KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Economics and
Applied Economics

Department of Economics

Increased paid maternity leave and children's development measured at age four to five. An empirical analysis

by

Catherine HAECK

Econometrics

Center for Economic Studies
Discussions Paper Series (DPS) 11.15
http://www.econ.kuleuven.be/ces/discussionpapers/default.htm

June 2011

DISCUSSION
PAPER

Increased paid maternity leave and children's development measured at age four to five. An empirical analysis.

Catherine Haeck*

Abstract

Parental leave policies are often enacted based on the premise that children will benefit from an extended period of time spent with their parent. A number of research studies have looked at the effect of maternal time investments on the early development of skills, behavioral well-being and health, but the results thus far are mixed and mainly based on multivariate analysis. This approach can often not eliminate selection bias and can rarely predict the sign and magnitude of the bias. In this paper, I evaluate the effect of extended maternal care on children's development at age 4 to 5 using observational data prior to and after the Canadian parental leave reform, which extended total paid leave from 25 to 50 weeks on December 31st, 2000. Previous research exploiting this labor supply shock found that mothers significantly increased their time at home in the first year, but generally found no significant effects on parent-reported measures of development between age 7 and 24 months. For the first time in this literature, children of mothers receiving maternity leave benefits are identified and compared with all other children. Using matching difference-in-differences, I find that the policy change had positive effects on cognitive development, measured using different standardized tests for children aged 4 and 5. Behavioral development effects are mixed and mainly not significant. Effects on the family environment and parent-reported health measures are positive and significant.

* *Catherine Haeck:* Catholic University of Leuven. Email: Catherine.Haeck@econ.kuleuven.be. I gratefully acknowledge financial support from the Belgian Federal Science Policy Office (Interuniversity Attraction Poles P5/26). This working paper was written while visiting the University of Québec in Montréal, whom I thank for their hospitality. I thank Bart Cockx, Marteen Goos, Pierre Lefebvre, Philip Merrigan, Anna Salomon, Johannes Van Biesebroeck, Frank Verboven, Xuelin Zhang 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. 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 center 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)]. Early interventions have been shown to be more effective [Knudsen et al. (2006)]. Brain development is non-linear, such that certain periods are more critical than others for the acquisition of different abilities, and early experiences (even prior to birth) have a decisive impact on the development of human capabilities through their impact on the brain architecture [Shore (1997)].

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. Most research on the subject uses multivariate analysis and generally suggests that maternal employment in the first year of life is detrimental to child development.[1] Since ability formation is a cumulative process, it is fairly complex to evaluate the determinants of early life ability. There is a bidirectional relationship between the choices parents and children make, and the environment in which they evolve, making it especially difficult for any researcher to identify causal relationships. Research studies using multivariate analysis deal with the family and children specific bias by incorporating a large number of control variables to reduce the bias. However, there is

[1] Using the same data set Desai et al. (1989), Belsky and Eggebeen (1991), Baydar and Brooks-Gunn (1991), Blau and Grossberg (1992), Brooks-Gunn et al. (2002a) and Ruhm (2004) find that maternal employment in the first year of life is detrimental to child development, while Brooks-Gunn et al. (2002a), using family fixed-effects, find neutral effects and Vandell and Ramanan (1992) find positive effects.

growing consensus that this approach can often not eliminate the bias and can rarely predict the sign and magnitude of the bias [Duncan et al. (2004)].

In this paper, I use a unique and relatively recent natural experiment that extended total available paid maternity[2] leave from 25 to 50 weeks across Canada starting December 31^{st}, 2000. This unanticipated shock in the labor supply of working mothers allows me to provide convincing evidence of the impact of extended maternal care in the first year of life while avoiding the possible bias of correlational studies. Previous research exploiting this external labor supply shock found significant effects on maternal labor supply and breastfeeding, but not on health or parent-reported measures of child development for children age 7 to 24 months [Baker and Milligan (2008b) and Baker and Milligan (2010)].

Using Statistics Canada National Longitudinal Survey of Children and Youth (NLSCY), I estimate the impact of the reform on the development of children aged 4 to 5 (or more precisely between 46 to 60 months old inclusively). This age group is particularly relevant. First, as previously mentioned, ability gaps measured as early as age 5 have been shown to persist later in life, yet no other research has investigated the impact of maternity leave reforms on children aged 4 to 5. Second, standardized measures of cognitive development for this age group are available in the NLSCY, in addition to parent-reported measures of behavioral development, health and family well-being. The NLSCY is a long-term biennial survey providing detailed information on the development and well-being of children. The sampling unit is the child and it is designed to provide estimates representative of the population of Canadian children. Third, for the first time in the literature exploiting maternity leave reforms, a precise distinction, without any inference, between children of mothers claiming

[2] From here on, I 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)]. Ruhm (2004), using a multivariate approach approach, finds that fathers' time investment may substitute for that of mothers. Although paternal time investment may also have positive benefits, the limited variation in paternal time investment makes it particularly difficult to estimate this effect.

paid maternity leave benefits and children of mothers who did not can be made for this age group in the NLSCY. Children of mothers not claiming benefits are used to control for underlying trends in the outcome variables in a matching difference-in-differences estimator. By comparison, Baker and Milligan (2010) use a before-after model where no distinction between children is made.

A limited number of recent research studies have also exploited different maternity leave reform in Western Europe to estimate the impact of maternal time investment in early life. Carneiro et al. (2010) is the only paper that partially identifies whether the mother benefited from the reform or not, but their classification is only imperfect. They use the 1977 Norwegian reform, that for the first time provided paid maternity leave of 18 weeks in combination with an extended unpaid leave period of 52 weeks, compared to 12 weeks prior to the reform. Carneiro et al. (2010) find a positive impact on high school completion rates. The authors use discontinuity design in combination with difference-in-differences and use only treated children to estimate the impact of the reform. The authors show that results are not significant if treated and non treated children are pooled together. Unfortunately, the authors do not directly observe how mothers' time allocation between work and home changed following the reform. They estimate (using income around birth) that, prior to and after the reform, mothers took 8 months of unpaid leave on average. The authors assume that all eligible mothers post reform took the entire 18 weeks of paid leave given the income replacement rate of 100%. Therefore, the extension they investigate is an extension from 8 to 12 months, which covers the same critical period as the Canadian expansion (prior to the reform, eligible mothers in Canada were taking on average 7.1 months of leave and post reform they take 10.6 months). The authors find that the high school dropout rate of treated children significantly decreased following the reform.

Compared to this paper, I can precisely classify treated and non treated children, and can accurately measure how long a mother stayed at home prior to returning to work. For the same maternity leave reform, Baker and Milligan (2010) have also demonstrated that a 55% income replacement rate for a median income mother resulted in no income effect, such that any estimated impact could be attributed to maternal time investment. In the case of the Norwegian reform with a 100% income replacement rate, Carneiro et al. (2010) assume that there were no income effects. The choice to work (instead of staying at home) implies that mothers have to pay for daycare and work related expenses. Both in Canada and Norway, the alternative to maternal care at the time was informal care. Given Baker and Milligan's finding, a 100% income replacement rate most likely implies that families have more disposable income if the mother stays at home. The proportion of work related expenditures to total income is likely greater for lower income families. Carneiro et al. (2010) find that children of low income families benefited more from the reform. They attribute this finding to the additional time spent with their mother, but it may also be in part due to the income effect.

The other three papers exploiting natural experiments focus on long term outcomes beyond age 12 and exploit reforms of a much shorter duration (6 weeks or less). Liu and Skans (2010) investigate the impact of the 1988 reform in Sweden that extended parental leave from 12 to 15 weeks. They find that, on average, the leave extension had no significant effects on outcomes measured at age 16, such as test scores, but that it strengthened the relationship between mother's education and children's school outcomes. Rasmussen (2010) looks at the impact of a parental leave expansion from 14 to 20 weeks in 1984 in Denmark. He finds a neutral effect on children's high school educational outcomes. Dustmann and Schönberg (2008) study three reforms that took place in Germany. The first one extended paid leave from 2 to 6 months in 1979, the second one extended paid leave from 6 to 10 months in 1986 and the third one extended unpaid leave from 18 to 36 months in 1992. They find that none of these reforms had significant impacts on child outcomes at age 13 to 14 as measured by the type of school attended.

In sum, only Carneiro et al. (2010) find significant positive effects, but it is also the only paper that identifies (though imperfectly) treated children and directly estimates the effect of the treatment on the treated and that exploits a reform of long duration in the same time window as the Canadian reform. All other papers found no significant effects on child outcomes. All five papers rely on a before-after approach and assume no underlying trend effects. In contrast to these papers, I use matching difference-in-differences and use non treated children as a control group.

My econometric results suggest that benefit recipient mothers increased their time away from work in the first year after their child's birth by about 3 to 4 months. I find that the policy change had positive effects on cognitive development measured using different standardized tests. Behavioral development effects are mixed and mainly not significant. Effects on the family environment and parent-reported health measures are positive and significant. These findings have important political implications, especially in the United States, where female labor force participation in the first year of life is high and maternal leave income replacement is not regulated by the government.

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

2 The reform

In Canada, maternity leave benefits are legislated and paid by the federal government through Employment Insurance, while maternity leave duration is regulated at the provincial level. More specifically, the federal government legislates the amount to be paid during 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 Employment Insurance Program (EI) 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 increased leave duration to at least 50 weeks. The requirements for job protection vary at the provincial level, but are typically lower than those required for leave benefits.[3] This implies that some mothers not eligible to increased leave benefits were, however, eligible to a longer (unpaid) leave duration.

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)]. This policy was enacted based (in part) on the premise that increasing the time spent at home with the parent would (1) promote child development, (2) balance work demands and care for young children and (3) be a short-term investment with long-term economic gains[4]. The first goal explicitly took into account research by Brooks-Gunn et al. (2002b) showing that the first year of life was a critical period for the development of children, and by Main (1990) showing that a "secured attachment" was critical for the emotional, social and cognitive development of the child. From the empirical literature review, I find that research on the subject at the time of the policy implementation provided mixed results, but, except for Vandell and Ramanan (1992), the results suggested

[3] In British Columbia, New Brunswick and Quebec, pregnant employees are immediately eligible to job protected leave through provincial legislation. In Ontario, pregnant employees must have been employed for at least 13 weeks, and 20 weeks in New Foundland, Prince Edward Island and Saskatchewan. In Manitoba, 7 months are required, while 12 months are required in Alberta and Nova Scotia.

[4] The other three goals were: (4) use employment insurance as an effective instrument, (5) promote gender equality and (6) allow businesses to retain valuable, experienced employees [HRSDC (2005)].

that maternal employment in the first year of life was either detrimental or at best neutral.

I focus my attention on developmental benefits, i.e. cognitive and non-cognitive abilities, 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. However, since family environment is critical to the development of children, the assessment of gains on children indirectly takes into account work-life balance effects impacting the child. Health related benefits, for both the mother and the child, are greatest for leave extensions in the first few weeks of life. Baker and Milligan (2008b) looked at the effect of the Canadian reform on breast-feeding. They found that, although the reform had a significant effect on breast-feeding duration, the health related benefits before age 3 were modest. Given the extension period of the reform (6 to 12 months), I only briefly report the effect of the policy on two parent reported measures of health at age 4 to 5 to assess the persistence of health related benefits.

I use a control group to account for common underlying trend effects, but self-selection into treatment or reforms impacting one group but not the other could still bias my estimate. The reform was announced on February 28, 2000. 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. I 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. From the Statistic Canada Vital Statistics Birth Database, I find weak evidence of mothers delaying their timing of birth. Effectively, for the period 1995 to 2005, I 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. The evidence for delayed conception due to the announcement date are weak, but I address this possibility in the empirical section.

Two reforms taking place at the same time as the Canadian expansion in maternity leave deserve special attention [Baker and Milligan (2010)]. The first reform was the subsidized daycare program implemented in the province of Québec on September 1^{st} 1997. Child eligibility was phased in from September 1997 to September 2000, based on age. As a result, children born prior to the reform were not eligible from the start and therefore had a different daycare experience than children born after the reform. Furthermore, not all children were in subsidized daycare due to capacity constraints and the proportion of children in subsidized daycare grew during that period. Research on the subject shows that children post reform were spending longer hours in daycare and that, overall, the reform had negative impacts on the development of children [Lefebvre and Merrigan (2008), Baker et al. (2008)]. The authors also show that the reform significantly increased the labor force participation of women. This finding likely implies that the composition of mothers eligible to leave benefits in Québec changed during the period of the maternity leave reform.

The second reform is the National Child Benefit Program (NCBP) that effectively increased the benefits per child paid to low-income families participating in the labor force. Milligan and Stabile (2007) find that this reform had a significant effect on the labor supply of low-income families and their overall income. Impacts were strong on single-parent families, but extremely modest on two-parent families. Following Baker and Milligan (2010), I exclude children born in the province of Québec and focus on children raised in two parent families. These restrictions are necessary to avoid possible confounding effects generated by those reforms.

3 The data set

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. The samples are constructed through the Labor Force Survey. As a result, children living in the 3 territories, on native reserves or in remote regions are excluded as well as children of individuals living in institutions or from military families. For each of the cycles, representative samples of the population of children living in Canada's 10 provinces are constructed: both cross-sectional and longitudinal data are accessible[5]. The target population at the time of selection is children aged 0 to 11, but longitudinal children may have been followed as late as age 25. 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. I first provide a brief overview of the subpopulation of children I focus on. Then I present the outcome measures that I use to estimate the effect of the maternity leave reform. Detailed summary statistics on children are presented in the empirical section.

I use the panel data of the subpopulation of 0 to 4 years old contained in cycles 4 to 7, inclusively. Cycle 4 provides data on children born in 1999 and 2000 (prior to the reform). These children are later observed at age 4 to 5 in cycle 6. Cycle 5 provides data on children born in 2001 and 2002 (after the reform). These children are later observed at age 4 to 5 in cycle 7.[6] I further restrict the data set to children first observed at age 0 to 12 months. For

[5] Cross-sectional and longitudinal weights, adjusted for total non-response matching known population count, are provided.

[6] Children are also observed at age 2 or 3 in cycle 5 or 6.

children born prior to the reform, whether the mother claimed maternity leave EI income replacement or not can only be identified accurately for mothers of children aged 12 months or less.[7]

Although maternity leave benefits are known for mothers of all children born after the reform, I also impose the restriction of 0 to 12 months at the first interview to this group. This restriction is necessary because of the age structure of the NLSCY. Children born in even years (2000 and 2002) are generally first observed at age 0 (less than 13 months) and later observed at age 4 (48 to 60 months). Children born in odd years (1999 and 2001) are generally first observed at age 1 (13 to 24 months) and later observed at age 5 (60 to 72 months). The timing of the first and third interview depends, therefore, on the child's birth date. Figure 1 shows the age distribution for children interviewed for the first time in their first year of life (0 to 12 months). The Kolmogorov-Smirnov test of equality of distribution functions rejects the null hypothesis of equality of distribution. Children of all ages are present in both groups such that it will be possible to control for age at the time of test.

More generally, figure 2 shows the average age distribution at the time of the third interview by year-month of birth for all children born between January 1^{st} 1999 and December 31^{st} 2002.[8] This figure shows two important aspects of the design of the NLSCY. First, the age pattern declines in time within a cycle. Each of the cognitive test scores follow the same pattern as average age and this can be attributed to the high age sensitivity of the tests. Similar, but somewhat weaker patterns are also identified for all four behavioral measures. Child development at such an early age is extremely age sensitive. A naive linear regression in cognitive scores would suggest that children are becoming less cognitively able as time

[7] In cycle 4, questions related to maternity leave only refered to the past 12 months. As of cycle 5, these questions were asked in absolute terms (i.e. were you on paid maternity leave after stopping work?). Prior to the reform whether mothers claimed benefits or not is known only for mothers of children interviewed in the first 12 months of life. Post reform, it is known for all children.

[8] Children born outside of Québec and living in a two-parent household, but including children first interviewed at age 13 to 24 months.

passes, while it is actually the age structure of the survey that is driving this effect. This first finding highlights the importance of properly accounting for age at the time of test.

Second, from this figure, it is clear that children born just before December 31^{st} 2000 typically took the developmental tests at an early age 4 ($\approx$ 46 months), while children born just after took them at a late age 5 ($\approx$ 69 months). Since children born shortly after the policy change took their tests two years after children born shortly prior to the policy change it is not possible to follow some of the prior research mentioned earlier and use discontinuity design to rule out underlying trend effects. Trend effects can occur for a variety of reasons. For example, for one of the tests I am using (Peabody Picture Vocabulary-Revised) the calibration of the level of difficulty in each question was done some years ago and as the English (and French) language evolves, drift in scores can be expected. Statistics Canada conducted such an analysis in Cycle 4 and did find evidence of drift in the level of difficulty of some questions. The score of each child relative to another within the same cycle were however not found to be under or over estimated. In sum, accounting for age at the time of test is critical and, although discontinuity design is attractive and has been used by other researchers looking at the effects of similar policies, the structure of the NLSCY and the outcome variables I am interested in prevents its use.[9] Given these findings, from here on, I exclusively focus on children interviewed prior to age 13 months, born between 1999 and 2002 outside of Quebec and living in two-parent households.

Table 2 shows the summary statistics of the outcome measures. Child (and environment) outcomes are divided into four distinct groups: maternal time at home, cognitive development, social development, family and health. Maternal time at home includes two variables: maternal time at home in the first year of life, and in the first four years of life. Table 2 shows that mothers stay on average 8.9 months with their child in the first year of life and

[9] Baker and Milligan (2008b) use a before-after model to assess the impact of the reform on breastfeeding. They also used the NLSCY, but focus on outcome variables based on retrospective questions not dependant on the timing of the interview (e.g. have you ever breastfed this child?).

17 months in the first four years of life. This table includes children of both working mothers and stay-at-home mothers.

At age 4, 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 measure is widely used in the literature. 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. All tests are, however, age sensitive, with older children scoring higher on average.

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 in all four measures.

Since adverse family environment is detrimental to child development, and the reform was explicitly designed to help families, not just children, I also estimate the impact of the reform on two measures of the environment of the family when the child is 4. The family functioning score is based on 12 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). Finally, given the potential effect on health, I 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).

In summary, I focus on 4 year-old children born 2 years prior to and after the reform, raised in two-parent families outside the province of Québec because (1) cognitive test scores are only available starting at age 4 in the NLSCY and haven't yet been studied in relation to this policy, (2) on average, differences in ability measured that early persist until adulthood, and (3) a control group can be identified.

Different characteristics Table 1 shows the summary statistics. I observe a total of 1,243 children, representing a population of 299,118 children born outside Québec and living in a two-parent family throughout their early childhood. From left to right, children are divided into four subgroups: (1) children born before the reform, (2) children born after the reform whose mother claimed maternity leave benefits, (3) children born before the reform, and (4) children born after the reform whose mother did not claim maternity leave benefits.

Looking at the child characteristics, Table 1 reveals that children in all four groups share similar characteristics: they are on average 52 months old at the time of the third interview and between 8 to 9 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. However, Baker and Milligan (2008b) have shown that the reform had a positive effect on the length of time a mother continued breastfeeding. I 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, but possibly slightly lower for post reform children in the treatment group.

Family characteristics are also comparable except for the number of siblings in the family at the time of birth. 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. There are two measures of parenting style, both based on a series of questions asked to the parents: one measures parent effectiveness and the other consistent parenting. The family functioning and social support measures are also based on a series of parent reported measures. More details on these measures are provided below.

Mothers characteristics are generally less stable across all four groups, suggesting that there might have been some selection. Mothers of children born after the reform claiming maternity leave benefits are typically more educated. This may be due to selection or sample design. Since maternal education is strongly related with child development, it is key to control for this difference. Mothers receiving maternity leave benefits are typically not recent immigrants, with less than 3% having immigrated in the last 4 years. 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.

Father's characteristics (education and income quartile) are fairly comparable, but again fathers appear to be more educated for children born after the reform whose mother claimed income replacement. The empirical approach specifically adresses the possibility of self-selection into treatment.

Work beyond the first year Baker and Milligan (2008a) find that longer leaves increase the probability that a women returns to work after the birth. This finding suggests that the reform is two fold: (1) mothers spend more time at home in the first year, and (2) they work more beyond the first year. Beyond the first year of infancy the estimated impacts of maternal labor supply on child development using multivariate analysis are mixed. Ruhm (2004) and Belsky and Eggebeen (1991) find that maternal employment has negative impacts, Desai et al. (1989) find no effects, and Blau and Grossberg (1992) and Brooks-Gunn et al. (2002a) find that work in the second and third year of life may have positive impacts on child development. Loeb et al. (2007) finds that the optimal entry age in center-based care is 2 to 3 years old, and that this type of care raises cognitive scores at age 5, but lowers behavioral scores. Belsky et al. (2007) also find that center-based care is associated with teacher reported behavioral problems.

Looking at work patterns of mothers who received maternity leave benefits at the time of the second interview (when the child is 2 to 3 years old) and the third interview (when the child is 4 to 5 years old), I find that mothers have increased their probability of working for most of the year full-time when the child is 4 to 5 years old, but also increased their probability of being full-time at home. The probability of working for most of the year part-time offset these two effects.[10] I find, however, no effects on work when the child is 2 to 3 years old.

In sum, it appears that their might have been a slight increase in the labor supply of mothers at age 4 to 5. Depending on the relationship between work at that age and child development the effect of maternal time investment may be slightly over or under estimated. Bias should be minor given the small increase and the late timing. The empirical approach addresses this possibility.

Type of care displaced The effect of the policy depends on the type of care displaced. Brooks-Gunn et al. (2010), using a multivariate approach, showed that full-time maternal

[10] Brooks-Gunn et al. (2010) find that part-time employment is positively linked to the behavioral development of the child, through the offsetting effects of greater maternal sensitivity and a better home environment.

employment in the first year was associated with lower cognitive test scores (although not all test scores)[11], but these effects were offset by greater maternal sensitivity and the use of center-based care. The NLSCY doesn't provide direct measures of maternal sensitivity, but does offer information on the type and quality of care used.

Table 3 shows the type of care used prior to the reform by mothers claiming maternity leave benefits in the child's first year of life. Only children born prior to the reform whose mother had already returned to work at the time of the first interview and who claimed maternity leave benefits are used to compute these statistics. I find that formal daycare was extremely rare (only 9%). About 36% of children were taken care of by a non relative in a family environment, either at the care provider's home (29%) or at the child's home (7%). A similar proportion (30%) were being taken care of by a relative. Surprisingly, 34% of children were in no formal type of care even if the mother had returned to work. A number of reasons can explain this fairly high number: the parent has not yet found its primary mode of care and claims using no care, the parent works from home and cares for the child at the same time, the mother and father work on different shifts or the father is not working and cares for the child.

This table also reveals that a small fraction of children (21%) are being taken care of by individuals who have been trained to care for young children and a similar fraction (20%) have a license. 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, on average 31 hours with a standard deviation of 17 hours. Even if all of these children are not older than 12 months old, many have already changed mode of care at least once, the average number of care providers being 1.7.

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

[11] 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.

spent with the mother versus time spent in informal care at a fairly high intensity (more than 30 hours per week).

4 Empirical strategy

My empirical approach differs from previous research. First, in contrast to Baker and Milligan (2010) and for the first time using the NLSCY, I identify children of mothers who claimed maternity leave benefits. As mentioned above, this question is precisely asked at the time of the first interview and no inference is required. Children in the treated group had mothers who claimed maternity leave benefits. Children in the control group had mothers who did not claim maternity leave benefits, either because they had never worked prior to birth, or they had not worked enough, or they were self-employed. For the first time, common trends impacting the outcome measures can be controlled for.

Second, my empirical strategy does not rely exclusively on children born around the discontinuity point. Buckles and Hungerman (2008) show that timing of birth is related to socioeconomic status, with a higher proportion of high-socioeconomic status mothers having babies in the summer. Discontinuity design estimates the impact of those born around the discontinuity point and can therefore be biased by mothers' self selection with respect to timing of birth. For example, in Carneiro et al. (2010), discontinuity design may have given more weight to children from high-socioeconomic background since the effective date of the Norwegian reform was July 1^{st}, 1977. In this case, the estimated impacts were most likely downward biased since larger effects were found on children born in less favored environments. Using the NLSCY, Baker and Milligan (2010) compare children born four years prior to and four years after the reform. I restrict my attention to a narrower time window to limit the impact of changes in maternal behavior. Effectively I compare children born in 2000 with children born in 2002. This has the benefit of reducing serial correlation in outcome variables.

Comparing the results of discontinuity design with my chosen approach to provide further input into Buckles and Hungerman (2008) findings is frustrated by the structure of the NLSCY. As shown earlier in Figure 2, children born just before December 31^{st} 2000 typically took the developmental tests at an early age 4, while children born just after took them at a late age 5. Age differential can possibly be accounted for through age-standardization of the outcome measures. However, unobserved shocks impacting the outcome variables (e.g. drift in overall score) occurring within this two year window are empirically more challenging.

Effectively, I implement the following difference-in-differences model:

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

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 was eligible to paid maternity leave (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 the drift in the PPVT score do not impact the consistency of the DID estimator, but uncommon shocks, such as the NCBP cannot be accounted for. 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. Assumption (2) implies that DID cannot accommodate for unobserved individual-specific transitory shocks influencing the participation decision. In the present context, divorce at the time of birth may influence the mother's participation decision and the development of the child.

Since I focus on intact families in the first 4 years of life, the NCBP and divorce examples do not jeopardize my results.[12]

To this simple model, two modifications can be made. First, I add age dummies at the time of test to account for the age sensitivity of certain outcomes, from 47 to 60 months old. Second, I also include a number of control variables to account for maternal, paternal, child and family characteristics. Equation 1 becomes:

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

where δ_a represents the age specific effect at the time of test ($a = 47, 48, 49, ..., 60$), $\mathbf{X}_i$,is a vector of maternal, paternal, child and family characteristics measured at the time of birth (or no later than the first interview), and $\mathbf{\Phi}$ is a vector of parameters. Standard DID assumes that $y_{i,by+t}$ is linear in $\mathbf{X}_i$, such that the estimated response to the reform β is also linear in $\mathbf{X}_i$.

In the present application, controls mainly include children of non working mothers (or stay-at-home mothers) and self-employed mothers whose response to the reform and to underlying trends may be very different from children of working and eligible mothers even if they share the same characteristics. Table 1 showed that children in the control and treated group had fairly different characteristics. If the response is non linear with respect to $\mathbf{X}_i$, then the standard DID approach provides biased estimates.

Furthermore, there may be compositional change between the control and the treated group. First, children of the same age prior to and after the reform are typically observed two years apart such that mothers would have had enough time to react to the reform.

[12] Donald and Lang (2007) suggested an approach to account for the small number of clusters and the possibility of unobserved cluster effects. In the present context, with only four clusters, this approach cannot be implemented. The bootstrap standard errors of the standard DID are already large enough however to deem any effects significant.

However, maternal employment trends of Canadian women (excluding Québec) with children below age 5 were generally stable around the maternity leave reform [Lefebvre and Merrigan (2008)].[13] Second, since the number of hours required for eligibility is lower post reform, mothers composition is likely different. Table 1 shows that the proportion of eligible mothers is indeed higher post reform (56% to 60% which corresponds more or less to the reported statistics from Statistics Canada's Survey of Employment Insurance Coverage in Baker and Milligan (2010)).

To address the possibility of compositional change and heterogeneity of response given $\mathbf{X}_i$, I also implement the matching difference-in-differences (MDID) estimator suggested by Heckman et al. (1997) and Heckman et al. (1998). With repeated-cross sections, the MDID estimator is [Blundell and Dias (2009)]:

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

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 $\mathbf{X}$ non-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 $\mathbf{X}$. This estimator also ensures group comparability prior to and after the reform and therefore limits the impact of compositional change on the outcome variables.

[13] 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) .

Effectively, I first estimate a probit model in which the dependant variable equals one if the mother claimed maternity leave benefits 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, I predict the propensity score of each child (or the probability that the child's mother claimed benefits or not). 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 $\mathbf{X}$. I implement kernel matching, local linear regression matching and nearest neighbor matching. Bootstrap standard errors are calculated for local linear regression and kernel matching to account for the underlying matching procedure (not consistent for nearest neighbor).

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, my preferred models (specification 5 and 6 below) include all of the variables I had access to that were measured at birth and likely influenced the child outcome and the decision of the mother. More specifically, I include the following set of variables measured at the time of birth: gender, prematurity, birth weight, multiple birth indicator, hospitalization at birth, breast-feeding at birth, age at which the child slept 6 continuous hours at night, marital status, number of siblings, maternal and paternal education, age of the mother at first child, age of the mother at birth, immigration status of the mother in the last 4 years, paternal income quartile, and also measures of the family functioning, social support and parenting skills at the time of the first interview. 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) 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. In some specifications, I also include the average provincial unemployment rate in the year preceding birth to better grasp the behavior of mothers.[14] Finally, since cognitive measures (and possibly behavioral measures) are sensitive to the age at which the child took the test, I 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.

In sum, I first implement the standard DID estimator. To account for heterogeneity of response and compositional change I also implement the matching DID estimator. Both estimators assess the impact of the reform on the treated only.

5 Estimated intention-to-treat effects of the reform

Table 5 presents the empirical results of the standard DID estimator (equation 1 and 2) and the MDID estimator (equation 3). From top to bottom, the first panel 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. Controls and matching variables are listed in the bottom panel and are further detailed in the table footnote. All measures have been converted such that a positive coefficient indicates a positive impact on the child or its family. The treated groups

[14] Using unemployment rate (3 months rolling average) at birth and 3, 6 and 9 months prior to birth instead of the annual average leads to the same conclusion.

(before and after the reform) are always restricted to children of mothers claiming maternity leave benefits, while the control groups include all other children.

The first three columns (specification 1 to 3) present the estimated impact using standard DID. The last five columns (specification 4 to 8) present the estimated impact using MDID. I present seven different specifications to show the robustness/sensitivity of the results. Specification 1 (first column) presents the DID estimates without controls. Specification 2 reports DID estimates with age at the time of test dummies, while specification 3 includes all controls except for unemployment rates prior to birth.

Specification 4 shows the MDID estimates when matching is performed exclusively on age at the time of test, and is therefore comparable to the second specification for DID estimates. Estimates of specification 5 rely on matching on all covariates except for unemployment rate, while specification 6 also includes unemployment rate. General equilibrium effects on the labor market due to the reform may lead to biased estimates when unemployment rate is included. However, I find that (1) results of both specifications are very similar, and (2) the percentage of women who would have been in the labor market prior to the reform and are now out of the labor market is fairly small (less than 1% of the active labor force) such that general equilibrium effects are unlikely. Specification 7 excludes parenting style and family functioning measures. Specification 8 also excludes paternal income quartile and age of the child when he/she slept through the night. Excluded measures in specifications 7 and 8 are taken shortly after birth, when the child is on average 8 months old. Since prior to the reform, mothers claiming maternity leave benefits were already taking on average more than 7 months at home, these measures were included as a proxy for family environment and wealth at the time of birth. However, it is possible that these measures were themselves impacted by the reform. For example, fathers may have changed jobs or worked longer hours to better support the family. Changing income quartiles in the short run remains fairly unlikely. In this case, estimates controlling for these measures would be downward biased. Results suggest that this was not the case.

I present the MDID estimates using nearest neighbors matching with 5 neighbors because these results exhibit the strongest consistency in terms of balancing property as defined by Rubin (2001). Balancing properties are further discussed after presentation of the estimated impacts on the various outcome measures.

Maternal time at home Figure 3 shows the discontinuity in time at home prior to returning to work. The top panel shows the time mothers who did not claim maternity leave benefits spent at home in the first year of life. The bottom panel shows the time mothers who claimed maternity leave benefits spent at home. The red line marks the timing of the reform. 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 bottom panel also suggests that benefit recipient mothers were spending more than 6 months at home prior to the reform (about 7.1 months). If mothers had increased their time at home by 6 months, this would have implied that post reform they were staying at home 13.1 months. For this reason, I estimate the impact of the reform not only in the first 12 months of life, but also in the first 4 years of life (48 months).

Table 5 shows the estimated effects of the reform on the time mothers stayed at home (before returning to work). In the first year of life, the DID estimator suggest that eligible mothers increased their time away from work by 2.8 to 2.9 months, while the MDID estimator finds an impact of 1.2 to 2.1 months. The DID estimates are comparable to Baker and Milligan (2010). In the first 4 years of life, the DID estimator suggests an impact of 1.7 to 2.9 months while MDID suggests an impact of 1.1 to 3.7 months. The more modest impact of 1.1 month is found when parenting style and family functioning are excluded from the model. Controlling for these characteristics appears essential to identify mothers whose behavior changed following the reform. Given these results, from here on I focus my attention mainly on specifications 5 and 6 in which the estimated reform impact averages 4 months. A positive effect on children's outcome variables would thus indicate that spending 4 more months with the mother in the critical period of 7 to 11 months is beneficial.

This finding highlight one important feature of the reform: not all eligible mothers increased their time at home by the full 25 weeks. In general, prior to the reform, I find that about 13% of eligible mothers were already taking more than 12 months and about 53% were taking short leaves of 6 months or less. Post reform, I find that 11% continued to take short leave of 6 months or less even if they were now eligible to 50 weeks of paid maternity leave. These mothers did not react to the reform. Mothers taking longer leaves (more than 18 months) prior to the reform may have continued to do so. Findings mentioned above on work beyond the first year suggests that some may also have re-entered the labor market faster. The higher bound estimate for the proportion of mothers reacting to the reform is thus 89%.

Excluding the bottom 11% in time spent at home prior to and after the reform, I find that mothers whose behavior changed were more educated and had patterns whose income at birth was higher. These mothers were also more likely to have no other children, be married, and to not work full-time the entire year when the child is 2 to 3 years old. Interpretation of the results should be in light of these findings: estimated effects are on children born in, arguably, more favorable family environments.

To get the estimated impact of being eligible (in the sense that the mother claimed maternity leave benefits) and actually reacting to the reform, estimated effects need to be scaled by 1.12 (1/0.89). For briefness and to further highlight the average effects of the reform on children of all mothers claiming maternity leave benefits, results to follow have not been scaled by 1.12. In other words, I present the intention-to-treat effects as opposed to the treatment effect on the treated.

Cognitive development The second panel of Table 5 presents the estimated impact of the policy on the PPVT, Who Am I? and Number knowledge scores. The DID estimator suggests that the reform did not have an impact on cognitive development since the coefficients are relatively small and the bootstrap standard errors are large.

The MDID estimators lead to different findings. In all specifications, the coefficients are positive and relatively large. Specification 5 suggests that the reform had an impact of 2.9 (or 17% of a std. dev.) on the PPVT, 1.1 (or 19% of a std. dev.) on the Who Am I? and 1.0 (or 21% of a std. dev.) on the Number knowledge. The results on all three tests are significant and of comparable magnitude. The results are similar for specification 6, but somewhat larger: 14%, 26% and 30% of a standard deviation for the PPVT, Who Am I? and Number knowledge measures, respectively. When only age is used for matching, the results are mixed: larger and strongly significant for the PPVT, similar for the Who Am I?, and smaller and not significant for the Number knowledge. Finally, results not estimated using parenting style, family functioning and paternal income quartile at birth are generally comparable, but weaker for the Number knowledge score. It is important to remember here that the effect on time at home was not significant for specifications 7 and 8. These findings suggest that response to the reform is non linear and controlling for a complete set of individual and family characteristics is important.

In sum, it appears that the reform had comparable positive effects of around 20% of a standard deviation across all three cognitive tests. These estimates were obtained using raw scores. Extremely similar findings are obtained using age standardized scores. 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.

Social development The third panel of Table 5 shows the estimated effects on behavioral measures. For all four measures of social development, the sign of the coefficients have been adjusted such that a larger score indicates better behavioral development. The DID estimates suggest that the reform did not have a significant impact on child social development. The signs of the coefficient suggests a positive impact on hyperactivity and indirect aggression and a negative impact on child conduct. The effects on emotional development are mixed.

The MDID estimates provide a more definitive picture. Focusing on specification 5, I find that the reform had a positive and weakly significant impact on hyperactivity (0.44 or 17% of a std. dev.) and a negative impact on indirect aggression (-0.21 or 16% of a std dev.). Effects on emotional and conduct disorder are not significant. Looking across specifications 4 to 8, I find that these findings generally persist.

It appears that the reform had some impact on child social development, mainly through increased indirect aggression (e.g. higher likelihood when mad at someone to influence others to dislike that person) and reduced hyperactivity (e.g. increased ability to concentrate on a task). The impacts on emotional development and conduct behavior are less clear, and generally not significant.

Family and health The fourth panel of Table 5 shows the estimated effects on the environment of the child at age 4 and on parent-reported child health measures. Focusing on the MDID estimates, I find that children's health reported by the parent improved according to both measures of health. The effect on general health ranges from 0.07 to 0.19 (or 16% to 45% of a std. dev.) and from 0.17 to 0.41 (or 25% to 60% of a std. dev.) on health recently.

Health related benefits may be due to increased length of breastfeeding as reported by Baker and Milligan (2008b). It may also be explained by reduced hyperactivity if parents associate poor health with hyperactivity. Reducing hyperactivity in young children has benefits in the short run in terms of better social behavior, improved health, and by extension drug use reduction and lower costs to the health care system. It also has long run benefits, since links between medication for hyperactivity and drug dependence in the long run have been found [Dafny and Yang (2006)].

Finally, Table 5 shows that the reform positively impacted family functioning. The impact ranges from 0.99 to 1.14 (or 19% to 22% of a std. dev.) and is significant. A few channels may explain this finding. 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.

In sum, maternity leave expansion had large and significant effects on child health, which can possibly be explained by both increased breastfeeding and reduced hyperactivity. Family functioning also benefited from the reform.

Balancing property Figure 4 shows the propensity score (pscore) distribution prior to (left panel) and after (right panel) matching for the estimated effects on the PPVT score. 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.

Rubin (2001) identifies three distributional conditions that must be met simultaneously. For the MDID estimator with repeated cross-sections, for each of the three groups (treated before treatment, and both control groups): (1) the difference between the mean propensity score with that of the treated after treatment group should be small, (2) the variance of the propensity score distribution should be similar to that of the treated after treatment, such that the variance ratio should be close to one, and (3) the variance of the residuals of each of the matching covariates should be similar (i.e. the ratio of the variance should be close to one). Table 4 provides the distributional conditions for the PPVT outcome, not only for nearest neighbor, but also for local linear regression and kernel matching.

Table 4 shows that the conditions with nearest neighbor matching are strongly met. More specifically, the mean differences are extremely close to zero, the propensity score variance ratios are nearly equal to one, and 93% 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, but a smaller proportion (90%) of the covariates are within the 0.8

to 1.2 bound. The conditions also find support for kernel matching, but the results for the variance of the residuals are less satisfactory. Similar balancing property results are found for all of the other outcome measures investigated.

The empirical results to follow are robust across all three techniques, but I focus on the results using nearest neighbor matching with 5 neighbors (Table 5). Estimates using local linear regression matching are presented in Table 6 and discussed in the robustness section below. Results are also robust to nearest neighbor with 4, 3, 2 and 1 neighbors. However, with more than 1 neighbor, estimates are more efficient and the sign of the coefficient remains the same.

Further robustness check To further confirm the above results, a variety of robustness checks were constructed. As previously discussed, specifications 4 to 8 were also implemented using local linear regression matching. Results are reported in the last five columns of Table 6. The impacts are generally more significant using this approach and in line with previous results.

I also implemented four variations of Table 5's specification 5. First, a very small number of observations in both control groups have propensity scores above 80% (Figure 4). To ensure that the results are not driven by these few observations, specification 5 is implemented while excluding those observations (specification 9, Table 6). Second, since the announcement date was more than 9 months prior to the effective date, specification 5 is implemented while excluding children born in November and December 2001 (specification 10, Table 6). Third, since maternal employment past the first year may have been influenced by the reform, specification 5 is implemented while including maternal work pattern at the time of the second and third interviews in the set of matching variables (specification 11, Table 6). Results are generally robust to these variations. Fourth, to validate that the effects are due to the reform and not to some underlying trend affecting only eligible mothers, specification 5 was implemented on children born post reform (from 2001 to 2004). Estimated impacts are generally not significant, or even show a reverse pattern if they are significant (specification 12, Table 6).

Finally, I also construct different control and treatment groups and include children age 5 at the time of the third interview. Children of mothers who worked prior to birth form the treatment group and children of mothers who did not form the control group. Again, for children born prior to the reform and interviewed beyond 12 months of age (generally born in 1999), I do not have information prior to birth. However, using work and income related questions over the past 12 months, I can identify whether the mother was likely to have worked prior to birth. Although this is not perfect, this increases the sample size from 1,243 to 3,897 observations (or from 299,118 to 828,641 weighted children) and allows me to estimate the impact of the reform not only on 4 year old children but also on 5 year old children. Post reform, maternal work before birth is identified for all children.

This sample warrants another point of caution. Mothers who work include mothers who were self-employed and therefore not eligible to income replacement along with mothers who did not work enough to be eligible. Working mothers are compared with non working mothers. Any trend specifically related to children of working mothers wouldn't be accounted for here since they are all grouped together.

These results are shown in Table 7. The table structure is exactly the same as that of Table 5. The MDID estimates for this group are extremely comparable to the more accurate but smaller 4 year-old group. Impacts on behavioral development are somewhat different: I no longer find significant and positive effects on hyperactivity and I find positive and significant effects on conduct behavior (physical aggression). The DID estimates are generally more in line with the MDID estimates, but still remain largely insignificant.

6 Conclusion

I investigate the effect of maternity leave expansion in Canada that formally increased paid maternity leave from 6 to 12 months on December 31^{st}, 2000. The magnitude of the reform and the recent timing of the reform are incomparable to reforms in other countries. This is also the first study that can control for underlying trend effects by accurately identifying children of mothers who claimed paid maternity leave benefits from children of mothers who did not. Finally, maternity leave reform effects on children age 4 to 5 had never been documented before, while the literature clearly shows that ability identified that early is a strong predictor of future adulthood socioeconomic success.

I find that the reform had a significant effect on the time a mother stays at home in the first year of about 3 months more. The type of care displaced was mainly informal and provided by individuals not specifically trained to care for young children. The MDID estimates suggest that maternal time at home as opposed to informal care is preferable on a variety of measures. Significant and positive effects on the order of 20% of a standard deviation are found on all three cognitive measures available for children aged 4. I also find that children whose mother benefited from extended maternal leave benefits are less hyperactive, but possibly more prone to indirectly behave in a forceful way (e.g. when mad at someone, influence others to dislike that person). Parents of these children report that their child is significantly more healthy. Two documented mechanisms may explain this last finding: increased breastfeeding and reduced hyperactivity. Finally, family functioning also improves following the reform, which suggests that parents are now better able to balance family and work demands. The estimated effects on child development depend on the type of care displaced. It may be possible that countries where no effects were found had better daycare in place.

This research could be extended in a number of ways. First, although I did find positive effects, this does not imply that this was the most efficient approach to generate these effects. A deeper understanding of the return on investment including not only child benefits, but also parent related benefits (and or losses) would bring considerable value.[15]

Second, mothers who benefited from and reacted to the reform were generally more educated, had fewer children and were less likely to have immigrated recently. Also, all children observed in this study were raised in two-parent families outside of Québec. The estimated impacts relate to these children only and may not hold true for all other children. Further research documenting the impact of the reform on children of single parent families would be valuable.

Third, I find that low educated mothers, most likely also low income mothers, had a lower take-up rate even if they claimed maternity leave benefits. Preliminary research by Carneiro et al. (2010) suggests that children of low educated mothers benefited the most from the Norwegian maternity leave reform, a combination of increased maternal time investment and possibly disposable income. Assuming the same holds true in Canada, increasing the income replacement rate of low income mothers might increase their take up rate and in turn the benefits to their children. 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. A more generous maternity leave policy for disavantaged families might increase their take-up rate, which would favor child development and help reduce the rising skill bias, most likely also prevalent in Canada.

7 References

Baker, Michael and Kevin Milligan, "How Does Job-Protected Maternity Leave Affect Mothers' Employment?," *Journal of Labor Economics*, 2008, *26*, 655–691.

[15] Canadian federal expenditures on maternal and parental benefits increased from 1.25 billion dollars to 2.15 billion between 2000/01 and 2001/02.

_ **and** _ , “Maternal employment, breastfeeding, and health: Evidence from maternity mandates,” *Journal of Health Economics*, 2008, *27*, 871–887.

_ **and** _ , “Evidence from Maternity Leave Expansions of the Impact of Maternal Care on Early Child Development,” *Journal of Human Resources*, 2010, *45* (1), 1–32.

_ **, Jonathan Gruber, and Kevin Milligan**, “Universal Child Care, Maternal Labor Supply, and Family Well-Being,” *Journal of Political Economy*, August 2008, *116* (4), 709–745.

Baydar, Nazli and Jeanne Brooks-Gunn, “Effects of Maternal Employment and Child-Care Arrangements on Preschoolers’ Cognitive and Behavioral Outcomes: Evidence from the Children of the National Longitudinal Survey of Youth.,” *Developmental Psychology*, 1991, *27*, 932–945.

Belsky, Jay and David Eggebeen, “Early and Extensive Maternal Employment and Young Children’s Socioemotional Development: Children of the National Longitudinal Survey of Youth,” *Journal of Marriage and Family*, 1991, *53* (4), 1083–1098.

_ **, Margaret Burchinal, Kathleen McCartney, Deborah Lowe Vandell, K. Allison Clarke-Stewart, and Margaret Tresch Owen**, “Are There Long-Term Effects of Early Child Care?,” *Child Development*, 2007, *78*, 681–701.

Blau, Francine D. and Adam J. Grossberg, “Maternal Labor Supply and Children’s Cognitive Development,” *The Review of Economics and Statistics*, August 1992, *74* (3), 474–481.

Blundell, Richard and Monica Costa Dias, “Alternative Approaches to Evaluation in Empirical Microeconomics,” *Journal of Human Resources*, 2009, *44* (3), 565–640.

Brooks-Gunn, Jeanne, Wen-Jui Han, and Jane Waldfogel, “The effects of early maternal employment on child cognitive development,” *Demography*, 2002, *39*, 369–392.

_ , _ , and _ , "Maternal Employment and Child Cognitive Outcomes in the First Three Years of Life: The NICHD Study of Early Child Care," *Child Development*, 2002, pp. 1052–1072.

_ , _ , and _ , "First-year maternal employment and child development in the first 7 years," *Monographs of the Society for Research in Child Development*, 2010, *75*, 1–147.

Buckles, Kasey and Daniel M. Hungerman, "Season of Birth and Later Outcomes: Old Questions, New Answers," NBER Working Papers 14573, National Bureau of Economic Research Inc, December 2008.

Carneiro, Pedro, Katrine V. Loken, and Kjell G. Salvanes, "A Flying Start? Long Term Consequences of Maternal Time Investments in Children During Their First Year of Life," IZA Discussion Papers 5362, Institute for the Study of Labor (IZA) Dec 2010.

Dafny, Nachum and Pamela B. Yang, "The role of age, genotype, sex, and route of acute and chronic administration of methylphenidate: A review of its locomotor effects," *Brain Research Bulletin*, 2006, *68* (6), 393–405.

Desai, Sonalde, P. L. Chase-Lansdale, and Robert Michael, "Mother or Market? Effects of Maternal Employment on Cognitive Development of Four-year-old Children," *Demography*, 1989, *26* (88-11), 545–561.

Donald, Stephen G. and Kevin Lang, "Inference with difference-in-differences and other panel data," *The Review of Economics and Statistics*, 2007, *89*, 221–233.

Duncan, Greg J., Katherine A. Magnuson, and Jens Ludwig, "The endogeneity problem in developmental studies," *Research in human development*, 2004, pp. 59–80.

Dustmann, Christian and Uta Schönberg, "The Effect of Expansions in Maternity Leave Coverage on Children's Long-Term Outcomes," Discussion Paper No. 3605, University College London July 2008.

Heckman, James and Salvador Navarro-Lozano, "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models," *Review of Economics and Statistics*, 2004, *86*, 30–57.

Heckman, James J., "Policies to Foster Human Capital," *Research in Economics*, 2000, *54* (1), 3–56.

_ , "Schools, Skills, And Synapses," *Economic Inquiry*, 07 2008, *46* (3), 289–324.

_ **, Hidehiko Ichimura, and Petra E. Todd**, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 1997, *64* (4), 605–654.

Heckman, James J, Hidehiko Ichimura, and Petra Todd, "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, April 1998, *65* (2), 261–294.

HRSDC, "Summative Evaluation of EI Parental Benefits," Technical Report SP-AH-674-01-05E, Human Resources and Skills Development Canada, Government of Canada, January 2005.

Knudsen, Eric, James J. Heckman, Judy Cameron, and Jack P. Shonkoff, "Economic, Neurobiological and Behavioral Perspectives on Building AmericaÃćâĆňâĎćs Future Workforce," *World Economics*, July 2006, *7* (3), 17–41.

Lefebvre, Pierre and Philip Merrigan, "Child-Care Policy and the Labor Supply of Mothers with Young Children: A Natural Experiment from Canada," *Journal of Labor Economics*, July 2008, *26* (3), 519–548.

Liu, Qian and Oskar Nordstrom Skans, "The Duration of Paid Parental Leave and Children's Scholastic Performance," *The B.E. Journal of Economic Analysis & Policy*, 2010, *10*, 1–33.

Loeb, Susanna, Margaret Bridges, Daphna Bassok, Bruce Fuller, and Russell W. Rumberger, "How much is too much? The influence of preschool centers on children's social and cognitive development," *Economics of Education Review*, February 2007, *26* (1), 52–66.

Main, Mary, *Touch: The foundation of experience* Clinical infant reports, Madison, CT, US: International Universities Press,

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

Milligan, Kevin and Mark Stabile, "The integration of child tax credits and welfare: Evidence from the Canadian National Child Benefit program," *Journal of Public Economics*, 2007, *91* (1-2), 305–326.

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

Rosenbaum, Paul R. and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 1983, *70* (1), 41–55.

Rubin, Donald B., "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation," *Health Services & Outcomes Research Methodology*, 2001, *2*, 169–188.

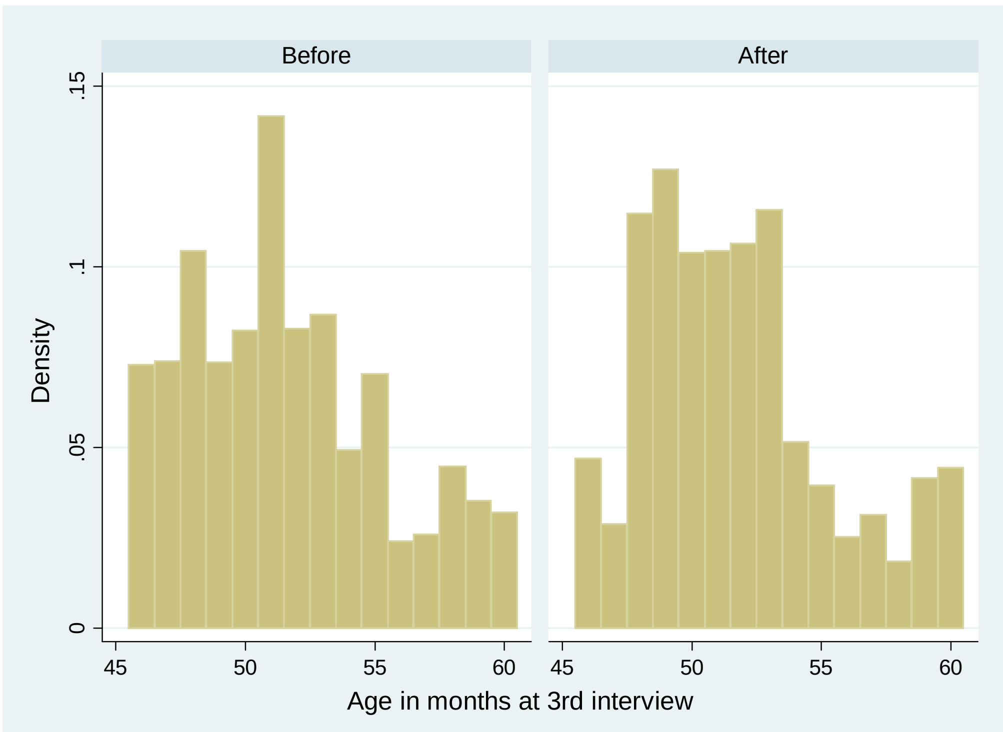
Ruhm, Christopher J., "Parental Employment and Child Cognitive Development," *Journal of Human Resources*, 2004, *39* (1), 155–192.

Shore, Rima, *Rethinking the Brain*, New York: Families and Work Institute, 1997.

Stuart, Elizabeth A., "Matching Methods for Causal Inference: A Review and a Look Forward," *Statistical Science*, 2010, *25*, 1–21.

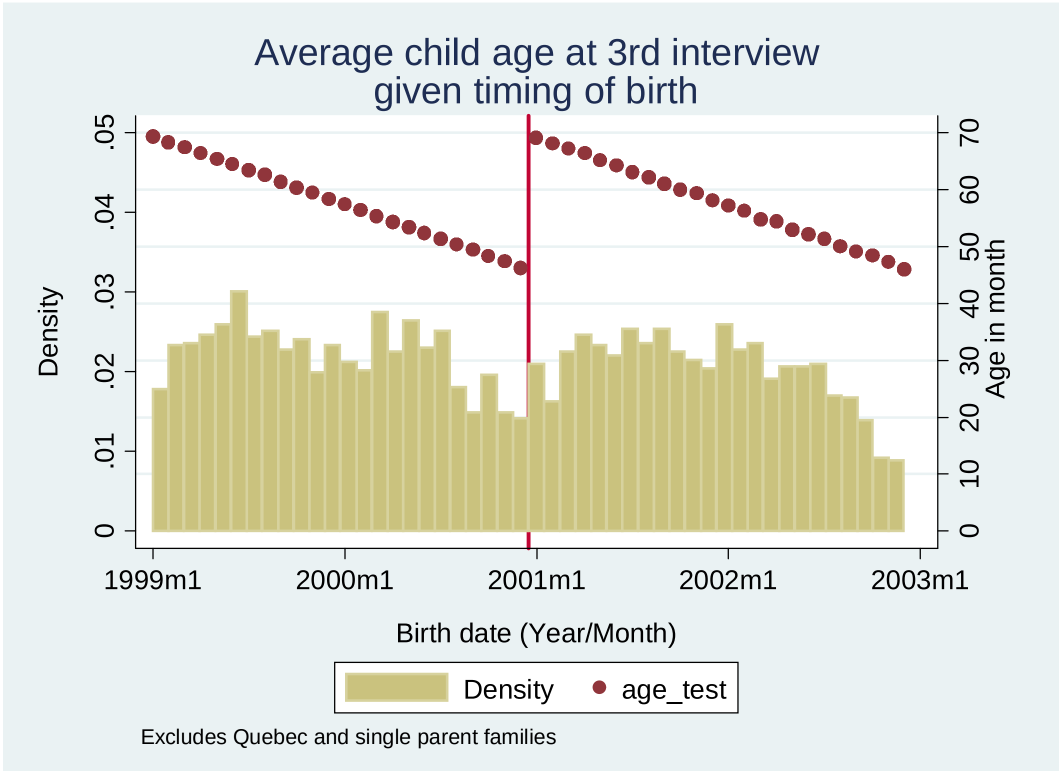
Vandell, Deborah Lowe and Janaki Ramanan, "Effects of Early and Recent Maternal Employment on Children from Low-Income Families," *Child Development*, 1992, *63* (4), 938–949.

Figure 1: Age distribution at the time of test



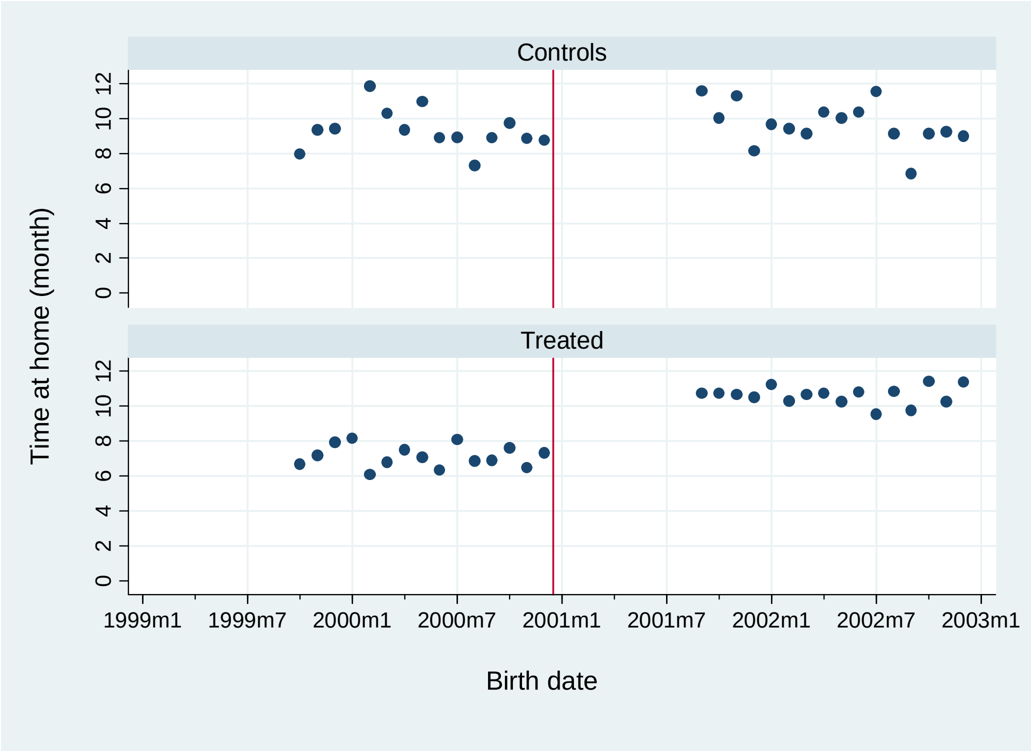
Note: Shows the average age in months at the time of the third interview before (left) and after (right) the reform.

Figure 2: Age distribution at the time of test by birth date



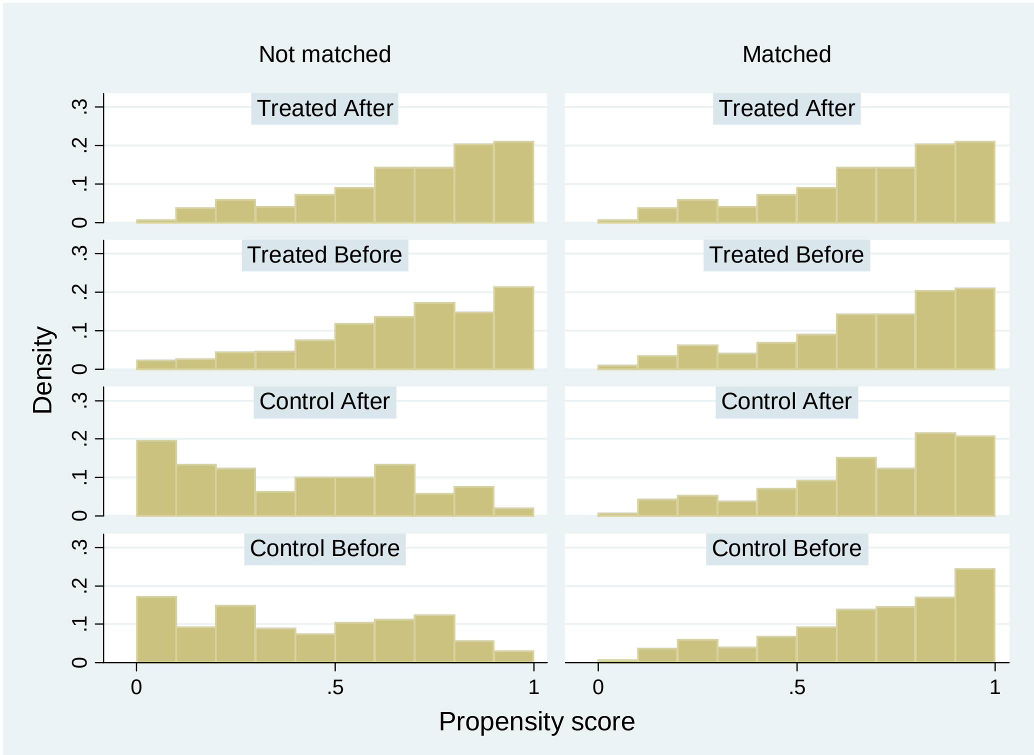
Note: Shows the average age in months at the time of the third interview and the density by month of birth. The vertical line indicates the timing of the reform. Children born just before the reform are on average 46 months old at the time of the third interview, while children born just after the reform are on average 69 months old.

Figure 3: 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 "Controls" includes mothers not claiming maternity leave EI income replacement. The bottom graph "Treated" includes only mothers who claimed maternity leave EI income replacement.

Figure 4: Pscore by work group



Note: Shows the propensity score distribution by groups before matching (left panel) and after matching (right panel). Matching is performed using nearest neighbor with 5 neighbors. The treated group includes children whose mother claimed maternity leave benefits. The control group include children whose mother did not claim maternity leave benefits. Groups labeled "Before" include children born prior to the reform. Groups labeled "After" include children born after the reform. The propensity scores are estimated using a probit model. All variables presented in Table 1 are included in the model. Dummies for province of residence and area of residence are also included. This specification corresponds to specification 5 in Table 5 presented later.

Table 1: Summary statistics at first interview

	Treated		Control	
	Before	**After**	**Before**	**After**
Child				
Age at 3rd interview (months)	51.71	51.91	51.58	51.57
Age at 1st interview (months)	8.09	8.90	8.29	8.63
Male	0.51	0.47	0.49	0.53
Premature	0.10	0.12	0.10	0.11
Low birth weight	0.07	0.04	0.05	0.03
Multiple births	0.05	0.03	na	na
Hospitalization at birth	0.23	0.21	0.18	0.10
Breastfed at birth	0.88	0.86	0.83	0.85
Sleep (age in months)	4.67	4.12	4.89	5.01
Family				
Married parents	0.90	0.90	0.89	0.86
No sibling	0.51	0.53	0.33	0.33
One sibling	0.36	0.40	0.36	0.40
Positive parenting	18.07	18.62	17.79	18.04
Ineffective parenting	1.36	1.63	1.37	1.39
Family functioning	8.51	9.38	8.73	8.62
Social support	18.52	18.32	17.98	18.43
Mother				
Education				
Less than college	0.42	0.29	0.59	0.50
College degree	0.31	0.28	0.14	0.21
University degree	0.26	0.42	0.26	0.24
Immigration				
0-4 year	0.03	0.03	0.17	0.15
5-9 years	0.06	0.07	0.12	0.09
Age				
at birth	29.94	30.60	29.26	30.19
at 1st baby	27.34	28.61	25.99	27.63
Father				
Education				
Less than college	0.43	0.34	0.41	0.45
College degree	0.33	0.32	0.29	0.22
University degree	0.21	0.31	0.29	0.25
Income quartile				
1st Q	0.19	0.21	0.27	0.29
2nd Q	0.29	0.26	0.27	0.17
3rd Q	0.23	0.26	0.17	0.23
4th Q	0.26	0.25	0.25	0.29
Number of weigthed children	75,037	86,102	74,555	63,424
Number of observations	369	351	291	232

Note: Shows the summary statistics of longitudinal children born between January 1^{st} 1999 and December 31^{st} 2002 outside of the province of Québec and living in a two-parent family throughout their early childhood. All variables are dummies, except for all age variables (including sleep age in months) and positive parenting, ineffective parenting, family functioning and social support. For positive parenting and social support, a greater score indicates a better outcome for the family. For ineffective parenting and family functioning, a lower score indicates a better outcome for the family.

Table 2: Outcome summary statistics

	Mean	Std.Dev.
Maternal time at home		
in the first year of life	8.9	3.6
in the first 4 years of life	17.0	16.2
Cognitive development		
PPVT	51.0	17.4
Who Am I?	21.0	5.9
Number knowledge	9.7	4.5
Social development		
Hyperactivity	4.1	2.6
Emotional disorder	1.8	1.9
Conduct disorder	1.6	1.8
Indirect aggression	0.5	1.3
Family and health		
General health	1.2	0.4
Health recently	1.4	0.7
Family functioning	7.7	5.2
Social support	19.4	3.6

Note: This table shows the average weighted outcome for all 1,243 observations (or 299,119 children) when available. For all three cognitive measures and for the social support measure a larger score indicates a better outcome (positive impact). For all four measures of social development, for the two health measures and for the family functioning score a larger score indicates a lower outcome (negative impact). 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 3: Type of care displaced

	Mean	Std. Dev.
Type		
Daycare	0.09	0.28
Non relative (outside)	0.29	0.45
Non relative (home)	0.07	0.26
Relative (outside)	0.18	0.38
Relative (home)	0.12	0.32
No care	0.24	0.43
Intensity		
Hours per week	30.96	17.09
Nbr. daycare	1.65	0.85
Quality		
Training	0.21	0.41
License	0.20	0.40
Number of children	44,935	
Number of observations	221	

Note: This table shows the type of care used in the first year of life by mothers claiming maternity leave benefits prior to the reform. Effectively, I include all children born prior to the reform whose mother had already returned to work at the time of the first interview and who claimed maternity leave benefits. 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. This category most likely refers to 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).

Table 4: Matching distributional conditions

	llr	kernel	nn(5)
Difference in mean propensity score			
Treated before vs treated after	0.0005	0.0003	0.0000
Control after vs treated after	0.0007	0.0153	0.0193
Control before vs treated after	0.0010	0.0120	0.0095
Propensity score variance ratio			
Treated before vs treated after	1.0023	1.0100	0.9989
Control after vs treated after	1.0002	1.0266	1.0513
Control before vs treated after	0.9977	1.0211	1.0197
Residuals variance ratio			
<1/2 and =<4/5	0.00	0.00	0.00
>4/5 and =<5/4	0.90	0.84	0.93
>5/4 and =<2	0.10	0.16	0.07

Note: Shows Rubin (2001) basic 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 5: ESTIMATED EFFECTS

Specification		DID brr 1	2	3	MDID nearest neighbor (5) 4	5	6	7	8
Time at home									
0 to 12 months	coef.	2.91***	2.89***	2.80***	2.07***	1.57***	1.37***	1.38***	1.16***
	std. err.	0.53	0.51	0.55	0.32	0.30	0.31	0.30	0.33
	obs	1,200			1,392				
0 to 48 months	coef.	1.71	1.76	2.86	2.09	3.72***	3.05**	1.87	1.12
	std. err.	2.75	2.65	2.48	1.49	1.25	1.25	1.19	1.35
	obs	1,243			1,392				
Cognitive development									
PPVT	coef.	-0.75	-0.31	0.09	9.99***	2.92**	2.38*	5.28***	2.68*
	std. err.	3.18	2.97	2.72	1.63	1.37	1.41	1.49	1.41
	obs	1,085			1,224				
Who Am I?	coef.	-0.05	-0.18	-0.46	0.96*	1.13**	1.54***	0.67	1.04**
	std. err.	0.98	0.91	0.81	0.56	0.52	0.54	0.53	0.51
	obs	1,058			1,216				
Number knowledge	coef.	0.36	0.30	-0.09	0.65	0.96**	1.37***	1.22***	0.58
	std. err.	0.79	0.71	0.68	0.49	0.47	0.46	0.45	0.45
	obs	1,085			1,224				
Behavioral development									
Hyperactivity	coef.	0.66	0.66*	0.60	1.29***	0.44*	0.32	1.08***	1.35***
	std. err.	0.41	0.40	0.40	0.22	0.23	0.23	0.24	0.25
	obs	1,225			1,384				
Emotional disorder	coef.	-0.08	-0.07	0.12	0.52***	-0.12	-0.05	-0.17	0.12
	std. err.	0.30	0.30	0.30	0.17	0.19	0.19	0.17	0.18
	obs	1,230			1,392				
Conduct disorder	coef.	-0.25	-0.26	-0.33	-0.01	-0.04	0.09	0.14	0.21
	std. err.	0.28	0.27	0.28	0.15	0.16	0.16	0.15	0.16
	obs	1,232			1,392				
Indirect aggression	coef.	0.31	0.28	0.25	-0.27***	-0.21***	-0.21***	-0.01	-0.08
	std. err.	0.30	0.29	0.24	0.08	0.08	0.08	0.08	0.08
	obs	1,214			1,380				
Health and family									
General health	coef.	0.06	0.06	0.05	0.07**	0.14***	0.19***	0.07	0.09**
	std. err.	0.06	0.07	0.07	0.04	0.04	0.04	0.04	0.04
	obs	1,240			1,404				
Health recently	coef.	0.02	0.04	0.04	0.17***	0.40***	0.41***	0.29***	0.32***
	std. err.	0.10	0.10	0.11	0.06	0.07	0.07	0.06	0.06
	obs	1,240			1,404				
Family functioning	coef.	1.04	0.94	0.59	0.99*	1.14**	1.14**	1.80***	1.02*
	std. err.	0.89	0.88	0.83	0.51	0.54	0.54	0.52	0.53
	obs	1,208			1,348				
Social support	coef.	0.66	0.59	0.63	0.05	0.46	0.12	0.06	0.08
	std. err.	0.57	0.56	0.52	0.34	0.35	0.36	0.35	0.35
	obs	1,207			1,352				
Controls									
Child age at test		no	yes	yes	yes	yes	yes	yes	yes
Child, parent, region		no	no	yes	no	yes	yes	yes	yes
Father income and sleep age		no	no	yes	no	yes	yes	yes	no
Family and parenting scales		no	no	yes	no	yes	yes	no	no
Unemployment rate		no	no	no	no	no	yes	no	no

Note: 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. Child age at test includes a set of age in month dummies (14), with the number of dummies included in parentheses. Child, parent and region include the following dummies: gender, prematurity (2), low birth weight, hospitalization, multiple births, breastfed, married, number of siblings (2), maternal education (5), paternal education (5), maternal age at first baby (16), maternal age at birth (16), immigration status in the last 4 years, province of residence (8) and area of residence (5). Father income and sleep age include father's income quartile at birth (4) and age in months when the child slept through the night (8). Family and parenting scales include positive parenting score (5), inefficient parenting score (4), family functioning scale (12) and social support scale (10). Finally, unemployment rate refers to the average provincial unemployment rate in the year preceding birth. The first three columns from the left show the results of the standard DID estimator. The other five columns shows the estimates of the matching DID estimator using nearest neighbor matching with 5 neighbors. Coefficient significance is denoted using asterisks: *** is $p<0.01$, ** is $p<0.05$, and * is $p<0.1$.

Table 6: Robustness checks

		MDID local linear regression					MDID nearest neighbor(5)			
Specification		**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Time at home										
0 to 12 months	coef.	2.68***	1.91***	1.91***	1.40***	1.39***	0.95***	1.13***	1.30***	-0.12
	std. err.	0.29	0.25	0.25	0.25	0.25	0.30	0.31	0.33	0.28
	obs	1,392					1,392			1,496
0 to 48 months	coef.	3.76***	3.79***	3.97***	2.20**	1.60	3.31**	3.09***	2.82***	-0.44
	std. err.	1.34	1.06	1.04	1.07	1.09	1.36	1.41	1.22	1.28
	obs	1,392					1,392			1,496
Cognitive development										
PPVT	coef.	4.29***	3.48***	3.43***	4.52***	3.58***	3.90***	3.52***	2.96***	-5.88***
	std. err.	1.33	1.24	1.23	1.33	1.25	1.39	1.47	1.28	1.76
	obs	1,224					1,224			1,308
Who Am I?	coef.	0.49	1.29***	1.49***	1.11**	1.38***	-0.18	1.22***	1.06***	-0.44
	std. err.	0.53	0.49	0.49	0.46	0.46	0.53	0.52	0.50	0.53
	obs	1,216					1,216			1,308
Number knowledge	coef.	-0.20	1.09**	1.16***	1.31***	1.00**	0.20	0.72	0.65	0.81*
	std. err.	0.44	0.44	0.44	0.44	0.43	0.44	0.45	0.45	0.42
	obs	1,224					1,224			1,320
Behavioral development										
Hyperactivity	coef.	1.37***	0.51**	0.55**	1.08***	1.20***	0.60***	0.62***	1.42***	-0.99***
	std. err.	0.21	0.23	0.23	0.21	0.21	0.21	0.24	0.23	0.26
	obs	1,384					1,384			1,484
Emotional disorder	coef.	0.55***	0.24	0.20	-0.11	0.11	0.92***	-0.30*	0.44***	-0.21
	std. err.	0.16	0.17	0.17	0.16	0.17	0.19	0.18	0.17	0.18
	obs	1,392					1,392			1,484
Conduct disorder	coef.	0.45***	0.12	0.19	0.27*	0.38***	0.44***	0.01	-0.04	0.09
	std. err.	0.14	0.15	0.15	0.14	0.14	0.16	0.15	0.16	0.17
	obs	1,392					1,392			1,472
Indirect aggression	coef.	0.01	-0.17**	-0.18**	0.00	-0.06	-0.19***	-0.07	-0.12*	0.12
	std. err.	0.08	0.07	0.07	0.07	0.07	0.07	0.08	0.07	0.08
	obs	1,380					1,380			1,460
Health and family										
General health	coef.	0.14***	0.12***	0.14***	0.11***	0.08**	-0.03	0.12***	0.12***	0.13***
	std. err.	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04
	obs	1,404					1,404			1,496
Health recently	coef.	0.21***	0.38***	0.38***	0.32***	0.29***	0.31***	0.37***	0.35***	-0.03
	std. err.	0.06	0.06	0.06	0.06	0.05	0.06	0.07	0.07	0.06
	obs	1,404					1,404			1,496
Family functioning	coef.	1.18**	1.63***	1.60***	1.40***	1.42***	1.31***	2.17***	2.33***	0.63
	std. err.	0.48	0.51	0.51	0.48	0.48	0.49	0.53	0.55	0.44
	obs	1,348					1,348			1,416
Social support	coef.	-0.23	0.30	0.33	0.31	0.38	1.77***	0.59*	0.75***	0.45
	std. err.	0.31	0.31	0.31	0.32	0.31	0.33	0.33	0.34	0.33
	obs	1,352					1,352			1,428
Restrictions										
Control with pscore<0.8		no	no	no	no	no	yes	no	no	no
Excl. Nov. & Dec. 2001 births		no	no	no	no	no	no	yes	no	no
Controlling for work pattern		no	no	no	no	no	no	no	yes	no
Birth year 2001 to 2004		no	no	no	no	no	no	no	no	yes
Controls										
Child age at test		yes	yes	yes	yes	yes	yes	yes	yes	yes
Child, parent, region		no	yes	yes	yes	yes	yes	yes	yes	yes
Father income and sleep age		no	yes	yes	yes	no	yes	yes	yes	yes
Family and parenting scales		no	yes	yes	no	no	yes	yes	yes	yes
Unemployment rate		no	no	yes	no	no	no	no	no	no

Note: This table shows the estimated intention-to-treat effects of the reform. The first five columns show the results of the matching DID estimator using local linear regression matching (comparable with specification 4 to 8 in Table 5). Specification 9 to 12 are slight variations of specification 5 in Table 5. Specification 9: observations in the control groups with propensity scores above or equal to 0.8 are excluded. Specification 10: children born in November or December 2001 are excluded. Specification 11: matching variables include maternal work pattern at the second and third interview. Specification 12: the birth time window is January 2001 to December 2004 (all post reform children). Coefficient significance is denoted using asterisks: *** is p<0.01, ** is p<0.05, and * is p<0.1.

Table 7: Estimated effects - Working mothers prior to birth

Specification		DID brr 1	2	3	MDID nearest neighbor (5) 4	5	6	7	8
Time at home									
0 to 12 months	coef.	3.17***	3.22***	3.11***	3.39***	3.67***	3.81***	3.86***	3.82***
	std. err.	0.27	0.27	0.27	0.15	0.15	0.14	0.15	0.15
	obs	3,720			5,372				
0 to 48 months	coef.	3.21**	3.52**	3.90***	8.28***	12.26***	12.88***	12.24***	12.32***
	std. err.	1.54	1.54	1.46	0.67	0.65	0.63	0.65	0.62
	obs	3,812			5,372				
Cognitive development									
PPVT	coef.	2.68	3.15*	0.92	2.19**	4.31***	3.70***	4.67***	2.12**
	std. err.	2.22	1.91	1.70	0.95	0.88	0.91	0.91	0.91
	obs	3,387			4,744				
Who Am I?	coef.	1.68**	1.93***	1.39***	1.04***	0.89***	0.91***	1.39***	0.74***
	std. err.	0.72	0.57	0.51	0.26	0.26	0.27	0.27	0.27
	obs	3,272			4,740				
Number knowledge	coef.	0.92*	1.10**	0.50	1.27***	0.94***	1.31***	1.01***	0.61***
	std. err.	0.53	0.45	0.44	0.23	0.22	0.22	0.22	0.21
	obs	3,387			4,752				
Behavioral development									
Hyperactivity	coef.	0.25	0.28	0.20	0.03	-0.28**	-0.35***	0.08	-0.02
	std. err.	0.26	0.26	0.25	0.12	0.12	0.12	0.12	0.12
	obs	3,766			5,340				
Emotional disorder	coef.	-0.01	-0.01	-0.02	0.24***	-0.07	-0.10	0.05	-0.04
	std. err.	0.20	0.20	0.20	0.09	0.08	0.08	0.09	0.09
	obs	3,780			5,372				
Conduct disorder	coef.	0.31	0.32	0.17	0.50***	0.18**	0.19**	0.27***	0.22***
	std. err.	0.20	0.20	0.19	0.08	0.08	0.08	0.08	0.08
	obs	3,781			5,364				
Indirect aggression	coef.	0.20	0.18	0.15	0.10**	0.01	-0.05	-0.07	0.01
	std. err.	0.18	0.17	0.15	0.05	0.05	0.05	0.04	0.05
	obs	3,698			5,252				
Health and family									
General health	coef.	0.10**	0.11**	0.06	0.01	0.10***	0.09***	0.12***	0.13***
	std. err.	0.05	0.05	0.04	0.02	0.02	0.02	0.02	0.02
	obs	3,802			5,400				
Health recently	coef.	0.27***	0.28***	0.21***	0.13***	0.10***	0.11***	0.07**	0.12***
	std. err.	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03
	obs	3,802			5,400				
Family functioning	coef.	1.94***	1.91***	1.96***	1.16***	0.86***	1.10***	0.90***	0.92***
	std. err.	0.51	0.50	0.49	0.25	0.24	0.24	0.24	0.25
	obs	3,711			5,232				
Social support	coef.	1.17***	1.16***	1.11***	0.89***	0.26	0.61***	0.45***	0.41**
	std. err.	0.34	0.34	0.32	0.18	0.17	0.17	0.17	0.17
	obs	3,699			5,256				
Controls									
Child age at test		no	yes	yes	yes	yes	yes	yes	yes
Child, parent, region		no	no	yes	no	yes	yes	yes	yes
Father income and sleep age		no	no	yes	no	yes	yes	yes	no
Family and parenting scales		no	no	yes	no	yes	yes	no	no
Unemployment rate		no	no	no	no	no	yes	no	no

Note: This table shows the estimated intention-to-treat effects of the reform. The table structure is exactly the same as that of Table 5. The estimated effects use observations on all children, not just those interviewed for the first time at age 0 to 12 months. Post reform the treatment and control group are accurately identified. Prior to the reform, the treatment and control group had to be infered using work related questions for children interviewed for the first time at age 13 months or more. Coefficient significance is denoted using asterisks: *** is p<0.01, ** is p<0.05, and * is p<0.1.