Research Group on Human Capital
Working Paper Series

Connecting the Dots: The Early Impacts of Increased Paid Maternity Leave on Child Development

Working Paper No. 15-01

Catherine Haeck

February 2015

http://grch.esg.uqam.ca/working-papers-series/?lang=en

Connecting the Dots: The Early Impacts of Increased Paid Maternity Leave on Child Development

Catherine Haeck*

Abstract

This paper evaluates the effect of extended maternal care on children's development at age 4 and 5 using observational data prior to and after the Canadian parental leave reform, which extended total paid leave from 25 to 50 weeks. In contrast with previous research on the Canadian parental leave reform, we estimate the impact of the reform while controlling for underlying trends in the outcome variables. We find that the policy change had positive effects on the cognitive development of children as well as parent-reported measures of child health and family well-being. Effects on behavioral development are mainly not significant. These results must be interpreted with respect to the effective treatment period and the type of care displaced. We find that mothers increased their time at home from 7 to 11 months and that the type of care displaced was mainly unregulated and provided by individuals without specific training.

Keywords: maternity leave reform, child development, family well-being, natural experiment

JEL Classification: J13, J18, J22, J24

* *Catherine Haeck:* Université du Québec à Montréal (haeck.catherine@uqam.ca). I gratefully acknowledge financial support from the Belgian Federal Science Policy Office (Interuniversity Attraction Poles P5/26). I thank Bart Cockx, Maarten Goos, Pierre Lefebvre, Philip Merrigan, Anna Salomons, Johannes Van Biesebroeck, Frank Verboven, Xuelin Zhang, Jason Gilmore, and the participants at the 2011 SCSE Conference, the 2011 CEA Conference and the CES-LICOS joint seminar in Leuven. I also want to thank the Quebec Inter-University Centre for Social Statistics (QICSS) for great support throughout this project. The analysis is based on Statistics Canada's National Longitudinal Survey of Children and Youth (NLSCY) restricted-access Micro Data Files available at the QICSS. Finally, I want to thank Samuel Paré and Martin Leblond-Létourneau for skillfull research assistance. All computations on these micro-data were prepared by the author who assumes responsibility for the use and interpretation of these data.

1 Introduction

Understanding the early life determinants of ability is at the centre of a large body of research. It has become clear that gaps in ability identified as early as age 5 are strong predictors of future adulthood skill level, and that low skill level in adulthood is associated with lower economic success (earnings and probability of employment) and with a number of socioeconomic problems, such as dropping out of high school, crime, and chronic health conditions (Heckman, 2008). Maternity leave policies are often enacted based on the premise that children will benefit from an extended period of time spent with their mother in their first year of life. While most OECD countries have government regulated paid maternity leave and generally offer at least 25 weeks, research on the link between maternal time investment and the early development of children provides inconclusive results.

Research on the link between maternal employment and child development mainly uses a multivariate approach, but also sibling fixed-effects, instrumental variable approach and structural modeling. Generally, these studies suggests that maternal employment during the first year of life is detrimental to child development. A growing number of research studies exploit changes in job protected parental leave reforms to assess the impact of time investment on child development. By changing the family budget constraint, paid parental leave reforms induce changes in the optimal allocation of time between work and home. These exogenously induced changes in the labor supply of mothers provide convincing evidence of the impact of extended maternal care. Carneiro et al. (2015), Danzer and Lavy (2013), Cools et al. (2011) and Liu and Skans (2010) find positive effects, but in most cases on a specific sub-group of children. In contrast, Dahl et al. (2013), Baker and Milligan (2008a, 2010, 2015), Rasmussen (2010) and Dustman and Schönberg (2012) find that parental leave reforms have no impact on child development.

The main objective of this paper is to provide new evidence on the link between maternal time investment in the first year of life and child development at age 4 to 5.[1] To do so, we exploit a unique and relatively recent natural experiment that extended total available paid maternity[2] leave from 25 to 50 weeks across Canada starting January 1^{st}, 2001. In contrast with previous research on the Canadian parental leave reform, we estimate the impact of the reform while controlling for underlying trends in the outcome variables. More specifically, children of mothers not eligible for paid maternity leave are used to control for underlying trends in the outcome variables in a matching difference-in-differences (MDID) estimator. Furthermore, while ability gaps measured as early as age 5 have been shown to persist later in life, no research has investigated the impact of maternity leave reforms on children aged

[1] This is the last point of observation prior to school entry and therefore estimated effects do not depend on the mediating effect of schooling.

[2] From here on, we focus on extended maternal leave benefits, as opposed to parental leave, because the data reveals that in the first few years after the implementation of the amendment, the take-up rate by fathers was still fairly limited (Marshall, 2008).

5 while controlling for underlying trends in the outcome measures. We estimate the impact of the reform on the time mothers stay at home with their children after birth, and on a number of standardized measures of cognitive development and parent-reported measures of behavioral development, health and family well-being.

Three main conclusions emerge. First, benefit recipient mothers increased their time away from work in the first year after their child's birth by about 3 to 4 months, up from 7 months on average prior to the reform. Second, the policy change had positive effects on some measures of children's cognitive development and parent-reported measures of child health and family well-being. Third, the evidence on behavioral development is mixed, except for conduct disorder which appears to have improved, but the effects are generally not significant.

This paper contributes to the literature in three ways. First, it extends the work of Baker and Milligan (2010, 2015) and provides new evidence on the effect of the Canadian parental leave reform using a different empirical approach. Second, it provides further evidence on the link between early childhood development at age 5 and later outcomes, by demonstrating the early life impact of a reform whose treatment is comparable to the Norwegian maternity leave reform found to have a positive impact on the probability of not dropping out of high school (Carneiro et al., 2015). Third, it shows the importance of controlling for underlying trends using a difference-in-differences approach and non linearity in the control variables using matching techniques.

The outline of the paper is as follows: Section 2 provides an overview of the literature on maternity leave reform. Section 3 describes the reform and Section 4 the data set. Section 5 outlines the framework and empirical strategy. The econometric results are presented and analyzed in Section 6. We conclude in Section 7.

2 Empirical Evidence

This section summarizes recent empirical research findings on the link between parental leave in the first year of life and child development. In Table A.1 and A.2, research studies linked to maternal time investments are classified into two groups. Specifically, we differentiate between factors affecting maternal time investment: (1) maternal employment and (2) parental leave reforms.

Maternal employment Research on the link between maternal employment and child development mainly uses a multivariate approach, but also sibling fixed-effects, instrumental variable approach and structural modeling (Bernal, 2008; James-Burdumy, 2005; Baum, 2003, Brooks-Gunn et al., 2002b, Blau and Grossberg, 1992). Generally, these studies suggests that maternal employment during the first year of life is detrimental to child develop-

ment (Brooks-Gunn et al., 2010; Lucas-Thompson et al., 2010; Berger et al., 2008; Bernal, 2008; Sherlock et al., 2008; Gregg et al., 2005; Hill et al. 2005; James-Burdumy, 2005; Rhum, 2004; Baum, 2003; Brooks-Gunn et al., 2002a; Brooks-Gunn et al., 2002b; Han et al., 2001; Barglow et al., 1998; Belsky et al., 1991, Blau and Grossberg, 1992; Desai et al., 1989), with a few finding minimal adverse effects (Verropoulou and Joshi, 2009; Harvey, 1999).

Findings on the impact of employment during the second and third year are less conclusive: they range from significantly negative to neutral, to significantly positive, even when the same data set is used (see, for example, studies using the National Longitudinal Survey of Youth (NLSY) summarized in Table A.1).While the inclusion of different control variables and the selection of different age groups and subgroups of the population may explain in part the disparity in the results, issues in fully accounting for the selection bias may also partially explain the lack of consensus.[3] The inclusion of an extensive set of control variables cannot often eliminate the bias and can rarely predict the sign and magnitude of the bias (Duncan et al., 2004). As shown by Bernal (2008), a mother's decision to return to work is influenced by her child's cognitive endowment, as well as the budget constraint she faces and her preferences. Using a structural approach, she finds that full-time maternal employment and childcare have a sizeable negative impact on cognitive development (around 0.13 std. dev. for one year of employment during the first 5 years). She also finds that this effect is greater for children with higher ability endowments.

Parental leave reforms In addition to structural modeling, a recent but growing number of research studies have exploited different paid maternity leave reforms in Western Europe and Canada to palliate the issue of selection bias. Paid maternity leave reforms induce exogenous shocks on the family budget constraints, and thereby influence the labor supply of mothers and the time they spend with their child during infancy.

In Canada, a number of research studies have been conducted using the January 1^{st}, 2001 reform. Baker and Milligan (2008a) find that the reform increased the time mothers stayed out of the labor market in the first year of life by about 2.3 months (up from 8.2 months) and significantly raised the proportion of children breastfed above the critical 6 month period. They found that parent-reported health benefits before age 3 were modest and generally not significant. Baker and Milligan (2008b) use the Canadian variations in leave entitlements since 1963 to show that short leaves (17-18 weeks) generally do not impact the average time mothers stay out of the labor market, while longer leave periods significantly raise the average time a mothers stays out of the labor market.[4] As paid parental leave may not only

[3]For example, Rhum (2004) shows that the inclusion of what he calls supplemental variables and maternal employment controls greatly impacts the estimated effect of maternal labor supply, but are often not included in most other comparable work.

[4]Baker and Milligan (2008b), section III, provides a review of the literature on the link between maternity leave and maternal labor supply.

impact the time a mother stays out of the labor market but also her income while on leave, Baker and Milligan (2010) carefully show that, at statutory income replacement rate of 55%, the income after-tax, transfer and childcare cost for a median income mother is equivalent whether she works or stays at home and claims maternity benefits. The income effect of parental leave policies is rarely documented in the literature, yet it is crucial to understand the treatment induced by the reform. In this paper, they also investigate the impact of the reform on child development at age 7 to 24 months. Using a variety of parent-reported measures of development the authors conclude that the effects are largely not significant. Baker and Milligan (2015) look at the impact on children age 4 and 5, and again find no significant effects. In both papers (2010, 2012), they use a before-after approach where each birth year cohort is pooled together and no distinction is made between treated and non-treated children.[5] To address the possibility of underlying trends in the outcome variables, Baker and Milligan (2015) mainly use polynomials in time defined at the quarter of birth level. In the empirical results section, the parallel between their work and this paper's finding is provided.

In Norway, Carneiro et al. (2015) study the 1977 reform that for the first time provided paid maternity leave for a duration of 18 weeks in combination with an extended unpaid leave period of 52 weeks, compared to 12 weeks prior to the reform. The income replacement rate is 100% for the entire 18 week period, and about two-thirds of the mothers are estimated to be eligible. Assuming that all eligible mothers take the full 18 weeks and using a predicted measure of unpaid leave duration (using income around birth), the authors estimate that mothers increased their time at home from 8 to 12 months on average. Using discontinuity design in combination with difference-in-differences, the authors find that the high school completion rate of children of eligible mothers significantly increased following the reform and that children of low educated mothers benefited most from the reform. Their results are significant when using the eligible group (children of mothers eligible for the reform[6]), but no longer significant when all children (eligible and not eligible) are pooled together. Their findings show the importance of being able to identify children of eligible mothers. Few studies are able to match children and mothers, and therefore estimate the impact of a reform on children of mothers who are eligible. The approach used in this paper also exploits the identification of children of eligible mothers. Our inference of eligibility is also imperfect, and as a robustness check children of mothers receiving benefits are used. These are perfectly identified, and likely match those that are eligible given the generosity of the

[5] More specifically, using children born between 1997 and 2004 they implement a two step procedure. First they regress the outcome variable on a set of controls and year of birth dummies (without a constant). Then they regress the estimated year of birth effects on a dummy variable (instrument) equal to one if the birth cohort was exposed to the reform. They also use an IV approach in which the exclusion restriction implies that the reform (post dummy) only affects children through time spent at home by the mother.

[6] Eligibility is inferred from observation of annual income.

program, especially in the first few weeks after birth.

In Western Europe, several other recent papers exploit parental leave reforms to assess the impact of maternal time investment on child development. Most papers rely on some form of research discontinuity design (RDD) and do not identify children of eligible mothers. Since RDD focuses on a fairly short time period, bias induced by underlying trends is unlikely. However, estimated impact using RDD should be interpreted in light of those most likely to give birth around the discontinuity point.[7] As pointed out by Buckles and Hungerman (2008), timing of birth is related to socioeconomic status, with a higher proportion of high-socioeconomic status mothers having babies in the summer. All papers from Western Europe focus on long term outcomes beyond age 12, and therefore bring considerable evidence of the long term effect of parental leave policies. They are summarized in Table A.2.

Overall, Carneiro et al. (2015), Danzer and Lavy (2013), Cools et al. (2011) and Liu and Skans (2010) find positive effects, but in most cases on a specific sub-group of children. In contrast, Dahl et al. (2013), Baker and Milligan (2008a, 2010, 2015), Rasmussen (2010) and Dustman and Schönberg (2012) find that parental leave reforms have no impact on child development. Compared to other studies, Rasmussen (2010) and Dahl et al. (2013) study reforms of short duration (6 weeks, and 2 to 4 weeks respectively), while Dustman and Schönberg (2012), except for income at age 28-29, use aggregated outcomes which may fail to identify impact at the top or bottom of the education distribution. While most studies rely on RDD, Baker and Milligan (2015) use a before-after approach and do not identify eligible children. More generally, the various results from studies exploiting changes in parental leave reform may be explained by six distinct sources of variation: (1) the wide variety of extension periods (e.g. 14 to 20 weeks in Rasmussen (2010) versus 12 to 24 months in Danzer and Lavy (2013)) and effective treatment periods, (2) the impact on disposable income (accounting for work related expenses which likely varies by country), (3) the type of care displaced, (4) the outcome measures used, (5) the empirical approach, and (6) the sub-sample used. A precise identification of these key components is a necessary condition to correctly characterize and compare the causal impact of these types of reforms. Unfortunately, due to data constraints, few papers actually document the effective treatment period (in terms of the age of the child when the treatment actually takes place), the impact on disposable income, and the type of care displaced. This makes comparison of results particularly difficult. To the extent that the information is available, section 7 position this reform with respect to others to facilitate the interpretation of the results.

While this study documents most of the key sources of variations, it has its own limitations with respect to its data set. First, given our sample size, sub-groups analysis is not possible and may be important to understand the differential impacts across children

[7] This study also focuses on a specific subset of mothers, those in dual-parent families. This restriction is carefully explained in the following section.

(e.g. Danzer and Lavy, 2013). Second, although we can perfectly identify mothers taking up leave (and their children), we can only infer eligibility through income related questions. Third, to obtain a reasonnable sample size, we have to use a fairly large time period, and as a result selection into the reform becomes a possibility. This paper addresses selection carefully by characterizing the extent to which selection may have occurred and by controlling for it using a matching approach. This study is a complement to previous research on the long term impacts of parental leave reform. More specifically, given the comparability of the Norwegian and Canadian reform in terms of effective treatment period, the positive early life impacts identified in this study provides further evidence on the link between early childhood development and later outcomes (Almond and Currie, 2011).

3 The reform

In Canada, maternity leave benefits are legislated and paid by the federal government through Employment Insurance (EI), while maternity leave duration is regulated at the provincial level. More specifically, the federal government legislates the amount to be paid during the leave, while the provincial governments set the time a mother may stay at home and preserve the right to return to her former job.

In the 1990's, the EI program provided mothers with a minimum of 700 hours of insurable employment in the 12 months preceding birth with 15 weeks of paid maternity leave. The Parental Benefit Program (PBP) also provided an additional 10 weeks of paid maternity leave that could be shared by both parents. At the time, a mother could therefore claim up to 25 weeks of paid leave. On December 31^{st}, 2000, the federal government passed an amendment to the EI Act increasing the PBP from 10 weeks to 35 weeks. This extension effectively resulted in a 6 month increase in paid maternity benefits. At the same time, the number of insurable hours required for eligibility was lowered from 700 to 600 hours. The rate of coverage remained unchanged at 55% of prior earnings. To protect working mothers while on paid maternity leave, provincial laws were also adjusted and leave duration increased to at least 50 weeks. The Canadian maternity leave extension was primarily designed to help "parents balance their work and family responsibilities and ensure that children get the best possible start in life" (HRSDC, 2005).

We focus our attention on children's outcomes, including assessments of their cognitive and behavioral abilities and their health status, and leave the assessment of the other anticipated gains (e.g. reduced work absenteeism by parents, more productive work force, increased employee retention, etc.) to future research. Since the family environment is critical to the development of children, the assessment of gains on children should take into account work-life balance effects. Therefore, we also report the effect of the policy on two parent reported measures of family well-being.

4 Data set and contextual setting

The Canadian National Longitudinal Survey of Children and Youth (NLSCY) is a long-term biennial survey providing detailed information on the development and well-being of Canadian children. The first survey was administered in 1994 and provided detailed information on children during the period 1994-1995. Since then, seven other cycles have been released by Statistics Canada. The last cycle (cycle 8) was released in November 2010 and covers 2008 and 2009. For each of the cycles, representative longitudinal samples of the population of children living in Canada's 10 provinces are constructed.[8] The NLSCY is an extremely rich data set that contains detailed information on the demographic situation of the family, education, labor force, and income of both parents, as well as a detailed account of pregnancy and birth conditions. It also contains a number of developmental measures for children aged 4 to 5. We provide a brief overview of the subpopulation of children on which the effects are estimated by first presenting the subsample used and then an overview of the outcome measures used.

Subsample definition In the NLSCY, information on maternity leave benefits and maternal labor force participation prior to birth is not fully consistent across cycles. In cycles 1 to 3, no information was collected on maternity leave benefits. In cycle 4, questions related to maternity leave and labor force participation only referred to the past 12 months (e.g. Since 12 months ago, how many weeks have you been on paid maternity or parental leave?). As of cycle 5, we can perfectly identify mothers working prior to birth, as well as those benefiting from maternity leave benefits.[9] Cycle 4 provides data on children born in 1999 and 2000 (prior to the reform). Cycle 5 provides data on children born in 2001 and 2002 (after the reform).

Given the information available in the different cycles, we decided to restrict our attention to children born between 1999 and 2002: two years before the reform and two years after the reform. Effectively, we use the panel data of the subpopulation of 0 to 5 year olds contained in cycles 4 to 7, inclusively. Furthermore, in line with the previous literature on the Canadian parental leave reform, we focus our attention on two-parent families outside of the province of Québec to avoid the cofounding effects of the National Child Benefits Program and the Quebec childcare reform (for more details on these reforms see Baker and Milligan, 2010). From here on, we exclusively focus on children born between 1999 and 2002 outside of Québec and living in two-parent households until age 5.

Table 1 shows the summary statistics for children observed in the main sample. This

[8] Longitudinal weights adjusted for total non-response matching known population count are provided.

[9] On work prior to birth, we only know if the mother worked or not, and therefore cannot, for example, exploit the discontinuity in hours of work to identify the impact of the reform around the discontinuity point of 600 hours.

sample includes all children born between 1999 and 2002, outside of Québec and living in a two parent family. From left to right, children are divided into four subgroups: children whose mother worked prior to birth born before the reform (1) and born after the reform (2), and children whose mother did not work prior to birth born before the reform (3) and born after the reform (4).

Looking at the child characteristics, Table 1 reveals that children in all four groups share similar characteristics: they are on average 58 months old at the time of the third interview and 15 months old at the time of the first interview. The share of males is slightly different, but close to 50%. All children share similar birth characteristics in terms of prematurity, birth weight, hospitalization and multiple births. They also tend to be equally breastfed at birth. As previously mentioned, research has shown that the reform had a positive effect on the length of time a mother continued breastfeeding. In the empirical section below, we only control for whether or not the mother breastfed her child at birth, to capture potential effects explained by increased length of breastfeeding (e.g. health related benefits). The age at which the child slept a minimum of 6 non interrupted hours is also comparable.

Family characteristics are slightly different between groups. Children of non working mothers (control) before and after the reform have more siblings than children of working mothers (treated). This is an important characteristic to control for since mothers with a higher number of children have less time to devote to the newborn baby. Generally, families with working mothers have better outcomes on all four family related scales before and after the reform. Both parenting style measures are based on a series of questions asked to the parents: one measures parent effectiveness (ineffective parenting) and the other consistent parenting (positive parenting). The family functioning and social support measures are also based on a series of parent reported measures. The family functioning score is based on 13 questions and indicates the presence of family dysfunction (e.g. drinking is a source of disagreement and tension in our family). The social support score is based on 8 questions and indicates the presence of a social network supporting the family (e.g. there are people I can count on in case of emergency). While a higher score indicates a worsening of the family functioning and ineffective parenting measures, a higher score indicates an improvement of the positive parenting and social support measures.

The mothers characteristics are slightly less stable across all four groups. Working mothers are more educated. Furthermore, post reform all mothers are more educated. This can be attributed to a change in the way Statistics Canada computed the number of years of education in cycle 5 (compared to cycle 4), but also possibly selection into and out of the labor force. We specifically discuss this possibility in the empirical section below. Working mothers are typically not recent immigrants, with about 5% having immigrated in the last 4 years compared to more than 17% to 20% in the control group. Since cognitive development tests are administered in English or French, the effect of time spent with an immigrant

mother in early life may not be well captured by these tests. The father's characteristics (education and income quartile) are fairly comparable, but again fathers appear to be more educated after the reform if their partner is working.

In sum, treated children are generally comparable to non treated children, except that they have mothers who are more educated, have fewer children and are less likely to have immigrated recently.

Outcome measures At age 4 and 5, three tests measuring cognitive development are administered: the Peabody Picture Vocabulary-Revised (PPVT) Test, the Who Am I? Test and the Number Knowledge Test. The PPVT measures receptive and hearing vocabulary. This cognitive test is widely used in the literature to measure school readiness. The Who Am I? includes two sets of tasks. The copying tasks measure the child's ability to visualize and reproduce geometric figures. The writing tasks measure the ability of the child to understand and use symbols, such as letters and numbers. A total of 10 questions are answered by children. Each question is given a score of 1 to 4, such that the overall score can range from 10 to 40. The Number Knowledge assesses a child's understanding of whole numbers. This test measures essential mathematical skills required for successful school learning. The test includes 30 questions. The overall score can range between 0 and 30. All three tests are well suited to measure the development of children aged 4 and 5. All tests are, however, age sensitive, with older children scoring higher on average. Following previous research, we use aged standardized scores.[10]

Four parent-reported measures of social (or anti-social) development are available. The emotional score (1) is based on six questions and indicates the presence of anxiety and emotional disorder (e.g. how often the child gets nervous, high strung or tense). The inattention score (2) is based on seven questions and indicates behaviors associated with hyperactivity and inattention (e.g. how often the child cannot focus on anything for more than a few moments). The conduct score (3) is based on six questions and indicates the presence of conduct disorder and physical aggression (e.g. how often the child gets into a fight). Finally, the aggression score (4) is based on five questions and indicates the presence of behaviors associated with indirect aggression (e.g. how often the child, when mad at someone, tries to get others to dislike that person). A higher score implies further evidence of behavioral disorder across all four measures.

Since adverse family environment is detrimental to child development, and the reform was explicitly designed to help families, not just children, we also estimate the impact of the

[10]More specifically, we use the PPVT score standardized by Statistics Canada using a normative sample of children from Cycle 1 to 5. For the Who Am I? and Number Knowledge tests we accounted for age differences using age in month dummies. Since these tests were first taken as of Cycle 4, it was not possible to create a stable normative sample of children prior to our observation period, but following the approach used by Statistics Canada we were able to use lowess smoothing to ensure that the expected score increases with age in months.

reform on two measures of the environment of the family when the child is 4 and 5. The family functioning scale and the social support scale are based on the same set of questions as the outcome measures discussed before (family functioning and social support) but asked at the time of the third interview, when the child is aged 4 to 5. Finally, given the potential effect on health, we also estimate the impact on two parent-reported measures of health: a general assessment of the child's health (excellent to poor, 5 levels), and the frequency at which the child has been in good health in the past few months (almost all the time to almost never, 5 levels).

The summary statistics for these outcome measures are presented at the bottom of Table 1. For the cognitive measures, we see that post reform children generally score lower, but this is even more the case for children of non working mothers. This would suggest that the reform had a positive impact on cognitive development. For the family and health measures the trend is more positive for children of working mothers than for other children (considering that for three of these measures a lower score indicates a better outcome, marked with (-) in Table 1). The patterns are less stable for the behavioral measures.

5 Empirical strategy

Our econometric approach is based on a difference-in-differences (DID) procedure. We observe children of working mothers, born before and after the policy change. These children include both children of mothers eligible to maternity leave and children whose mother was not eligible (e.g., self-employed mothers). Our comparison group will be children of non working mothers born during the same period. Non working mothers were not eligible for paid maternity leave before and after the reform. This grouping has the advantage of preserving our group composition relative to the change in the number of hours worked prior to birth required for eligibility (from 700 to 600). Indeed, post reform, some mothers who were not eligible prior to the reform because they had not worked enough hours were now eligible because of the lower threshold.

An alternative source of selection is that of mothers who may have self-selected into or out of the labor force or may have delayed their conception. Both are most likely extremely modest for three reasons. First, the percentage of mothers working prior to birth is stable over the observation period. Indeed, using the Statistics Canada Employment Insurance Coverage Survey (EICS), we find that the participation rate is 73.0% prior to the reform and 74.8% after. Using the EICS, we find no evidence of selection into treatment around the policy change.[11] This suggests that observed differences in our subsamples are most likely random sample differences. Furthermore, evidence from maternal employment trends of Canadian women (excluding Québec) with children below age 5 does not suggest that

[11] Further details are presented in our Web Appendix.

mothers labor market participation changed drastically around the maternity leave reform (Lefebvre and Merrigan, 2008).[12] Prior Canadian evidence by Phipps (2000) suggests that women do not adjust their labor-supply behavior to benefit from parental leave benefits. Second, using the Canadian Vital Statistics Birth Database we find that mothers did not delay conception to benefit from the reform during our observation period (see Appendix for more details). Nonetheless, as discussed above, Table 1 shows that both treated and control mothers are more educated post reform (by 1.4 and 0.6 years respectively). Since maternal education is strongly related to child development, it is key to control for this difference.

In a first step, we implement a simple before-after (BA) model to allow comparison of our results to those of Baker and Milligan (2015). Effectively, for each of the outcomes listed above, the average outcome of children born in 1999 and 2000 is compared with that of children born in 2001 and 2002. Given that the BA approach fails to control for underlying trends, in a second step, the following difference-in-differences (DID) model is implemented:

$$y_{i,by+t} = \alpha + \theta I(by \geqslant 2000) + \gamma T_{i,by} + \beta T_{i,by} I(by \geqslant 2000) + \varepsilon_{i,by+t} \tag{1}$$

where by is the birth year of infant i and t is the number of years between the birth of the child and the interview at age 4 or 5. $I(by \geqslant 2000)$ is an indicator function equal to one if the child was born after the policy change and zero otherwise. $T_{i,by}$ is the treatment status of the mother and is equal to one if the mother worked prior to birth (prior to and post reform) and equal to zero otherwise. $\varepsilon_{i,by+t}$ is an error term. The estimated effect of the policy reform is β. The DID estimator can be consistently estimated using OLS under the following assumptions: (1) common trend, and (2) no selection on transitory shocks. Assumption (1) implies that common shocks such as a drift in the PPVT score[13] do not impact the consistency of the DID estimator. Under assumption (2), the DID estimator is consistent even in the presence of selection on unobservable individual fixed effects. More explicitly, eligible mothers may have permanent differences when compared with non eligible mothers and these differences can influence the outcome variables. The same holds for permanent differences between children.

To this simple model, two modifications can be made to assess the sensitivity of the DID estimator to different covariates. First, we add age dummies at the time of test to further account for the age sensitivity of certain outcomes, from 47 to 60 months old. Second, we also include a number of control variables to account for maternal, paternal, child and family

[12] Figures 1 to 5, in Lefebvre and Merrigan (2008), show the trends in labor force participation, annual weeks worked, annual hours worked, annual earned income and proportion in full-time employment for mothers of young children in the Rest of Canada (RofC) .

[13] In Cycle 4, Statistics Canada conducted an analysis which provides evidence of drift in the level of difficulty of some questions in the PPVT Test.

characteristics. Equation 1 becomes:

$$y_{i,by+t} = \alpha + \theta I(by \geqslant 2000) + \gamma T_{i,by} + \beta T_{i,by} I(by \geqslant 2000) + \sum_{a=47}^{60} {}_a\delta_a d_{i,by+t} + \Phi X_i + \varepsilon_{i,by+t} \quad (2)$$

where δ_a represents the age[14] specific effect at the time of test ($a = 47, 48, 49, ..., 60$), X_i, is a vector of maternal, paternal, child and family characteristics measured at the time of birth (or no later than the first interview), and Φ is a vector of parameters. This estimator is limited in a number of ways. First, it assumes a linear relationship between y_{by+t} and X. If the response is non linear with respect to X, this approach provides biased estimates of β. Second, it does not accommodate compositional change. Controlling for X non linearly may be important. As mentioned above, although there is no evidence of mothers changing their labor market decision post reform, treated mothers pre and post reform have slightly different characteristics which may not be well captured by standard DID.

In a third step, to address the possibility of compositional change and non linearity of response with respect to X, the MDID estimator suggested by Heckman et al. (1997) is also implemented. With repeated-cross sections, the MDID estimator is (Blundell and Dias, 2009):

$$\hat{\beta}^{MDID} = \sum_{i \in T_1} \left\{ \left[y_{it_1} - \sum_{j \in T_0} \tilde{w}_{ijt_0} y_{jt_0} \right] - \left[\sum_{j \in C_1} \tilde{w}_{ijt_1} y_{jt_1} - \sum_{j \in C_0} \tilde{w}_{ijt_0} y_{jt_0} \right] \right\} w_i \quad (3)$$

where individual j can either be part of the treatment group prior to the reform T_0, the control group prior to the reform C_0 or the control group after the reform C_1. The outcome variables are measured at time t_0 (prior to the reform) for individuals in T_0 and C_0. The outcome variables are measured at time t_1 (after the reform) for individuals in T_1 and C_1. Each individual j when compared to individual i is attributed a specific weight $\tilde{w}_{ijt}$ that depends on the matching technique used, and w_i stands for sampling weights. The MDID estimator controls for X semi-parametrically by ensuring that children in each group (control prior to treatment, control after treatment and treated prior to treatment) all share the treated group after treatment distribution for each of the characteristics contained in X. This estimator also ensures group comparability prior to and after the reform and therefore limits the impact of compositional change on the outcome variables. Given the large number of control variables at our disposition, this approach is well-suited to assess the robustness of our DID results to compositional change. For each of the three approaches, we calculate bootstrap standard errors using the 1,000 bootstrapped weights provided by Statistics Canada to account for the underlying matching procedure and the sampling design of the

[14] Cognitive test scores are highly sensitive to age. We use age standardized score and account for age using age in month dummies for none standardized measures. Our results are not sensitive to using age standardized score versus raw scores with age in month dummies.

NLSCY. Bootstrapping increases standard errors by a factor of 1.5 on average, which implies that estimated effects are rarely significant at $p < 0.05$, while most are when bootstrap is not used.

Ideally, to confirm the robustness of our results and compare them with previous work in the literature, we would like to also use a regression discontinuity approach (RDD). Two main reasons limits us in doing so convincingly. First and foremost, our sample size is already fairly small, it is even more so if we focus on children born around the discontinuity point (less than 350 children between October 1st, 2000 and March 31st 2001). Second, as mentioned above, children born just before January 1^{st} 2001 were typically assessed at early age 4 ($\approx$ 46 months), while children born just after were assessed at late age 5 ($\approx$ 69 months). Although they were born a few days apart, they were observed two years apart at ages 4 and 5. As a result, one can no longer assume away trend effects as is typically the case in RDD framework.

In sum, we first implement the standard BA and DID estimators. To account for non linearity in X and compositional change we also implement the MDID estimator. Both the DID and MDID estimators assess the impact of the intention to treat of the reform.

6 Estimated intention-to-treat effects of the reform

Table 2 presents the empirical results of the standard BA and DID estimators (equation 2) and the MDID estimator (equation 3) using the main sample. The first two columns (Age only and All) present the estimated impact using the BA model. Results using standard DID are presented in columns 3 and 4 (Age only and All). The last two columns present the estimated impact using MDID with local linear regression matching (All llr) and nearest neighbor matching with five neighbors (All nn5). These results exhibit the strongest consistency in terms of balancing properties as defined by Rubin (2001). The balancing properties are presented and discussed in the Appendix. Results (not reported here) are also robust to kernel matching and to nearest neighbor with 4, 3, 2 and 1 neighbors. However, with more than 1 neighbor, estimates are more efficient.

From top to bottom, the first panel of Table 2 presents the effect of the policy on maternal time at home. The second panel shows the impact on cognitive measures and the third on behavioral measures. The fourth panel shows the impact on child health and family measures. All measures have been converted such that a positive coefficient indicates a positive impact on the child or its family. The treated groups (before and after the reform) are always restricted to children of working mothers, while the control groups include all other children.

Controls and matching variables are listed at the bottom of the table. Child age at test includes a set of age in month dummies. Child and family (at birth) include the following variables measured at the time of birth: breastfed at birth (yes,no), premature (yes,no),

birth weight (normal, low, very low), multiple births (yes, no), hospitalization at birth (yes, no), male, number of siblings (0, 1, 2 or more), maternal and paternal education (less than high school, high school degree, more than high school but no college degree, college degree, university degree), mother immigrated in the last 4 years (yes, no), age of the mother, age of the mother at the time of her first baby, urban area (rural to more than 500,000 inhabitants), and province of birth. Columns 1 and 3 present the BA and DID estimates controlling for age at the time of test dummies only, while columns 2, 4, 5 and 6 also include all exogenous covariates discussed above. We now discuss our empirical results.

Maternal time at home Figure 1 shows the discontinuity in time at home prior to returning to work. The top panel shows the time working mothers (treated) spent at home in the first year of life. The bottom panel shows the time non working mothers (control) spent at home. The vertical line marks the timing of the reform and also marks the break between cycle 4 and 5 in the NLSCY. This figure clearly shows that mothers in the control group did not change their behavior following the reform, while mothers in the treated group did. The top panel suggests that benefit recipient mothers were spending 7 months at home prior to the reform compared to 10 months post reform, while the bottom panel suggests that non working mothers generally spent 11 months with their child on average. We observe slightly more noise in the last 6 months of a cycle (2000m7 to 2000m12, and 2002m7 to 2002m12). This is due to the uneven sampling of birth months in the NLSCY, with extremely young children (those born near the end of a cycle) being under represented.

Table 2 shows the estimated effects of the reform on the time mothers stay at home (before returning to work). The BA estimates of 2.2 to 2.3 months are comparable to those of Baker and Milligan (2010). The DID estimator, which controls for underlying trends and measures the impact within the treated group, suggests a slightly larger effect of 3.2 months. In this case, underlying trends do not play an important role (the trend is flat, as seen in Figure 1), but when pooled together, the average time at home for non working mothers pulls the mean prior to the reform more than the mean after the reform, which explains the smaller effect measured using the BA estimator. In fact, because there are no trend effects, the DID estimate can be restored by dividing the BA estimate by the fraction of working mothers, i.e. 2.3/0.74=3.1. The MDID estimator suggests a slightly larger impact of 3.6 to 3.8 months using local linear regression and nearest neighbor matching. Controlling for X non linearly seems to be playing a small role. Within the group of working mothers, about 80% were eligible for paid maternity leave (Marshall, 2008). This implies that the impact of the reform on the treated is of the order of 4.4 months (i.e. 3.6/0.8).[15]

[15] To get the estimated impact on the treated, estimated effects need to be scaled by 1.25 (1/0.8). For conciseness and to further highlight the average effects of the reform on children of all working mothers, results to follow have not been scaled by 1.25. In other words, we present the intention-to-treat effects as opposed to the treatment effect on the treated.

This finding highlights one important feature of the reform: since working mothers prior to the reform took 7 months of leave on average, the estimated impacts relate to an increase from 7 to 11.4 months.[16] This result is different from Baker and Milligan (2010) who document an increase from 6 to 9 months. To find the mean time at home pre-reform (i.e. 6 months) the authors calculate the average time at home for mothers who returned to work within 12 months after the birth. This group excludes mothers taking more than 12 months of leave. Instead, in this paper, the estimated mean time at home prior to the reform is calculated for working mothers irrespective of the time they spent at home post birth.

Cognitive development The second panel of Table 2 presents the estimated impact of the policy on the PPVT, Who Am I? (WAI) and Number knowledge (NK) scores. To ensure comparability of the estimated effects, percentages of a standard deviation are also discussed and presented in Appendix, Table A.5.

The BA estimators suggest that the reform had a negative impact on child development, of the order of -2.0 for the PPVT, -1.5 for the WAI and -0.3 for the NK. These estimates are in line with those of Baker and Milligan (2015).[17] These authors attribute their findings to the fact that children post reform entered daycare when separation anxiety is at its peak (6 to 8 months) (Barglow et al., 1985; Schaffer and Emerson, 1964). We argue the opposite. First, our findings on maternal time at home suggest that children pre reform entered daycare in the peak anxiety period (at about 7 months), while post reform they enter daycare beyond the peak period (at about 10 to 11 months). Second, both the DID and MDID estimators are generally positive and significant for all three measures of cognitive development.

Indeed, standard DID without covariates shows positive effects for all three measures. Underlying trends, not accounted for by the BA estimator, play an important role. Once we control for covariates, the DID estimates remain positive but they are smaller and not significant (except for the WAI). This may be in large part attributed to the variation in maternal education within both groups before and after the reform. The MDID estimates (columns 5 and 6) are generally larger and more significant than that of the DID with all covariates (column 4). DID with covariates imposes a linear relationship. As such, unless interactions between different covariates are specifically included, the impact of a given covariate is assumed to be identical for all children. MDID allows for such interactions to be taken into account. MDID suggests that the reform had an impact of 3.0 to 3.7 (or 20% to 25% of a std. dev.) on the PPVT, 1.1 to 1.2 (or 17% to 18% of a std. dev.) on the WAI

[16] Note that we also get a pre-reform estimate of 7 months if we restrict our attention to mothers who actually claimed maternity leave benefits. This sample is much smaller and will be discussed in further detail below.

[17] The estimated impact on the NK are not comparable because the authors use the 4 point standardised score on the NK. At age 4 to 5, most children have either 2 or 3 out of 4. Instead, in this paper, the 30 point raw score is used as it provides a more refine picture on math skills. For cycles 4 and 5 these scores are not readily available in the NLSCY. They can be obtained on demand through Statistics Canada.

and 0.7 to 0.8 (or 14 % to 17% of a std. dev.) on the NK. The results on all three tests are generally significant and of comparable magnitude.

For reasons detailed above, while we used the Statistics Canada aged-standardized score for the PPVT, we used our own age-adjusted raw scores for the other measures. To make sure that we accounted for age differences properly, we re-estimated all of our specifications using the PPVT raw score, and obtained extremely similar estimates. This suggests that the difference in age distributions between the control group and the benefit recipients group have been well balanced (as also suggested by the balancing conditions) and that the tests' age sensitivity are not driving the results. To ensure that our MDID estimates were also not the results of maternal education not being properly matched, we also estimated the model using exact matching on maternal education. The results are again extremely similar.[18]

In sum, it appears that the reform had comparable positive effects of around 18% of a standard deviation across all three cognitive tests. These positive effects may be in part due to the fact that prior to the reform children were entering daycare within the critical anxiety period, and after they were not. The DID and MDID estimates are different from the BA estimates in both sign and significance. This shows the importance of controlling for underlying trends. The DID with covariates and MDID estimates are slightly different, which suggest that response is non linear in X and controlling for a complete set of individual and family characteristics is important.

Social development The third panel of Table 2 shows the estimated effects on behavioral measures. For all four measures of social development, the sign of the coefficients has been adjusted such that a larger score indicates better behavioral development.

The BA estimates suggest that the reform had a positive but generally not significant impact on behavioral development.[19] On all measures, except for conduct disorder, the effects are of a small magnitude (1% to 4% of a std. dev.). For conduct disorder, the effect is positive and significant, at 0.19 (i.e. 10% of a std. dev.). This suggests that spending more time with the mother in the first year of life helped children to better control themselves at age 4 and 5, by for example getting into a fights less often.

The DID and MDID estimates are, however, not significant across all specifications and for all four measures. Furthermore, while the DID estimates are positive, the MDID estimates are generally negative. The only effect that appears to be constant across all models is the positive impact on conduct disorder. The estimated impact is also of the order of 0.19 (or 10% of a std. dev.), but is not statistically different from zero.

In sum, it appears that the reform did not significantly impact the social development of the child, except possibly in helping to improve his overall conduct at age 4 and 5. Parent-

[18] These estimates can be obtained from the author on request.

[19] This is in line with Baker and Milligan (2011) who find non significant results.

reported measures are subject to a number of biases and therefore may not be extremely well suited to capture the development of the child.

Family and health The fourth panel of Table 2 shows the estimated effects on the environment of the child and on parent-reported child health measures.[20] The BA results controlling for covariates suggest an improvement in child's health (7% and 8% of a std. dev.), but a worsening of the social support of the family (10% of a std. dev.).

The positive effects on health persist once we control for underlying trends through DID, but the negative effect on social support disappears. The DID estimates with and without covariates are similar. The effect on general health ranges between 0.09 and 0.12 (17% to 22% of a std. dev) and between 0.22 and 0.28 (29% to 37% of a std. dev) on health recently. These effects are large, but they are not robust to controlling for X non linearly.

Using MDID, we find that children's health reported by the parent improved according to both measures of health, but the effects are only weakly significant. The magnitude is also smaller, ranging between 13% and 19% of a std. dev. for general health and 9% and 14% of a std. dev. for health recently. Health related benefits may be due to increased length of breastfeeding as reported by Baker and Milligan (2008a)[21] and improved birth outcomes (Rossin, 2011).

On family functioning and social support, the DID estimates are also larger and more significant than the MDID estimates, but both estimators suggest a positive impact on family functioning and social support. The estimated impact is of the order of approximately 20% of std. dev. (using MDID) for both measures. More specifically, we find that the impact of the reform on family functioning is positive between 0.8 to 1.3 (or 17% to 25% of a std. dev.). The estimate is significant only when using nearest neighbor matching. On social support, we find significant impacts of the order of 0.6 to 0.8 (or 17% to 25% of a std. dev.). In practical terms, these effects imply that the mother would have answered differently on one of the subquestions[22] of the family functioning measure and on no more than one of the subquestions of the social support measure (from strongly agree to agree, for example). Thus it is a fairly modest effect. A few channels may explain these findings. First, if the child is better off, so are the parents. Second, more time away from work in the first year may allow parents to better organize life as a family and foster a better network, possibly by getting involved in parent-child activities while on maternity leave.

[20] Baker and Milligan (2015) do not estimate the impact of the reform on these measures.

[21] Baker and Milligan (2008a) use a before-after model to assess the impact of the reform on breastfeeding. These outcome variables are based on retrospective questions not dependent on the age of the child at the time of the interview (e.g. have you ever breastfed this child?).

[22] The subquestions are provided in the Web Appendix.

Robustness checks In Table 3 we test the sensitivity of our results to (1) additional matching variables, (2) the reform announcement date, (3) alternative matched samples, and (4) future trends. We use MDID with local linear regression matching, but similar results are obtained using nearest neighbor matching (with 5 neighbors). For all outcomes measures, the DID and MDID estimates lead to comparable results (at least in terms of the sign of the coefficient). For convenience, we repeat the base specification in column 1 (also reported in column 5 of Table 2).

First, in column 2, we add the provincial unemployment rate in the year preceding birth. While the provincial unemployment rate prior to birth may influence the labor market participation of mothers prior to birth, it does not influence age 4 and 5 outcomes such as cognitive test scores. The estimated impacts in columns 1 and 2 are extremely similar, except that the impact on the NK score is now weakly significant. This further suggests that compositional changes between our treatment and control group (from not working to working) are unlikely to drive our results. Second, in column 3 we add a set of possibly endogenous variables measured shortly after birth, namely marital status (married or not), ineffective parenting scale, positive parenting scale, family functioning scale, social support scale, paternal income quartile (measured at the provincial level), and age in months at which the child slept a full night. Estimated impacts on all outcome variables are extremely similar, except for the PPVT and family functioning. In both cases, the impact is larger and more significant when we match with the full set of variables.

In column 4 we test the sensitivity of our results to the reform announcement date. The official announcement date was February 28th, 2000. Babies conceived in March and April 2000 were due for November and December 2000, before the implementation of the reform. Under the assumption that there were no delays in conception, we should obtain similar results whether we include November and December 2000, and January and February 2001 births or not. We find that our results are not sensitive to the exclusion of babies born around the implementation date of the reform.

In the last two columns of the table, we test the sensitivity of our results to the matching procedure. Matching procedures rely on specific decision rules which often imply that certain observations will be matched more often than others because they have characteristics that are more similar to the observations they are being matched with. Figure 2 illustrates this idea. This figure shows the propensity score (pscore) distribution prior to (left panel) and after matching (right panel) for each of the four groups. Looking at the treated groups (before and after) we see that they have fairly similar pscore distributions. Since the matching techniques used in this paper are based on pscore, most observations are used in fairly equal proportions in this case (the distributions are already almost identical). However, when the observations of the control group are matched to those of the treated group after the reform, observations with fairly high pscores are given higher weights than other observations to

ensure that the pscore distribution of the control groups converges to that of the treated groups. We test the sensitivity of our results to the exclusion of these highly matched observations in two ways. First, we exclude the 5% highest pscore in the reference group (treated after the reform), such that the matched samples also exclude observations with comparable pscores (column 5, Table 4). Second, we exclude observations that are matched more than 2% of the time (column 6, Table 4). Results are generally robust to these restrictions.

Finally, Blundell and Dias (2009) recommend comparing trends of the treated and controls using historical data prior to the reform in a period with comparable macro trends. Such information is not available in the NLSCY. However, we have future information, up to children born in 2004. To ensure that we have comparable trends in both groups, we estimate our base model on a sample of children born exclusively after the reform. We use children born between 2001 and 2004 (the latest year of observation in the NLSCY) and set the reform date to December 31^{st}, 2002. The indicator function marking the reform date becomes $I(by \geqslant 2002)$. This function equals one if the child was born after 2002 and zero otherwise. Since all children in this sample were born post reform, the estimated impact of the reform β should not be significantly different from zero for all outcome variables. This will be true if both groups, treated and control, share comparable macro trends. Children born in 2001 to 2004, are observed in 2005 to 2008. This corresponds to an economic downturn in Canada. While this may impact family related measures, it likely did not impact the cognitive and behavioral development of children, at least not significantly. Again our treatment group includes all children of mothers who worked before birth, and our control group includes all other children.

We find that except for the PPVT, family functioning, and social support, all estimated coefficients are closer to zero and are not statistically different from zero (column 7 of Table **??**). For the PPVT, the estimated impact is negative and significant. This suggests that we may not be properly controlling for the underlying trend. Since the estimated coefficient is negative, this implies that the trend used in our estimates may be more negative than it is. This further implies that our estimated impact, using our main sample, may be even more positive. For the family functioning and social support measures, we have the opposite effects: i.e. the estimated impacts are positive, which may imply that the underlying trend we account for in our main model is not sufficiently positive, and therefore our estimated impacts may be less positive then they actually are. For these measures, we can however think of an alternative explanation. As mentioned above, between 2005 and 2008, the economy slowed down in Canada. Single income households (with non working mothers) may have been impacted more negatively than dual income households (with working mothers), which would explain the positive (or less negative) effect measured for families with working mothers.

In sum, it appears that our empirical strategy can capture the effect of the parental leave reform on time at home, and on cognitive, behavioral and health measures. For the family

functioning and social support measures, our falsification exercise is incorrectly specified due to important variations in macro trends over the period.

7 Discussion

Why would maternal time investment lead to better outcomes? There are multiple channels by which children may be impacted including, but not limited to[23], (1) the formation of a secure attachment, (2) type of care displaced, and (3) changes (or not) in disposable income. Before going into further details, it is important to keep in mind the age at which treatment occurs when interpreting the results because of non linearity in brain development. Above results have already shown that the effective treatment period was from 7 to 10 months (at a minimum) for the Canadian reform. This treatment period is closest to the Norwegian reform studied by Carneiro et al. (2015) which finds an effective treatment period of 8 to 12 months. Unfortunately, most other papers either fail to document the effective treatment period or had a strikingly different treatment period.

Theories in psychology and recent empirical evidence in neurosciences suggest that increasing the time spent with the mother allows the formation of a more secure attachment (Bowlby, 1958; Bell and Ainsworth, 1972; Ainsworth et al., 1978; Schore, 1994, 2001). This is especially true in the 7 to 11 months window, since specific attachment forms when the child is around 8 months (Schaffer and Emerson, 1964). According to Bowlby, the failure to develop attachment (with the mother or its replacement) may be linked with delinquency, depression, increased aggression and reduced cognitive skills.

The importance of the type of care has been shown by Brooks-Gunn et al. (2010). Using a multivariate approach, they showed that full-time maternal employment in the first year was associated with lower cognitive test scores (although not all test scores)[24], but these effects were offset by greater maternal sensitivity and the use of centre-based care. The NLSCY does not provide measures of maternal sensitivity, but it does provide measures on the type and quality of care displaced. Table 4 shows the type of care used prior to and after the reform for children aged less than 12 months.[25] We find that prior to the reform formal daycare was extremely rare (only 6%). About 22% of children were taken care of by a non relative in a family environment, either at the care provider's home (16%) or at the child's home (6%). A slightly smaller proportion (19%) were being taken care of by a relative, while 53% of children were in the care of their parents. Prior to the reform, mothers were entitled

[23] Other channels discussed previously include increased breastfeeding duration, and aditionnal time to organize life as a family and foster a better network.

[24] Previous research by Brooks-Gunn *et al.* (2002b) using the same data set more specifically showed that maternal employment by the ninth month was associated with lower cognitive scores at age 3.

[25] Baker and Milligan (2010) also document the impact of the reform on the type of care displaced, but they use the sample of children aged 7 to 12 months inclusively, and do not focus on treated children. Their findings are directionnally similar.

to 6 months of paid maternity leave, which explains the large fraction of children aged less than 12 months being in care of their parents. Post reform, we see that most children (73%) are in the care of their parents. Children in other types of care are mainly taken care of by a relative (13%) or a non relative in a family environment (8%), while a very small fraction is in formal daycare (1%). Comparison of the pre and post reform childcare usage allows us to conclude that the main type of care displaced was that of a non relative in a family environment. This is important because in Canada, this type of care typically has a 6 to 1 child to educator ratio, while the ratio is generally one to one when the child is in the care of the mother.

This table also reveals that prior to the reform only a small fraction of children in care were being taken care of by individuals trained to care for young children, while post reform this fraction is slightly higher (27% compared to 20%). We can also observe that prior to the reform only 17% of children in care were being taken care of by individuals who had a license. This statistic is however not available post reform. These are the only measures available to proxy for the quality of care. Children in care spend a fairly high number of hours in care prior to the reform, on average 27 hours with a standard deviation of 17 hours, and this is also true post reform (24 hours, with std. dev. of 16). In sum, the type of care displaced is mainly informal and performed by individuals who have not been trained to educate young children. Estimated effects therefore compare time spent with the mother versus time spent in informal care at a fairly high intensity (27 hours per week). To our knowledge, papers on reforms other than the Canadian reform, do not document the type of care displaced, but generally describe the main type of care prior to the reform. In Denmark, publicly subsidized universal day care and other childcare facilities are widely available, while in Germany and Norway, it appears that the main form of care is generally informal.

On disposable income, the choice to work (instead of staying at home) implies that mothers have to pay for daycare and work related expenses. As discussed above, Baker and Milligan (2010) showed that the 55% replacement rate results in no income effect for a median income mother once taxes, deductions and childcare costs are accounted for. Since work related expenditures, including childcare costs, may vary by country, it is not possible to determine the impact on disposable income elsewhere, but rates much larger than 55% are likely to increase the disposable income of mothers taking paid leave, while rates much lower than 55% are likely to reduce their disposable income. In studies cited above, rates vary from 0% to 100%, the likely impact of many of these reforms is unlikely to be a pure maternal time investment effect. For example, Dustmann and Schönberg (2012), accounting for daycare costs, find a negative income effect (for the 1992 reform) which according to the authors could explain the negative impact of the extension on type of school attended and test scores.

While the Canadian and Norwegian reforms appear to be most comparable, in terms of

effective treatment period and type of care displaced, their respective replacement rate was fairly different, at 55% in Canada and 100% in Norway. Carneiro et al. (2015) argue, but do not provide evidence, that there was no income effect in Norway, while Baker and Milligan (2010) show evidence that for median income mothers there was no income effect in the Canadian context, while for mothers at the 25th percentile of the income distribution there was a positive income effect. As such, in the Canadian context, a 100% replacement rate would result in a positive impact on disposable income for low and median income mothers. Without formal evidence of the impact in Norway, it is hard to determine whether both reforms were indeed comparable or not with respect to the impact on disposable income. Carneiro et al. (2015) find larger positive effects for children of less educated mothers, which could be due to the dual effect of maternal time investment and increased disposable income. Recent evidence suggests that increased disposable income raises the achievement of students, and that these gains are larger for children from disadvantaged families (Dahl and Lochner, 2012). More generally, Carneiro et al. (2015) find a positive impact on the probability of completing high school, which suggests that effects identified in this study are not only positive but possibly long lasting. Their study of the potential mechanisms find limited evidence that increased cognitive ability (measured using IQ at age 18-19) explains the positive impact on high school completion. The authors suggest that non cognitive skills may be a better explanation, but their data does not allow them to provide evidence on non-cognitive skills. This paper generally finds positive although not significant effects on conduct behavior, but also positive and significant effects on cognitive development, health and family functioning when the child is aged 4 to 5. There is considerable evidence between the link between early life outcomes and later life outcomes. Given the important similarities between the Canadian and Norwegian reforms, it appears that the impact on cognitive skills identified in this paper fades over time, but the early life positive impacts on health and family functioning, along with possibly conduct behavior, may be associated with a stronger predisposition to complete high school. Combined evidence from both studies are in line with Heckman's (2008) findings on the persistence of ability gaps identified at age 5. This study focuses on dual-parent families, and therefore may be missing important impacts on single-parent families. The Canadian parental leave is however not as generous and therefore may not produce similar positive effects at the lower end of the education distribution.

8 Conclusion

This paper investigates the effect of maternity leave expansion in Canada that formally increased paid maternity leave from 6 to 12 months on December 31^{st}, 2000. While the literature clearly shows that ability identified as early as age 5 is a strong predictor of future adulthood socioeconomic success, effects of paid maternity leave reforms controlling

for underlying trends have never been documented before for this age group.

Results suggest that the reform had a significant effect on the time a mother stays at home in the first year of about 3 to 4 months more, up from 7 months on average prior to the reform. The type of care displaced was mainly informal and provided by individuals not specifically trained to care for young children. Comparisons of BA and DID estimates show the importance of controlling for underlying trends, while comparisons of DID and MDID estimates show that controlling for compositional change and non linearity of response with respect to X produces directionnally similar results. The DID and MDID on a variety of outcome measures suggest that maternal time at home, as opposed to informal care, is preferable. Significant and positive effects of the order of 20% of a standard deviation are found for two of the three cognitive measures available for children aged 4 and 5. We also find that children whose mother benefited from extended maternal leave benefits possibly have better conduct (e.g. gets into fewer fights, is less mean to others), but are otherwise not different in terms of parent-reported behavioral measures. Finally, child health, social support and family functioning also improve following the reform.

This study is not without its limitations. First, the type of care displaced was mainly informal. The effect documented in this study may not be generalizable to children and families in other countries where the type of care is more formal and possibly of higher quality. Second, estimated effects on child development may depend on the type of children treated. This study focused on children raised in two-parent households, and showed that mothers who benefited from the reform were generally more educated, had fewer children and were less likely to have immigrated recently. As mentioned above, Liu and Skans (2010) found that the effects of a similar reform on test scores measured at age 16 were neutral on average, but positive for children of highly educated mothers. Further research using a different data set should focus on documenting the impact of the Canadian reform on children of single parent families and more generally on children from different family backgrounds, which is not possible using the current data set. Heckman (2000) finds that the rising skill gap in the United States can be in part attributed to a rise in the proportion of children born in less favorable family environments. Understanding how to better address the needs of these children should be a priority.

A number of policy implications can be drawn from this study. First, documentation of the effective treatment period shows the importance of considering the current behavior of mothers when thinking about reforming parental leave policies. Different policies may trigger similar treatment periods as shown by the comparison of the Canadian reform and the Norwegian reform. Second, this study showed that with a fairly low income replacement rate, mothers strongly reacted to the reform by increasing their time away from the labor market in their child's first year of life. Policy makers therefore do have to choose the replacement rate wisely in order to contain the cost of such a program. Third, since the

type of care displaced was mainly informal, one might argue that providing more formal and possibly better quality care might be less expensive and provide stronger positive effects for children. The cost of the parental leave program for a median income mother was $264.42 per week for the extension period at the time of the reform. This is more or less equivalent to the cost of formal daycare for children aged less than 18 months after the implementation of the Québec daycare reform. Studies on the Québec reform do not find that child development improved following the reform (Haeck et al., 2015; Lefebvre and Merrigan, 2008). Providing care of higher quality may very well be more expensive than paying mothers to stay at home with their child in the first year, in part because the educator to child ratio is smaller for this age group. As a result, policy makers should carefully assess the costs and the likely benefits of paid parental leave versus daycare before choosing which option is more favorable to the development of the child and the well-being of the families within the first year of life.

In summary, this paper shows that the expansion of the Canadian parental leave policy had positive effects on children and their families, in addition to the benefits previously documented on improved job continuity for mothers and increased duration of breastfeeding. Policy makers interested in improving the well-being of children and families should consider a similar reform, after careful consideration of the current behavior of mothers during the first year of life, the type and quality of care available for children of working mothers and the overall cost of the program.

9 References

Ainsworth, M. D. S. (1972) 'Attachment and dependency: A comparison', In J. L.Gewirtz (Ed.) *Attachment and dependency.* Washington, D. C.: V. H. Winston & Sons. pp. 97-137.

Ainsworth, M.D., Blehar, M, Waters, E, & Wall, S. (1978) 'Patterns of Attachment: A Psychological Study of the Strange Situation', Hillsdale, NJ: Lawrence Erlbaum, p. 167

Almond, D. & Currie, J. (2011) 'Human Capital Development before Age Five', *Handbook of Labor Economics*, David Card et Orley Ashenfelter (eds.), Elsevier, 1315-1486.

Baker, M. & Milligan, K. (2008a) 'Maternal employment, breastfeeding, and health: Evidence from maternity mandates', *Journal of Health Economics*, vol. 27, pp. 871–87.

Baker, M. & Milligan, K. (2008b) 'How does job-protected maternity leave affect mothers' employment?', *Journal of Labor Economics*, vol. 26, pp. 655–91.

Baker, M. & Milligan, K. (2010) 'Evidence from maternity leave expansions of the impact of maternal care on early child development', *Journal of Human Resources*, vol. 45(1), pp. 1–32.

Baker, M. & Milligan, K. (2015) 'Maternity leave and children's cognitive and behavioral development', *Journal of Population Economics* vol. 28(2), pp. 373-391.

Baum, C.L. (2003) 'Does Early Maternal Employment Harm Child Development? An Analysis of the Potential Benefits of Leave Taking', Journal of Labor Economics, vol. 21(2), pp. 409-448.

Barglow, P., Conteras, J., Kavesh, L., & Vaughn, B. (1998) 'Developmental follow-up of 6-7 year old children of mothers employed during their infancy', *Child Psychiatry and Human Development*, vol. 29(1), pp. 3-20.

Barglow, P., Vaughn, B, & Molitor. (1987) 'Effects of Maternal Absence Due to Employment on the Quality of Infant-Mother Attachment in a Low-Risk Sample', *Child Development*, vol. 58(4), pp. 945-954.

Bell, S. & Ainsworth, M. (1972) 'Infant crying and maternal responsiveness', *Child Development*, vol. 43, pp. 1171.1190.

Belsky, J. & Eggebeen, D. (1991) 'Early and extensive maternal employment and young children's socioemotional development: Children of the national longitudinal survey of youth', *Journal of Marriage and Family,* vol. 53(4), pp. 1083–98.

Belsky, J., Lowe Vandell, D., Burchinal, M., Clarke-Stewart, K.A., McCartney, K., & Tresch Owen, M. (2007) 'Are There Long-Term Effects of Early Child Care', *Child Development,* vol. 78(2), pp. 681-701.

Berger, L., J. Brooks-Gunn, C. Paxson & J. Waldfogel, (2008) First-year maternal employment and child outcomes: Differences across racial and ethnic groups, Children and Youth Services Review, vol. 30, pp. 365-387

Bernal, R. (2008) The Effect of Maternal Employment and Child Care on Children's Cognitive Development, International Economic Review, vol. 49(4), pp. 1173-1209

Blau, F. D. & Grossberg, A. J. (1992). 'Maternal labor supply and children's cognitive development', *The Review of Economics and Statistics*, vol. 74(3), pp. 474–81.

Blundell, R. & Dias, M. C. (2009). 'Alternative approaches to evaluation in empirical microeconomics', *Journal of Human Resources*, vol. 44(3), pp. 565–640.

Bowlby, (1958) The nature of the child's tie to his mother, International Journal of Psychology Analysis, XXXIX, 1.23.

Brooks-Gunn, J., Han, W.-J., & Waldfogel, J. (2002a). 'Maternal employment and child

cognitive outcomes in the first three years of life: The NICHD study of early child care', *Child Development*, vol. 73(4), pp. 1052–72.

Brooks-Gunn, J., Han, W.-J., & Waldfogel, J. (2002b). 'The effects of early maternal employment on child cognitive development', *Demography*, vol. 39, pp. 369–92.

Brooks-Gunn, J., Han, W.-J., & Waldfogel, J. (2010). 'First-year maternal employment and child development in the first 7 years', *Monographs of the Society for Research in Child Development*, vol. 75, pp. 1–147.

Buckles, K. & Hungerman, D. M. (2008). 'Season of birth and later outcomes: Old questions, new answers', NBER Working Papers 14573, *National Bureau of Economic Research* Inc.

Carneiro, P., Loken, K. V., & Salvanes, K. G. (2015). 'A Flying start? long term consequences of maternal time investments in children during their first year of life', forthcoming at *Journal of Political Economy*.

Cools, S., Fiva, J.H., & Kirkebøen, L. J. (2011). 'Causal Effects of Paternity Leave on Children and Parents', CESifo Working Paper Series 3513, CESifo Group Munich.

Dahl, G. B. & Lochner, L. (2012). 'The impact of family income on child achievement: Evidence from the earned income tax credit', *American Economic Review*, vol. 102(5), pp. 1927–56.

Dahl, G, B., Løken, K. V., Mogstad, M. & Salvanes, K. V. (2013). 'What Is the Case for Paid Maternity Leave?', NBER Working Papers 19595, *National Bureau of Economic Research* Inc.

Danzer N. & Lavy, V. (2013).'Parental Leave and Children's Schooling Outcomes: Quasi-Experimental Evidence from a Large Parental Leave Reform', NBER Working Papers 19452, National Bureau of Economic Research, Inc.

Desai, S., Chase-Lansdale, P. L., & Michael, R. (1989). 'Mother or market? effects of maternal employment on cognitive development of four-year-old children', *Demography*, vol. 26(88-11), pp. 545–61.

Duncan, G. J., Magnuson, K. A., & Ludwig, J. (2004). 'The endogeneity problem in developmental studies', *Research in human development*, pp. 59–80.

Dustmann, C. & Schönberg, U. (2012). 'The effect of expansions in maternity leave coverage on children's long-term outcomes', *American Economic Journal: Applied Economics*, 4,

190-224

Engster, D. & Stensöta, H.O. (2011) 'Do Family Policy Regimes Matter for Children's Well-Being?', *Social Politics*, vol. 18(1), pp. 82-124.

Greenstein, T. (1995). 'Are the "Most Advantaged" Children Truly Disadvantaged by Early Maternal Employment?: Effects on Child Cognitive Outcomes', *Journal of Family Issues*, vol. 16(2), pp. 149-169.

Gregg, P., Washbrook, E., Propper, C. & Burgess, S. (2005) The effects of a mother's return to work decision on child development in the UK, *The Economic Journal*, vol. 115(501): pp. F48-F80.

Haeck, C., Lefebvre, P. & Merrigan, P. (2015), 'Canadian Evidence on Ten Years of Universal Preschool Policies: The Good and the Bad', *Labour Economics*, forthcoming

Harvey, E., (1999) Short-Term and Long Term Effects of Early Parental Employment on Children of the National Longitudinal Survey of Youth, Developmental Psychology, vol. 35(2), pp. 445-59.

Han, W., Waldfogel, J. & Brooks-Gun, J. (2001) The Effects of Early Maternal Employment on Later Cognitive and Behavioral Outcomes, Journal of Marriage and Family, vol. 63(2): pp. 336-354

Heckman, J. J. (2000). 'Policies to foster human capital', *Research in Economics*, vol. 54(1), pp. 3–56.

Heckman, J. J. (2008). 'Schools, skills, and synapses', *Economic Inquiry*, vol. 46(3), pp. 289–324.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). 'Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme', *The Review of Economic Studies*, vol. 64(4), pp. 605–54.

Heckman, J. & Navarro-Lozano, S. (2004). 'Using matching, instrumental variables, and control functions to estimate economic choice models', *Review of Economics and Statistics*, vol. 86, pp. 30–57.

Hill, J., Waldfogel, J., Brooks-Gunn, J. & Han, W. (2005) Maternal Employment and Child Development: A Fresh Look Using Newer Methods, *Developmental Psychology*, vol. 41(6): pp. 833-850.

HRSDC (2005). 'Summative evaluation of EI parental benefits', Report No. SP-AH-674-01-05E, Human Resources and Skills Development Canada, Government of Canada.

James-Burdumy, S., (2005) The Effect of Maternal Labor Force Participation on Child Development, Journal of Labor Economics, University of Chicago Press, vol. 23(1), pp. 177-211.

Lefebvre, P. & Merrigan, P. (2008). 'Child-care policy and the labor supply of mothers with young children: A natural experiment from canada', *Journal of Labor Economics*, vol. 26(3), pp. 519–48.

Liu, Q. & Skans, O. N. (2010). 'The duration of paid parental leave and children's scholastic performance', *The B.E. Journal of Economic Analysis & Policy*, vol. 10, pp. 1–33.

Lucas-Thompson, R.G., W.A. Goldberg & J. Prause, (2010); Maternal Work Early in the Lives of Children and Its Distal Associations With Achievement and Behavior Problems: A Meta-Analysis, Psychological Bulletin, vol. 136(6): pp. 915-942

Marshall, K. (2008). 'Father's use of paid parental leave', *Perspectives on Labour and Income*, Statistics Canada Catalogue no. 75-001-X.

National Institute of Child Health and Human Development Early Child Care Research Network (2003) 'Does Amount of Time Spent in Child Care Predict Socioemotional Adjustment during the Transition to Kindergarten?', *Child Development*, vol. 74(4), pp. 976-1005.

Phipps, S. A. (2000). 'Maternity and parental benefits in canada: Are there behavioural implications?', Canadian Public Policy, vol. 26(4), pp. 415–36.

Rasmussen, A. W. (2010). 'Increasing the length of birth related leave: The effect on children's long-term educational outcomes', *Labour Economics*, vol. 17, pp. 91–100.

Rosenbaum, P. R. & Rubin, D. B. (1983) 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, vol. 70(1), pp. 41–55.

Rossin, M.(2011) 'The effects of maternity leave on children's birth and infant health outcomes in the United States', Journal of Health Economics, vol. 30(2), pp. 221-239.

Rubin, D. B. (2001) 'Using propensity scores to help design observational studies: Application to the tobacco litigation', *Health Services & Outcomes Research Methodology*, vol. 2, pp. 169–88.

Rhum, C. J. (2004) 'Parental employment and child cognitive development', *Journal of Human Resources*, vol. 39(1), pp. 155–92.

Schaffer, H.R.& Emerson, P.E. (1964) 'The Development of Social Attachments in Infancy', *Monographs of the Society for Research in Child Development*, vol. 29(3), pp. 1-77.

Schore, A.N. (1994) Affect Regulation and the Origin of the Self : The Neurobiology of Emotional Development. L. Erlbaum Associates

Schore, A.N. (2001) 'Effects of a secure attachment on the right brain development, affect regulation, and infant mental health', *Infant Mental Health Journal*, 22, p. 7-67

Sherlock, R. L., A.R. Synnes & M. Koehoorn (2008) Working mothers and early childhood outcomes: Lessons from the Canadian National Longitudinal study on children and youth, Early Human Development, vol. 84, pp. 237-242

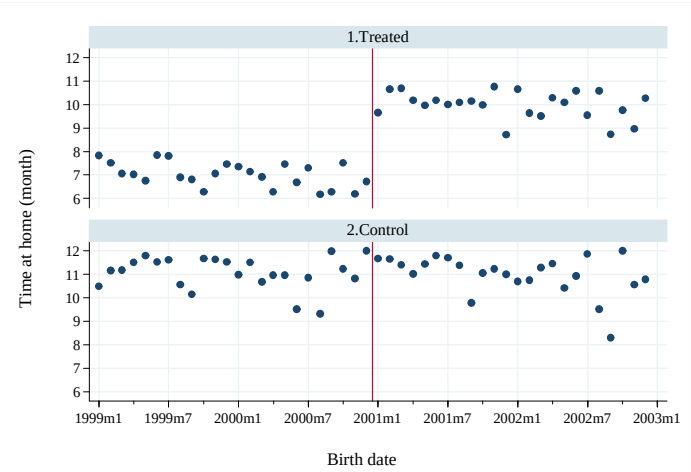
Stuart, E. A. (2010). 'Matching methods for causal inference: A review and a look forward', *Statistical Science*, vol. 25, pp. 1–21.

Verropoulou, G. & Joshi, H. (2009) Does mother's employment conflict with child development? Multilevel analysis of British mothers born in 1958, Journal of Population Economics, vol. 22, pp. 665-692.

Vandell, D. L. & Ramanan, J. (1992). 'Effects of early and recent maternal employment on children from low-income families', *Child Development*, vol. 63(4), pp. 938–49.

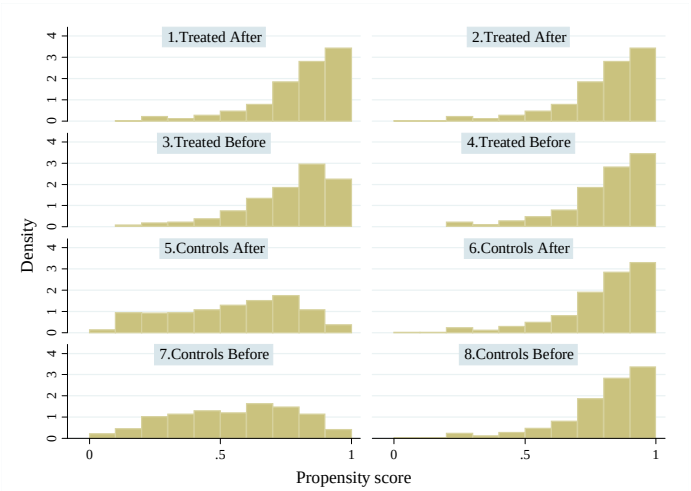
10 Figures

Figure 1: Maternal time at home in the first 12 months of life



Note: This figure shows the average number of months spent at home by the mother in the first year of life given the birth date (year-month) of the child. The vertical line indicates the timing of the reform. The upper graph "Treated" includes working mothers. The bottom graph "Controls" includes only mothers who did not work.

Figure 2: Pscore by work group



Note: Shows the propensity score distribution by groups prior to matching (left panel) and once matched (right panel). The treated group includes children whose mother worked prior to birth, while the control group includes children whose mother did not. Matching is performed using local linear regression. The specification used corresponds to the specification "All" further defined in Table 2. All observations with a PPVT score are included. Similar figures are found for all of the other outcome measures studied.

11 Tables

Table 1: Summary statistics

	Treated before		Treated after		Control before		Control after	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Child and family characteristics								
Child								
Age at 1st interview (months)	15.38	6.97	14.63	6.05	15.95	7.08	15.29	6.20
Age at 3rd interview (months)	58.09	6.66	57.87	6.53	58.28	6.80	58.60	6.53
Male	0.52	0.50	0.49	0.50	0.46	0.50	0.50	0.50
Premature	0.09	0.29	0.11	0.32	0.11	0.31	0.08	0.27
Birth weight - low	0.05	0.21	0.04	0.21	0.07	0.25	0.06	0.25
Birth weight - very low	0.01	0.10	0.00	0.07	na	na	0.01	0.11
Multiple births	0.03	0.17	0.03	0.17	0.03	0.17	0.03	0.18
Hospitalization at birth	0.16	0.36	0.18	0.38	0.14	0.35	0.14	0.34
Breastfed at birth	0.86	0.35	0.89	0.31	0.89	0.31	0.85	0.36
Sleep (age in months)	5.01	4.66	5.16	4.53	5.52	5.05	5.95	5.06
Family								
Parents not married	0.11	0.31	0.10	0.30	0.08	0.28	0.10	0.30
Number of siblings	0.80	0.93	0.66	0.77	1.08	1.12	1.32	1.22
Positive parenting scale (+)	17.90	2.05	18.25	1.89	17.66	2.53	17.99	2.07
Ineffective parenting scale (−)	2.21	1.71	2.35	1.74	2.46	2.01	2.26	1.87
Family functioning scale (−)	8.70	4.65	8.57	4.94	9.13	4.84	8.94	4.51
Social support scale (+)	18.56	3.58	18.73	3.45	17.64	3.58	18.38	3.21
Mother								
Age at birth	29.93	4.83	30.42	4.71	29.86	5.88	29.61	5.49
Age at first baby	27.04	4.88	28.59	5.07	26.33	5.07	26.29	5.66
Years of education	13.47	2.34	14.83	1.80	12.84	2.76	13.42	2.44
Immigration - 0 to 4 years	0.05	0.22	0.04	0.20	0.17	0.37	0.20	0.40
Father								
Years of education	13.23	2.35	14.66	2.04	13.32	2.83	13.93	2.59
Income	47.1	44.4	50.6	39.6	49.3	54.0	48.4	35.1
Outcome measures (age 4 and 5)								
Maternal time at home								
in the first year of life	6.89	3.27	9.99	3.26	11.03	2.52	10.94	2.81
Cognitive measures								
PPVT (standardized) (+)	103.84	14.59	103.32	15.06	98.72	15.46	95.59	16.33
Who Am I? (+)	25.28	6.53	24.55	6.73	26.50	6.37	24.05	6.34
Number knowledge (+)	12.26	4.88	12.34	4.99	11.78	4.99	10.95	4.45
Behavioral measures								
Hyperactivity (−)	4.00	2.73	3.89	2.50	3.73	2.66	3.86	2.58
Emotional disorder (−)	1.92	1.89	1.93	1.86	1.89	1.88	1.90	2.04
Conduct disorder (−)	1.64	1.90	1.41	1.63	1.71	2.00	1.78	1.94
Indirect aggression (−)	0.54	1.12	0.47	0.99	0.56	1.09	0.69	1.58
Family and health								
General health (−)	1.20	0.54	1.13	0.41	1.12	0.41	1.16	0.41
Health recently (−)	1.50	0.76	1.39	0.62	1.42	0.66	1.57	0.73
Family functioning scale (−)	7.98	5.05	7.35	5.04	7.52	4.79	8.93	5.24
Social support scale (+)	19.58	3.54	19.71	3.51	19.46	3.39	18.35	3.51
N	1,498		1,352		489		455	

Note: Shows the summary statistics of our main sample. Unless noted otherwise, the child and family characteristics are measured at the time of the first interview, shortly after birth. For the different scales, beside the label, we indicate whether a larger score implies a better outcome using "(+)" or a lower score implies a better outcome using "(−)". For simplicity, in the empirical section, the sign of these measures has been reversed such that a larger score always indicates a better outcome.

Table 2: Estimated effects

		Before-After		DID		MDID		Number of
		Age only	All	Age only	All	All llr	All nn5	observations
Time at home								
1st year of life	coef.	2.19***	2.33***	3.24***	3.22***	3.88***	3.81***	3,704
	s.e.	(0.15)	(0.16)	(0.27)	(0.26)	(0.27)	(0.26)	
Cognitive measures								
PPVT	coef.	-0.95	-1.79**	2.80	0.57	2.99*	3.70**	3,370
	s.e.	(0.77)	(0.72)	(1.72)	(1.55)	(1.54)	(1.68)	
Who Am I?	coef.	-1.23***	-1.53***	1.99***	1.25**	1.09*	1.16	3,256
	s.e.	(0.24)	(0.26)	(0.57)	(0.52)	(0.61)	(0.57)	
Number knowledge	coef.	-0.10	-0.32	1.10**	0.52	0.70	0.81*	3,370
	s.e.	(0.20)	(0.21)	(0.45)	(0.44)	(0.44)	(0.48)	
Behavioral measures								
Hyperactivity	coef.	0.04	0.04	0.27	0.14	-0.09	-0.11	3,748
	s.e.	(0.12)	(0.13)	(0.26)	(0.26)	(0.26)	(0.28)	
Emotional disorder	coef.	-0.03	0.08	0.00	0.01	-0.15	-0.38	3,762
	s.e.	(0.09)	(0.09)	(0.20)	(0.20)	(0.19)	(0.20)	
Conduct disorder	coef.	0.15*	0.19**	0.32	0.16	0.19	0.19	3,763
	s.e.	(0.08)	(0.09)	(0.20)	(0.19)	(0.17)	(0.20)	
Indirect aggression	coef.	0.02	0.00	0.18	0.12	-0.06	-0.05	3,681
	s.e.	(0.06)	(0.06)	(0.17)	(0.15)	(0.08)	(0.08)	
Health and family								
General health	coef.	0.03	0.04*	0.12**	0.09**	0.10**	0.07*	3,784
	s.e.	(0.02)	(0.02)	(0.05)	(0.04)	(0.04)	(0.04)	
Health recently	coef.	0.03	0.06**	0.28***	0.22***	0.11*	0.07	3,784
	s.e.	(0.03)	(0.03)	(0.07)	(0.07)	(0.07)	(0.07)	
Family functioning	coef.	0.00	-0.03	2.00***	1.94***	0.84	1.26**	3,693
	s.e.	(0.22)	(0.24)	(0.50)	(0.50)	(0.57)	(0.55)	
Social support	coef.	-0.26*	-0.36**	1.23***	1.09***	0.64*	0.75***	3,681
	s.e.	(0.15)	(0.16)	(0.34)	(0.32)	(0.40)	(0.39)	
Controls								
Child age at test (months)		yes	yes	yes	yes	yes	yes	
Child and family (at birth)		no	yes	no	yes	yes	yes	

Note: Children of working mothers are compared with children of non working mothers. The top panel shows the estimated intention-to-treat effects of the reform on child development and family well-being. The bottom panel specifies which control/matching variables are included in the estimate. Columns 1 and 2 show the results using the standard BA estimator, columns 3 and 4 the results using the DID estimators, and columns 5 and 6 using MDID with local linear regression (llr) and nearest neighbor with 5 neighbors (nn5). Child and family (at birth) refers to the following control variables: breastfed at birth (yes,no), premature (yes,no), birth weight (normal, low, very low), multiple births (yes, no), hospitalization at birth (yes, no), male, number of siblings (0, 1, 2 or more), maternal and paternal education (less than high school, high school degree, more than high school but no college degree, college degree, university degree), mother immigrated in the last 4 years (yes, no), age of the mother, age of the mother at the time of her first baby, urban area (rural to more than 500,000 inhabitants), and province of birth. Matching is performed using psmatch2, version .4.0.4, E. Leuven and B. Sianesi (November 10, 2010). Coefficient significance is denoted using asterisks: *** is $p<0.01$, ** is $p<0.05$, and * is $p<0.1$.

Table 3: Robustness checks

Sample		Main sample (MS)			MS excl. Nov to Feb births	MS excl. top 5% pscore	MS excl. matched $\geqslant$ 2%	Post reform 2001-2004 births
Specifications		Base	Unemp.	Full set	Base	Base	Base	Base
Time at home								
1st year of life	coef.	3.88***	3.92***	3.76***	4.00***	3.85***	3.78***	0.40
	s.e.	(0.27)	(0.27)	(0.27)	(0.33)	(0.27)	(0.31)	(0.32)
Cognitive measures								
PPVT	coef.	2.99*	2.92*	4.20**	2.93*	2.99*	3.36*	-2.77*
	s.e.	(1.54)	(1.58)	(1.66)	(1.63)	(1.55)	(1.94)	(1.63)
Who Am I?	coef.	1.09*	1.05*	1.00	1.00	1.08*	1.08	0.19
	s.e.	(0.61)	(0.64)	(0.62)	(0.63)	(0.62)	(0.78)	(0.57)
Number knowledge	coef.	0.70	0.73*	0.86*	0.72*	0.73*	0.96*	-0.60
	s.e.	(0.44)	(0.43)	(0.45)	(0.43)	(0.44)	(0.58)	(0.50)
Behavioral measures								
Hyperactivity	coef.	-0.09	-0.07	-0.28	-0.01	-0.05	-0.14	0.12
	s.e.	(0.26)	(0.26)	(0.28)	(0.26)	(0.26)	(0.32)	(0.29)
Emotional disorder	coef.	-0.15	-0.15	-0.03	-0.06	-0.13	-0.17	-0.27
	s.e.	(0.19)	(0.18)	(0.18)	(0.19)	(0.19)	(0.23)	(0.19)
Conduct disorder	coef.	0.19	0.20	0.18	0.16	0.17	0.04	0.02
	s.e.	(0.18)	(0.17)	(0.17)	(0.17)	(0.18)	(0.23)	(0.16)
Indirect aggression	coef.	-0.06	-0.05	-0.03	-0.06	-0.06	-0.05	-0.03
	s.e.	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
Health and family								
General health	coef.	0.10**	0.10**	0.11***	0.12***	0.10**	0.03	-0.01
	s.e.	(0.04)	(0.04)	(0.04)	(0.04)	(0.07)	(0.05)	(0.05)
Health recently	coef.	0.11*	0.10	0.12*	0.09	0.11	0.03	-0.04
	s.e.	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.09)	(0.06)
Family functioning	coef.	0.84	0.80	1.10*	0.44	0.82	1.22**	0.93*
	s.e.	(0.52)	(0.51)	(0.04)	(0.57)	(0.52)	(0.60)	(0.51)
Social support	coef.	0.64*	0.66*	0.62**	0.40	0.60	0.90**	1.29***
	s.e.	(0.37)	(0.37)	(0.37)	(0.40)	(0.37)	(0.46)	(0.36)
N min		4,674	4,682	4,643	4,299	4,606	4,575	4,884
N max		5,299	5,338	5,321	4,889	5,246	5,302	5,678
Controls								
Child age at test (months)		yes	yes	yes	yes	yes	yes	yes
Child and family (at birth)		yes	yes	yes	yes	yes	yes	yes
Child and family (age 0 to 1)		no	no	yes	no	no	no	no
Unempl. rate before birth		no	yes	no	no	no	no	no

Note: This table shows the estimated intention-to-treat effects of the reform. All estimated coefficients are obtained using MDID with local linear regression. All estimates rely on the main sample (MS). The treatment group includes children of working mothers and the control group includes all other children. Column 1 shows the base specification. Column 2 presents the base specification when average unemployment rate is included. Column 3 presents the base specification when all possibly endogenous variables are included. Child and family (at birth) includes the same set of variables as in Table 2. Child and family (age 0 to 1) further includes the following variables measured at the time of the first interview: marital status (married or not), ineffective parenting scale, positive parenting scale, family functioning scale, social support scale, paternal income quartile (measured at the provincial level), and age in months at which the child slept a full night. In column 4, children born between November 2000 and February 2001 are excluded. In column 5, the top 5% highest pscores are excluded from the reference group. In column 6, observations matched at least 2% of the time are excluded. In column 7, we use only children born post reform, between 2001 and 2004 and the treatment groups include children of working mothers. Coefficient significance is denoted using asterisks: *** is $p<0.01$, ** is $p<0.05$, and * is $p<0.1$.

Table 4: Type of care displaced

	Before		After		
	Mean	Std. Dev.	Mean	Std. Dev.	Diff.
Type					
Daycare	0.06	0.23	0.01	0.11	-0.04***
Non relative (outside)	0.16	0.36	0.05	0.22	-0.11***
Non relative (home)	0.06	0.24	0.03	0.18	-0.02
Relative (outside)	0.13	0.33	0.09	0.28	-0.04
Relative (home)	0.06	0.24	0.04	0.19	-0.03
Own parent	0.53	0.50	0.73	0.44	0.20***
Intensity					
Hours per week	27.04	17.83	24.10	16.24	-2.94
Quality					
Training	0.20	0.40	0.27	0.44	0.07
License	0.16	0.37	na	na	na
Number of observations	472		396		

Note: This table shows the type of care used in the first year of life prior to (columns 1 and 2) and after (columns 3 and 4) the reform for treated children from the main sample aged less than 12 months. Non relative (outside) means that the child is taken care of by a non relative outside of the home, but not in a daycare center (e.g. family based daycare). Non relative (home) means that the child is taken care of by a non relative in his own home. The same logic applies for the categories relative (outside) and relative (home).

Web Appendix to Connecting the dots: The early impacts of increased paid maternity leave on child development

Catherine Haeck

August 2013

In this Appendix, we first provide summary tables of research studies on the link between maternal time investment and child development. Second, we provide the underlying questions to the family functioning and social support measures. Third, we present the balancing properties of the MDID estimator using kernel, local linear regression and nearest neighbor matching. This explains ou focus on local linear regression and nearest neighbor matching in Table 2 of our paper. Fourth, we discuss self-selection based on the timing of birth. Finally, to ensure comparability of the estimated effects of the reform, percentages of a standard deviation are also presented.

1 Literature

In this literature review, we differentiate between research focused on (1) maternal employment (and daycare use) and (2) parental leave reforms, both of which are factors affecting maternal time investment.

Table A. 1: Maternal employment and child outcomes

Study and data	Approach and outcomes	Main findings
First-Year Maternal Employment and Child Development in the First Seven Years (Brooks-Gunn, Han, Waldfogel (2010)) NICHD, N= 900 (non-Hispanic White) + 113 (African-American), age 3, 4.5 and 7 or 8	Mediating model including mother's earning, home environment, type and quality of care as mediating factors. Outcomes include: Bracken School Readiness scores: age 3 (cognitive ability), Preschool Language Scale: Age 4.5, Woodcock-Johnson Psycho-Educational Battery-Revised, Social Skills Questionnaire, Social Skills Rating System, Current School Performance reported by teachers (age 7), Child Behavior Checklist (Mother Reported, Teacher Reported)	Full-time maternal employment in the first 12 months of life (but not part-time) is associated with lower cognitive scores (but not all measures). This association is offset by greater maternal sensitivity and the use of center-based care.
Maternal Work Early in the Lives of Children and Its Distal Associations With Achievement and Behavior Problems: A Meta-Analysis (Lucas-Thompson, Goldberg, Prause (2010))	Meta-analysis of 69 studies using a random effects model. Focus on two main outcomes: acheivement and behavioral problems.	Maternal employment during Year 1 is associated with negative outcomes, while during Year 2 and 3 it is associated with higher acheivement. Early employment is most beneficial for children of single parent or welfare status.
Does mother's employment conflict with child development? Multilevel analysis of British mothers born in 1958 (Verropoulou, Joshi (2009)) NCDS, N=1714, age 4 to 17	Controls include mother's own childhood test scores on behavioral adjustment and cognitive development, child health, age and gender, family composition and status. Outcomes include the PIAT (reading and math), External and Internal behavioral adjustment	Maternal employment in the first year of life is associated with slightly lower reading . Employment beyond the first year is not associated with any of the outcomes studied.

Table A. 1: Maternal employment and child outcomes (Continue)

First-year maternal employment and child outcomes: Differences across racial and ethnic groups (Berger, Brooks-Gunn, Paxson, Waldfogel (2008)) Fragile Families and Child Wellbeing (FFCW), N=1483, born between 1988 and 1990	Multivariate regression using propensity score matching to address selection bias. Large number of controls and mediating variables. Outcomes : Child Behavior Checklist and PPVT-R at 36 months	Maternal employment in the first year of life is associated with lower vocabulary scores for White children (but Black and Hispanic), and with higher behavioral problems for Hispanic children (but not Black and White). Mediating factors such as type of childcare, maternal depressive symptoms and parenting skills do not mediate these relationships.
The effect of maternal employment and child care on children's cognitive development (Bernal (2008)) NLSY	Structural model jointly modeling the decision to work and the choice of childcare. Outcomes include PPVt and PIAT (reading and math).	Working full-time and using child care during one year is associated with a reduction in ability test scores (-0.13 SD).
Working mothers and early childhood outcomes: Lessons from the Canadian National Longitudinal study on children and youth (Sherlock, Synnes, Koehoorn (2008)), NLSCY 98-99, N=6664	Logistic regression assessing the association between duration of maternity leave and impaired performance (>-1 SD below the mean). Outcome : Motor and Social Development (MSD) scale. Controls include maternal age, gender, and socioeconomic status.	Maternity leave of less than 2 years is associated with lower MSD score, and the effect is stronger if less than 1 year. The association is stronger for males and and children of yunger mothers. Higher socio-economic status and reduced number of siblings mediate (in part) these associations.
Are There Long-Term Effects of Early Child Care? (Belsky, Burchinal, McCartney, Vandell, Clarke-Stewart, Owen (2007)) NICHD SECCYD, N=1364 (max), age 4.5 to 12	Multivariate regression with a large number of covariates (e.g. type, quantity and quality of childcare, family income, maternal mental health, parenting quality) combined with multiple imuptation for missing data. Outcomes variables include Child Behavior Checklist Teacher Report Form, Student-Teacher Relationship Scale, Woodcock-Johnson Psycho-Educational Battery-Revised, Social Skills Questionnaire, Social Skills Rating System, Teacher Checklist of Peer Relations (socioemotional functioning)	Two persistent relationships are established: (1) higher quality care is positively correlated with vocabulary test score in grade 5 and (2) time spent in center-based care in early childhood is associated with teacher reported behavioral problems in grade 6.
Maternal Employment and Child Development: A Fresh Look Using Newer Methods (Hill, Waldfogel, Brooks-Gunn, Han (2005)) NLSY, N=6114, born between 1982 and 1993	Multivariate regression using propensity score matching and multiple imputation. Outcomes include PPVT (age 3-4), PIAT (math and reading) (age 5 to 8), Behavioral Problem Index Internalizing and Externalizing (age 5 to 8)	Full-time maternal employment in the first year, compared to employment beyond the first year, is associated with small but significant negative impact on cognitive skills and with smal but not statistically significant adverse behavioural effect.
The effects of a mother's return to work decision on child development in the UK (Gregg, Washbrook, Propper, Burgess (2005)) Avon Longitudinal Study of Parents and Children (ALSPAC), N=4607 to 6792	Multivariate regression with a large number of controls including type of childcare used, child's age in months, gender, ethnicity, birth weight and admission to a special care unit immediately after the birth, mother's age at the start of pregnancy and educational attainment, labour market outcomes of the spouse. Cognitive outcomes include Entry Assessment (4 or 5 years), Key Stage 1 (6 or 7 years), ALSPAC literacy (7 years)	Full-time maternal employment within the first 18 months of life is associated with small adverse consequences on cognitive development when maternal care is substituted by informal care by a relative or a friend or a neighbor. These effects are concentrated amongst children of highly educated mothers. Maternal employment is not associated with children's cognitive development when using formal care (centre based care or child minder).
The Effect of Maternal Labor Force Participation on Child Development (James-Burdumy (2005)) NLSY, N=1775	Sibling fixed-effect model and IV with the percentage of the county labor force employed in services as instrument. Outcomes : PPVT and PIAT (math and reading).	FE: Maternal employment during the first year of life has a negative impact on the PIAT-math (weeks worked and hours) and PIAT-reading (weeks worked) scores, while weeks worked during the 3rd year has a positive effect on the PIAT-math score.

Table A. 1: Maternal employment and child outcomes (Continue)

Parental Employment and Child Cognitive Development (Rhum (2004)), NLSY 1986 and 1996, N=3042, age 3 to 6	Multivariate analysis using a very large number of control variables (e.g. maternal education, ASVAB score and age, marital status, race, gender,but also low birth weight and short gestation indicators, hospital stay at birth, substance used during pregnancy, maternal occupation and average hours worked prior to birth, height-adjusted weight prior to pregnancy, religious affilition among others). Outcomes include the standardized scores on PPVT (vocabulary) and on PIAT (math, reading).	Maternal employment (first 3 years) has a deleterious effect on verbal ability of 3-4 yr-old and has an even larger effect on reading and mathematics achievement of 5-6 yr-old.
Does Early Maternal Employment Harm Child Development? An Analysis of the Potential Benefits of Leave Taking (Baum (2003)) NLSY, N=3103, born between 1988 and 1993	Multivariate regression and IV using and large set of control variables. Instrument maternal labour supply after birth using locallabour market conditions and using a more homogeneous group of mothers all working prior to birth. Maternal employment is actually define as a mother working as opposed to a mother being employed (and possibly at home on maternity leave). Crucial distinction often not done prior to this research. Outcomes include PPVT and PIAT (reading and math)	Maternal work in the first year of life has negative effects on child development partially offset by higher family income.
Does Amount of Time Spent in Child Care Predict Socioemotional Adjustment During the Transition to Kindergarten? (NICHHDECCRN (2003)) NICHD, N=1058	Multivariate regression models with increasing number of control variables: (gender, ethnicity, mother's education, maternal depression (intercept, slope) 6–54 months average income-to-needs ratio, 6-month temperament, average quality of care, proportion of center care, proportion of peer group exposure, instability of care, maternal sensitivity (intercept, slope)). Outcomes include Social Skills Rating System (SSRS), Child Behavioral Checklist (CBCL), Student-Teacher Relationship Scale (STRS), Preschool Social Competency Scale, parent-reported variables.	More time spent in nonmaternal care (across the first 4.5 years) is significantly related to a variety of behavioral problems on a continuous scale, with no apparent treshold of quantity effect.
Maternal Employment and Child Cognitive Outcomes in the First Three Years of Life: The NICHD Study of Early Child Care (Brooks-Gunn, Han, Waldfogel (2002)) NICHD, N=900, European American children	Multivariate regression. Controls include: type, intensity and quality of childcare (every 3 months), maternal depression, sensitivity, PPVT score, age and education, HOME, family income and area of residence. Child cognitive outcomes measures include Bracken School Readiness scores: age 15, 24, 36 months (cognitive ability), and Bayley MDI: age 15 and 24 months (cognitive ability)	Maternal employment by the ninth month was found to be linked to lower Bracken School Score at age 36 months, with stronger effects for 30 hours of work or more per week. Children of less sensitive mothers, boys, children with married parents were also more adversly impacted by maternal employment in early childhood (<= 36 months).
The Effects of Early Maternal Employment on Child Cognitive Development (Brooks-Gunn, Han, Waldfogel (2002)) NLSY, N=903 (White) + 582 (Black) + 387 (Hispanic), birth to age 7 or 8	Family fixed effects and multivariate regression. Controls include: mother and family characteristics, including breast-feeding and type of childcare used. Outcomes include: PPVT-R, PIAT at age 5 or 6, PIAT at age 7 or 8.	First year maternal employment is associated with lower cognitive scores, while employment in 2nd and 3rd year of life is associated with higher cognitive outcomes for some measures (for non-Hispanic White).
The Effects of Early Maternal Employment on Later Cognitive and Behavioral Outcomes (Han, Waldfogel, Brooks-Gunn (2001)) NLSY, N= 462, age 3 to 4 in 1986	Multivariate regression using a large number of control variables. Outcomes include PPVT (age 3-4), PIAT (math and reading) (age 5 to 8), Behavioral Problem Index Internalizing and Externalizing (age 5 to 8)	Maternal employment in the first year is associated with lower cognitive score for White children (but not Black) and this association persists over time.
Short-Term and LongrTerm Effects of Early Parental Employment on Children of the National Longitudinal Survey of Youth (Harvey (1999)) NLSY 1994	Multivariate regression approach with a clear distinction between selection variables and mediating variables. Outcomes : Compliance - Temperament Scale, Behavioral Problem Index (BPI), Self-Perception Profile, PPVT-R and PIAT.	Maternal employment is minimally associated with child's development: Increased work hours is associated with slightly lower cognitive score through age 9 and academic achievement scores before age 7, but not to behavioral problems, compliance, or self-esteem.
Developmental Follow-Up of 6-7 Year Old Children of Mothers Employed During Their Infancies (Barglow, Contreras, Kavesh, and Vaughn (1998)), N=113	Multivariate regression with limited controls. Outcomes studied at age 6-7: Stanford Binet I.Q. test, Coddington Inventory of Life Events, Child Behavor Checklist (CBC), Preschool Interpersonal Problem Solving Test (PIPS), Personality Inventory of Children (PIC)	Maternal employment during early childhood is not associated with I.Q. but is positively associated with behavioral problems in male children.

Table A. 1: Maternal employment and child outcomes (Continue)

Are the "Most Advantaged" Children Truly Disadvantaged by Early Maternal Employment?: Effects on Child Cognitive Outcomes (Greenstein (1995)) NLSY, N=2040 age 4 to 6	Multivariate regression with a large number of control variables including gender, low birth weight, age, mother's education, AFQT score, marital status and past employment history.	No effects are found for maternal employment over the first 4 years of a child's life.
Early Parental Work, Family Social Capital, and Early Childhood Outcomes (Parcel, Menaghan (1994)) NLSY 1986, N=786	Multivariate regression including numerous controls, including occupational complexity, wage and work hours for both parents. Outcomes: PPVT-R and Behavioral Problem Index (BPI).	Maternal employment within the first 3 years of life is minimally positively associated with cognitive test score. Non-linear association between working hours and child development, with overtime and less than part-time work being associated with lower outcomes. Paternal part-time work associated with elevated behavioral problem.
Effects of Early and Recent Maternal Employment on Children from Low-Income Families (Vandell, Ramanan (1992)) NLSY 1986, N=189, in grade 2 and from low-income family	Multivariate regression. Controls include: maternal education, self-esteem and ASVAB score, family income, HOME and values. Outcomes include the PPVT, Behavior Problem Index, and Peabody Acheivement Test.	Positive effect of maternal employment on math (first year employment) and reading (recent employment) acheivement for children of mothers who chose to be employed in low-income families.
Early and Extensive Maternal Employment and Young Children's Socioemotional Development: Children of the National Longitudinal Survey of Youth (Belsky, Eggebeen (1991)) NLSY 1986, N=1248, age 4 to 6	Multivariate analysis using five measures of socioemotional functioning: compliance, inhibition, attachment insecurity, sociability, and behavior. Controls include maternal education, ASVAB score and age, birth order, gender race, porverty status.	Full-time maternal employment in the first year or second year of life is associated with lower levels of compliance than more limited maternal employment.
Maternal Labor Supply and Children's Cognitive Development (Blau, Grossberg (1992)) NLSY, N=876, 3-4 year-old in 1986	OLS and IV to accounted for the self-selection of mothers into the labor market. Controls include maternal education, ASVAB score, family income, paternal education, race, gender.	Maternal employment in the first year has a negative impact on the child's PPVT score, with potentially offsetting positive effects when it occurs beyond the first year. Overall, employment over the first 3 to 4 years would have a neutral effect.
Mother or Market? Effects of Maternal Employment on the Intellectual Ability of 4-Year-Old Children (Desai, Chase-Lansdale, Michael (1989)) NLSY 1986, N=503, age 4	Multivariate regression including numerous controls such as maternal education, ASVAB score and age, birth order, race, family income, paternal education.	Maternal employment in the first year of life has detrimental effects on cognitive development, but only for boys in higher income families, while maternal employment after the first year was not found to be detrimental
Effects of Maternal Absence Due to Employment on the Quality of Infant-Mother Attachmentin a Low-Risk Sample (Barglow, Vaughn and Molitor (1987)), Data from Joffe, Vaughn, Barglow, Benveniste (1985), N=110	Multivariate regression with limited controls. All substitute care was provided in the child's home by someone not in the family. Outcome studied: Ainsworth Strange Situation (12-13 months)	Full-time maternal work between age 8 to 11 months (at least) is associated with a higher probaility of an insecure infant-mother attachment

Table A. 2: Parental leave reforms and child outcomes

Study and data	Approach and outcomes	Main findings
A flying start? Long term consequences of maternal time investments in children during their first year of life (Carneiro, Løken, Salvanes (forthcoming))	Norwegian paid parental leave reform from 0 to 18 weeks, 100% income replacement rate, and unpaid leave from 12 to 52 weeks (effective July 1st 1977). Using children of eligible mothers (measured using income before birth), they use a regression discontinuity approach with diff-and-diff using 3 different control groups: (1) children born in 1975 of eligible mothers; (2) children born in 1979 of eligible mothers; and (3) children born in 1977 of ineligible mothers.	Using children of eligible mothers, they find that the reform had a positive impact on children through an increase in the probability of completing high school (2.7%), with a larger effect for children of less educated mothers (5.2%).
What Is the Case for Paid Maternity Leave? (Dahl, Loken, Mogstad, Salvanes, 2013) Administrative records, N>122,000	Norwegian paid parental leave reforms (series of expansions between 1987 to 1992) leading to an overall increase of 17 weeks (from 18 to 35 weeks) with a 100% income replacement rate. Regression discontinuity design around each reform (measuring the effect of an increase of 2 to 4 weeks). Sample of mothers with a certain earning level (likely eligible). Outcomes include children's schooling outcomes, parental earnings and participation in the labor market, completed fertility, marriage or divorce.	Effective treatment period is beyond age 10 to 11 months. Results suggest that the reform had little effects. The authors further suggest that this reform was a pure leisure transfert tomiddle and upper income families.
Parental Leave and Children's Schooling Outcomes: Quasi-Experimental Evidence from a Large Parental Leave Reform (Danzer, Lavy, 2013)) PISA OECD, N=1480 (boys: 752, girls: 728)	Austrian paid parental leave reform from 12 to 24 months (effective July 1st 1990), replacement rate about 40% of net median earnings. Long term outcomes using PISA 2003 and 2006. Regression Discontinuity Design (children born 2 months before, 2 months after) with diff-and-diff (control group is same months, but in year 1987 prior to the policy change with outcomes in 2003)	On average no impact at age 15 on PISA test scores, but positive effects on test scores of boys of highly educated mothers and negative effects on boys of less educated mothers.
Maternity Leave and Children's Cognitive and Behavioral Development (Baker and Milligan, 2012) NLSCY, N=9950 to 10971, ages 4 and 5	Canadian paid parental leave reform effectively increasing paid leave from 25 to 50 weeks (effective January 1st 2001), 55% income replacement rate. IV approach, where time at home is instrumented using (1) timing of birth (pre versus post reform) and (2) changes in provincial legislation on job-protection after child birth. Outcomes include the PPVT score, Who Am I? score, Number Knowledge 4 points standardized score, and behavioural scores (4).	No positive impact, and small negative impact on cognitive scores.
The Effect of Expansions in Maternity Leave Coverage on Children's Long-Term Outcomes (Dustman, Schonberg (2012)) Administrative data on school choice and labor market outcomes, N>15000 for the 1986 and 1992 reforms, and N>42000 for the 1979 reform.	3 expansion in leave coverage in Germany: 1979 (2 to 6 months, replacement rate about 1/3 of average pre-birth earnings), 1986 (6 to 10 months, replacement rate), 1992 (18 to 36 months unpaid leave). Long-run education and labor market outcomes are type of school attended, graduation from highest track at age 20, highest level of school completed (3 categories) and income at age 28-29. Regression Discontinuity Design (children born one month before, one month after) with a diff-and-diff (control group is same months, but in years prior to the policy change)	1979 reform, no impact on level of education completed or income at age 28-29. 1986 reform, no impact on the probability of completing education from the highest track at age 20. 1992 reform, no impact on the type of school attended.
Causal effects of paternity leave on children and parents (Cools, Fiva, Kirkebøen (2011)) Administrative records, N=28797	Norwegian paid parental leave reform from 35 to 42 weeks (effective April 1st 1993), including 4 weeks exclusively reserved to the father. Outcomes include maternal and paternal labor supply, fertility, divorce rate, and children's test score at age 16. Instrumental variable approach in which paternity leave is instrumented with the timing of birth (before or after the reform).	Adverse effect on maternal labour market outcomes (child age 2 to 9) and positive and significant impact on female children's school performance (0.38 SD) in families with highly educated fathers. No significant effects are detected on male children or other outcomes.
Do Family Policy Regimes Matter for Children's Well-Being? (Engster, Olofsdotter Stensöta (2011))	Multivariate regression across 20 OECD countries to identify the association between family policies and child well-being, including poverty, mortality, educational attainement and achievement (PISA tests)	Paid parental leave is associated with lower child poverty rate and mortality, but not educational attainement and achievement

Table A. 2: Parental leave reforms and child outcomes (Continue)

The effects of maternity leave on children's birth and infant health outcomes in the United States (Rossin (2011))	USA non paid maternity leave reform from 0 to 12 weeks (effective August 5th, 1993). Diff-and-diff (DD) and also DDD, using states and firm sizes pre-existing rules to identify controls groups.	Small increases in birth weight and decreases in the likelihood of a premature birth, and substantial decreases in infant mortality for children of college-educated and married mothers.
Evidence from Maternity Leave Expansions of the Impact of Maternal Care on Early Child Development (Baker and Milligan, 2010) NLSCY, N=4447 (max), aged 7 to 24 months.	Canadian paid parental leave reform, see Baker and Milligan (2012). Before-After (3 years prior, 3 years after) approach combined with IV using quarter of birth as intruments. Also compares mother who returned to work within 12 months with all other mothers. Outcomes include a wide variety of parent reported behavioral mesure and the Motor and Social Development scale.	The reform mainly displaced unlicensed care by non-relative. Generally, no effects are found with the power to rule out changes of greater than 10 to 15 percent of a standard deviation.
The Duration of Paid Parental Leave and Children's Scholastic Performance (Liu, Nordstrom Skans (2010)) IFAU database, N=252393, age 16 and born between 1987-89	Swedish paid parental leave reform from 12 to 15 months (effective August 1st, 1988), replacement rate of 90%. Outcomes include tests scores at age 16 (Swedish, English, Math, overall GPA) and number of hospital admittance. The approach used is OLS with controls for timing of birth (reform was progressive RDD not possible).	The reform had on average no effects, except for children of well-educated mothers who were better off post-reform. The reform had no impact on hospital admittance for children at age 3, 6 and 16.
Increasing the length of parents' birth-related leave: The effect on children's long-term educational outcomes (Würtz Rasmussen (2010)) Administrative register of children born between January to May 1983 and January to May 1984, N=8600	Danish parental leave reform from 14 to 20 weeks after birth (effective March 26, 1984). Long-term educational outcomes include high school enrollment, high school completion, reading score at age 15 and GPA at age 21. Regression Discontinuity Design (children born 2 months before, 2 months after) with diff-and-diff (control group is same months, but in year (1983) prior to the policy change)	No effects are found.
Maternal employment, breastfeeding, and health: Evidence from maternity leave mandates (Baker and Milligan, 2008) NLSCY, N=5708 (max), 13 to 29 months	Canadian paid parental leave reform, see Baker and Milligan (2012). Also diff-in-diff with older cohorts for health outcomes. Outcomes include breastfeeding and a large number of parent-reported health outcomes.	Find that the reform had a postive impact on the probability of reaching the 6 months critical period (between 7.7 and 9.1 p.p.), but generally not impact on health reported outcomes.

2 Outcomes

In this section, we provide the questions that are used to form the family functioning score and the social support score.

Family functioning score The total score varies between 0 and 36, a high score indicating family dysfunction. Points for answers: 0 Strongly agree; 1 Agree; 2 Disagree; 3 Strongly disagree (the value were reversed for answers starting with a R)

1 Planning family activities is difficult because we misunderstand each other (R)

2 In times of crisis we can turn to each other for support

3 We cannot talk to each other about sadness we feel (R)

4 Individuals, in the family, are accepted for what they are.

5 We avoid discussing our fears or concerns (R)

6 We express feelings to each other
7 There are lots of bad feelings in our family (R)
8 We feel accepted for what we are.
9 Making decisions is a problem for our family (R)
10 We are able to make decisions about how to solve problems.
11 We don't get along well together (R)
12 We confide in each other.
13 Drinking is a source of tension or disagreement in our family (R)

Social support The total score varies between 0 and 24, a high score indicating the presence of social support. Points for answers: 0 Strongly agree; 1 Agree; 2 Disagree; 3 Strongly disagree (the value were reversed for answers starting with a R).

Do you strongly disagree, disagree, agree or strongly agree with the following statement :

1 ...if something went wrong, no one would help me? (R)
2 ...I have family and friends who help me feel safe, secure and happy?
3 ...there is someone I trust whom I would turn to for advice if I were having problems?
4 ...there is no one I feel comfortable talking about problems with? (R)
5 ...I lack a feeling of closeness with another person? (R)
6 ...there are people I can count on in an emergency?
7 ...I feel part of a group of people who share my attitudes and beliefs?
8 ...there is no one who shares my interests and concerns? (R)

3 Selection into treatment using the EICS

Using the Statistics Canada Employment Insurance Coverage Survey (EICS) we analyze selection into treatment. More specifically, we assess whether mothers changed their working behavior prior to birth in order to be entitled to the new parental leave benefit. It is important to remember that prior to the reform, mothers were already entitled to paid maternity and parental leave for up to 25 weeks. In this senses, we want to know whether adding an additional 25 weeks of paid parental leave, as enticed mothers who were not participating in the labour market prior to the reform to participate after the reform given the additional benefit.

We use the 2000 and 2002 EICS.[26] Mothers of children aged 12 months or less in 2000 gave birth prior to the reform, while mothers of children aged 12 months or less in 2002 gave birth after the reform. Table A.3 shows marginal effects of a probit model of maternal

[26] In the 2001 EICS, mothers of children aged 12 months or less may have given birth before and after the reform. Since we are unable to classify mothers, we decided not to use this wave of the EICS.

labour force participation in the past 12 months on post, a dummy variable equal to one if the mother gave birth after the reform and zero otherwise.

Table A. 3: Variation in labour force participation and eligibility status

	Work	Work	Eligible	Eligible
Post	0.02	0.00	0.12***	0.09***
	(0.65)	(0.00)	(4.00)	(3.53)
Maternal age (25 to 44)		-0.01		0.14***
		(-0.37)		(3.82)
Province: Ontario		-0.04		0.01
		(-1.29)		(0.29)
Province: Manitoba and Saskatchewan		-0.06*		0.02
		(-2.11)		(0.53)
Province: Alberta		-0.01		0.00
		(-0.37)		(0.01)
Province: Colombie-Britannique		-0.08*		-0.03
		(-2.12)		(-0.63)
Immigrant		-0.03		-0.04
		(-1.11)		(-1.16)
2 parent, 1 earner		-0.50***		-0.40***
		(-17.76)		(-12.97)
2 parent, no earner		-0.59***		-0.36***
		(-8.12)		(-4.85)
Other type of family		-0.16		0.02
		(-0.82)		(0.08)
Family size: 4		-0.07*		-0.14***
		(-2.56)		(-4.40)
Family size: 5 and more		-0.07*		-0.24***
		(-2.38)		(-6.74)
Maternal education: High school		0.04		0.01
		(0.95)		(0.20)
Maternal education: Post secondary		0.04		0.05
		(1.62)		(1.50)
Maternal education: University		0.05		0.07
		(1.51)		(1.94)
N	1,939	1,923	1,939	1,923

Note: This table shows the estimated marginal effects of a probit model of labour force participation in the last year and of eligibility to paid maternity leave for mothers giving birth prior to (Post=0) and after (Post=1) the reform. Estimates are based on EICS data for 2000 and 2002. Significance is denoted using asterisks: *** is p<0.01, ** is p<0.05, and * is p<0.1.

Results in columns 1 and 2 clearly show that maternal labour force participation hasn't changed post reform: the *Post* coefficient is both economically and statistically not significant. Given that the reform was generally not anticipated and the generosity of paid leave prior to the reform, it is not surprising to find that few mothers reacted to the change in parental leave benefits around the policy change. In columns 3 and 4, we look at the probability of claiming maternity leave benefit prior to and after the reform. We find that post

reform, the probability of claiming maernity leave benefit increased by about 8 p.p. on average. Since the number of hours required to be eligible was reduced post reform, this change was expected, and is also observed in the NLSCY data if we focus on mothers claiming maternity leave benefits instead of working mothers.

4 MDID and balancing properties

Effectively, to implement the MDID estimator, we first estimate a probit model in which the dependent variable equals one if the mother worked at the time of birth and equals zero otherwise. The control variables are fixed at the time of birth (or shortly thereafter) and influence the mothers' labor supply decision as well as the outcome of the child. Using this model, we predict the propensity score of each child (or the probability that the child's mother was working prior to birth). Matching is then performed using the propensity score. Rosenbaum and Rubin (1983) show that if observations in the treated and control groups have the same propensity score distribution, the underlying characteristics used to calculate the propensity score are also distributed equally. Children in the treatment group prior to the reform are first matched with children in the treatment group after the reform. Then children in the control group after the reform are matched with children in the treatment group after the reform. Finally children in the control group prior to the reform are also matched with treated children post reform. This ensures that all four groups share similar distributions for each of the characteristics contained in X. We implement kernel matching, local linear regression matching and nearest neighbor matching.

Stuart (2010) recommends the inclusion of a large number of variables to estimate the propensity score (as large as 100 covariates). While including a non influential variable has a minor impact on the propensity score model and thus on the estimated effect, omitting an influential variable can seriously bias the result. Ideally, variable selection should be done without knowledge of the impact on the estimated outcome. In this spirit, in our base model we include all of the variables we had access to that were measured at birth and likely influenced the child outcome and the decision of the mother. More specifically, we include the following set of variables measured at the time of birth: gender, prematurity, birth weight, multiple birth indicator, hospitalization at birth, breastfeeding at birth, number of siblings, maternal and paternal education, age of the mother at first child, age of the mother at birth, and immigration status of the mother in the last 4 years. All of these characteristics are known to be related to child development. Province of residence and area of residence (5 categories from rural to more than 500,000 inhabitants) at the time of birth are also included to control for regional differences. All variables are categorical dummies. Heckman and Navarro-Lozano (2004) recommend the inclusion of variables influencing the participation of the decision maker (here the mother) but not the outcome. When we conduct our sensitivity

analysis, we also include the average provincial unemployment rate in the year preceding birth to better understand the behavior of mothers. Finally, since cognitive measures (and possibly behavioral measures) are sensitive to the age at which the child took the test, we include age in months (dummies) at the time of test in the set of matching variables to ensure the equality of the age distribution prior to and after the reform.[27] These measures can be considered exogenous to the parental leave reform.

We have at our disposition a number of other measures related to child development, but they are measured at the time of the first interview (at age 0 to 1): the age at which the child slept 6 continuous hours at night (max. 6 months), marital status, paternal income quartile (by province), family functioning scale, social support scale and parenting skills scales. These measures influence child development as they reflect both child and parental behavioral differences. However, since they are measured between the age of 0 to 24 months, it is possible that the parental leave reform impacted these early measures such that they cannot be considered fully exogenous. Not controlling for these characteristics may however be problematic if they determine selection into treatment more than they are influenced by the reform itself. Measures of family functioning, social support and parenting skills are likely correlated with maternal sensitivity which has been shown to be correlated with child development (Brooks-Gunn et al., 2010). The empirical section checks the robustness of our results to their inclusion.

When using matching, Rubin (2001) identifies three distributional conditions that must be met simultaneously. First, the difference between the mean propensity score of the matched groups (i.e. control before, control after, and treated before) and that of the treated after treatment group should be small. Second, the variance of the propensity score distribution of the matched groups should be similar to that of the treated after treatment, such that the variance ratio should be close to one. Third, the variance of the residuals of each of the matching covariates in the matched groups should be similar to that of the treated after treatment group (i.e. the ratio of the variance should be close to one). For the MDID estimator with repeated cross-sections, each condition must be met for each of the three groups (treated before treament, and both control groups). Table A.4 provides the distributional conditions for the PPVT outcome for nearest neighbor, local linear regression and kernel matching.

[27] In the NLSCY, each cycle covers two birth years, but since all children are surveyed within a few months, children born early in the cycle are surveyed when they are much older than children born near the end of the cycle. As such, the age in month pattern declines in time within a cycle. As discussed above, since cognitive test scores are highly sensitive to age, each of the cognitive test scores follows the same pattern as average age at the time of test. Plotting average test score by birth month over multiple cycles results in a series of downward trending slopes. As such, a naive regression on average raw score by birth month would have a negative slope, mistakenly suggesting a negative impact of the reform on child development. Similar, but somewhat weaker, patterns are also identified for all four behavioral measures. We use age standardized score and account for age using age in month dummies for none standardized measures. Our results are not sensitive to using age standardized socre versus raw scores with age in month dummies.

Table A. 4: MATCHING DISTRIBUTIONAL CONDITIONS

	llr	kernel	nn(5)
Difference in mean propensity score			
Treated before vs treated after	0.0002	0.0037	0.0005
Control after vs treated after	0.0347	0.0614	0.0382
Control before vs treated after	0.0236	0.0504	0.0250
Propensity score variance ratio			
Treated before vs treated after	1.0004	1.0526	1.0028
Control after vs treated after	1.0096	1.0671	1.0202
Control before vs treated after	1.0058	1.0646	1.0095
Residuals variance ratio			
<1/2 and =<4/5	0.01	0.01	0.01
>4/5 and =<5/4	0.81	0.72	0.92
>5/4 and =<2	0.15	.023	0.07
>2	0.04	0.04	0.01

Note: Shows Rubin (2001) distributional conditions for the PPVT estimates. The first column presents the results for local linear regression matching (llr), the second column kernel matching (kernel), and the third nearest neighbor matching with 5 neighbors (nn(5)). Only observations for which the PPVT outcome is available are included. Similar findings are obtained for all other outcomes.

Table A.4 shows that the conditions with nearest neighbor matching (with 5 neighbors) are strongly met. More specifically, the mean differences are extremely close to zero, the propensity score variance ratios are nearly equal to one, and 92% of the covariates have a ratio of variance residuals between the optimal bound of 0.8 to 1.2. The conditions are also strongly supported for local linear regression. Compared to nearest neighbor matching, the difference in mean is closer to zero, the pscore variance ratio is closer to one, but a smaller proportion of the covariates are within the 0.8 to 1.2 bound. The conditions also find support for kernel matching, but the results are less satisfactory. Similar balancing property results are found for all of the other outcome measures investigated.

Figure 2 in the paper further confirms the above results. It shows the propensity score (pscore) distribution prior to (left panel) and after (right panel) matching for the estimated effects on the PPVT score using local linear regression matching. Effectively, the pscore is calculated using all matching variables described earlier (except for unemployment rate) for each observation with a non missing PPVT result. Prior to matching, the pscore distribution of both treated groups are similar, but very different from the pscore distributions of both control groups. After matching, all distributions are extremely similar. Similar figures are found for all of the other outcome measures studied.

5 Self-selection based on the timing of birth

The reform was announced on February 28^{th}, 2000.[28] Babies conceived around the announcement date were expected around November 20^{th}, 2000, prior to the reform. Mothers trying to conceive at the time of the announcement date may have delayed conception in order to be eligible for the expanded leave. The necessary conditions for delayed conception are that (1) births in November and December 2000 should be lower than expected, and (2) births in January, February and possibly March and April should be higher than expected. We use a larger time window post reform since the day a mother starts trying often does not result in conception. The results to follow are robust to a smaller or wider birth time window post reform. Using Statistics Canada Vital Statistics Birth Database, we find weak evidence of mothers delaying their timing of birth. Effectively, for the period 1995 to 2005, we regress monthly birth count on year dummies, month dummies, D_{2000} a dummy equal to one if birth occurs in November or December 2000, and D_{2001} a dummy equal to 1 if birth occurs in January to April 2001. D_{2000} is negative but not significant, and D_{2001} is positive but not significant. To address the possibility that the reform was anticipated prior to February 28, 2000, D_{2000} was set equal to one if births occurred between July and December 2000, and D_{2001} was set to one if births occurred between January and June 2001. D_{2000} is negative and significant, but D_{2001} remains not significant. More generally, birth counts regressed on year dummies shows that the 2001 year dummy is not significantly different from the 2000 year dummy.[29] As such, the evidence for delayed conception due to the announcement date is weak, but to further validate the robustness of the results this possibility is addressed in the empirical section.

[28] In the Speech from the Throne in the fall of 1999, the possibility of increasing the number of weeks of paid parental leave was covered, but it was only on February 28^{th}, 2000 that the government committed to increasing parental leave. Prior to this period, the subject was not covered in newspapers across the country.

[29] These results can be obtained from the author on request.

6 Estimated effects of the reform, percentages of a standard deviation

Table A. 5: Ratio of the estimated effects to the standard deviation

		Before-After		DID		MDID		Number of
		Age only	All	Age only	All	All llr	All nn5	observations
Time at home								
1st year of life	coef/sd	0.67***	0.71***	0.99***	0.98***	1.09***	1.17***	3,704
Cognitive measures								
PPVT	coef/sd	-0.06	-0.12**	0.19	0.04	0.20*	0.25**	3,370
Who Am I?	coef/sd	-0.19***	-0.23***	0.30***	0.19**	0.17*	0.18	3,256
Number knowledge	coef/sd	-0.02	-0.07	0.23**	0.11	0.14*	0.17*	3,370
Behavioral measures								
Hyperactivity	coef/sd	0.01	0.01	0.10	0.05	-0.03	-0.04	3,748
Emotional disorder	coef/sd	-0.02	0.04	0.00	0.01	-0.08	-0.20	3,762
Conduct disorder	coef/sd	0.08*	0.10**	0.17	0.08	0.10	0.10	3,763
Indirect aggression	coef/sd	0.02	0.00	0.16	0.11	-0.05	-0.04	3,681
Health and family								
General health	coef/sd	0.06	0.07*	0.22**	0.17**	0.19**	0.13*	3,784
Health recently	coef/sd	0.04	0.08**	0.37***	0.29***	0.14	0.09	3,784
Family functioning	coef/sd	0.00	-0.01	0.40***	0.38***	0.17	0.25**	3,693
Social support	coef/sd	-0.07*	-0.10**	0.35***	0.31***	0.18*	0.21***	3,681
Controls								
Child age at test (months)		yes	yes	yes	yes	yes	yes	
Child and family (at birth)		no	yes	no	yes	yes	yes	

Note: Shows the ratio of the coefficient to the standard deviation (treated before). See notes Table 2 for more details on the different specifications. Coefficient significance is denoted using asterisks: *** is p<0.01, ** is p<0.05, and * is p<0.1.