have the opposite effects of previous benefits expansions: it would reduce fertility, possibly by a substantial amount (Treasury & DWP 2015). Yet our results suggest a relatively small impact on fertility. This underscores the need for robust causal evidence specific to the policy in question during the policymaking process.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11113-025-09935-5.

Funding We acknowledge funding from the Nuffield Foundation from grant FR-23208. The authors have no relevant or material financial interests that relate to the research described in this paper. This paper uses confidential data maintained by the Office for National Statistics. The data can be obtained by filing a request directly with the ONS. We thank Mike Brewer, Kitty Stewart, Kate Andersen, Aaron Reeves, our Nuffield advisory board, and attendees of a CASE Researchers' Workshop and LSE Social Policy Quantitative Reading Group for helpful comments on previous drafts.

References

Andersen, K., Patrick, R. & Reeves, A. (2023). Needs Matter How the two-child limit and the benefit cap harm children. *Benefit Changes and Larger Families.*

Becker, G. (1960). *An Economic Analysis of Fertility, Demographic and Economic Change in Developed Countries, a Conference of the Universities*. Princeton University Press.

Bradshaw, J. (2020). The two-child limit: impact on abortion. Retrieved from Feb 03 2022 https://bit.ly/3GxNwb9.

Brewer, M., Ratcliffe, A., & Smith, S. (2012). Does welfare reform affect fertility? Evidence from the UK. *Journal of Population Economics, 25*(1), 245–266.

British Pregnancy Advisory Service (2020). Forced into a corner: the two-child limit and pregnancy decision-making during the pandemic. Retrieved from Feb 03 2022 https://bit.ly/35FXzhz.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2021). Optimal bandwidth choice for robust biascorrected inference in regression discontinuity designs. *The Econometrics Journal, 23*(2), 192–210.

Camasso, M. J., & Jagannathan, R. (2016). The future of the family cap: Fertility effects 18 years post-implementation. *Social Service Review, 90*(2), 264–304.

Card, D. (2020). Reforming the financial incentives of the welfare system. Retrieved from Feb 03 2022 https://davidcard.berkeley.edu/papers/finan-incen-welfare.pdf.

Cattaneo, M., Idrobo, N., & Titiunik, R. (2019). *A practical introduction to regression discontinuity designs: Foundations, Elements in quantitative and computational methods for social science*. Cambridge University Press.

Child Poverty Action Group. (2022). Has the two-child limit affected how many children families have? Retrieved from Jun 20 2022.

Chzhen, Y., and Bradshaw, J. (2024). The Two-child Limit and Child Poverty in the United Kingdom. *International Journal of Social Welfare.*

Cohen, A., Dehejia, R., & Romanov, D. (2013). Financial incentives and fertility. *The Review of Economics and Statistics, 95*(1), 1–20.

Cooper, K. & Stewart, K. (2017). Does Money Affect Children's Outcomes? An Update. *Centre for Analysis of Social Exclusion* .

Department of Health and Social Care (2021). Abortion statistics 2020: freedom of information request and other data releases. Retrieved from Feb 03 2022 https://bit.ly/3guQDWH.

Department of Work and Pensions and Her Majesty's Revenue and Customs (2024). Universal Credit and Child Tax Credit claimants: statistics related to the policy to provide support for a maximum of 2 children, April 2022 and previous versions. Retrieved from 19 February 2025 https://www.gov.uk/government/statistics/universal-credit-and-child-tax-credit-claimants-statistics-related-to-the-policy-to-provide-support-for-a-maximum-of-2-children-april-2024/universal-credit-and-child-tax-credit-claimants-statistics-related-to-the-policy-to-provide-support-for-a-maximum-of-two-children-april-2024.

Department for Work and Pensions (2022). Universal Credit. Retrieved from Feb 03 2022 https://www.gov.uk/universalcredit/what-youll-get.

Dobkin, C., Finkelstein, A., Kluender, R., & Notowidigdo, M. J. (2018). The economic consequences of hospital admissions. *The American Economic Review, 108*(2), 308–352.

Dwyer, P., Scullion, L., Jones, K., McNeill, J., & Stewart, A. B. R. (2022). *The impacts of welfare conditionality: Sanctions*. Policy Press.

Dyer, W. T., & Fairlie, R. W. (2004). Do family caps reduce out-of-wedlock births? Evidence from Arkansas, Georgia, Indiana, New Jersey and Virginia. *Population Research and Policy Review, 23*(5/6), 441–473.

Gaffney, D. & Portes, J. (2013). Conservative claims about benefits are not just spin, they're making it up. Retrieved from Feb 03 2022 https://bit.ly/3ox0Rds.

Gaffney, D. (2015). Retrenchment, reform, continuity: Welfare under the coalition. *National Institute Economic Review, 231*(231), R44–R53.

González, L., & Trommlerová, S. K. (2021). Cash transfers and fertility: How the introduction and cancellation of a child benefit affected births and abortions. *Journal of Human Resources, 58*, 220.

Gorman, E., Leyland, A., McCartney, G., White, I., Katikireddi, S. V., Rutherford, L., Graham, L., & Gray, L. (2014). Assessing the representativeness of population-sampled health surveys through linkage with administrative data on alcohol-related outcomes. *American Journal of Epidemiology, 189*(9), 941–948. https://doi.org/10.1093/aje/kwu207

Grice, A. (2015). Budget 2015: Tax credits and housing benefit to be cut for families with more than two children. Retrieved from Feb 03 2022 https://bit.ly/3ryWHUB.

Hyytinen, A., Meriläinen, J., Saarimaa, T., Toivanen, O., & Tukiainen, J. (2018). When does regression discontinuity design work? Evidence from random election outcomes. *Quantitative Economics, 9*(2), 1019–1051.

Jensen, T., & Tyler, I. (2015). "Benefits broods": The cultural and political crafting of anti-welfare commonsense. *Critical Social Policy, 35*(4), 470–491.

Joyce, T., Kaestner, R., Korenman, S., & Henshaw, S. (2004). Family cap provisions and changes in births and abortions. *Population Research and Policy Review, 23*(5/6), 475–511.

Kearney, M. S. (2004). Is there an effect of incremental welfare benefits on fertility behavior? A look at the family cap. *Journal of Human Resources, 39*(2), 295–325.

Laroque, G., & Salanié, B. (2014). Identifying the response of fertility to financial incentives. *Journal of Applied Econometrics, 29*(2), 314–332.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*(2), 281–355.

Millar, J., & Bennett, F. (2017). Universal credit: Assumptions, contradictions and virtual reality. *Social Policy & Scoiety, 1*(2), 1–14.

Milligan, K. (2005). Subsidizing the stork: New evidence on tax incentives and fertility. *The Review of Economics and Statistics, 87*(3), 539–555.

OECD (2021). Purchasing power parities. Retrieved from Feb 03 2022 https://bit.ly/3JbkH5Z.

Office for National Statistics (2016). Birth registrations, England and Wales Microdata Metadata. Retrieved from Feb 03 2022 https://bit.ly/3rAkYd1.

Office for National Statistics. (2022). Births in England and Wales: 2021. Retrieved from Feb 03 2022.

Olden, A., & Møen, J. (2022). The triple difference estimator. *The Econometrics Journal, 25*(3), 531–553. https://doi.org/10.1093/ectj/utac010
Patrick, R. (2017). *For whose benefit? The Everyday realities of welfare reform*. Policy Press.
Patrick, R., & Andersen, K. (2022). The two-child limit and fertility decision making: When policy narratives and lived experiences collide. *Social Policy & Administration*. https://doi.org/10.1111/spol.12877
Reader, M., & Andersen, K. (2022). Size matters: The experiences of larger families on a low income during COVID-19. In K. Garthwaite, R. Patrick, A. Tarrant, & R. Warnock (Eds.), *COVID-19 Collaborations: Researching poverty and low-income family life during the pandemic.* Policy Press.
Reed, H. & Portes, J. (2015). Cumulative impact assessment of tax and welfare reforms. *Equality and Human Rights Commission.* Retrieved from Feb 03 2022.
Riphahn, R. T., & Wiynck, F. (2017). Fertility effects of child benefits. *Journal of Population Economics, 30*(4), 1135–1184.
Sandner, M., & Wiynck, F. (2023). The fertility response to cutting child-related welfare benefits. *Population Research and Policy Review, 42*(2), 25.
Slack, J. (2015). Now the taxpayer won't fund big families: Tax credits to be limited to two children to cut bill that's soared to £30billion. Retrieved from Feb 03 2022 https://bit.ly/3J9OxIg.
Social Mobility Commission (2022). https://www.gov.uk/government/publications/state-of-the-nation-2022-a-fresh-approach-to-social-mobility/state-of-the-nation-2022-chapter-2-mobility-outcomes
Stewart, K., Patrick, R., & Reeves, A. (2023). A time of need: Exploring the changing poverty risk facing larger families in the UK. *Journal of Social Policy, 54*, 1–25.
Stewart, K., Patrick, R., & Reeves, A. (2023). The sins of the parents: Conceptualising adult-oriented reforms to family policy. *Centre for Analysis of Social Exclusion*. https://doi.org/10.1177/09589287241290739
Stone, L. (2020). Pro-natal policies work, but they come with hefty price-tag. Retrieved from Jul 11 2022 https://bit.ly/3z5yMjv.
Treasury, H. M. (2015). Summer budget 2015, hc 264, paras 1.141–1.150.
Treasury, H. & DWP (2015). Welfare reform and work bill: Impact assessment of tax credits and universal credit, changes to child element and family element. Retrieved from Feb 03 2022 https://bit.ly/3HD99bg.
UK Government (2021). PM's speech on Welfare Reform Bill. Retrieved from Feb 03 2022 https://bit.ly/3gm9Pps.
Whitehouse, H. (2015). Budget 2015: Child tax credits to be limited to two children after 2017. Retrieved from Feb 03 2022 https://bit.ly/3rBvMYm.
Work and Pensions Select Committee (2019). The two-child limit: Third report of session 2019. Retrieved from Feb 03 2022 https://bit.ly/3uwvTWX.