I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18193

The Effects of Maternity Leave Benefits on Mothers and Children. A Reexamination

Otto Sevaldson Lillebø
Simen Markussen
Knut Røed

OCTOBER 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18193

The Effects of Maternity Leave Benefits on Mothers and Children. A Reexamination

Otto Sevaldson Lillebø
Nordic Institute for Studies in Innovation, Research, and Education

Simen Markussen
The Ragnar Frisch Centre for Economic Research and IZA

Knut Røed
The Ragnar Frisch Centre for Economic Research and IZA

OCTOBER 2025

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 Effects of Maternity Leave Benefits on Mothers and Children. A Reexamination*

We provide a full reexamination of the effects of a maternity leave extension implemented in Norway in 1977. Previous research reporting large favorable long-term effects on mothers' health and on offspring's educational and labor market outcomes relied on an incorrect description of the reform and an invalid identification strategy. In the present paper, we show that the previously reported results are misleading. Building on an accurate description of the reform and its implementation, we document that it had no noticeable long-term effects on mothers' health or on offspring's education and labor market outcomes.

JEL Classification: C52, J13, J18
Keywords: family policies, maternity leave, replication

Corresponding author:
Knut Røed
The Ragnar Frisch Centre for Economic Research
Gaustadalléen 21
0349 Oslo
Norway

E-mail: knut.roed@frisch.uio.no

* Thanks to Bernt Bratsberg, Anna Godoy, Erik Grönqvist, Sverre Kittelsen, Oddbjorn Raaum, and Yuejun Zhao for good discussions and valuable comments to previous drafts. We acknowledge financial support from the Research Council of Norway, grant number 300917. Administrative registers made available by Statistics Norway have been essential.

1 Introduction

There is now an extensive research literature on the impacts of parental leave policies. A consensus view seems to be that while the very first weeks of paid parental leave have large favorable long-term effects on both maternal health and offspring's educational and labor market outcomes, there are small or no additional gains from extensions beyond those first weeks (Rossin-Slater, 2017; Almond, Currie, & Duque, 2018; Regmi & Wang, 2023; Dahl & Loken, 2024).[1] But while the failure to find significant long-term effects of parental leave *extensions* rests on the analysis of several different policy reforms (Rasmussen, 2010; Dustmann & Schönberg, 2012; Baker & Milligan, 2015; Dahl et al., 2016; Danzer & Lavy, 2018; Behaghel & Pinto, 2024), the reported large favorable effects of its *first weeks* essentially rely on two papers, both based on the same reform implemented in Norway in 1977 (Carneiro, Løken, & Salvanes, 2015; Bütikofer, Riise, & Skira, 2021). In both papers, the 1977-reform is studied under the presumption that it *introduced* paid maternity leave in Norway, with a sharp discontinuous increase in the number of paid weeks from 0 to 18.

The findings of the two papers indicate large favorable effects that likely have significant policy implications. According to Carneiro, Løken, and Salvanes (2015), the children of the treated mothers became 1.9 percentage points less likely to drop out of high school and 2 percentage points more likely to attend college. They also earned 5 percent higher wages at age 30. According to Bütikofer, Riise, and Skira (2021), measured at age 40, the mothers themselves became 3 percentage points less likely to be obese, 5 percentage points less likely to smoke, and 3 percentage points less likely to experience hypertension.

In a recent comment article, published in the Journal of Political Economy, Lillebø et al. (2024) document that the Norwegian 1977-reform did *not* introduce paid leave, but merely extended it from 12 (or 13) to 18 weeks. The perception in the literature of favorable effects from the introduction of paid leave is thus largely unfounded.[2] Lillebø et al. (2024) also document that the treatment-control differential, used by both Carneiro, Løken, and Salvanes (2015) and Bütikofer, Riise, and Skira (2021) did not exist. The reform was not implemented as a discontinuous change on July 1, but instead introduced in a

[1] There is also an extensive literature focusing on effects on parents' mental health, indicating favorable short-term effects for mothers; see Heshmati, Honkaniemi, and Juárez (2023) for a recent review.

[2] The only paper we are aware of that examines effects of the introduction of paid parental leave (from a level of zero weeks) on offspring education and employment is Regmi and Wang (2025). Based on exogenous variations in the US state provision of paid maternity leave resulting from the Pregnancy Discrimination Act of 1978, they present evidence indicating a positive effect on college graduation for the first affected offspring cohort. However, they find no significant effects on high school graduation or employment, and the effect on college graduation also appears to fade out for subsequently treated cohorts.

staggered fashion over several months, depending on the mother's sector of employment. In reality, the main treatment and control groups thus had the exact same maternity leave entitlements.[3]

In a reply, Carneiro, Løken, and Salvanes (2024) acknowledge the errors but claim that they are not important for the conclusions of their paper. They show that they can find similar effect estimates when using a model that takes into account the staggered implementation of the reform. And, as a response to several errors documented by Lillebø et al. (2024), they show that their original results are also robust to a large number of modifications based on updated and corrected data.[4] However, *none* of the many new results reported by Carneiro, Løken, and Salvanes (2024) are based on a model that *both* takes the staggered reform implementation into account *and* uses the corrected samples and outcomes, which is arguably the only valid approach.

Nevertheless, the conclusion in Carneiro, Løken, and Salvanes (2024) is that the "flying start" identified for the children of treated mothers is valid and "still flying". Together with the fact that the reported finding of health effects for mothers (Bütikofer, Riise, & Skira, 2021) has neither been retracted nor challenged, the impression left in the research literature thus seems to be that the Norwegian 1977-reform had large favorable long-term effects both for mothers and their children.

The confusion around the interpretation of these findings also seems to persist. For example, in the chapter on "Families, public policies, and the labor market" in the new Handbook of Labor Economics, the apparent favorable effects of the Norwegian 1977-reform are still attributed to the reform representing the *introduction* of parental leave as opposed to a moderate extension (Dahl & Loken, 2024, p. 608).

In the present paper, we provide a full reexamination of the effects of the Norwegian 1977-reform using a comprehensive set of labor market and health outcomes covering both the mothers and their children. Our reassessment is based on an empirical strategy explicitly taking into account that the reform was introduced at different points in time for mothers working in different sectors of the economy (state sector, municipal sector, and private sector) and also phased in retroactively for mothers who gave birth up to 12 weeks before the respective implementation dates. To facilitate such an analysis, we have collected data from the 1970 and 1980 censuses to identify the mothers' sector of employment. We

[3]As explained in Lillebø et al. (2024), the authors were informed about these errors in February/March 2020.

[4]Lillebø et al. (2024) document errors both in the construction of treatment and control groups (eligibility criteria for paid leave) and in the definition of the earnings outcome (with earnings measured at different ages for the different cohorts included in the analysis.

have also used updated administrative register data to provide a more comprehensive set of long-term outcomes, reflecting health consequences and labor market success for both mothers and offspring. Given the absence of sharp discontinuities, our causal analysis is mainly based on a difference-in-differences (DiD) strategy. In essence, we compare the differences in outcomes for eligible mothers (and their children) giving birth (or being born) just after versus just before the relevant eligibility cutoff dates in the year of reform with the corresponding differences in previous non-reform years.

Our results do *not* confirm that the Norwegian parental leave reform in 1977 had long-term effects on health and labor market outcomes for either the mothers or their children; nor do we find any effects on offspring's educational attainment. We estimate effects on future employment and earnings (during age 50–60 for mothers and age 30–40 for offspring) that are very close to zero and consistently statistically insignificant. The degree of precision allows us to rule out large favorable effects of the scale reported in previous studies. Similar zero-effects are found when we instead look at the incidence of disability or death before age 60 (for mothers) or age 40 (for offspring). For offspring, we also estimate the impacts on educational attainment (number of non-compulsory education years), again finding zero effects.

To investigate the sources of the discrepancy between our results and those reported by Carneiro, Løken, and Salvanes (2015) and subsequently confirmed by Carneiro, Løken, and Salvanes (2024), we perform a replication and sensitivity analysis. We show that while the results in Carneiro, Løken, and Salvanes (2015) are largely robust with respect to *either* correcting the model *or* the data, they are not robust to correcting both at the same time - the one robustness analysis missing in Carneiro, Løken, and Salvanes (2024).

To ensure transparency, as part of our replication-package we publish online a version of the data used in the present paper where individual observations are aggregated into week-of-birth cells. While ensuring anonymity, this version of the data preserves the essential identifying information, facilitating a close replication of our results, as well as critical reassessments based on alternative empirical strategies.

Our findings align well with the existing literature on other extensions of parental leave, including the analysis of a similar reform in Denmark (Rasmussen, 2010) and of several further extensions of parental leave in Norway (Dahl et al., 2016). Given the true content of the evaluated reform, this does, of course, not rule out that there could be favorable effects of parental leave *introduction*, combined with decreasing returns to additional weeks. But on that matter, the results provided by Carneiro, Løken, and Salvanes (2015) and Bütikofer, Riise, and Skira (2021) carry no information.

2 The Norwegian maternity leave reform and the revised causal analysis

On July 1, 1977, the legislated entitlement to paid leave for Norwegian mothers was extended from 12 weeks (13 weeks in the public sector) to 18 weeks, with no change in the replacement rate.[5] In practice, the 5–6 week extension was rolled out over three different periods during 1976-1977, depending on the mother's sector of employment (Lillebø et al., 2024). The "statutory" implementation dates were January 1, 1977, in the state sector and in the municipality of Oslo, May 1, 1977, in the rest of the municipal sector, and July 1, 1977, in the private sector. In all three cases, the entitlement to extended paid leave was provided retroactively to mothers who had recently given birth. In the first two cases (that is, in the public sector), the reform was introduced so that mothers who were already on paid leave at the time of implementation were also entitled to the extra weeks. Although it is difficult to establish exactly how this rule was practiced almost 50 years ago, it appears to imply that entitlement to the extra weeks could depend on exactly when the mothers had chosen to start their maternity leave period. As it was common to start the leave up to six weeks prior to expected delivery, we reckon that the effective birth cutoff dates for entitlement to the extended benefits were around November 13, 1976, in the state sector, and March 13, 1977, in the municipal sector. In the private sector, the reform was introduced so that all mothers who gave birth in the last 12 weeks before July 1 were eligible for the extension. Hence, in this case, the effective reform date with respect to the timing of birth was April 9, 1977.

Both before and after the reform, the requirement for eligibility was that the mother had been employed at least six of the last 10 months prior to expected delivery. Due to the difficulty of verifying this precisely at the individual level with today's data, eligibility must in practice be proxied using observed annual earnings. However, we know from official statistics (e.g., Statistics Norway, 1979, p. 327) that approximately 45 percent of all mothers who gave birth in 1977 were eligible for paid maternity leave.[6]

A point of some importance in this context is that also ineligible mothers (mothers without sufficient employment prior to delivery) were subjected to an important reform

[5]For most workers the replacement rate was close to 100 percent, but the legislated minimum level was considerably lower. One year later (July 1, 1978) it was raised to 100 percent for everyone.

[6]This contrasts with the eligibility criteria used by Carneiro, Løken, and Salvanes (2015), Bütikofer, Riise, and Skira (2021), and Carneiro, Løken, and Salvanes (2024), which implied the inclusion of approximately 65 percent of the mothers of 1977 in the treatment group. This discrepancy was originally caused by a coding error, whereby the eligibility threshold for earnings the year before birth was inadvertently set to a level corresponding to approximately 4% of the average full-time annual earnings; that is, well below earnings levels predictive of at least six months of employment.

in 1977, namely an increase in the one-time cash support provided to *ineligible* mothers from 6 to 20 percent of the Basic amount (BA), implying a raise from NOK 804 to NOK 2680.[7] In contrast to the paid maternity leave reform for eligible mothers, this reform *was* implemented in the form of a sharp discontinuity; that is, it applied for mothers who gave birth from July 1 only.

In their new analysis, Carneiro, Løken, and Salvanes (2024) use the correct April 9 cutoff date to provide revised estimates of the reform's effect on the children of eligible mothers, again based on a sort of difference-in-regression-discontinuity approach. They argue that some mothers might have misunderstood the reform in the same way as they themselves did, and thus failed to claim the extra six weeks of paid leave to which they were entitled. Consequently, they exclude from the sample all births occurring between April 9 and June 30. Given that they do not have information about the mothers' sector of employment, they disregard the fact that public sector employees were unaffected by the April 9 - July 1 roll-out, arguing (correctly) that this error can only attenuate the true reform effects. In combination with the fact that their eligibility proxy implies that at least 30 percent of their treatment group must have been ineligible due to insufficient employment, it is clear that the April 9 date cannot have been relevant for more than half of their treatment group. Their findings essentially confirm the previously reported results, although all point estimates are slightly smaller than those obtained when the wrong cutoff date was used.

A challenge with the analysis of the 1977 parental leave reform is that there are no data on actual leave behavior during the relevant time period; hence, there is no "first stage". Carneiro, Løken, and Salvanes (2015) argue that we can still be sure that paid leave did not substitute for unpaid leave, so that the time with the child increased in line with paid leave. Their primary basis for this claim is a set of regressions indicating that the reform did not affect the mothers' income in the years surrounding the reform. They argue that a reduction in unpaid leave should have led to an increase in observed income in the year after birth, since maternity leave benefits were part of their income measure. The income measure they used is the gross pension-qualifying earnings, which is in fact the only register-based income measure available for the period in question. However, in 1977, this measure did *not* include maternity leave benefits. Maternity leave benefits became part of pension-qualifying earnings (and taxable) from July 1, 1978; that is, one year later. Hence, if the reform led to an increase in total time spent with the child, we

[7] Inflated to todays value, this would amount to an increase from NOK 7800 to NOK 26000 (or from approximately 780 to 2600 USD). The Basic amount is an important parameter in the Norwegian social insurance system. It is adjusted each year according to the average wage growth. 1 BA is now (1 May 2025–30 April 2026) equal to NOK 130,160, which corresponds approximately to 17% of the average full-year full-time earnings in Norway.

should have seen a *decline* in incomes captured by pension-qualifying earnings in the year of birth and/or the year after.[8] The fact that Carneiro, Løken, and Salvanes (2015) did not find such a negative effect can, of course, not be taken as evidence *against* a positive effect on time spent with the child, since their analysis in any case was misspecified. However, a revised "first-stage" analysis, based on a correct interpretation of available income measures, seems to be an important ingredient of a full reexamination.

3 Data and empirical strategy

To reexamine the impacts of the 1977 reform, we make use of updated data on earnings and education, implying that we can evaluate children's outcomes up to a higher age. For the mothers, we do not have access to the survey-based health outcomes used by Bütikofer, Riise, and Skira (2021). Instead, we investigate health outcomes by exploiting register-based information on disability and mortality. In addition to administrative register data, we also include records from the 1970 and 1980 censuses in Norway to provide information about sector of employment, such that we can exploit the three different roll-out dates to efficiently estimate the effects of the reform.

A first task is to identify the mothers who were eligible for paid maternity leave and, therefore, were treated with the 5–6-week extension. We lack explicit information on the length of individual employment spells in the 1970s, preventing us from directly verifying if mothers met the employment criterion of having worked at least six of the ten months before their expected delivery. As in previous work, we use a proxy based on an annual earnings threshold in the last calendar year before birth. Since the probability of belonging to the group of eligible mothers is likely to be monotonically increasing in observed pre-birth earnings, we choose our threshold such that we match the fraction known from publicly available statistics to have been eligible (45%). This establishes an earnings threshold equal to 1.62 Basic amounts, corresponding to approximately 27% of average full-time earnings.

A second task is to identify the time at which eligible mothers became entitled to the extra 5-6 weeks, as determined by their sector of employment. Since the sector of employment is inferred from data collected in 1970 and 1980, it will not cover all workers and will also be identified with some error. We predict the sector of employment at the time of birth based on a combination of industry codes and occupation codes reported in the two censuses. Such codes will, of course, be missing for people who were not employed in the

[8] Anecdotal evidence suggests that maternity leave benefits were sometimes paid out in the form of regular wages in the public sector, implying that they became pension-qualifying after all. Our own findings presented in Section 5 appear to be consistent with such a practice being widespread.

year of the census. Based on census information, we assign the employment sector for mothers in our data in the following way:

- If the sector of employment is the same in 1970 and 1980, or if a sector is reported for one of the years but missing for the other year, we assign that sector as the sector of employment.
- If no sector of employment is reported for either 1970 or 1980, or if different sectors are reported for the two years, we interpret sector information as missing.

Using this strategy, we identify the sector of employment for 70% of eligible mothers. These mothers (and their children) constitute the main sample that we use to estimate the effects of extended paid leave.[9]

Given that extended parental leave was phased in without a sharp discontinuity, our preferred empirical strategy for the evaluation of this part of the reform is based on a difference-in-differences (DiD) approach rather than on regression discontinuity. The first difference is the difference in outcomes between eligible mothers/children who gave birth/were born just before and just after implementation of the 5–6 week extension. Since implementation was phased in with retroactive elements, we operate with two alternative before-after-definitions, which we refer to as "statutory" and "actual" implementation, respectively. Under "statutory implementation", we compare births from periods after the 1977 statutory cutoff dates (January 1, May 1, July 1) with those from more than 12 weeks prior (that is, before October 9, 1976, February 6, 1977, or April 9, 1977). With "actual implementation", we take the retroactive implementation into account and compare births occurring in a period after the predicted birth dates for actual entitlement to the extra weeks (November 13, 1976, March 13, 1977 or April 9, 1977) with the same before-treatment periods.

It seems plausible that the reform's influence on total leave length (including unpaid leave) may have been greater for mothers who knew about the extension well before they started their leave. As we document in Lillebø et al. (2024), the reform was announced well ahead of its implementation, yet its exact implementation dates and eligibility requirements appear to have been undecided until short before. We are not able to ascertain exactly when the mothers in the different sectors were aware of their own extended entitlements

[9] As an auxiliary exercise, we have used more recent register data, where we have access to annual information about sector of employment, to shed some light on the quality of our sector assignment strategy. Specifically, we have assigned sector of employment to eligible mothers giving birth in 2000-2002, based on data in 1995 and 2005, following the procedure outlined above, and then compared the results with the correct sector of employment identified in same-year data. We then find that the correct sector is identified in 86% of the cases.

and how that affected their choice of total leave length. By estimating the model based on both the statutory and actual cutoff dates, we may be able to capture differences in the way the reform affected the overall length of leave.

In the baseline version of the model, we follow Carneiro, Løken, and Salvanes (2015, 2024) and Bütikofer, Riise, and Skira (2021), in that we include 12 weeks on either side of the respective cutoff dates. We call this period – from 12 weeks before to 12 weeks after the relevant cutoff dates – the "Reform window". To eliminate any influence from week-of-birth effects, these differences are then compared with the corresponding differences applying for mothers/children subjected to the same cutoff dates in neighboring non-reform years. We call these periods the "Control windows".[10] In our baseline specification, we use two control windows occurring exactly one and two years prior to the reform window and estimate equations of the following form[11]

$$y_{wsi} = \alpha_{ws} + \beta\, Post_s + \gamma\,(Post \times Reform\ window) + \sum_g \delta_g X_{gi} + \varepsilon_{wsi}, \qquad (1)$$

where α_{ws} is a window-by-sector specific effect and $Post_s$ is a sector-specific (and $Post$ a sector-independent) indicator for births occurring after the respective cutoff dates within each window. The control variables X_{gi} include the mother's log(earnings) the year before childbirth, her age at childbirth (dummy-coded), the number of children in age brackets 0-1 (not including the new child), 2-5 and 6-12, respectively, and the focal child's sex. The coefficient of interest is γ.

We also present in Section 6 a number of robustness analyzes where we look at mothers working in the private and public sectors separately and use some alternative strategies for sector identification. Given the uncertainty regarding the correct classification of sectors, we also present results from a difference-in-differences model where we use the complete sample of eligible mothers (not only the 70% with identified sector) and compare births that occurred strictly after *all* sectors had been treated with births that occurred strictly before *any* sector had been treated. In the robustness section, we also include models with alternative choices of eligibility income thresholds, window sizes, and control periods.

For ineligible mothers, the reform implied a 235 percent increase in the level of the one-

[10] The reason why we use the terms "Reform window" and "Control windows" rather than "Reform year" and "Control years" is that some of the windows stretch into more than one calendar year.

[11] We avoid using post-reform control windows, both because the reform may have changed behavior in a more dynamic fashion and because there was a new important reform July 1, 1978, whereby the replacement ratio in the parental leave program was increased to 100 percent for everyone up to a relatively large threshold.

time cash transfer. This increase did not directly translate into more time with the child; hence it can be used to evaluate the impacts of improved economic conditions around the time of childbirth more generally. The increase in the transfer was small in absolute terms, however, implying that we also expect any long-term effects to be small, and thus potentially undetectable with our data. Since this part of the reform was implemented as a sharp discontinuity for births occurring after July 1, we use the difference in regression discontinuity design (DiRD) to study the impact of this reform. We continue to use a reform window of 12 weeks on both sides of the cutoff date, and allow day-of-year-slopes to vary arbitrarily on each side of the cutoffs. For the DiRD model, we thus estimate equations of the following kind:

$$
\begin{aligned}
y_{wi} = \alpha_w &+ \beta\, Post + \gamma\,(Post \times Reform\ window) \\
&+ Day(\pi + \phi\, Post) + \sum_g \delta_g X_{gi} + \varepsilon_{wi},
\end{aligned}
\tag{2}
$$

where Day is the running variable measured as the day of birth minus July 1. Equation (2) is estimated with triangular weights to give more weight to observations closer to the cutoff.

Also for this analysis, we present a number of robustness exercises in Section 6, including alternative eligibility thresholds and window sizes and some alternative model specifications (including regression discontinuity and difference-in-differences).

4 Outcomes and descriptive statistics

For eligible mothers, we first construct an outcome variable designed to capture the effect of the reform on the total length of leave; that is, on time spent with the child. If we think of any effects on long-term outcomes as being mediated by the length of leave (and not by the extra money as such), this is the closest we get to a first-stage analysis. As a proxy for the length of leave, we use the sum of recorded incomes in the year of birth and the following year. We note that each additional week of leave corresponds to a reduction in registered pension-qualifying income over these two years approximately equal to 1%. Hence, if total leave length increased in line with paid leave, we expect a negative effect on *registered* income for the year of birth and the year after amounting to approximately 5–6 percent. On the other hand, if the total length of the leave did not increase at all, we expect an effect close to zero. In the latter case, any long-term effects of the reform can be attributed to the improved economic conditions experienced by the household; similar to any effects identified for ineligible mothers (who received the increased lump-sum cash transfer).

To assess the long-term effects for the mothers, we construct the following three main outcomes:

- Log of total labor earnings during age 50–60 (conditional on survival to age 60 and at least one year of employment in the outcome period).
- Number of years employed during age 50–60 (conditional on survival to age 60).[12]
- Indicator for having died or received some form of disability insurance benefit before age 60.[13]

Aggregating labor earnings over several years, rather than focusing on a specific age, has the advantage that it becomes a better indicator of overall long-term labor market success. In addition, it reduces the fraction with zero earnings that creates difficulties for the log-specification of earnings. Still, in our data, there are 4% of mothers alive at age 60 who are not employed throughout the age 50–60 period and therefore cannot be used to examine earnings effects. This motivates the use of the number of employed years as a separate outcome. Finally, to capture any health effects on mothers, we use an indicator for disability insurance receipt or death before age 60.

For the children, we use similar outcomes as for the mothers for labor market success and health, assessed at a 20 year lower age, and add an outcome capturing educational attainment:

- Log of total labor earnings during age 30–40 (conditional on survival to age 40 and at least one year of employment in the outcome period).
- Number of years employed during age 30–40 (conditional on survival to age 40).
- Indicator for having died or received some form of disability insurance before age 40.
- Number of non-compulsory education years obtained before age 40.

Among the children, approximately 3% of the sample is "lost" in the analysis of earnings effects due to non-employment throughout the age 30-40 period.

Table 1 provides descriptive statistics for the samples of eligible mothers and their children, divided into treatment and control windows, and further into the three periods defined by the cutoff dates corresponding to actual and statutory implementation. The

[12]We define a person as employed in a given year if labor earnings that year exceeded 1 Basic amount.

[13]The disability insurance (DI) outcome includes both temporary and permanent DI. To receive a DI benefit in Norway, the work capacity must be reduced by at least 50% due to impaired health, and the impaired health status must be confirmed by a certified physician and approved by the social insurance administration.

table shows that the predetermined outcomes are balanced across the different periods, and long-term outcomes also seem to develop similarly over the treatment and control windows. The earnings of mothers in the year of birth and the year after are higher for births occurring in the post periods, most likely reflecting that mothers giving birth later in a calendar year have more time to work and earn income *prior* to delivery. It is notable, however, that this particular pattern is somewhat different in the treatment and control windows, possibly indicating an effect on total leave length.

For eligible mothers, we note that approximately 30% were disabled or deceased before age 60. Although death is a very rare outcome in our data (only 3.3% of the sample), we have chosen to merge it with the disability outcome, since mortality is closely related to health. For those still alive, average employment during the 11-year period from age 50 to 60 was relatively high, with around 9.5 employment years in all data windows. For the offspring, these outcomes are measured at a 20 year lower age. 11-12% of the offspring were disabled or deceased before age 40 (approximately 1.3% were dead), and the average number of years employed from age 30 to age 40 was approximately equal to 10 in all data windows.

Table 2 provides similar statistics for the samples of ineligible mothers; that is, mothers that were exposed to the increased lump-sum transfer. Again, the predetermined outcomes appear to be balanced, and the patterns of long-term outcomes are similar for the treatment and control windows. It is notable that ineligible mothers are quite strongly negatively selected in terms of health and employment. More than 40% were dead or disabled before age 60 and the average number of employment years age 50-60 for the survivors was around 8; that is, 1.5 years less than for eligible mothers. This adverse selection also spilled over to the offspring, where approximately 16% were disabled or deceased before age 40.

Figure 1 gives a more detailed account of the pattern of long-term outcomes for eligible and ineligible (not employed) mothers who gave birth during the periods surrounding the actual and statutory cutoff dates, which are marked with vertical dashed lines. Figure 2 provides the corresponding long-term outcomes for their children. A visible inspection of these figures reveals no clear indication of reform-related shifts in any of the long-term outcomes around the cutoff dates – for either mothers or offspring.

Table 1: Main sample of eligible mothers and their children: Descriptive statistics based on 12-week long pre-treatment and post-treatment windows.

	Treatment window			Control windows		
	I. Before	II. Actual	III. Statutory	IV. Before	V. Actual	VI. Statutory
A. Mothers, pre-birth information						
Number of observations	3,392	3,567	3,420	7,037	7,594	7,197
Age at time of childbirth	26.872	26.960	26.921	26.400	26.617	26.433
Earnings year before child birth (BAs)	3.652	3.672	3.685	3.493	3.574	3.551
Log earnings year before child birth	1.232	1.238	1.245	1.191	1.214	1.208
Fraction private sector	0.643	0.625	0.608	0.663	0.630	0.621
Number of children prior to current birth	0.518	0.519	0.492	0.427	0.464	0.447
B. Mothers, outcomes						
Average earnings birth year and year after (BAs)	2.106	2.158	2.192	1.874	2.042	2.087
Log earnings, birth year and year after	0.292	0.415	0.495	0.105	0.306	0.391
Disabled or dead, before age 60	0.303	0.296	0.296	0.308	0.298	0.309
Years employed, age 50-60	9.668	9.646	9.618	9.485	9.537	9.477
Average labor earnings, age 50-60	4.368	4.341	4.299	4.196	4.260	4.205
Log average labor earnings, age 50-60	1.284	1.298	1.264	1.234	1.251	1.247
C. Children, outcomes						
Disabled or dead, before age 40	0.116	0.121	0.116	0.113	0.117	0.121
Completed years of non-comp. education	4.902	4.866	4.889	4.770	4.784	4.783
Years employed, age 30-40	9.989	9.906	9.911	9.868	9.874	9.805
Average labor earnings, age 30-40	5.644	5.696	5.702	5.589	5.593	5.531
Log average labor earnings, age 30-40	1.558	1.535	1.546	1.525	1.534	1.506

Notes: The "actual" treatment/control windows cover the 12-week period after entitlement to the extra paid weeks became available for those who had already started their leave (November 13, March 13, or April 9). The "statutory" treatment/control windows cover the 12-week period after full implementation (January 1, May 1, or July 1. The "before" treatment/control windows cover the 12-week period just before anyone had become entitled to the extension. The number of observations vary slightly between outcomes, as employment is measured conditional on survival to age 60 (mothers) or 40 (children) and log earnings are measured conditional on having positive earnings in at least one of the years included. Observation numbers used in the analysis for each outcome are specified together with the estimation results reported in Table 3.

Table 2: Main sample of ineligible mothers and their children: Descriptive statistics based on 12-week long pre-treatment and post-treatment windows.

	Treatment window		Control windows	
	I. Before	II. After	III. Before	IV. After
A. Mothers, pre-birth information				
Number of observations	6,353	5,654	14,910	13,126
Age at time of childbirth	26.332	26.039	26.253	25.923
Earnings year before childbirth (BAs)	0.308	0.320	0.298	0.303
Number of children prior to current birth	1.127	1.103	1.157	1.136
B. Mothers, outcomes				
Disabled or dead, before age 60	0.400	0.417	0.413	0.418
Years employed, age 50-60	8.331	8.190	8.093	8.023
Average labor earnings, age 50-60	3.178	3.105	3.059	3.063
Log average labor earnings, age 50-60	0.853	0.828	0.787	0.789
C. Children, outcomes				
Disabled or dead, before age 40	0.165	0.157	0.161	0.166
Completed years of non-comp. education	4.100	3.994	3.961	3.897
Years employed, age 30-40	9.499	9.488	9.543	9.460
Average labor earnings, age 30-40	5.019	4.974	5.006	4.944
Log average labor earnings, age 30-40	1.371	1.368	1.382	1.371

Notes: The before/after treatment/control windows cover the 12-week periods just before and just after implementation of the reform (July 1). The number of observations vary slightly between outcomes, as employment is measured conditional on survival to age 60 (mothers) or 40 (children) and log earnings are measured conditional on having positive earnings in at least one of the years included. Observation numbers used in the analysis for each outcome are specified together with the estimation results reported in Table 3.

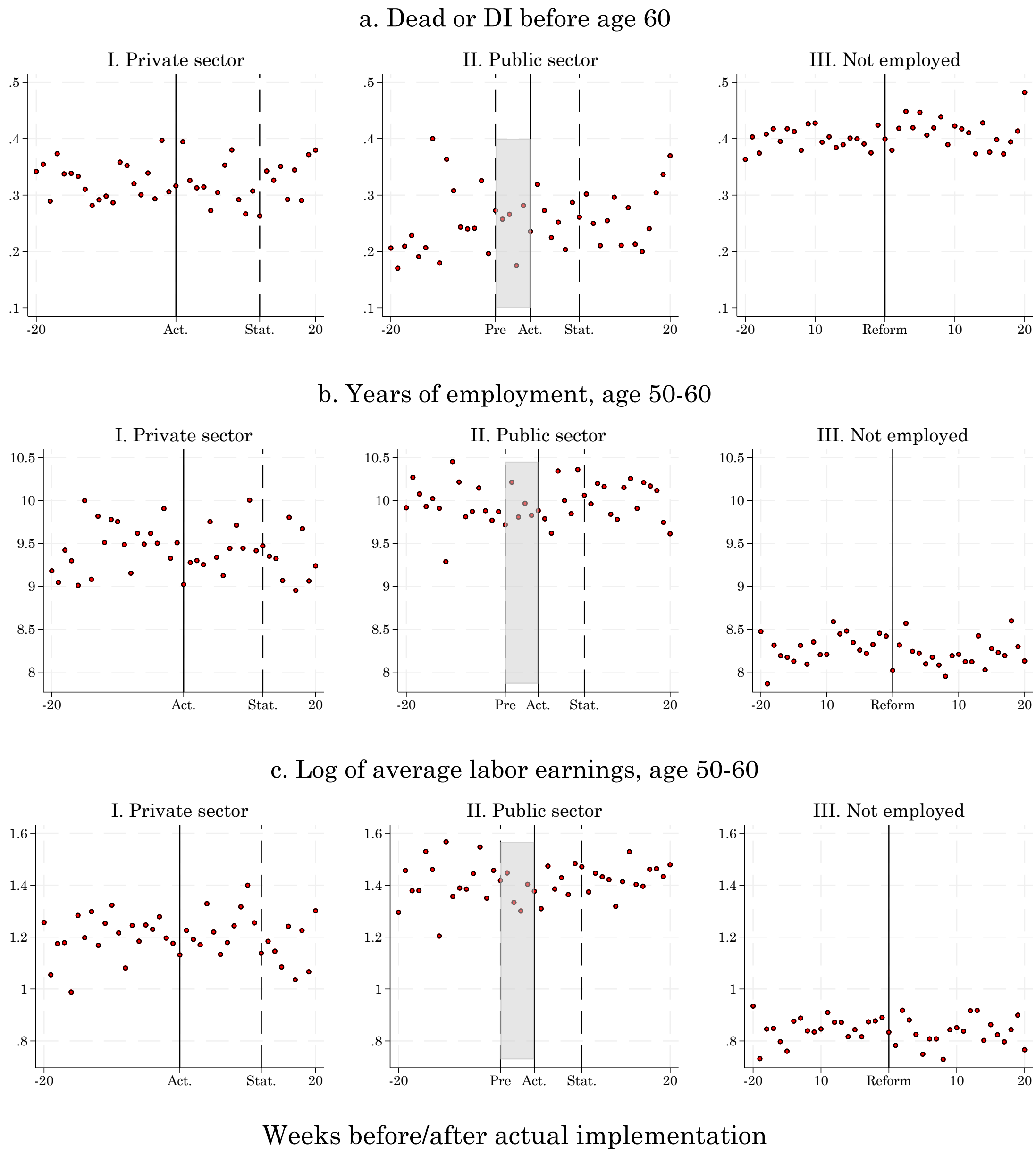

Figure 1: Long-term outcomes for mothers. Treatment window.
Note: The figures show average long-term outcomes by week of giving birth. The vertical dashed lines mark the week of statutory implementation (the week containing July 1, 1977 in the private sector and January 1 or May 1 in the public sector), whereas the vertical solid lines mark the week with actual implementation (April 9, 1977 in the private sector and November 13, 1976 or March 13, 1977 in the public sector). The gray areas mark the period of uncertain eligibility in the public sector, due to its dependence on the timing of the start of the leave. For the non-employed, the times of actual and statutory implementation coincide on July 1, 1977.

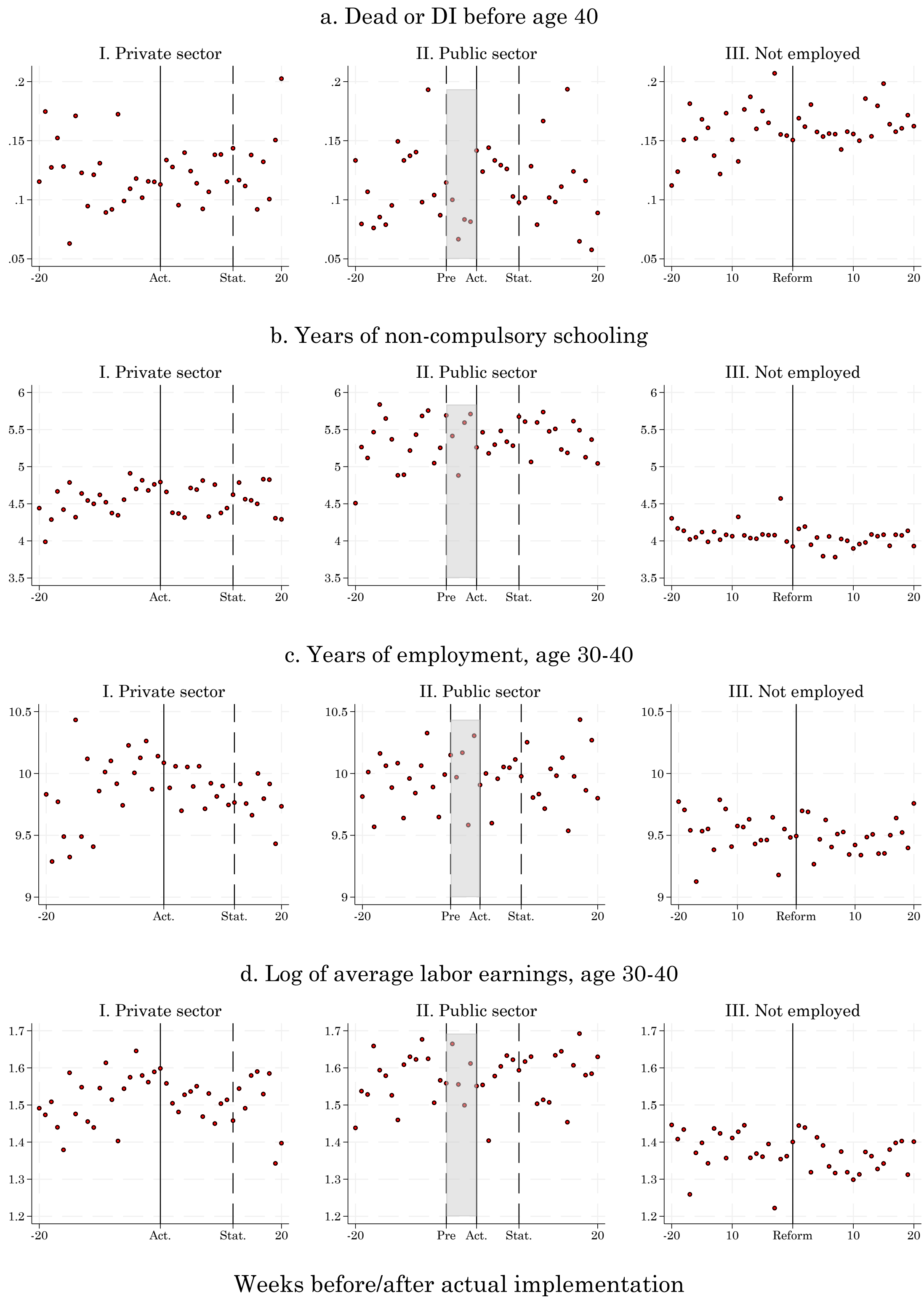

Figure 2: Long-term outcomes for children. Treatment window.

Note: The figures show average long-term outcomes by week of birth. The vertical dashed lines mark the week of statutory implementation (the week containing July 1, 1977 in the private sector and January 1 or May 1 in the public sector), whereas the vertical solid lines mark the week with actual implementation (April 9, 1977 in the private sector and November 13, 1976 or March 13, 1977 in the public sector). The gray area marks the period of uncertain eligibility in the public sector, due to its dependence on the timing of the start of the leave. For the non-employed, the times of actual and statutory implementation coincide on July 1, 1977.

5 Estimation results

Our main estimation results are presented in Table 3. The estimated effects of extended paid maternity leave based on Equation (1) are presented in the first two columns, using actual and statutory implementation dates as cutoffs, respectively, whereas the estimated effects of increased lump-sum transfer based on Equation (2) are shown in the third column.

For mothers (Panel A), the results suggest that longer paid leave did cause total time away from work to increase, as indicated by reduced earnings in the birth year and the year after. For mothers giving birth after statutory implementation, the point estimate indicates a 6.8% reduction in earnings over those two years, which is close to what we would expect if the total length of leave increased in line with paid leave. For mothers subjected to retroactive entitlement, the estimated effect is only 3% (and not statistically significant), perhaps suggesting less scope for planning. We interpret these estimates as evidence for the existence of a first stage effect and that total maternity leave was extended roughly in line with paid leave. The difference between the estimates for statutory and actual entitlement is not statistically significant at conventional levels (t-value=-1.47). Yet, taken at face value, it may suggest that although there was no entitlement discontinuity at July 1, the reform's influence on total leave increased as mothers became more accustomed to the extended duration of payment.

However, for long-term outcomes, we find no indications of any favorable effects of extended maternity leave, regardless of which of the two cutoff dates that are used to define the treated group. For mothers, the estimated effect on the probability of dying or becoming disabled before age 60 is almost exactly zero. The estimated effects on employment and earnings during age 50-60 (for those still alive) are also close to zero and, in most cases, on the negative side. Also for the children (panel B), the estimated effects on long-term outcomes are close to zero and statistically insignificant. If anything, the estimates point in a negative direction for both education, employment, and earnings. Based on these estimates and their confidence intervals, we can rule out any noticeable positive effects on either children's educational attainment or their adult employment and earnings.

The estimated long-term effects of the increase in the one-time benefit payment to non-employed (ineligible) mothers are also insignificant and close to zero.

Table 3: Main regression results

	Extended maternity leave		Increased one-time benefit
	I. Actual	II. Statutory	III.
Mothers' outcomes			
Log earnings, years t and t+1	-0.030	-0.068**	
	(0.030)	(0.029)	
	[21,326]	[20,824]	
DI or dead, age 60	-0.002	-0.010	0.011
	(0.013)	(0.013)	(0.012)
	[21,952]	[21,372]	[40,299]
Years employed, age 50–60	-0.058	-0.008	-0.017
	(0.084)	(0.087)	(0.099)
	[21,236]	[20,654]	[38,658]
Log labor earnings, age 50–60	-0.015	-0.004	0.002
	(0.013)	(0.013)	(0.013)
	[19,420]	[18,855]	[30,674]
Children's outcomes			
DI or dead, age 40	0.002	-0.006	-0.013
	(0.009)	(0.009)	(0.009)
	[21,732]	[21,160]	[40,128]
Years of non-comp. schooling	-0.010	-0.010	-0.008
	(0.073)	(0.073)	(0.063)
	[21,446]	[20,890]	[39,432]
Years employed, age 30–40	-0.079	-0.020	0.067
	(0.076)	(0.077)	(0.077)
	[21,446]	[20,890]	[39,432]
Log labor earnings, age 30–40	0.004	0.001	0.002
	(0.013)	(0.013)	(0.012)
	[20,176]	[19,608]	[35,977]

Note: Year t and year t+1 refer to the calendar year of birth and the subsequent year. Columns I and II show estimation results based on Equation (1), using the actual and the statutory cutoff dates for reform implementation, respectively. Column III shows results based on Equation (2). Standard errors are displayed in parentheses and observation numbers are shown in brackets. */**/*** indicate statistical significance at the 10/5/1% level.

6 Robustness

In this section, we present the results from a number of robustness exercises. For the effects of the paid maternity leave extension, the previous section reported estimates based on a difference-in-differences framework. To compute the estimates reported in Table 3, we used 12-week windows on both sides of the relevant cutoffs and included all employed mothers identified in at least one census by their sector of employment. In this section, we report two types of robustness analysis, one in which we use alternative configurations of the data, specifically related to information about the sector of employment, and one in which we use alternative eligibility thresholds or model specifications. The robustness exercises regarding sector of employment are motivated by the fact that the sector of employment is measured with some error and is completely missing for 30% of the mothers. These robustness checks are designed as follows:

- Using only employees with same sector of employment in both the 1970 and the 1980 censuses (strict sector definition).
- Using private sector employees only
- Using public sector employees only
- Using a wider definition of private sector, assuming that mothers with missing sector information worked in the private sector
- Using all eligible mothers regardless of sector information, by comparing births during the 12-week period after July 1 with the period just before the first group of mothers potentially became entitled to the extra paid weeks (that is, before October 10, 1976)

The results are presented graphically in Figure 3. Looking first at the effects on the overall length of maternity leave in panel A, it is notable that the estimated "first stage" effect appears to be primarily driven by responses in the private sector. This pattern does not necessarily indicate lack of responses by mothers in the public sector. It may be explained by a practice that appears to have been common in the public sector at the time, namely to pay maternity leave benefits in the form of regular wages. These wages qualified for pensions and were thus included in our earnings measure.

Moving on to the long-term effects for mothers (panels B–D) and children (panels E–H), it appears that almost all causal estimates are close to zero, and only 3 of 70 estimates are statistically significant at the 5% level, very close to what we would expect under the null hypothesis of zero effects on all long-term outcomes.

The remaining robustness exercises comprise an alternative eligibility threshold (making the eligibility assumption almost certain to hold), alternative window sizes (shorter or longer), and alternative choices of control periods. They can be described as follows:

- Using a stricter earnings threshold for eligibility (earnings above 3 Basic amounts)
- Using 6-week windows instead of 12-week windows (short window)
- Using 18-week windows instead of 12-week windows (long window)
- Using only windows lagged two years (1975) as controls
- Using only windows lagged one year (1976) as controls

The results are presented graphically in Figure 4. Again, all estimates of long-term effects are centered around zero, and none of them are statistically significant at the 5% level. It is notable, however, that almost all point estimates on earnings effects for children are negative. Although we do not interpret this as sufficient evidence for a negative earnings effect, it does make it possible to robustly rule out positive effects of the sizes previously reported.

For the effects of increased one-time payment for those who were not eligible for paid leave (the non-employed), we apply the following robustness exercises:

- Using a stricter criterion for *not* being eligible (earnings below 0.5 Basic amounts)
- Using 6 week windows instead of 12-weeks windows (short window)
- Using 18 week windows instead of 12-week windows (long window)
- Using a more flexible DiRD model (allowing for different trend slopes in each window)
- Using a regression discontinuity design (RDD) instead of DiRD
- Using a difference-in-differences design (DiD) instead of DiRD

The results are reported in Figure 5. All the exercises confirm that the true effects are close to zero and that large effects in any direction seem implausible. However, for children, all the point estimates indicate positive employment and earnings effects, perhaps suggesting that erroneous inclusion of many ineligible in the eligibility group could be a source of positive bias in the analysis of paid leave in models relying on the July 1 discontinuity.

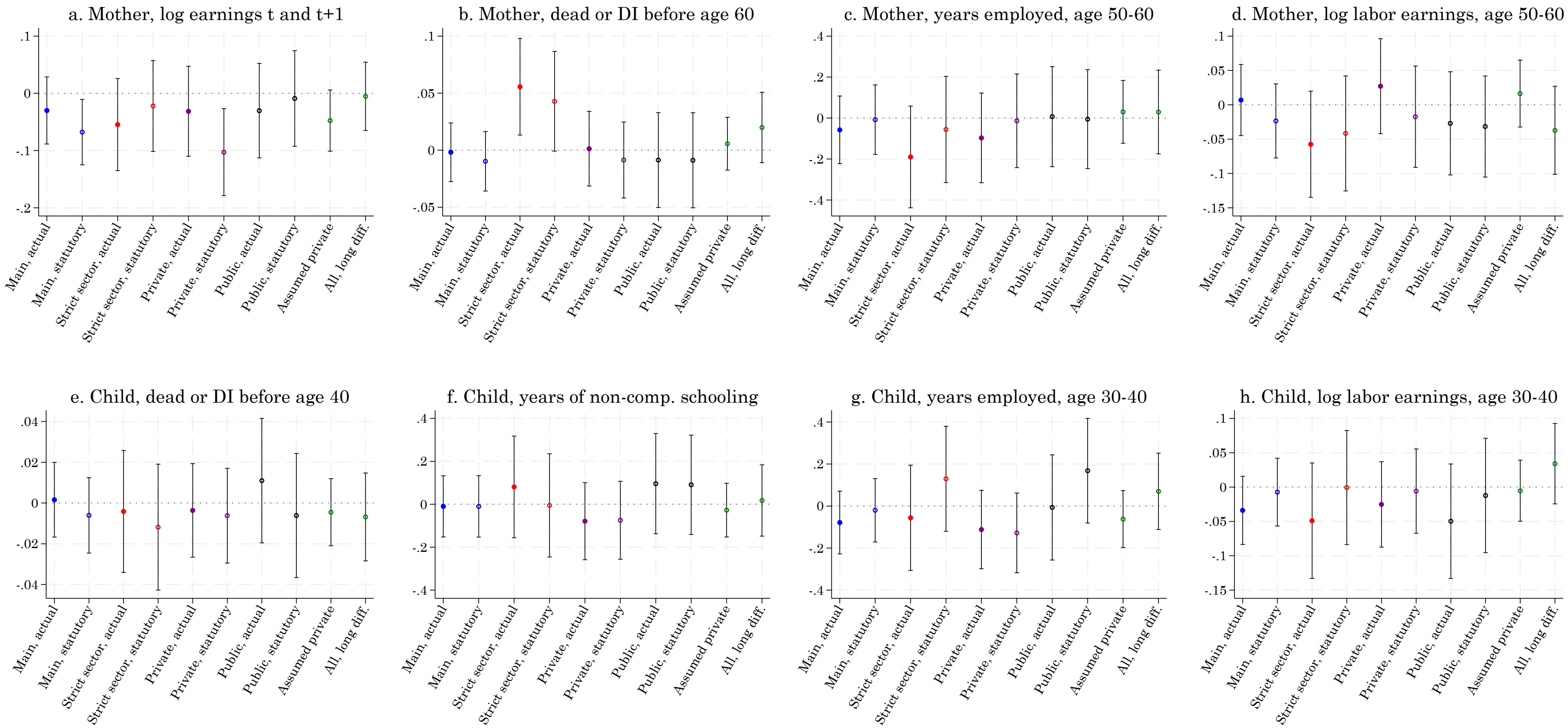

Figure 3: Effects of extended parental leave (with 95% CI). Robustness for sector of employment

Note: The panels show point estimates with 95% confidence intervals under alternative sample configurations. “Main actual” and “Main statutory” are repeated from Table 3. “Strict sector” indicates that only mothers with same sector information in 1970 and 1980 are used in the analysis. “Private” refers to a separate analysis for mothers in the private sector whereas “Public” refers to a separate analysis for mothers in the public sector (in both cases based on the original sector definition). “Assumed private” refers to a model for private sector only, assuming that all mothers with missing sector information worked in the private sector. “All” refers to a model with all mothers/offspring included with the post period defined as the period after July 1, 1977 and the pre-period defined as the period before October 10, 1976.

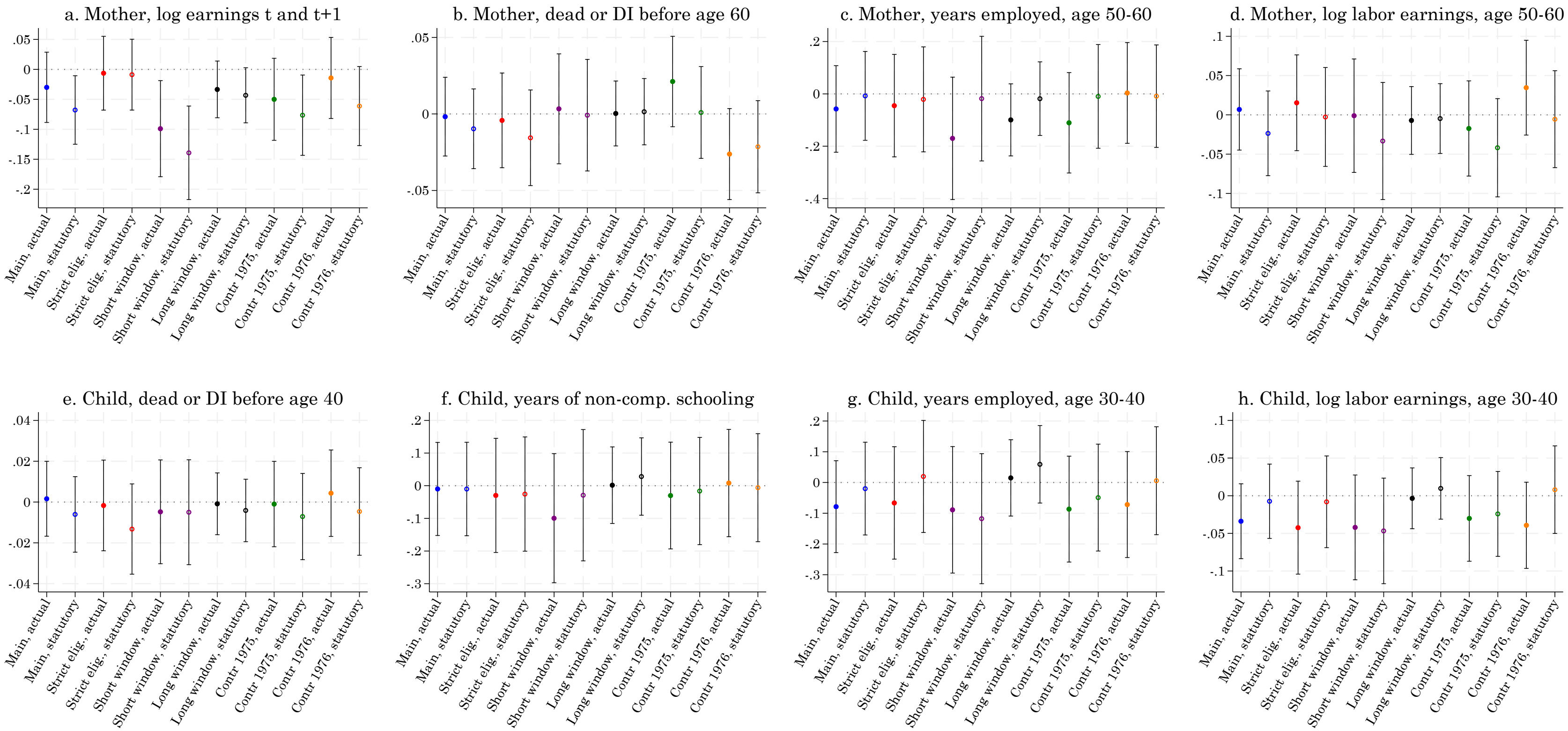

Figure 4: Effects of extended parental leave (with 95% CI). Robustness for model specification

Note: The panels show point estimates with 95% confidence intervals under alternative eligibility thresholds or model specifications. "Main actual" and "Main statutory" are repeated from Table 3. "Strict eligibility" refers to a model based on a reduced sample where the threshold for eligibility set to 3 Basic amounts. "Short window" refers to a model where the pre- and post-windows only include 6 weeks, whereas "long window" refers to a model where they include 18 weeks. "Contr 1975" refers to a model where we have used windows lagged two years as control windows whereas "Contr 1976" refers to a model where we have used windows lagged only one year.

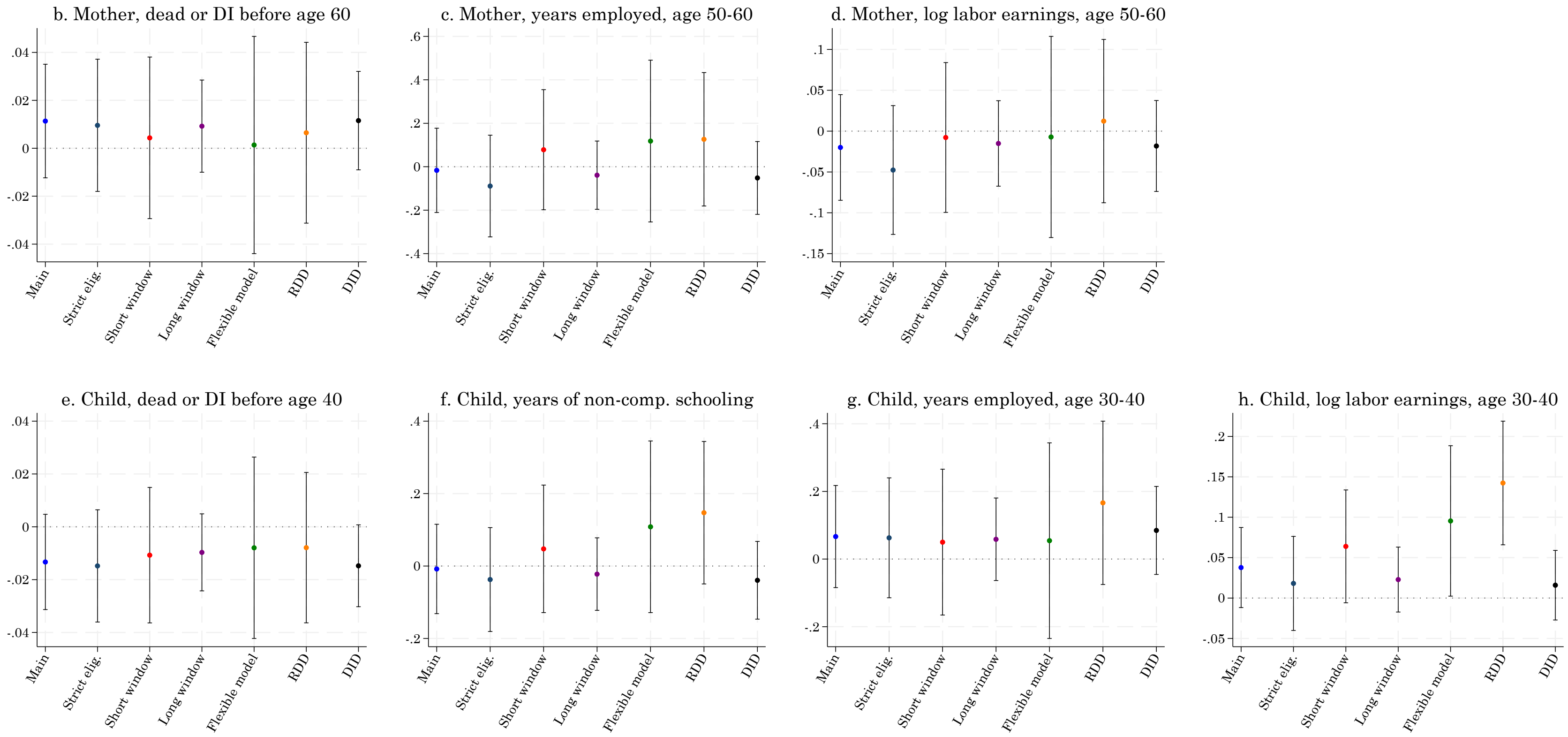

Figure 5: Effects of increased one-time benefits (with 95% CI). Robustness for model specification

Note: The panels show point estimates with 95% confidence intervals under alternative eligibility thresholds or model specifications. “Main” estimates are repeated from Table 3. “Strict eligibility” refers to a model based on a reduced sample where the threshold for not being eligible is set to 0.5 Basic amounts. “Short window” refers to a model where the pre- and post-windows only include 6 weeks, whereas “long window” refers to a model where they include 18 weeks. “Flexible model” refers to a less restrictive DiRD model, where day-of-year slopes are allowed to vary across windows as well as on each side of the cutoff dates. “RDD” refers to a regression discontinuity design based on the July 1 discontinuity, with different day-of-year slopes on each side of the cutoff date. “DiD” refers to a difference-in-differences model based on Equation (1), comparing the periods just after and just before July 1.

In the appendix, we report main results for models without the inclusion of individual covariates and for models based only on aggregated data (aggregated by weak of birth); see Appendix Tables A1 and A2. Both sets of results are similar to those reported in Table 3. This illustrates that the identifying information is preserved in the aggregated data and suggests that these data can be used as a basis for replications and estimation of alternative models. As the micro-data are subjected to restricted access due to privacy concerns, this has motivated us to make the aggregated-by-weak-of-birth-cell data openly accessible, including data collected from potential control periods not used by us. Details on how to access both the restricted individual data and the unrestricted aggregated data are provided in the replication package accompanying this paper.

7 Relation to previously reported effect estimates

Given our robust finding of reform effects close to zero, a natural question to ask is how the previously reported large effect estimates came about, and how some of them apparently could be confirmed by the new analysis in Carneiro, Løken, and Salvanes (2024). To shed some light on this, we start by replicating the main results reported by Carneiro, Løken, and Salvanes (2015) and Carneiro, Løken, and Salvanes (2024), which comprise estimated effects on the following child outcomes:

- Income at age 30
- High school dropout
- College attendance

Our replications and subsequent modification of the analyzes in question are based on administrative registers adapted to be as similar as possible to the data used in the previous analyzes. As these analyzes are based on data without information on the sector of employment of mothers, we also disregard this information and estimate models *as if* all mothers worked in the private sector. We start by seeking to replicate previously reported results and then move on to correct some data errors and sample restrictions.

Figure 6 presents our results. The upper part of the figure (part A) shows the original estimates (with 95% confidence intervals) from Carneiro, Løken, and Salvanes (2015), together with the authors' replication (based on data that are available today) presented in Carneiro, Løken, and Salvanes (2024), as well as our own replication. While our replication gives almost exactly the same results as Carneiro, Løken, and Salvanes (2024) for the effect on adult income, we get slightly smaller estimates for the two educational outcomes. We have not been able to identify the source of this small discrepancy.[14] Note that all these estimates are based on the original difference-in-regression-discontinuity (DiRD) design (with triangular weights), assuming – *incorrectly* – the existence of a sharp discontinuity in eligibility on July 1, 1977. They are also based on the original eligibility threshold (now known to have been too low) and the original choice of control years, including two post-reform years (1978 and 1979) and excluding 1976 (for reasons we questioned in Lillebø et al. (2024)).

Part B of the figure reports estimates when all births occurring between April 9 and July 1 are dropped from the analysis, yet still based on a DiRD model (though the discontinuity now "jumps" directly from April 9 to July 1; hence, it is no longer a real discontinuity). Line B.i first shows the revised estimates provided by Carneiro, Løken, and Salvanes (2024), while line B.ii reports our replication of these new results. Line B.iii then reports the estimates obtained when we use an earnings threshold for eligibility consistent with the correct fraction of eligible mothers (reducing the fraction of assumed eligible mothers from 65% to 45%). We note that all three point estimates become smaller and that the confidence intervals become larger (due to the smaller sample sizes). In line B.iv, we also correct an error in the age at which incomes were measured, such that incomes are measured at age 30 for all the children (in the original analysis, income was by accident measured at age 28-29 for children of mothers born in 1978 and 1979). In row B.v we correct the the two educational outcomes, such that they build on an international classification standard (ISCED) and such that education obtained outside Norway is included (in the original analysis, education obtained abroad was in most cases excluded). Finally, in row B.vi, we remove from the sample mothers registered to have worked in the public sector, such that they were unlikely to have been affected by the April 9 -July 1 rollout. After all these corrections have been implemented, there are no longer any signs of favorable effects, and all three point estimates have flipped sign (though none of them are statistically significant).

[14]One potential explanation is that the data provided to us by Statistics Norway are slightly different from the data used by Carneiro, Løken, and Salvanes (2024). In some cases, Statistics Norway updates data for education retroactively when they obtain more information.

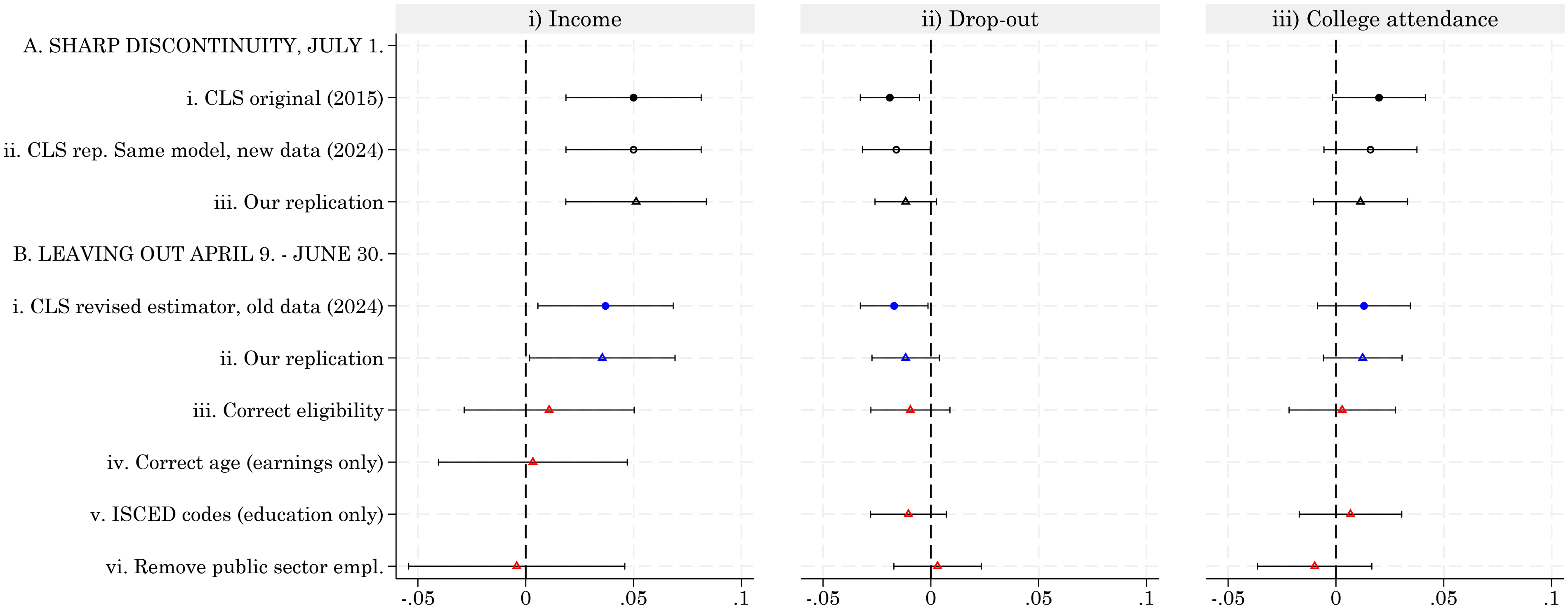

Figure 6: Robustness of previously reported results

Note: The panels show point estimates with 95% confidence intervals. CLS (2015) and CLS (2024) refer to Carneiro, Løken, and Salvanes (2015) and Carneiro, Løken, and Salvanes (2024), respectively. In section A we first display the findings in CLS (2015). Below we display the CLS (2024) replication of their original findings (same model and reform dates, but with updated data), and our attempt to replicate the same (using our data). In section B we first present CLS (2024), which uses the original CLS (2015) dataset (including erroneous outcomes and eligibility criteria) but a revised identification strategy, leaving out births occurring in April 9th - June 30th. We then present our replication (ii). In (iii) we revise the eligibility criteria such that the number of eligible mothers matches historical statistics for users of maternity leave in 1977. In (iv) we also measure earnings at age 30 for all, not partly at age 28 and 29 as done in the original CLS (2015) paper, an error which CLS (2024) have chosen to maintain. In (v) we also revise the educational outcomes by using the International Standard Classification of Education (ISCED) codes for highest completed education. As dropouts, we consider those with values $0, 1, 2$ and as college those with values $5, 6, 7, 8$. In (vi) we also remove mothers we classify as public sector employees, as the treatment differential is not relevant for them.

8 The responsibility of journals

Research matters. One of the reasons we have written this article is that we have found it deeply disturbing that proving existing research wrong ends up prompting a highly questionable authorization of the results that were proven wrong in the first place. This is particularly troubling given that the case in question involves one of the most prestigious journals within the field of economics and relates to a finding with clear policy implications.

It is inevitable that published research from time to time contains errors. As in this case, information about institutional details from the past can be hard to find and easily misunderstood. However, when serious errors are discovered so that published research is invalidated, it is imperative to promptly make readers aware of them and to be transparent about their consequences. In the present case, it took three years after JPE was informed of the errors in the published article until the readers were informed in the form of a comment article (Lillebø et al., 2024).[15] No expression of concern was published. Meanwhile, other researchers built their work on a false description of a reform apparently constituting a unique opportunity to study the effects of the introduction of paid parental leave.

Among the journals commonly referred to as "Top 5", only the American Economic Review (AER) has an editorial policy posted on its web page that explicitly explains how it deals with comments and replies. Yet, given the prominent status of AER, this may perhaps serve as a sort of "industry standard". The policy states that "Comments may be summarily rejected by the editor or the assigned coeditor or sent for review to the author of the original article and likely to other reviewers. If the author of the original paper submits a Reply, it will be sent to the author of the Comment and generally to other reviewers."

In the present case, the critical comment (Lillebø et al., 2024) was sent to the original authors as well as to external reviewers, whereas the original authors' reply (Carneiro, Løken, & Salvanes, 2024) was neither sent to the authors of the comment nor to external reviewers. The readers are not informed about this asymmetry.

This procedure is not unique to JPE, as many journals treat replies as editorial correspondence rather than new research articles. However, when substantial new analyzes are introduced, the absence of external peer review should raise concerns about the ro-

[15]The first submission documented the errors in the form of relevant legal texts, regulations and collective agreements, both in their original (Norwegian) language and in English translation.

bustness of published claims. The unbalanced use of peer review implies that the quality control of the original authors' reply may have been less stringent than for the comment. In the present case, there are at least two aspects of the results presented in the original authors' reply that should arguably have raised serious concerns about their validity, and which we believe would have been challenged under regular peer review.

First, despite that Carneiro, Løken, and Salvanes (2024) acknowledge and correct errors *both* in the data (construction of treatment and control groups as well the earnings outcome) *and* in the identification strategy, *none* of the results in their reply implement both corrections at the same time. The only valid approach is to do both; correct all known errors *and* use an appropriate identification strategy. This should be a minimum requirement for claiming that the original results are "still flying". In this paper, we have shown that, when doing so, the estimated effects are no longer distinguishable from zero.

Second, the ultimately *undisputed* content of the reform also requires a new interpretation of the magnitudes of the estimated effects, leaving them highly implausible. For example, according to the revised estimate provided by Carneiro, Løken, and Salvanes (2024), the reform increased offspring earnings at age 30 by 3.7%.[16] Under the assumption that the reform provided 18 weeks of paid leave, this translates into an average effect of 0.21% per week. However, given that the actual extension was merely 5–6 weeks, the implied per week-effect rises to 0.62%. Furthermore, at the time the revised estimates were published by JPE, it was also known (and acknowledged by the authors) that the treatment group included a substantial share of mothers who were not affected by the evaluated treatment, either because they were ineligible or because they were treated at another point in time. As we documented in Section 2, this degree of misclassification implies that the estimated effects must be scaled by a factor of approximately two. The implied effect on offspring is therefore around 1.25% higher earnings at age 30 per additional *week* of parental leave 30 years prior; an effect 4.5 times larger than suggested by the original interpretation.

At face value, this suggests that a few weeks of paid leave for the mother produce the same long-term earnings gain for the offspring as one full *year* of offspring schooling. If true, an effect of this magnitude would have dramatic policy implications. Yet, this is not discussed or given any attention in the published reply, nor is the estimate subjected to any plausibility considerations.

As a consequence, the reader is left with the impression that, despite the misunderstandings of the reform's content and implementation, the original results are still valid. This impression is further strengthened by the fact that the JPE made the comment by Lillebø

[16] The original estimate in Carneiro, Løken, and Salvanes (2015) was an earnings effect of 5%.

et al. (2024) available only behind a paywall, while the original authors' reply (Carneiro, Løken, & Salvanes, 2024) was published open access. As a result, the reply has more than ten times the number of views than the comment.

Although we will argue that authors have a personal responsibility for taking on the duty of notifying the research community upon discovering flaws in their published work, we think it is important to also raise a discussion about the responsibilities of journals. Imagine that a journal has published a paper showing that some costly new treatment has large positive effects, dramatically reducing mortality. The finding has important policy consequences, and governments, based on these findings, spend hugely on this wonder-treatment. Then, it turns out that the findings were based on an invalid empirical strategy. The positive effect could still be there, but there is no evidence in either direction. We believe that the answer to this imaginary journal's dilemma is that they should immediately inform their readers. If not, governments will continue to risk wasting resources on this undocumented treatment and the credibility of research in general will suffer. The case at hand has similarities with this example. The treatment, paid maternity leave, is expensive. The papers published by JPE and AEJ:EP have found large positive effects on mothers' health and offspring's labor market careers. Whether or not this has influenced policy is hard to tell, but given the attention these papers have received, it is clearly something that cannot be ruled out.

9 Conclusion

Existing influential research (Carneiro, Løken, & Salvanes, 2015; Bütikofer, Riise, & Skira, 2021) has indicated large favorable long-term effects on mothers' health and on offspring's education and labor market outcomes of a 5–6 week extension of paid maternity leave implemented in Norway in 1977. Lillebø et al. (2024) documented that this evidence was built on an invalid identification strategy and a wrong interpretation of the reform's content. Despite that, the favorable effects were confirmed by Carneiro, Løken, and Salvanes (2024) in a new analysis formulated as a reply to Lillebø et al. (2024). In the present paper, we have shown that also the new analysis builds on models that either use an invalid identification strategy or rely on data with errors in the construction of treatment groups and the definition of outcomes, and that the reported result therefore are misleading. These previously reported results should therefore be discarded in any future discourse on the effects of paid maternity leave.

In an attempt to replace the invalid results, we have provided a full reexamination of the effects of the Norwegian 1977 maternity leave reform. Equipped with updated data that facilitate efficient utilization of the staggered and sector-wise implementation of extended paid leave, we apply a modified and valid identification strategy to estimate long-term effects on both mothers and offspring. The estimated effects are all close to zero and large favorable effects of the magnitudes previously reported can be ruled out. However, the results are consistent with previous studies evaluating similar moderate extensions to already generous parental leave policies in Denmark and Norway (Rasmussen, 2010; Dahl et al., 2016).

Supplementary material

Supplementary material, including replication package and aggregated data, available at: https://frischsenteret.no/publication-details/?skrift_id=1787

References

Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature*, *56*(4), 1360–1446. https://doi.org/10.1257/jel.20171164

Baker, M., & Milligan, K. (2015). Maternity leave and children's cognitive and behavioral development. *Journal of Population Economics*, *28*(2), 373–391. https://doi.org/10.1007/s00148-014-0529-5

Behaghel, L., & Pinto, M. F. (2024). Extended maternity leave and children's long-term development. *The Scandinavian Journal of Economics*, *126*(2), 224–253.

Bütikofer, A., Riise, J., & Skira, M. M. (2021). The impact of paid maternity leave on maternal health. *American Economic Journal: Economic Policy*, *13*(1), 67–105. https://doi.org/10.1257/pol.20190022

Carneiro, P., Løken, K., & Salvanes, K. G. (2024). Still flying: Reply to "Not a flying start after all?" by Lillebø et al. *Journal of Political Economy*, *132*(12), 4213–4222. https://doi.org/10.1086/732220

Carneiro, P., Løken, K. V., & Salvanes, K. G. (2015). A flying start? Maternity leave benefits and long-run outcomes of children. *Journal of Political Economy*, *123*(2), 365–412. https://doi.org/10.1086/679627

Dahl, G. B., & Loken, K. V. (2024). Families, public policies, and the labor market. In *Handbook of labor economics* (pp. 581–617, Vol. 5). Elsevier.

Dahl, G. B., Løken, K. V., Mogstad, M., & Salvanes, K. V. (2016). What is the case for paid maternity leave? *Review of Economics and Statistics*, *98*(4), 655–670. https://doi.org/10.1162/REST_a_00602

Danzer, N., & Lavy, V. (2018). Paid parental leave and children's schooling outcomes. *The Economic Journal*, *128*(608), 81–117. https://doi.org/10.1111/ecoj.12493

Dustmann, C., & Schönberg, U. (2012). Expansions in maternity leave coverage and children's long-term outcomes. *American Economic Journal: Applied Economics*, *4*(3), 190–224. https://doi.org/10.1257/app.4.3.190

Heshmati, A., Honkaniemi, H., & Juárez, S. P. (2023). The effect of parental leave on parents' mental health: A systematic review. *The Lancet Public Health*, *8*(1), e57–e75. https://doi.org/10.1016/s2468-2667(22)00311-5

Lillebø, O. S., Markussen, S., Røed, K., & Zhao, Y. (2024). Not a flying start after all? A comment. *Journal of Political Economy*, *132*(12), 4205–4212. https://doi.org/10.1086/732218

Rasmussen, A. W. (2010). Increasing the length of parents' birth-related leave: The effect on children's long-term educational outcomes. *Labour Economics*, *17*(1), 91–100. https://doi.org/10.1016/j.labeco.2009.07.007

Regmi, K., & Wang, L. (2023). Maternity leave. *Handbook of labor, human resources and population economics*, 1–40.

Regmi, K., & Wang, L. (2025). Baby steps to success? the impact of paid maternity leave on children's long-term outcomes in the united states. *Journal of Population Economics*, *38*(2), 37.

Rossin-Slater, M. (2017). *Maternity and family leave policy* (No. w23069). National Bureau of Economic Research. https://doi.org/10.3386/w23069

Statistics Norway. (1979). *Statistical Yearbook of Norway 1979*. Statistics Norway, Oslo. https://www.ssb.no/en/befolkning/artikler-og-publikasjoner/statistisk-arbok-1979

Appendix

Table A1: Main regression results, without individual controls

	Extended maternity leave I. Actual	Extended maternity leave II. Statutory	Increased one-time benefit III.
Mothers' outcomes			
Log earnings, years t and t+1	-0.062*	-0.076**	
	(0.034)	(0.033)	
	[21,326]	[20,824]	
DI or dead, age 60	-0.000	-0.010	0.012
	(0.013)	(0.013)	(0.012)
	[21,952]	[21,372]	[40,299]
Years employed, age 50–60	-0.060	-0.036	-0.037
	(0.086)	(0.089)	(0.102)
	[21,236]	[20,654]	[38,658]
Log labor earnings, age 50–60	0.002	-0.034	-0.030
	(0.027)	(0.028)	(0.034)
	[20,696]	[20,117]	[36,025]
Children's outcomes			
DI or dead, age 40	0.001	-0.008	-0.011
	(0.009)	(0.009)	(0.009)
	[21,732]	[21,160]	[40,128]
Years of non-comp. schooling	-0.030	-0.022	-0.011
	(0.075)	(0.075)	(0.065)
	[21,446]	[20,890]	[39,432]
Years employed, age 30–40	-0.074	-0.010	0.055
	(0.076)	(0.077)	(0.077)
	[21,446]	[20,890]	[39,432]
Log labor earnings, age 30–40	-0.029	0.006	0.029
	(0.026)	(0.026)	(0.026)
	[21,068]	[20,525]	[38,420]

Note: The table displays estimation results using the same data and same models as Table 3, but without using individual controls. Columns I and II show estimation results based on Equation (1), using the actual and the statutory cutoff dates for reform implementation, respectively. Column III shows results based on Equation (2). Standard errors are displayed in parentheses and observation numbers are shown in brackets. */**/*** indicate statistical significance at the 10/5/1% level.

Table A2: Regression results using aggregated data for sharing

	Extended maternity leave		Increased one-time benefit
	I. Actual	II. Statutory	III.
Mothers' outcomes			
Log earnings, years t and t+1	-0.067	-0.075*	
	(0.041)	(0.040)	
	[216]	[216]	
DI or dead, age 60	0.002	-0.008	0.011
	(0.014)	(0.013)	(0.010)
	[216]	[216]	[72]
Years employed, age 50–60	-0.076	-0.034	-0.055
	(0.093)	(0.091)	(0.083)
	[216]	[216]	[72]
Log labor earnings, age 50–60	-0.005	-0.034	-0.025
	(0.028)	(0.027)	(0.026)
	[216]	[216]	[72]
Children's outcomes			
DI or dead, age 40	0.002	-0.008	-0.012
	(0.009)	(0.010)	(0.008)
	[216]	[216]	[72]
Years of non-comp. schooling	-0.047	-0.014	-0.025
	(0.076)	(0.075)	(0.062)
	[216]	[216]	[72]
Years employed, age 30–40	-0.086	-0.012	0.071
	(0.077)	(0.081)	(0.068)
	[216]	[216]	[72]
Log labor earnings, age 30–40	-0.033	0.005	0.015
	(0.026)	(0.026)	(0.026)
	[216]	[216]	[72]

Note: The table displays estimation results using the aggregated data made available for sharing. Columns I and II show estimation results based on Equation (1), using the actual and the statutory cutoff dates for reform implementation, respectively. Column III shows results based on Equation (2). Standard errors are displayed in parentheses and observation numbers are shown in brackets. */**/*** indicate statistical significance at the 10/5/1% level.