

Not a flying start after all?

A reexamination of the effects of maternity leave benefits on long-run outcomes of children*

Otto Sevaldson Lillebø^a, Simen Markussen^b, Knut Røed^b, and Yuejun Zhao^{b,c}

^a*Nordic Institute for Studies in Innovation, Research and Education, Norway*

^b*The Ragnar Frisch Centre for Economic Research, Norway*

^c*Department of Economics, University of Gothenburg, Sweden*

Abstract

In a much-cited paper published in the *Journal of Political Economy*, Carneiro, Løken, and Salvanes (2015) examine the effects of a reform that they claim introduced paid maternity leave in Norway. To identify causality, they exploit an alleged implementation discontinuity implying that only mothers giving birth after a specific cutoff date were entitled to paid leave. They report large favorable effects on the children's educational attainment and adult earnings. We show that their analysis was built on an incorrect description of the evaluated reform, rendering its reported effects meaningless. The reform did not introduce paid maternity leave in Norway, but extended it by five to six weeks. The extension was implemented at different dates in different sectors of the labor market and was also offered to mothers giving birth during the last months prior to the respective implementation dates. Thus, the discontinuity used to identify the effects never existed, and the alleged treatment and control groups were in reality subjected to the same maternity leave conditions. We explore the origins of the erroneously identified effects, and conclude that they most likely came about through a combination of unintended data construction errors, coincidence, and model selection.

Keywords: Replication, Model evaluation, Family policies, Maternity leave

JEL codes: C52, J13, J18

* This paper reveals some critical errors in two research papers published in the *Journal of Political Economy* and *American Economic Journal: Economic Policy*, respectively. The errors arise due to a flawed description of a maternity leave reform in Norway 1977, and imply that the reported effects were identified based on a non-existent treatment-control differential. We informed the Norwegian authors of the two papers that we suspected these errors in the beginning of March 2020. We have later verified the errors by collecting the relevant documentation from the primary sources, and communicated these documents to the two author teams. In Online Appendix A to this paper, we provide documentation for all our claims regarding the reform's content and the way it was implemented. We wish to thank Sarujan Mathimaran for assistance with finding out how the reform was actually implemented and Katrine Løken for assistance to our replication exercise. Also thanks to Oddbjørn Raaum and Astrid Marie Jorde Sandsør for valuable comments. This research has received support from the Norwegian Research council (grant # 300917). Micro data leased from Statistics Norway have been essential.

1 Introduction

In an influential paper published in the *Journal of Political Economy*, Carneiro, Løken, and Salvanes (2015), hereafter CLS, examine the effects of a reform in 1977 that they claim introduced paid maternity leave in Norway by replacing entitlement to 12 *unpaid* weeks with 18 *paid* weeks. To identify the effects of the reform, CLS exploit an asserted sharp discontinuity at the implementation date of July 1, 1977, claiming that only mothers who gave birth after this specific date were entitled to paid leave. The main conclusion of the paper is that the children of treated mothers got “a flying start” and as a result attained more education and earned as much as 5% higher wages at age 30. In the present paper, we show that their results build on an erroneous description of the reform and the way it was implemented. The reform did *not* introduce paid leave, but instead extended the mandated paid leave from 12 to 18 weeks. Moreover, the six-week extension was implemented at three different dates during 1977 (January 1, May 1, and July 1), depending on the mothers’ sector of employment (state, municipality, or private), and was phased in such that it also applied to births that took place the last two to three months before the relevant cutoff dates. The alleged discontinuity used by CLS to identify the causal effects never existed, and the analysis thus builds on an invalid identification strategy as well as an erroneous interpretation of the evaluated treatment.

The CLS paper has been much cited in the literature (with more than 330 citations recorded in Google scholar). Other research on extensions of pre-existing paid leave schemes has consistently failed to identify favorable effects on offspring outcomes (Rasmussen, 2010; Dustmann and Schönberg, 2012; Baker and Milligan, 2015; Dahl et al., 2016; Danzer and Lavy, 2018), and the CLS findings have therefore been interpreted as evidence of favorable effects associated with the very first weeks of maternity leave, but then diminishing returns to further extensions (Rossin-Slater, 2017; Almond, Currie, and Duque, 2018). CLS refer to Rasmussen (2010), who analyzed an extension of paid maternity leave in Denmark in 1984, and argue that her failure to find any favorable effect on child outcomes may be related to the fact that the Danish reform extended an already-generous maternity leave program—presumably in contrast to the Norwegian reform. In fact, the Danish 1984-reform was very similar to the Norwegian 1977-reform, as it also extended the mandated maternity leave by six weeks, only from a starting point of 14 as opposed to 12 weeks.

Unfortunately, the incorrect description of the 1977-reform provided by CLS has also been reiterated and adopted by other researchers, most recently by Bütikofer, Riise, and Skira (2021), hereafter BRS, in a paper published in the *American Economic Journal: Economic Policy*. Building on the same invalid identification strategy and the same erroneous interpretation of the reform, they examine the impact of paid maternity leave on maternal health, and conclude that the reform improved a range of maternal health outcomes.

Our own interest in this reform was also inspired by the CLS paper, and our original plan was to adopt their research strategy to examine impacts of maternity leave on intergenerational economic mobility. However, when we started looking into the details of the reform’s imple-

mentation, it became clear that CLS’s description of it could not be correct and that the basis for identification of the reported effects simply did not exist. Hence, what started out as an attempt to use a well-documented quasi-experiment and an already peer-reviewed empirical strategy to study effects on a novel outcome has instead ended up as a re-assessment of the original findings and an attempt to understand how it has been possible to identify significant and robust causal effects from data where the alleged treatment and control groups in reality were subjected to the same maternity leave conditions.

Our paper is structured as follows: We first provide a brief description of the 1977 maternity leave reform and the way it was implemented (with documentation provided in Online Appendix A), and explain why the actual features of the reform completely invalidate the empirical strategy used by both CLS and BRS. We then replicate the results in CLS and seek to identify the sources of the apparently significant and robust effects. Our findings indicate that the origin of the false positive effects is a blend of unintended data construction errors, coincidence, and model selection. Finally, in Online Appendix B, we re-estimate the effects of the 1977-reform based on a correct description of its implementation. Unfortunately, the true implementation does not provide the same novel and quasi-experimental setting for empirical evaluation. The reform was neither that fundamental nor that different from reforms already evaluated in other countries and there was no sharp discontinuity. Data on mothers’ sector of employment back in 1977, which could provide a precise timing of their eligibility for the five-to-six-week paid leave extension, do not exist. Instead, we use information on the mothers’ employment 15 years later and their educational attainment to (imperfectly) predict their sector of employment in 1977. In accordance with the existing literature, we do not find any significant effects on children’s long-term outcomes.

2 The maternity leave reform and the CLS description

CLS describe the 1977 maternity leave reform in Norway as follows:

- A: Mandatory paid maternity leave increased from 0 to 18 weeks (pp. 365, 366–367, 371–372).
- B: Mothers on leave were compensated 100% of their full pre-birth salary (p. 372).
- C: Unpaid maternity leave increased from 12 to 52 weeks, which came on top of the 18 paid weeks (pp. 366, 372).
- D: These changes applied to births taking place from July 1, 1977, but not to births taking place just prior to that date (i.e., there was a sharp discontinuity) (pp. 367, 372).
- E: The new legislation was not covered by the media until just before its implementation, hence the reform was largely unexpected at conception (pp. 373–374).

Every part of this description is incorrect.

A: The reform did not introduce paid maternity leave in Norway. Prior to the 1977 reform, all female employees were entitled to at least 12 weeks of *paid* maternity leave (13 weeks in the

public sector), provided they had been employed six out of the last 10 months prior to expected delivery.¹ Hence, paid maternity leave increased from 12 (or 13) to 18 weeks, not from 0 to 18. Paid maternity leave in Norway dates back at least to the 1930s, when it covered six weeks. This was further extended to 12 weeks already in the health insurance act of 1956 (NOU, 1996, p. 214).²

B: The reform did not raise the compensation to 100% in 1977. The new legislation did indeed raise the level of mandatory maternity leave benefits to the full pre-birth salary (up to a threshold). However, for reasons related to the integration of maternity leave benefits into the sick-leave insurance system, this part of the reform was implemented exactly one year later, that is, on July 1, 1978.³ Moreover, many workers, including all employees in the public sector, enjoyed this compensation level (for 13 weeks) already before the 1977 reform.⁴

C: The reform increased the *total* leave period (unpaid plus paid) to 52 weeks; i.e., the unpaid leave did not come on top of the paid leave.⁵

D: There was not a sharp entitlement discontinuity on July 1, 1977 for either the paid or the unpaid leave extensions. The extension of paid leave was first implemented for employees in the state and in the municipality of Oslo on January 1, 1977 as part of the tariff agreement, then for employees in the rest of the public sector on May 1, 1977, and finally, for workers in the private sector on July 1, 1977.⁶ Most importantly, the reform was phased in such that the extra five to six weeks were also offered to mothers who had given birth during the last months prior to the respective cutoff dates. For the January and May implementation dates (i.e., the tariff agreements), this was done such that those who were already on leave at the time of implementation were granted the extra weeks, implying that it also applied to births taking place 7–13 weeks before these dates (depending on how many weeks before the delivery the

¹ Depending on the timing of work experience within the 10-month window, entitlement could also be secured with as little as three months of employment.

² Law on health insurance (Lov om syketrygd) of June 6, 1930, § 13, item 7, and the same revised law of March 2, 1956, § 55. The regulations were later (in 1966) incorporated into the law on social insurance § 3–13 (see Online Appendix A, Item 1).

³ The delayed rise in the replacement ratio follows from Law modification (Endringslov) 84, passed by the Parliament on June 10, 1977, § 3–5 and § 3–21, with the date of implementation specified in the law’s part II; see Online Appendix A, Item 3.

⁴ For the municipal sector, the pre-reform maternity leave conditions are described in the tariff agreement (Hovedoverenskomst) 1968–70 (p. 18), which states that female employees with at least one year of tenure are entitled to the full wage for six weeks before and seven weeks after childbirth; see Online Appendix A, Item 2. Similar rules for state employees are described in a circular letter from the ministry (Rundskriv P-65/76 fra Forbruker- og administrasjonsdepartementet § 10); see Online Appendix A, Item 7. In the private sector, 90 percent replacement rate was common, and in total, it was stipulated in a Government White Paper from 1973/74 (St. meld. 51 1973–74, p. 129) that approximately 75–80 % of all employees had at least 90% coverage. For the remaining employees, the mandated pay was NOK 4 plus 0.1% of annual earnings *per day* (up to a ceiling).

⁵ Law on work environment (Arbeidsmiljøloven) of 1977, § 31, No. 3 (Law modification 4 1976–77); see Online Appendix A, Item 4.

⁶ The staggered implementation is described in the new joint regulations for state employees (Rundskriv P-38/76 fra Forbruker- og administrasjonsdepartementet: Nye fellesbestemmelser i lønnsregulativet med departementets kommentarer, § 10) and in the minutes from the tariff negotiations in the municipal sector in 1977 (A-rundskriv nr. 6-1977 Endringer i Hovedoverenskomsten pr. 1.5.1977); see Online Appendix A, items 5 and 6.

mother had begun her leave).⁷ For the July implementation date (i.e., the legislated extension), the extra weeks were granted to those who had given birth the last 12 weeks prior to the implementation date.⁸ The extension of unpaid leave to one year also applied to mothers who had given birth prior to July 1, extending leave for the days left until the child’s one-year birthday.⁹ Hence, there was no discontinuity in maternity leave conditions for eligible mothers in relation to births occurring around July 1.¹⁰

E: The reform was not poorly covered by the media and it was not unexpected. The extension of paid maternity leave from 12 to 18 weeks was proposed by a majority in the parliament’s social policy committee already in 1970, it was put forward by the Gender Equality Council (Likestillingsrådet) in 1973, and subsequently discussed in a White Paper from the Government (St. Meld. 51 (1973/74, p. 129)). It was also a central election promise given by the ruling Labor party (Arbeiderpartiet) during the local election campaign in 1975. By searching for the string “12 til 18 uker” (12 to 18 weeks) in the national library database and limiting the search to 1970 through 1976, we find a large number (289 in total) of reports in different local and national newspapers about the suggested extension of paid maternity leave from 12 to 18 weeks.¹¹

3 Where did the reform effects identified by CLS come from?

CLS estimate the impact of the maternity leave reform on a range of child outcomes, namely, educational attainment (high-school dropout, college attendance, and completed years of education), adult earnings, teenage pregnancy (females) and IQ (males). They report favorable effects on educational attainment as well as on adult earnings, of magnitudes that are both economically and statistically significant.

The absence of the alleged discontinuity on July 1, together with the fact that public sector workers gained access to the extended entitlement earlier, implies that the identification strategy used by CLS (and reused by BRS) is invalid. In reality, eligible mothers giving birth in June and July 1977 were subjected to exactly the same (paid and unpaid) maternity leave durations and received exactly the same compensation. How, then, could CLS (and BRS) find significant and

⁷ For the state sector, this is described in a circular letter from the ministry (Rundskriv P-65/76 fra Forbruker- og administrasjonsdepartementet: Rett til utvidet førdelspermisjon); see Online Appendix A, Item 7. For the municipal sector, it follows from the tariff negotiation minutes; see Online Appendix A, Item 6.

⁸ This follows from the Royal Decree (Kongelig resolusjon) of June 10, 1977 No. 1747; see Online Appendix A, Item 8. Eligibility for mothers who gave birth up to 12 weeks prior to the cutoff date is also described in a newspaper article explicitly referred to by CLS (Verdens Gang, June 30, 1977).

⁹ The transitional rules are explained in an ad inserted in a large number of Norwegian newspapers by the Government’s information agency on June 30, 1977; see Online Appendix, Item 9.

¹⁰ However, two related reform-elements did generate small discontinuities at this date; see Online Appendix, Item 9. The first was that the one-time payment provided to ineligible mothers (the control group in the CLS analysis) was increased from NOK 800 to NOK 2680, and the second was that eligible mothers could share up to 12 of the 18 weeks with the father. The latter reform element could potentially have had a non-negligible effect on offspring outcomes. However, as noted by both CLS and BRS with reference to Rønsen and Sundström (2002), very few fathers used this opportunity.

¹¹ See the search result here: <https://www.nb.no/search?q=%2212%20til%2018%20uker%22&mediatype=aviser&fromDate=19700101&toDate=19761231>.

robust effects of a non-existent treatment? In this section, we first replicate the CLS findings, and then seek to identify the origins of the reported significant effects.¹²

CLS employ a wide range of regression techniques to identify the effects of the reform; see Table 1, panel A. They start with a single-difference (SD) estimator, essentially comparing children of presumed eligible mothers born in June and July 1977. To circumvent the potential problem of seasonal differences in maternal characteristics (date-of-birth effects), they move on to a difference-in-differences (DD) strategy, comparing the single-difference in 1977 with the corresponding difference in other years. They also apply a regression discontinuity (RD) design, essentially comparing offspring born just before and just after July 1, 1977. Finally, they apply a difference-in-regression discontinuity (RD-DD) design, comparing the July 1 discontinuity in 1977 with the corresponding discontinuities in other years. The specifications are given by

$$Y_i = \alpha + \beta D_i^+ + \varepsilon_i, \quad (\text{SD})$$

$$Y_i = \alpha + \beta D_i^+ + \sum_t \gamma^t W_i^t + \delta D_i^+ W_i^{1977} + \varepsilon_i, \quad (\text{DD})$$

$$Y_i = \alpha + \beta D_i^+ + \tau^+(X_i - c)D_i^+ + \tau^-(X_i - c)(1 - D_i^+) + \varepsilon_i, \quad (\text{RD})$$

$$Y_i = \alpha + \beta D_i^+ + \sum_t \gamma^t W_i^t + \delta D_i^+ W_i^{1977} + \tau^+(X_i^t - c^t)D_i^+ + \tau^-(X_i^t - c^t)(1 - D_i^+) + \varepsilon_i, \quad (\text{RD-DD})$$

where Y_i is the child outcome, D_i^+ indicates births after July 1, W_i^t denotes births in year t , X_i represents the birth date, c marks the reform date, τ^- identifies the trend before the reform, τ^+ characterizes the trend after the reform, X_i^t designates birth date in year t , and c^t specifies the corresponding reform date in year t . Under SD and RD, the parameter of interest is β ; under DD and RD-DD, it is δ . As an alternative to using children born around July 1 in non-reform years as controls in the DD and RD-DD analyses, CLS also use children born in 1977 to ineligible mothers. Their results appear to be robust.

In the DD and RD-DD analyses, CLS use births in 1975, 1978, and 1979 to control for date-of-birth effects. They consistently drop births in 1976 from the analysis, arguing that the new and more liberal abortion law that was implemented in Norway in January 1976 may have affected the sorting into June and July births differently (p. 374). We find this argument a bit surprising. The 1976-law formalized the then already established practice that abortions could be authorized on the basis of the social situation of the mother. It does not appear to have had any effect either on the authorization practice or on the number of induced abortions. While the number of abortions in Norway rose sharply from the late 1960s and up to around 1974/1975, it actually fell slightly in 1976 (Statistics Norway, 1980).¹³ A more fundamental change in the abortion law was implemented in January 1979, introducing full self-determination during the first 12 weeks of pregnancy. Hence, if anything, it is probably the inclusion of 1979-controls that may be questioned. Yet, even this reform appears to have had little effect on the number of induced abortions (Statistics Norway, 1982).

In the process of replicating the CLS analysis, we discovered that CLS made an error in the

¹² We are not able to replicate the BRS findings as we do not have access to the required health data.

¹³ The number of authorized abortions in Norway was 15,132 in 1975 and 14,754 in 1976. Abortion applications authorized as percentage of applications considered increased from 95.0 in 1975 to 96.2 in 1976.

Table 1

IMPACT OF THE ALLEGED REFORM ON CHILDREN'S OUTCOMES

ESTIMATE	Dropout Rate (1)	College Attendance (2)	Ln(Earnings) Age 30 (3)	Completed Years of Schooling (4)	Teen Pregnancy (Females) (5)	IQ (Males) (6)	Birth Weight (7)
A. CLS							
SD	-0.020*	0.017	0.045**	—	—	—	—
	(0.011)	(0.014)	(0.022)	—	—	—	—
Observations	4966	4966	4733	—	—	—	—
DD	-0.032**	0.036**	0.072***	—	—	—	—
	(0.013)	(0.016)	(0.026)	—	—	—	—
Observations	20783	20783	19845	—	—	—	—
RD	-0.022*	0.027	0.062**	0.152	0.004	0.200**	-10.54
	(0.012)	(0.019)	(0.028)	(0.093)	(0.013)	(0.092)	(22.0)
	[0.18]	[0.47]	[12.6]	[13.0]	[0.054]	[5.4]	[3518]
Observations	15025	15025	14348	15025	7194	6838	14979
RD-DD	-0.019**	0.020*	0.050***	0.116**	-0.001	0.084	.429
	(0.007)	(0.011)	(0.016)	(0.053)	(0.007)	(0.054)	(12.8)
	[0.19]	[0.44]	[12.5]	[12.8]	[0.051]	[5.4]	[3505]
Observations	63571	63571	60732	63571	30737	29075	63388
B. REPLICATION							
SD	-0.013	0.017	0.038*	0.113	0.005	0.094	—
	(0.010)	(0.014)	(0.022)	(0.069)	(0.009)	(0.069)	—
	[0.16]	[0.49]	[12.4]	[14.3]	[0.045]	[5.4]	—
Observations	4965	4965	4797	4965	2488	2491	—
DD	-0.021*	0.031*	0.059**	0.192**	0.006	0.177**	—
	(0.012)	(0.016)	(0.025)	(0.081)	(0.010)	(0.080)	—
	[0.17]	[0.47]	[12.4]	[14.2]	[0.048]	[5.4]	—
Observations	20776	20776	20062	20776	10421	10195	—
RD	-0.018	0.029	0.051*	0.187**	0.005	0.140*	—
	(0.011)	(0.021)	(0.030)	(0.091)	(0.012)	(0.084)	—
	[0.15]	[0.50]	[12.4]	[14.4]	[0.048]	[5.4]	—
Observations	15046	15046	14553	15046	7450	7626	—
RD-DD	-0.012*	0.016	0.038**	0.103**	0.002	0.047	—
	(0.007)	(0.012)	(0.016)	(0.052)	(0.006)	(0.052)	—
	[0.16]	[0.48]	[12.4]	[14.2]	[0.046]	[5.4]	—
Observations	63626	63626	61504	63626	31763	31436	—

Table 1**IMPACT OF THE ALLEGED REFORM ON CHILDREN'S OUTCOMES**

ESTIMATE	Dropout Rate (1)	College Attendance (2)	Ln(Earnings) Age 30 (3)	Completed Years of Schooling (4)	Teen Pregnancy (Females) (5)	IQ (Males) (6)	Birth Weight (7)
C. REPLICATION WITH 1976 INCLUDED							
DD	-0.016 (0.012) [0.17]	0.024 (0.016) [0.48]	0.059** (0.024) [12.4]	0.147* (0.078) [14.2]	0.007 (0.010) [0.048]	0.156** (0.077) [5.4]	— — —
Observations	25774	25774	24880	25774	12930	12660	—
RD-DD	-0.009 (0.007) [0.16]	0.009 (0.011) [0.48]	0.040** (0.016) [12.4]	0.067 (0.051) [14.2]	0.002 (0.006) [0.046]	0.029 (0.050) [5.4]	— — —
Observations	78728	78728	76048	78728	39313	38879	—
D. REPLICATION WITH 1976 INCLUDED AND CORRECT ELIGIBILITY							
SD	-0.008 (0.011) [0.15]	0.005 (0.015) [0.51]	0.024 (0.027) [12.4]	0.068 (0.074) [14.4]	0.003 (0.008) [0.039]	0.035 (0.074) [5.4]	— — —
Observations	4420	4420	4168	4420	2219	2215	—
DD	-0.010 (0.012) [0.16]	0.015 (0.017) [0.49]	0.052* (0.030) [12.4]	0.111 (0.083) [14.3]	0.003 (0.010) [0.043]	0.083 (0.082) [5.4]	— — —
Observations	22878	22878	21399	22878	11421	11288	—
RD	-0.012 (0.013) [0.14]	0.013 (0.022) [0.51]	0.029 (0.041) [12.4]	0.127 (0.098) [14.4]	-0.002 (0.011) [0.046]	0.096 (0.087) [5.5]	— — —
Observations	13324	13324	12595	13324	6585	6763	—
RD-DD	-0.005 (0.008) [0.16]	0.007 (0.012) [0.49]	0.029 (0.021) [12.4]	0.053 (0.055) [14.3]	-0.002 (0.006) [0.042]	0.011 (0.050) [5.5]	— — —
Observations	69811	69811	65447	69811	34766	34513	—

Notes: The effects of the alleged reform on children's outcomes using SD, DD, RD, and RD-DD. Panel A restates the findings of CLS. Panel B replicates their findings. Panel C additionally includes 1976 as a control year, and panel D introduces the correct eligibility criteria and includes 1976 as an additional control year. The SD sample consists of births in June and July 1977, and the DD sample contains births in June and July in 1975, (1976), 1977, 1978, and 1979. The RD sample comprises births within 182 days of the discontinuity, while RD-DD samples involve births within 182 days of the respective discontinuities in 1975, (1976), 1977, 1978, and 1979. SD regresses the outcomes on a July indicator. DD runs regressions on year indicators, a July indicator, and a July 1977 indicator. RD and RD-DD report local linear regression estimates with triangular kernels, a bandwidth of 91 days, and separate trends on each side of the discontinuity. For column (3), panels A to C use pensionable income (which includes benefits), while panel D adopts earnings (which exclude benefits). Children born to ineligible mothers are excluded. For Panel D, the correct eligibility criterion (based on nominal earnings) is imposed. The analytical sample is further restricted to residents of Norway only. Means of the prereform sample are in brackets. Following the practice of CLS, we cluster the standard errors, reported in parentheses, at the date of birth for RD and RD-DD. *** p<0.01. ** p<0.05. * p<0.1.

way they defined treatment and control groups. In 1977, the main eligibility criterion for paid leave was that the mother had worked at least six of the 10 months immediately prior to giving birth. Because of limitations in the data (no direct records of employment), CLS constructed a proxy for eligibility by requiring that annual earnings in the year prior to birth exceeded NOK 10,000, pointing out that this earnings level corresponded to the lowest level of income providing pension points in the Norwegian social security system in 1977. However, when CLS imposed this eligibility criterion, they did it based on earnings inflated to 1998-value. Due to the large wage inflation during this period (239%), the eligibility requirement became much milder than intended, and most likely included a number of mothers with incomes inconsistent with the required work experience. If the NOK 10,000 threshold had been measured in 1977-value as appears to have been intended, approximately 15% of the members in the treatment group had been moved to the control group.

To replicate the CLS results, we start by making the same error. As can be seen from Table 1, panel B, we are then able to replicate their main results quite precisely.¹⁴ We find the same (or very similar) positive effects on college attendance, completed years of schooling, IQ (for males) and log earnings at age 30, and negative effect on high-school dropout.¹⁵ However, when we reintroduce 1976 as a control year in the DD and RD-DD analyses, the significant effects disappear for all outcomes except for log earnings at age 30 (DD and RD-DD), completed years of schooling (DD), and IQ (DD); see Table 1, panel C. Hence, the (in our view questionable) decision to drop 1976-births from the analyses appears to have been important for the apparent consistency and robustness of the estimated reform effects. We then repeat the analysis based on a more appropriate income threshold in the definition of treatment and control groups (NOK 10,000 measured in 1977-value); see panel D. This makes the relevant point estimates a bit smaller, and eliminates statistical significance for the remaining parameters, with the DD-estimate for log-earnings at age 30 as the only exception.

Next, we take a closer look at the one remaining significant effect on offspring earnings. Figure 1 reports the estimated effects on log earnings measured at a range of different ages (25–40), based on the RD and RD-DD designs. In panels B.i and B.ii we first show the results based on our replication of the CLS data and method. It is notable that when we use the RD-DD design, age 30 is actually the only age at which the estimated effect is statistically significant at the 5% level. Yet almost all point estimates are on the positive side. Reintroducing 1976 as a control year in the RD-DD-analysis reduces most of the point estimates slightly, but leaves significance statements unchanged; see panel C.ii. Correcting the definition of treatment also does little to modify the results; see panels D.i and D.ii. Hence, the consistent pattern of positive point estimates appears robust with respect to the corrections of eligibility and the inclusion of the additional control year.

¹⁴ In order to match the number of observations in CLS, in panels B and C, we dropped all offspring born to mothers with immigrant background, as well as those who were not residents of Norway at the time the outcomes were measured. In panel D and subsequent analysis, we removed the immigration filter. These choices have little consequence for our results.

¹⁵ Since we do not have access to birth weight data, we are unable to replicate the results under column (7).

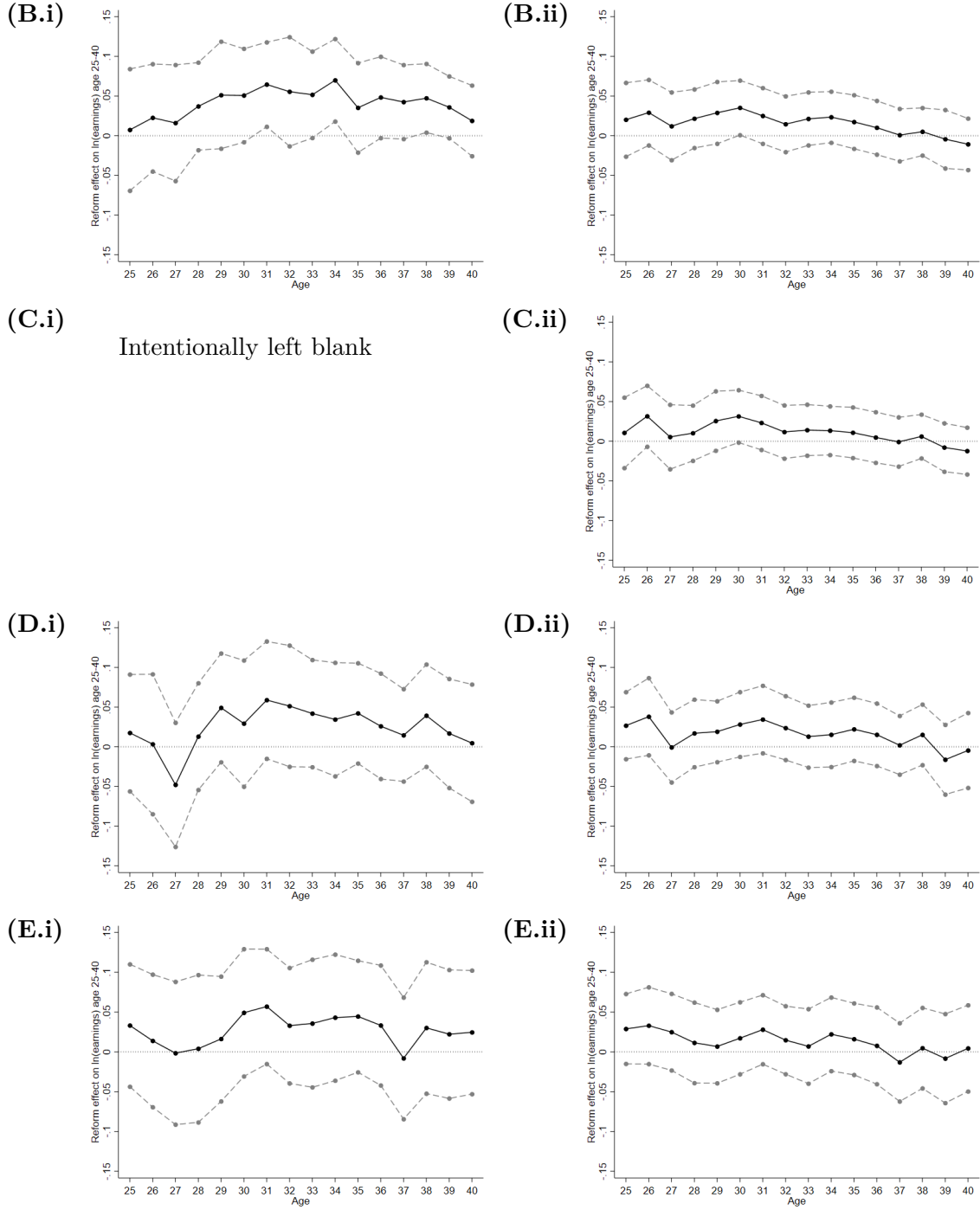


Figure 1

The effects of the alleged reform on children's log earnings at ages 25–40 using *i*, RD and *ii*, RD-DD. Earnings are CPI-adjusted to 1998 values. *B*, Replication. *C*, Replication with 1976 included. *D*, Replication with 1976 and correct eligibility. *E*, Replication with 1976 and correct eligibility, adding 10,000 NOK to all earnings prior to taking the logarithm. Panels B and C use pensionable income (which includes benefits), while panels D and E adopt earnings (which exclude benefits). The analytical sample is restricted to residents of Norway only. Estimates are drawn from local linear regressions with triangular kernels, a bandwidth of 91 days, and separate trends on each side of the discontinuity. Ages are shown on the *x*-axes. The solid black line traces the estimate at each age. The dashed gray lines mark the 95 percent confidence intervals. The dotted horizontal line denotes the significance cutoff.

Given that the alleged treatment-control differential in maternity leave conditions did not exist, the remaining positive effects on adult earning still constitute a sort of puzzle. We now inspect whether they may be driven by the specific choice of functional form. CLS use a linear log earnings specification, standard in earnings regressions, yet known to be sensitive with respect to the inclusion of very low earnings and the exclusion of zero earnings. CLS choose to drop zero earnings from the analysis, implying that 3.2% of the observations are lost from the total sample. We have done the same in our replication and correction exercises. In order to check whether this choice has been decisive for the results, we employ an alternative strategy, where instead of excluding observations with zero earnings, we add a small positive number to all earnings before applying the log function, making it possible to include all observations and also modify the influence of variation across very low earnings. In Figure 1, panels E.i and E.ii show the results from this exercise where we have used NOK 10,000 (measured in 1998-value) as the “small number” added to all observed earnings before the log function is applied (average earnings at age 30 is approximately NOK 300,000 in our data, and NOK 10,000 corresponds roughly to the fifth percentile in the distribution of positive earnings). The resulting earnings estimates are all centered on zero.

Another and perhaps more satisfactory way of dealing with low and zero earnings in relation to the log specification is to aggregate earnings data for many years. We thus use the log of total earnings over the complete period from age 25 to age 40 as an alternative outcome. With this outcome, the fraction with zero earnings drops to 1%. Table 2 shows the results for this model both with zero-earnings observations excluded and with the outcome defined as total earnings plus NOK 10,000 and all observations included. The reform effect is in any case estimated to be approximately zero.

Table 2

IMPACT OF THE ALLEGED REFORM ON CHILDREN’S LOG EARNINGS

	SD	DD	RD	RD-DD
Ln(\sum Earnings, Ages 25–40)	0.007 (0.028) [15.2]	0.027 (0.031) [15.1]	0.022 (0.032) [15.2]	0.014 (0.020) [15.1]
Observations	4401	22719	13270	69400
Ln(\sum Earnings + Δ , Ages 25–40)	0.018 (0.032) [15.2]	0.031 (0.035) [15.1]	0.041 (0.037) [15.2]	0.017 (0.023) [15.1]
Observations	4446	22963	13403	70125

Notes: The effects of the alleged reform on children’s earnings (excluding benefits), summed between ages 25 and 40 and then logged. Alternatively, we add a small amount ($\Delta = 10,000$ NOK) to the sum before taking the logarithm. Earnings are CPI-adjusted to 1998 values. The SD sample consists of births in June and July 1977, and the DD sample contains births in June and July from 1975 to 1979. The RD sample comprises births within 182 days of the discontinuity, while RD-DD samples involve births within 182 days of the respective discontinuities from 1975 to 1979. SD regresses the outcomes on a July indicator. DD runs regressions on year indicators, a July indicator, and a July 1977 indicator. RD and RD-DD report local linear regression estimates with triangular kernels, a bandwidth of 91 days, and separate trends on each side of the discontinuity. Children born to ineligible mothers under the correct eligibility criteria are excluded. Also excluded are children who are residents of Norway for less than 10 years between age 25 and 40. Means of the prereform sample are in brackets. Following the practice of CLS, we cluster the standard errors, reported in parentheses, at the date of birth for RD and RD-DD. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Our conclusion from this exercise is that the seemingly significant and robust reform effects reported by CLS reflect a mixture of coincidence and the authors' specific choices of models and outcomes.

4 What were the true effects of the 1977 maternity leave reform?

The finding of a zero-effect in an econometric analysis exploiting a non-existent discontinuity says of course nothing about the reform's actual effects.

In Online Appendix B, we seek to identify the true effects of the five-to-six-week extension of parental leave based on a correct description of its implementation. Unfortunately, the foundation for robust identification is not nearly as sound as it would have been had the CLS description of the reform been correct. The timing of the extension and the way it was implemented varied across sectors of employment (state, municipality, private), and data linking employed mothers directly to correct sector are not available for the time period studied in this paper. Such data are available from 1992, however; hence, our strategy is to use observed employment in 1992, in combination with data on educational attainment and age, to predict sector of employment in 1977 (and neighboring years). We then apply regression discontinuity (RD) and difference in regression discontinuity (RD-DD) models to estimate the effects of the reform, taking into account the specific implementation practice in each sector.

It is difficult to assess the quality of our sector-of-employment predictions. What we can do is to examine how well we *would have* predicted had we used the same prediction algorithm to guess sector in 1992 from data on employment and education in 2006. The result from this exercise indicates that we predict correct sector of employment for approximately 70% of the mothers.

In the online appendix, we present estimation results for nine different offspring outcomes; i.e., dropout probability, college attendance, completed years of schooling, teen pregnancy, IQ (males), log earnings at age 30 (with and without a small number added) and log sum of all earnings age 25–40 (with and without a small number added). We estimate three different models; i.e., conventional RD, bias-corrected RD, and RD-DD. And we estimate models for each of the three sectors separately, as well as jointly for all sectors. In total, this leaves us with 108 estimated parameters. We obtain nine estimates that are statistically significant at the 5% or 10% level, six indicating favorable effects and three indicating adverse effects of the reform. The number of significant coefficients is lower than what we would expect from purely random data, and with no particular pattern in point estimates. When we involve the DD specification using non-reform years or ineligible births as controls, none of the estimated parameters is statistically significant. Hence, we conclude that we are unable to verify any effects of the reform on offspring outcomes.

5 Conclusion

In a highly cited paper published in the *Journal of Political Economy*, Carneiro, Løken, and Salvanes (2015) (CLS) presented empirical evidence apparently indicating that the introduction of paid maternity leave in Norway had large and favorable effects on offspring outcomes. In a more recent paper published in the *American Economic Journal: Economic Policy*, Bütikofer, Riise, and Skira (2021) (BRS) present evidence that the same reform improved a range of maternal health outcomes.

In the present paper, we have shown that both results were built on an incorrect description of the evaluated reform and on an invalid identification strategy. Instead of introducing paid maternity leave with a direct shift from 0 to 18 weeks, the reform extended mandated paid leave by six weeks—from 12 to 18. And instead of being introduced with a sharp discontinuity at a specific cutoff date, the reform was phased in gradually. Consequently, the alleged treatment and control groups used to identify the maternity leave effects in CLS and BRS were in reality subjected to the same maternity leave conditions.

To identify the origins of the effects reported by CLS, we have replicated their results, and then introduced slight modifications to their modeling choices, to which the results reported by CLS are not robust. We conclude that their findings may have come about through a combination of unintended data construction errors, coincidence, and model selection.

Finally, we build on a correct description of the reform’s implementation to estimate the true effects of the six-week paid maternity leave extension on offspring outcomes, admittedly with a relatively weak foundation for causal identification. In line with the existing literature, we find no significant effects. However, whether or not there are favorable effects of the very first weeks of paid maternity leave cannot be answered based on the reform evaluated by CLS and reexamined in the present paper. At this point, the jury is still out.

In this case, it is clear that two highly skilled research groups have reported significant, robust, and large effects of a treatment-control differential that we now know never existed. Both papers appear extremely convincing, and they have been published in top journals. We believe there is an important lesson to be learned from this “placebo treatment” experience, in which there are the researchers—and not the agents—that happen to (erroneously) believe some specific treatment has occurred. Given the richness of today’s data sources and the wide range of alternative empirical tools, a sufficiently thorough search for significant and robust effects in a particular direction may turn out successful even if no such effects exist in reality. As most researchers are well aware that publication in a top journal requires seemingly watertight evidence behind a good story, the incentives for communicating any doubts or nuancing results through the submission and refereeing processes are weak. This accidental placebo treatment experiment should perhaps be food for thought for all of us working in the field of empirical research.

References

- Almond, Douglas, Currie, Janet, and Duque, Valentina (2018). “Childhood Circumstances and Adult Outcomes: Act II”. *Journal of Economic Literature* 56.4, pp. 1360–1446. DOI: 10.1257/jel.2017.1164.
- Baker, Michael and Milligan, Kevin (2015). “Maternity Leave and Children’s Cognitive and Behavioral Development”. *Journal of Population Economics* 28.2, pp. 373–391. DOI: 10.1007/s00148-014-0529-5.
- Bütikofer, Aline, Riise, Julie, and Skira, Meghan M (2021). “The Impact of Paid Maternity Leave on Maternal Health”. *American Economic Journal: Economic Policy* 13.1, pp. 67–105. DOI: 10.1257/pol.20190022.
- Carneiro, Pedro, Løken, Katrine V, and Salvanes, Kjell G (2015). “A Flying Start? Maternity Leave Benefits and Long-Run Outcomes of Children”. *Journal of Political Economy* 123.2, pp. 365–412. DOI: 10.1086/679627.
- Dahl, Gordon B, Løken, Katrine V, Mogstad, Magne, and Salvanes, Kari Veia (2016). “What Is the Case for Paid Maternity Leave?” *Review of Economics and Statistics* 98.4, pp. 655–670. DOI: 10.1162/REST_a_00602.
- Danzer, Natalia and Lavy, Victor (2018). “Paid Parental Leave and Children’s Schooling Outcomes”. *The Economic Journal* 128.608, pp. 81–117. DOI: 10.1111/ecoj.12493.
- Dustmann, Christian and Schönberg, Uta (2012). “Expansions in Maternity Leave Coverage and Children’s Long-Term Outcomes”. *American Economic Journal: Applied Economics* 4.3, pp. 190–224. DOI: 10.1257/app.4.3.190.
- NOU (1996). “Offentlige overføringer til barnefamilier”. Norges Offentlige Utredninger 1996:13. <https://www.regjeringen.no/contentassets/5a5c94a53b474bc3bedf00d9f836eee4/no/pdfa/nou199619960013000dddpdfa.pdf>.
- Rasmussen, Astrid Würtz (2010). “Increasing the Length of Parents’ Birth-Related Leave: The Effect on Children’s Long-Term Educational Outcomes”. *Labour Economics* 17.1, pp. 91–100. DOI: 10.1016/j.labeco.2009.07.007.
- Rønsen, Marit and Sundström, Marianne (2002). “Family Policy and After-Birth Employment Among New Mothers: A Comparison of Finland, Norway and Sweden”. *European Journal of Population/Revue Européenne de Démographie* 18.2, pp. 121–152. DOI: 10.1023/A:1015532305179.
- Rossin-Slater, Maya (2017). “Maternity and Family Leave Policy”. NBER Working Paper 23069. DOI: 10.3386/w23069.
- Statistics Norway (1980). “Helsestatistikk 1980”. Statistisk sentralbyrå. https://www.ssb.no/a/histstat/nos/nos_b277.pdf.
- (1982). “Helsestatistikk 1982”. Statistisk sentralbyrå. https://www.ssb.no/a/histstat/nos/nos_b465.pdf.