The Long-Run Impacts of a Universal Child Care Program

By MICHAEL BAKER, JONATHAN GRUBER, AND KEVIN MILLIGAN*

ACCEPTED, AEJ: ECONOMIC POLICY
THIS VERSION: NOVEMBER 8, 2018

Past research documents the persistence of positive impacts of early-life interventions on non cognitive skills. We test the symmetry of this finding by studying the persistence of a sizeable negative shock to non cognitive outcomes arising with the introduction of universal child care in Quebec. We find that the negative effects on non cognitive outcomes persisted to school ages, and also that cohorts with increased child care access had worse health, lower life satisfaction, and higher crime rates later in life. Our results reinforce previous evidence of the central role of the early childhood environment for long-run success. (JEL I1, J13, K42)

* Baker: Department of Economics, University of Toronto, 150 St. George St., Toronto ON M5S 3G7, and NBER (e-mail: baker@chass.utoronto.ca); Gruber: Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142, and NBER (e-mail: gruberj@mit.edu); Milligan: Vancouver School of Economics, University of British Columbia, 6000 Iona Drive, Vancouver BC V6T 1L4, and NBER (e-mail: kevin.milligan@ubc.ca). We thank Hélène Durocher, Pablo Gutierrez Cubillos, and Timea Molnar for outstanding research assistance and seminar participants at McMaster, the Norwegian Research Council workshop "Interventions during childhood and subsequent child development", Simon Fraser, University of Victoria, and the World Bank. Much of the analysis for this paper was conducted at the British Columbia Interuniversity Research Data Centre, which is part of the Canadian Research Data Centre Network (CRDCN). The services and activities provided by the CRDCN are made possible by the financial or in-kind support of the SSHRC, the CIHR, the CFI, Statistics Canada and participating universities whose support is gratefully acknowledged. The views expressed in this paper do not necessarily represent the CRDCN's or that of its partners. Baker gratefully acknowledges the research support of SSHRC (Grant, #410-2011-0724) and a Canada Research Chair at the University of Toronto.

An enduring question about child development is the persistence of early childhood interventions. The debate on Head Start in the United States, for example, has focused in part on whether positive effects of the program found at younger ages fade through time. For example, Gibbs et al. (2013), Bitler et al. (2014), and Kline and Walters (2016) use data from a randomized study of Head Start and generally find the initial cognitive impact fades by first grade. Kline and Walters (2016) argue, however, that even that initial impact can have long-run impact on earnings. Using a regression discontinuity approach, Carneiro and Ginja (2014) found sizeable and persistent long-run benefits from Head Start program participation. With a similar focus on persistence, the re-examination of evidence on the Perry Preschool Project by Heckman et al. (2013) distinguished between positive cognitive effects which faded and positive non cognitive effects which persisted. These latter long-run impacts led to improved economic outcomes and lower incidence of criminal behavior, such that Heckman et al. (2010) find the measured annualized rate of return to the Perry Preschool investment is 6-10%.

These findings leave untouched an important question of symmetry: are there equally persistent and important negative long-run impacts of interventions that had an initial negative impact? This question is important because an affirmative answer would buttress the general case for the importance of the early childhood environment. When a supportive developmental environment is present, children may benefit in the long run. This case is made stronger with evidence that a deficient early environment has a long-run detrimental impact. An example of this latter type of evidence is Bertrand and Pan (2013) who found that childhood non cognitive deficits contribute to the gender difference in teenage disruptive behavior. Also, investigations into the symmetry of early-life health and human

capital events have also reinforced the importance of looking at both positive and negative shocks in related early-life contexts.[1]

In this paper, we develop a causal estimate of the long-run impact of a child care intervention on long-run later-life outcomes. To do this we study the largest experiment with universal child care in North America in recent years: an introduction of very low-cost child care for children aged 0-4 in Quebec in 1997. In an earlier paper (Baker, Gruber, and Milligan 2008; henceforth BGM) we documented large increases in maternal labor supply and in the placement of children in child care in Quebec relative to the rest of Canada, where child care services remained unchanged. (See also Lefebvre and Merrigan 2008 and Lefebvre et al. 2009.) At the same time, there was a large, significant, negative shock to the preschool, non cognitive development and health of children exposed to the new program, with little measured impact on cognitive skills. Subsequent research (Kottelenberg and Lehrer 2013) has confirmed that the negative contemporaneous impact of the program on young children's non cognitive development has persisted as the program has matured.

Our analysis extends the study of the impact of this universal child care program to children at older ages, looking for any persistence of the contemporaneous impacts in four spheres: non cognitive development, cognitive development, health, and crime. We begin our investigation by replicating previous results showing that exposure to the Quebec program increased use of child care among children age 0-4 and led to lower non cognitive outcomes at those ages. We then proceed to our examination of persistence. We provide new evidence for children aged 5-9, showing the program's negative effect on non cognitive skills do not appear to have faded by those ages—and in some cases are

[1] For example, studies of early-life or in-utero infections are reviewed in Almond and Currie (2011), while negative nutrition shocks from fasting are reviewed in Almond, Currie, and Duque (forthcoming). For human capital, Almond, Edlund, and Pålme (2009) find negative long-run human capital implications from an early-life negative shock to child development caused by the Chernobyl nuclear accident.

even stronger. In this way, our results are a mirror-image of the Perry Preschool and Head Start evidence. We also extend previous evidence of the heterogeneous contemporaneous impact of the program. At older ages the primary, negative impact of the program in on children who are already struggling.

We next explore the impacts of this child care intervention on outcomes in the preteen and teenage years. The oldest children who were eligible for the Quebec program are currently in their early 20s, and those who received full treatment are now in their teens. We therefore focus, as the data allow, on their cognitive achievement and health at these ages. We ask whether any early negative impacts of the Program on these children's cognitive development fade out, as has been found in evaluations of other early-years programs? The program also led to earlier exposure to infections and diseases and increases in early childhood anxiety and aggression, and we investigate whether these provide any immunity against maladies or have any consequences for life satisfaction and mental health, respectively, at older ages.

Finally, we also look for any longer-term impacts of the early childhood deficits in non cognitive development on criminal activity. This line of inquiry is guided by the evidence from programs such as Perry Preschool which have shown how positive, early childhood, impacts on non cognitive development result in more positive interactions with various social institutions including the criminal justice system.

Our results find no consistent evidence of a lasting impact of the Quebec program on cognitive test scores; the available data give opposing answers for math scores and show little effect on English or science scores. We do, however, find a significant decline in self-reported health and in life satisfaction among teens. Most strikingly, we find a sharp and contemporaneous increase in criminal behavior among the cohorts exposed to the Quebec program, relative to their peers in other provinces. We illustrate graphically a monotonic increase in crime

rates among cohorts with their exposure to the child care program, and we show in regression analysis that exposure led to a significant rise in overall crime rates. We also report that these effects are primarily for boys, who also see the largest deterioration in non cognitive skills.

By showing that early shocks to childhood development can have persistent effects, our results reinforce previous research emphasizing the importance of early development for later-life outcomes and also provide an important input to the current debate over child care policy. The rapid growth in female labor force participation has led policy makers around the world to consider increased public entitlement to child care for two-worker families. As one example, New York City is implementing universal pre-kindergarten for three and four-year olds.[2] The evidence presented in this paper suggests that measurement of the near-term impact of these policies can serve as a key indicator of the likely long-run success or failure of similar programs.

Our paper proceeds as follows. Part I provides a summary of the extant literature on child care and child outcomes. Part II discusses the Quebec reform. Part III then introduces the wide variety of data sources that we will use for the analysis and discusses our empirical strategy. Part IV updates the results on the contemporaneous impact of the program, and Part V presents our results on the persistence of the impact. Part VI concludes.

I. Background

There is now an enormous literature on the impacts of child care and preschool on the outcomes of young children, and a smaller literature that examines any longer-run impacts as the children age. Two important distinctions have emerged interpreting the evidence. First is whether the child care intervention being studied

[2] See http://www1.nyc.gov/office-of-the-mayor/news/258-17/mayor-de-blasio-3-k-all#/0

was targeted at children in more disadvantaged families or was universal and targeted at all families. Second is the impact on cognitive or non cognitive outcomes, both in the short and long run. We review this literature to place our results in context, with an emphasis on recent research.[3]

A recent view of effects of previous child care exposure on outcomes in adolescence suggest that more hours in child care in general does not affect test scores, but has a negative effect on non cognitive outcomes, such as impulsivity and risk-taking (Vandell et al., 2010). That study, typical of many in the literature, relies on parental choice of child care mode, raising the question of whether any estimated impacts of child care mode are causal or due to selection by parent type. Similar problems affect the interpretation of the large existing literature in economics on maternal work and child outcomes.

A growing body of evidence comes from the use of experimental and quasi-experimental methods to examine the impacts of child care. Perhaps best known are programs targeted toward at-risk children; for example, the experimental variation embedded in the evaluations of the Abecedarian and Perry Preschool interventions. These randomized trials from the 1960s and 1970s have shown that high quality pre-school targeted to low-income children has substantial positive effects. For example, Heckman et al. (2010) estimate a statistically significant annual return of between 7 and 10 percent for the Perry Preschool intervention. Carneiro and Heckman (2003) summarize the evidence from these programs as improving motivation and social skills, while reducing crime and related

[3] See a review of the literature up to 2008 in BGM, and in Baker (2011), Cascio (2015), and Almond, Currie, and Duque (forthcoming). After completing our paper, we became aware that Haeck et al. (2018) have published a paper investigating long term impacts of the Quebec program on health, behavior, and motor-social development. They report that the contemporaneous program effects on non cognitive skills persisted as the program matured and that effects for "emotional disorder" and anxiety persist at school ages, although with smaller magnitudes.

behavior. Heckman et al. (2013) also argue that the non cognitive improvements were pivotal to the long-run impact.

Unlike the experimental evaluations of model programs, our paper focuses on a universal program that services a more economically and socially diverse group of children. In contrast to the literature on programs targeting at-risk children, the evidence on broader programs is mixed (see Baker 2011 and Cascio 2015 for recent overviews). As examples we summarize the findings of evaluations of programs in Denmark, Norway, Spain and Germany.

Exploiting variation in access to center-based preschool (versus a family-based alternative) in Denmark, Datta Gupta and Simonsen (2010) report little effect on non cognitive outcomes at age 7, and a negative impact of family-based child care for boys of parents with low education. Black et al. (2014) utilize a discontinuity in the price of child care in Norway, reporting that while neither child care utilization nor parental labor supply is sensitive to price, they observe a positive impact on children's junior high school outcomes, presumably through a disposable income effect. Havnes and Mogstad (2011) explore an expansion of the Norwegian system, reporting positive impacts. The public system led to higher educational attainment (primarily for children of low education mothers) and earnings (mostly for girls) at ages 30–40. In a related paper Havnes and Mogstad (2014) find that the earnings gains are primarily for children of low-income parents and that children of upper-class parents experience an earnings loss. Felfe et al. (2015) exploit variation across states in the expansion of the Spanish child care system, finding improvements in reading skills at age 15 of 0.15 standard deviations, driven by the impacts for girls and children from disadvantaged families. Finally, Cornelissen et al. (forthcoming) explore a policy reform of the German child care system which entitles every child to a place on their third birthday. They find child care attendance has a more positive effect for children from more disadvantaged backgrounds. While there are clearly studies

here that report positive impacts of universal children programs, in many cases these impacts are primarily enjoyed by less advantaged children. There is a little clear evidence that these programs provide significant benefits more broadly.

Universal preschool has also been a focus of recent research in the United States. Many of these studies exploit age cutoffs for preschool enrollment comparing the youngest children in a preschool cohort to the children just a little bit younger who had to wait an additional year before enrolling. Perhaps the best-known program is in Oklahoma. Gormley and Gayer (2005) document positive impacts for Hispanics and blacks, but not for whites, which is correlated with eligibility for free school lunch. Using a different cognitive measure Gormley et al. (2005) report more broadly-based gains. A study of New Mexico's program (Hustedt et al. 2008) finds positive effects on math achievement and literacy in a sample that over represents Hispanics and Native Americans. Taking a wider view, Wong et al. (2007) examine preschool programs in five states (a mix of targeted and universal programs) on a variety of outcomes. They record positive impacts on a little more than half of the outcomes investigated. Finally, Fitzpatrick (2008) studies the introduction of pre-K program in Georgia, finding positive impacts for disadvantaged children in small towns and rural areas. As with the European studies, the recent American evidence mostly fits the pattern that the positive impact of universal programs is concentrated in more at-risk children.

Most relevant to the current paper is research on the introduction of universal child care in Quebec. The initial evaluation of this policy in BGM found striking negative impacts of the program on children's non cognitive scores and family outcomes. In a series of papers Kottelenberg and Lehrer show that most of these contemporaneous negative effects of the program on young children and family outcomes measured shortly after it was introduced have persisted on newer cohorts of children as the program has matured (2013), that the negative impacts

on child outcomes are larger the younger the age the child entered the program (2014), that the impacts vary by the sex of the child (2018). They also find (2017) evidence of heterogeneous impacts. For PPVT and Motor-Social Development scores they find more positive outcomes for children in single-parent families and in the bottom quintiles of the tests score distribution. These are offset at the mean by negative impacts for children from two-parent families. Haeck et al. (2015) present evidence that the program had negative effects on children's cognitive development at age 5. Finally, Brodeur and Connolly (2013) and Molnar (2017) focus on heterogeneity by education, with the former paper looking at life satisfaction and the latter time allocation. Our work is distinguished from these previous studies by focusing on longer-run outcomes at older ages.

To summarize, the literature on child care and preschool seems to indicate that high-quality interventions for low-income populations deliver both short and long-run benefits, particularly through non cognitive channels. But, universal child care expansions do not appear to provide broadly-based short-term benefits, with mixed evidence on long-term effects.

II. The Quebec Universal Child Care Policy

Introduced in September 1997, the Quebec child care policy aimed to provide regulated child care places to all children aged 0-4 in the province at a price of $5 per day, with the rest of the cost covered by government subsidy. This program raised child care subsidies to almost 80 percent on average, which can be compared to subsidies of roughly one-third in other provinces.[4] Children were eligible for the program whether or not their parents worked. Child care under the

[4] BGM report that the program raised subsidies for two-parent households from 50% to 80%; and 60% to 80% in single-parent households. In the pre-policy period these subsidies were tax incentives rather than direct price subsidies. A full description of child care subsidization in Canada in these years is presented in Baker et al. (2005).

program was primarily provided in two venues. The first were child care centers (centres de la petite enfance--CPE) created out of existing nonprofit child care centers. The second was home-based care staffed by regulated providers and organized into networks affiliated with a local CPE. Existing for-profit child care centers could provide subsidized spaces as well. Typically, older children enrolled in the CPE center-based care and younger children were enrolled in family home-based care. The fee was raised to $7 a day in 2004 and to $7.30 in 2014. Haeck et al. (2015) report that regulated child care places in the province rose from 78,864 in 1997 to 245,107 in 2012, while total provincial subsidies rose from 288 million dollars in 1996/97 before the program to 2.2 billion dollars in 2011/12.[5]

The introduction of the Quebec program was accompanied by some important structural reforms of child care provision. Formal qualifications for caregivers were raised and operational regulations modified.[6] There was also an expansion of voluntary full-time kindergarten and subsidized after-school care for children aged 5-12. The government introduced new higher wage policies, pressured in part by strikes by the unionized child care providers. In our analysis, we cannot distinguish the impacts of these supply-side interventions on the quality of care from the subsidization of fees which happened at the same time.

The program was first introduced to four-year olds in September 1997. So, children born before 1993 were not eligible. In 1998 three-year olds were

[5] There appears to have been queues for subsidized places at the start of the program. The magnitude of excess demand is hard to estimate, however, because waiting lists included children not yet eligible, children in subsidized care but wanting to change providers and children on multiple wait lists.

[6] The proportion of staff required to have a college diploma or university degree in early childhood education rose from one-third to two-thirds. To facilitate this goal the government provided financial aid for staff enrolled in college-level early childhood education. Home-based providers faced increased training (24–45 hours) and annual professional development (6 hours) requirements. While maximum center size increased from 60 to 80 places, staff/child ratios remained unchanged, except the ratios for 4- and 5-year-olds rose from 1 : 8 to 1 : 10. There was also an increase in parental participation in governance as their representation on the board of directors rose from 51% to two-thirds of members.

included, followed in 1999 by two-year olds. Finally, in 2000, both zero- and one-year old children were included. So, birth cohorts from 2000 onward were eligible from ages zero to four, while those born from 1993 to 1999 were eligible for part of their early lives. Appendix Figure 1 depicts this cohort eligibility pattern.

There has been a number of reviews of the program's quality since its inception, which we review in the Appendix. In Japel et al. (2005), the quality of care was judged to be at "minimal quality" in just over 60 percent of the venues evaluated, and just over one-quarter provided services that were judged good, very good or excellent. This is comparable to the quality of care provided in many other developed countries, and better than the quality of care in for-profit or unregulated care in Quebec.

III. Data and Empirical Strategy

We make use of four types of data (consisting of six data sets) for our analysis to trace the long-run impact of the Quebec program from the period of treatment through to young adulthood, covering a variety of relevant outcomes. For all the data sources, our sample selection decisions are guided by how each source covered the cohorts exposed to program treatment. Below we describe each of the four data sources in turn.

A. Child Care Enrollment and Child Outcomes: NLSCY and SYC

Our first dataset is the National Longitudinal Study of Children and Youth (NLSCY), which was the primary dataset in BGM. The NLSCY is a nationally representative survey of children, conducted biannually between 1994-95 (cycle 1) and 2008-09 (cycle 8). While the NLSCY has a longitudinal component, we

use it cross-sectionally.[7] About 2,000 children ages 0 to 5 are available in each cycle, with some coverage of children at older ages. We also add the Survey of Young Canadians (SYC) conducted in 2010-11 as a cross-section survey of children aged 1-9 with similar child development content as the NLSCY. We pool the data from the NLSCY (cycles 1 to 8) and SYC (which we denote as cycle 9) together to produce a time series of cross sections.[8]

We use the NLSCY/SYC for two purposes. First, we re-examine the contemporaneous impact of the Quebec Family Plan on child outcomes at age 5 or younger. We benchmark our estimates to BGM using cycles 1 through 5 (excluding the transitional cycle 3, as in BGM). We then extend this evidence by adding NLSCY cycles 6-8 and the SYC. Second, we want to see if the estimated contemporaneous impacts of the program persist into grade school. We do this by selecting cycles that provide a cross-section view of the children at ages 5 through 9. This includes cycles 1 and 2 for the 'pre' period which we compare to observations for this same age group in cycle 7 and cycle 9 (the SYC).[9]

At the younger ages we focus on the same outcome measures examined in BGM and many subsequent studies of the Quebec program. We use a binary indicator for any type of non parental care while the parent works or is at school, a set of parent-reported non cognitive scores, the child's score on the well-known Peabody Picture Vocabulary Test (PPVT), and finally a parent report of how well the child gets along at school with her/his teacher.

[7] The longitudinal component started with an initial cohort of children age 0-11 in the first cycle. These children have been followed in each subsequent cycle, with non longitudinal kids at other ages added to fill in the gaps.

[8] By focusing on restricted age intervals (e.g., ages 2-3 or 4-5) children appear only once in our pooled samples.

[9] Cycle 7 is the only post-treatment NLSCY cycle with children at each age between 5 and 9. The other cycles (4-6 and 8) have holes at some or all ages. Our results are similar using all available cycle 4 to cycle 8 observations within the age 5 to 9 range. Our sample restricted to cycle 7 and the SYC is a more conservative approach.

The non cognitive scores are built up from a menu of questions about the behavior and development of their children. The questions are based on best practices in the relevant fields. Each question solicits a response on a three-point scale, and the score is constructed as the sum of responses on all questions for a given behavior or skill. An overview of the questions that make up each measure is provided in the Appendix.[10] At ages 2 and 3 we observe scores for Hyperactivity, Anxiety, Separation Anxiety, and Aggression.[11] For the 5-9 year olds we have scores for Hyperactivity, Anxiety, Aggression, Indirect Aggression and Prosocial Behaviour. While some of the indices for the older age group have the same names as corresponding indices for the younger children, they are based on a different set of age-appropriate questions.

B. Test Scores: SAIP/PCAP and PISA

To measure the impact of the Quebec program on test scores of older children, we turn to two different sources. The first source combines data from the School Achievement Indicators Program (SAIP) and the successor Pan Canadian Assessment Program (PCAP), which are initiatives of the Council of Ministers of Education. The SAIP assesses the performance of 13 and 16 year old students across the country in the core subjects of math, reading and science. The tests were conducted 9 times between 1993 and 2004, each focusing on one of the core subjects. The PCAP succeeded the SAIP and has been conducted triennially starting in 2007; we use 2007, 2010, and 2013 tests. Each PCAP focuses on one of the core subjects; just like SAIP. Unlike SAIP, however, a smaller sample of students writes tests in the other non focal subjects. So, scores for each subject are

[10] A more detailed discussion of the NLSCY non cognitive measures was published in the online appendix to BGM.

[11] For the non cognitive outcomes we focus on 2-3 year olds within the 0-4 age group, because the measures do not exist for children ages 0-1 and the non cognitive indices for 4 year olds are based on different questions.

available in each PCAP wave. We construct time-series cross-section samples, pooling data from SAIP and PCAP for the test scores in each subject area. In addition to the PCAP data, from SAIP we have math scores from 1997 and 2001, reading scores from 1998 and science scores from 1996.

The second data set comes from the Programme for International Student Assessment (PISA), which is a triennial test of 15-year olds conducted by the OECD around the world. This testing program was initiated in 2000, and covers the core subject areas of math, reading and science. Our analysis sample includes the Canadian test scores from 2000-2015.

C. Health and Well-Being: CCHS

To assess the impact of the child care intervention on the health of older children, we use the Canadian Community Health Survey (CCHS). The CCHS offers biannual national cross-section data for 2001, 2003, and 2005 of approximately 130,000 observations; followed by annual cross-section surveys of around 65,000 observations starting in 2007. We use all surveys up to and including 2015. We examine questions on self-assessed health, life satisfaction, and mental health. We take a sample of 12 through 20 year olds, which in the chosen years contains both individuals who were and were not exposed to the child care program at younger ages.

D. Criminal Behavior: UCRS

We combine special tabulations of crime accusations and convictions from Statistics Canada's Uniform Crime Reporting Survey (UCRS) with single-age population counts to construct crime rates by age, sex, province, year cells. The

UCRS is a survey of police-reported crime.[12] This means that the crime incident has been substantiated by the police, and therefore the survey misses crimes that are never detected and/or not reported to the police. The accused includes those ultimately charged as well as cases dealt with through extrajudicial measures. We examine rates (separately) for crimes against persons, crimes against property, "other criminal code violations", and drug violations; as well as an aggregate crime rate based on these four categories.[13] For our age groups most "other criminal code violations" involve failures to appear in court and breaches of probation.[14]

Our data is annual for the years 2006 through 2014. We start the analysis in 2006 to avoid any impact of the introduction of the Youth Criminal Justice Act in 2003. We discuss this choice in depth below. As in our analysis of the CCHS, we construct our sample for 12- through 20-year olds. We exclude the data from 2010 because of missing data for that year from Montreal, the largest city in Quebec.

E. Empirical Strategy

Our basic empirical strategy is a straightforward difference-in-difference analysis that follows BGM. The base observation is of children *i* in province *p* in year *t*. The analysis of a given outcome is for a specific age interval (e.g., 2-3 year olds). We assign a binary indicator $EXPOSURE_{pt}$ equal to one for children who live in Quebec, in the years that children in that age interval would be exposed to

[12] Responding to the coverage is mandatory and survey compliance is reported as "virtually 100 percent" (http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3302).

[13] We omit the traffic crime category as it is less relevant to our age range. We also omit a residual 'other federal statute violations' that includes violations of legislation such as the Bankruptcy Act and the Competition Act.

[14] Other prevalent youth crimes are theft under $5000, assault, mischief, breaking and entering, cannabis possession, uttering threats and possession of stolen property. See Zhang (2014) for a recent comparison of youth and adult crime rates by offence.

the Quebec child care program.[15] So, for example, 2-3 three year olds in Quebec in 1996-97 (wave 2 of the NLSCY) would not be exposed to the Quebec program, while 2-3 year olds who live in Quebec in 2001-03 (cycle 5 of the NLSCY) would. For the NLSCY/SYC and SAIP/PCAP/PISA datasets, all children in a given year and province have the same value of $EXPOSURE_{pt}$, so we estimate basic difference-in-difference models of the form:

$$(1) \quad Y_{ipt} = \alpha + \beta EXPOSURE_{pt} + \pi PROV_p + \delta YEAR_t + \lambda X_{ipt} + \epsilon_{ipt}.$$

We control for a set of province dummies *(PROV*) and year dummies (*YEAR*), as well as variables *X* that differ (according to availability as shown in Appendix Table 2) by data set but can include gender, child's age, mother's age and education, the number of older and younger siblings, and mother's /father's/family's immigrant status and ethnicity. We focus on the estimation of β, the coefficient on exposure to the Quebec child care program. The core identifying assumption is that the time trend across years is common between Quebec and other provinces. We assess this assumption graphically later in the paper.

For the health and crime analysis, we have data that covers a larger number of cohorts over a larger number of years. This allows us to estimate a more flexible version of equation (1) with a more extensive set of controls. Our samples span a larger interval of ages so that the *EXPOSURE* variable now varies by age (i.e., birth cohort) within year. The base equation estimated for the crime rate *CR* is

[15] See the complete cohort map in Appendix Figure 1. We assume that children observed in Quebec at older ages also lived there from age 0-4. In the 2016 Census, the proportion of Quebec residents who were resident in another province five years previous was 1.1 percent. (Statistics Canada Catalogue no. 98-400-X2016313.)

(2)

$$CR_{aspt} = \tau + \rho EXPOSURE_{apt} + \varphi PROV_p + \kappa YEAR_t + \omega SEX_s + \theta AGE_a + \epsilon_{aspt}$$

where *a* indexes age, *s* indexes sex, *p* indexes province, *t* indexes year, *SEX* is a 0/1 indicator for males and *AGE* is a vector of single-year age dummy variables. We also estimate a variant of equation (2) in which we add a full set of second-order interactions between *PROV*, *AGE*, *SEX*, and *YEAR* except *YEAR*PROV*. In a third specification we add a set of province-specific linear trends. For samples in which we pool the crime rates for different offences together, we add a full set of offence fixed effects and (as noted) their interactions with *AGE*, *PROV*, *SEX* and *YEAR*.

F. Threats to Identification

The major threat to our identification strategy comes from pre-existing time trends that could confound our policy inference. We use three approaches to building the case for our policy inference. First, we present graphs showing the trends in our outcome variables in the before and after periods, comparing the rest of Canada to Quebec. Second, we include robust controls in our regression specifications, with province-specific time trends in our crime analysis along with time interactions with age and sex. Finally, for the crime analysis in the appendix we report results from regressions using leads and lags of the policy variable as a type of placebo test to examine whether our policy variable is picking up underlying trends.

G. Scaling Reduced-Form Results

As discussed in BGM, our modeling of outcomes is a reduced form of an underlying process through which the Quebec policy impacts maternal labor supply and child care utilization. To interpret the results structurally, BGM either scaled the estimated effects by the impact of the Quebec policy on maternal labor supply (a 7% rise) or by its impact on use of child care (a 14% rise). Haeck et al. (2015) show that between the mid-1990s and 2008 the proportion of children aged 1-4 who were in center-based care as their primary arrangement rose in Quebec from under 10 percent to close to 60 percent, while in the rest of Canada it rose from about 10 percent to just under 20 percent. The proportion in parental care fell from around 55 percent to roughly 25 percent in Quebec over this same period, while the similar proportion in the rest of the country fell from just under 60 percent to about 50 percent, where it has stabilized since 1998. By this metric the proportion of treated children in Quebec is much higher than the proportion that moved into non parental care with the advent of the program.

There are therefore a wide variety of "first stage" estimates one could apply to the longer-run reduced-form impacts we estimate here. As a result, we are reticent here to interpret any of our longer run results in a structural way and focus instead on the sign and significance of our reduced-form findings.

H. Additional Factors

In our previous study of the Quebec program we limited the analysis to children in two-parent families. This choice minimized any possible confounding effects of concurrent changes to Canada's National Child Benefit on our sample of 0-4-

year olds.[16] As we turn our focus to children at older ages, the restriction to two-parent families makes less sense. Due to family dynamics, at older ages children currently living in single-parent families may have lived in two-parent families when they were young. Likewise, children currently in two-parent families may have been born into single-parent households. We therefore sample children in all family types.

Another factor relevant to our analysis of teenage criminal activity is that the Youth Criminal Justice Act (YCJA) came into effect on April 1, 2003. The YCJA is a federal act governing the prosecution of youth crimes across the country. Quebec has a history of taking a more rehabilitative approach to youth criminal activity. One of the impacts of the YCJA was to make the rest of Canada more like Quebec, in that it encouraged the use of extrajudicial remedies instead of the courts for less severe crimes.[17] Correspondingly there appears to be a sharp drop in the proportion of youth offenders charged in most provinces in 2003 and corresponding uptick in the proportion chargeable but not charged (Carrington and Scholenberg 2005). An exception is Quebec, no doubt reflecting the province's pre-existing proclivity for extrajudicial measures for youth crime. As evidenced in Bala et al. (2009), this impact appears mostly discrete to the year the Act was implemented, and the rates of charged and otherwise cleared youth crimes "settled" into new post YCJA levels by about 2005. As a result we use crime data starting in 2006 to stay clear of this impact of the YCJA. We also note that by examining both data on accusations and convictions we provide evidence

[16] BGM show the main findings extend to the children of single-parent households. Milligan and Stabile (2011, Figure 1) show that child benefits saw about a threefold increase between 1990 and 2005 for families under CAD25,000 income, but did not increase after 1994 for families with CAD50,000 or more of income—which is a large majority of two-parent families. Including children from single-parent families in our sample may attenuate the estimated impacts if the child benefits push child development in the opposite direction.

[17] As argued by Trepanier (2004) the YCJA also put some limits on Quebec's rehabilitative approach (for example, no rehabilitation while accused is remanded in custody) and was perceived as a triumph of the principle of proportionality over rehabilitation and reintegration.

of criminal activity both before and after the application of any extrajudicial remedies.

IV. Contemporaneous Impacts

In this section we model the contemporaneous impact of exposure to the Quebec Family Plan on child care use and non cognitive skills. We begin by replicating earlier analysis showing a substantial increase in the use of non parental care. We then reproduce our earlier results on the negative effects on contemporaneous non cognitive skills of young children.

[Insert Figure 1 Here]

We begin by graphing in Figure 1 the unconditional standardized means of our contemporaneous dependent variables for Quebec and for the rest of Canada by cycle of the NLSCY. We standardize the score-based dependent variables to have mean zero and unit standard deviation. For each outcome, we indicate the onset of the policy with a vertical line at cycle 3 in 1998-99. For the first five outcomes, the solid line for the rest of Canada is almost flat, indicating little trend. This stability in the untreated provinces is an important part of the case for our identification strategy. Also, with perhaps the exception of separation anxiety, the scores in cycles 1 and 2 in Quebec are visually parallel to the scores in the rest of Canada. The more notable exception is the PPVT score, which shows an upward trend starting in cycle 4 for the rest of Canada. Furthermore, for Quebec, there is a large, anomalous downward spike in PPVT scores in cycle 3, which almost completely dissipates in cycle 4. Post policy, being in care increases sharply starting in cycle 3 and continuing through cycle 9. The behavioral scores also each show a distinct relative increase after cycle 3, with varying patterns in later cycles. The time trends for the PPVT in Quebec are less clear.

In Table 1 we report regression results to quantify the impacts seen in the figure. These regressions use a similar specification to the original BGM analysis but differ by the inclusion of both single-parent and two-parent families. We exclude observations from cycle 3, as the program is in transition. In our regressions, we report intent-to-treat estimates unscaled by proportion of the population treated, for the reasons discussed above. We show standard errors both clustered by province-year of birth and estimated following the method of Bester et al. (2011) allowing dependence over time and within region.[18] Finally the reported levels of statistical significance are adjusted for multiple testing following Anderson (2008).

[Insert Table 1 Here]

In the first row are estimates of the program effect on the probability of the child being enrolled in child care at ages zero to four. At just under 15 percentage points, the estimate using waves 1, 2, 3 and 4 almost exactly matches the estimate in our previous paper (0.146). Extending the sample by the additional cycles of the NLSCY and the SYC leads to a larger estimate, which might be expected as the supply of subsidized child care spaces expanded as the program matured. This latter estimate represents an increase in child care use of just over 40 percent of the baseline rate.

In the next four rows are the estimates of the program impact on standardized non cognitive outcomes at ages 2-3. In each case we have created the variable so that a higher score indicates a poorer outcome. The components of each score are listed in the Appendix. The estimates using cycles 1, 2, 4, and 5 show statistically significant estimates for Anxiety and Aggression but not for Hyperactivity and

[18] For the Bester et al. (2011) method, we cluster using four regional groups and base inference on a t -distribution with three degrees of freedom.

Separation Anxiety. They are marginally smaller in magnitude than the estimates in our previous paper, although not enough to qualitatively change our inference.[19] Adding the additional waves of data leads to a statistically significant estimate for Hyperactivity and maintains the inference for Anxiety and Aggression. Also, the estimates are generally stable in magnitude across the two samples, except for Aggression, which is just under 50 percent larger.

In the last row is the estimated impact of the program on a measure of cognitive development—PPVT—at ages 4 and 5. We report a statistically-significant decline of 11 to 14 percent of a standard deviation depending on sample. This result, while consistent with Haeck et al. (2015), is different than the estimate in BGM which used a different age range.[20]

The results in Table 1 demonstrate that the main conclusions of BGM for young children of two-parent families extend to the full sample of young children from all family types and persist as the program has matured. The Quebec program led to a substantial increase in the use of child care and increases in children's levels of anxiety and aggression. We do not pursue analysis of heterogeneity for these contemporaneous outcomes, but Kottelenberg and Lehrer (2017) provide evidence of a positive boost to child development for children from disadvantaged single-parent families, with more negative outcomes from two-parent families. In addition, Kottelenberg and Lehrer (2018) find significant differences between boys and girls.[21]

[19] In BGM the estimates are 9 percent (anxiety) and 12 percent (aggression) of a standard deviation.

[20]The addition of five-year olds here explains the difference to BGM. With an age 4 sample, we obtain an insignificant estimate of -0.250 (0.843), consistent with the insignificant result of 0.36 (0.75) in BGM.

[21] In the appendix we do examine the policy impact using the Firpo, Fortin, and Lemieux (2009) unconditional quantile regression approach described below. The results indicate that the impact is close to the mean impact across deciles for three of the four non cognitive outcomes. The notable exception is aggression for which we find a larger impact at higher deciles—that is for children with elevated scores.

V. Persistence of Impacts at Older Ages

We now turn to the examination of the persistence of the impacts at older ages. The analysis begins with the use of several behavioral non cognitive scores at ages 5-9, and then moves on to cognitive test scores at early teen ages. Following that we examine health outcomes, and finally criminal behavior.

A. Non Cognitive Outcomes

In Table 2 we present the results for non cognitive outcomes at ages 5-9. The details on each measure's construction are provided in the Appendix. We again normalize the non cognitive scores to have zero mean and unit standard deviation. The first four measures show that the Quebec program's negative effects on non cognitive skills appear to strongly persist into school years, and in some cases are larger than at younger ages. For Anxiety the impact is now just over one quarter of a standard deviation, which is more than twice as large as for 2-3 year olds, while for Aggression the estimate is very similar to the result for the younger age group. Hyperactivity shows an increase of 13 percent of a standard deviation. For the two new indices we see a statistically significant impact on Indirect Aggression of 19 percent of a standard deviation, while the result for Prosocial behavior is very close to zero.

[Insert Table 2 Here]

For the older children we also have an alternative measure of behavior: a parent-reported indication of how the child gets along with his/her teacher at school. The variable is coded 0/1, where one indicates the child gets along very well with his/her teacher (there are no problems). The estimate for this variable is in the last row of Table 2. It is consistent with the results for the non cognitive

indices, indicating exposure to the Quebec program leads to a statistically significant worse outcome.

We extend this analysis by examining heterogeneity in the policy impact using the Firpo, Fortin, and Lemieux (2009) unconditional quantile regression approach based on the recentered influence function (RIF). These RIF estimates show how the policy variable affects individuals at the decile cutoffs of the unconditional distribution of the dependent variable. We graph in Figure 2 the RIF results by decile cutoff for the first four dependent variables analyzed in Table 2.[22] All four non cognitive scores show larger impacts at higher deciles than at lower deciles. The impact on aggression is much larger at the 7th, 8th, and 9th decile cutoffs. This suggests that the primary impact of the program was to increase aggression scores for those who already had high scores.[23]

[Insert Figure 2 Here]

In the appendix, we provide further analysis of these outcomes by performing the analysis separately by gender. The results indicate that the impact on boys is generally stronger, especially in the cases of hyperactivity and aggression.

Taken together, the negative impact of the Quebec program on the non cognitive outcomes of young children appears to persist and in some cases increase as they reach school ages. The impact is stronger on those who already had elevated non cognitive scores and for boys.

[22] We leave out the prosocial score since it was not statistically significant and also the binary indicator for getting along with the teacher, since RIF analysis is only useful with continuous dependent variables.

[23] Heckman, Smith, and Clements (1997) note that impacts on quantiles need not correspond to impacts on individuals who would otherwise be at a particular quantile.

B. Cognitive Outcomes

We now turn to the study of cognitive outcomes. Unfortunately, there is no parallel verbal cognitive measure available in the NLSCY at older ages to follow up on the PPVT result in Table 1. A math score is available, but we do not use it here due to data issues that render it unreliable.[24] Instead we use data from periodic standardized testing of Canadian teens through SAIP/PCAP and PISA. Note that the 2009 PISA scores for 15-year olds are likely to capture both teenagers in Quebec who were and were not exposed to the child care program. We consider different coding of the EXPOSURE dummy for the 2009 scores to discover how the estimates vary on this margin.

The estimates are presented in Table 3. The standard deviations of the scores are approximately one so the point estimates can be read directly as proportions of a standard deviation. In the first row are the results for the PCAP/SAIP tests. The estimates indicate negative but statistically insignificant impacts of exposure to the Quebec program on math scores, reading, and science scores. In the next two rows are the results for the PISA tests alternatively viewing the 2009 scores as capturing Quebec children who are not or who are exposed to the child care program. If we view the 2009 scores as pre-program, we obtain a statistically significant positive impact of exposure on math scores of just over 30 percent of a standard deviation, a marginally significant increase for reading of 10 percent and a result for science which is statistically insignificant and close to 0. If instead we view the 2009 scores as post-program, the impact on math is still positive but smaller at 20 percent of a standard deviation while the impacts on reading and science are both close to zero and statistically insignificant.

[24] A disproportionate number of perfect scores (particularly in Quebec) was detected in the first two cycles. Also, the response rate to the math test was low and variable in cycles 1-3.

[Insert Table 3 Here]

On balance the results in Table 3 do not provide unambiguous evidence of a persistent negative impact of the Quebec program on cognitive ability. For math, the estimates show exposure to the Quebec program leading to a decrease in scores (insignificant) in the SAIP/PCAP but an increase in PISA scores.[25] The results for reading and especially science provide a more consistent story of no impact of the Quebec program. Overall the evidence on the long-run impact of the Quebec Family Plan on test scores is mixed, and there is no evidence that the initial negative impact on PPVT scores persist as evidenced by the reading scores we examine.

C. Health and Life Satisfaction

We next study the impact of exposure to the Quebec program on health status and on life satisfaction using the CCHS survey. The motivation is to see if the negative impacts on physical health and behavior at early ages have a persistent impact. We use a sample with ages 12-20 for this analysis. We examine the outcomes self-perceived health, satisfaction with life in general, and finally self-perceived mental health. These variables are coded between 1 and 5, running from better to worse, and we normalized them by their standard deviations for our regressions. As noted in the Introduction our hypotheses here are that the treated children may enjoy higher immunity due to earlier exposure to germs, and/or the negative shocks to their non cognitive development lead to lower levels of (self-perceived) wellbeing.

[25] The relatively stronger performance of Quebec students on PISA measured math testing is a matter of some public debate in Canada. In Quebec a teaching certificate requires a 4 year degree in education instead of a 1-2 year certificate as in other provinces. Also, Quebec students make the transition to high school courses taught by subject specialists in grade 7 while in most other provinces it is in grade 9.

[Insert Table 4 Here]

The estimates in Table 4 indicate that exposure to the Quebec program is associated with some worsening of self-reported health and overall life satisfaction, but not in self-reported mental health. For youths exposed to the program, the poor health indicator rises by 7.3 percent of a standard deviation. The estimate for life satisfaction is smaller, but still statistically significant, and again indicates a poorer outcome from exposure to the program. Finally, the estimate for mental health is negative but very small and statistically insignificant. This may indicate that mental health is not affected by negative shocks to non cognitive development at earlier ages, some of the developmental impacts (e.g., hyperactivity) fade as the children age, or that the children are not self aware and a clinical report would show a different result.

D. Youth Crime

Our final measure of longer-run outcomes is youth criminal activity. In evaluations of Perry Preschool, the long-run impact on crime was a vital component of the analysis.[26] Our aim here is to investigate whether the link between non cognitive development and crime holds up in a symmetric case where there is a *decrease* in measured non cognitive development. We have two measures of criminality—rates of accused and convictions. We focus on four crimes (personal, property, other criminal code convictions and drugs), as well as an aggregate measure of the incidence of all these crime categories.

[26] Belfield et al. (2006) find that crime reduction by males provides most of the long-run financial benefit of the Perry Preschool program. Heckman et al. (2013, p. 2070) find in their study of Perry that "…the evidence from this paper suggests that reducing early externalizing behavior reduces crime."

[Insert Figure 3 Here]

To begin, we examine time trends by year of birth cohorts with different exposure to the policy. We graph the difference between Quebec and the rest of Canada across ages separately by year-of-birth cohort in Figure 3. We group the cohorts by the number of years of eligibility, ranging from 0 before 1993 birth year to 5 for those born in 2000 or later. The differences are negative, indicating lower accusation rates in Quebec than in the rest of Canada. The pattern across cohorts is striking—the gap between Quebec and the rest of Canada shrinks for the cohorts most exposed to eligibility for the Quebec program. In the appendix we repeat the analysis for convictions and separately for four specific crime types with similar monotonic shifts across birth cohorts consistent with the degree of exposure.

In Table 5 we formalize this inference with regression estimates. Column (1) presents the simple difference-in-differences results, controlling for fixed effects for province, year, age, and gender. In addition, we include a set of dummies for crime type in the pooled regression for all crime types. In column (2), a richer specification includes the full set of second-order interactions between province, age and gender (and crime type in the aggregate rate regressions). Finally, in column (3) we add controls for province*year trends (and crime type*province*year in the aggregate rate regressions) to allow for province-specific trends.

[Insert Table 5 Here]

The estimates are generally consistent with the graphical evidence: exposure to the Quebec program leads to higher rates of crime. Looking first at all crime counts, the estimates from the simple difference-in-differences specification indicate increases in both the rates of accused and convictions that is statistically

significant. The coefficient of 514 for accused is a 27 percent increase on the overall mean of 1872 accusations per 100,000. The coefficient of 208 for convictions is of a similar magnitude relative to the mean conviction rate. This estimate for rates of accused does not change much when we add the second order province/age/gender interactions in the second column, but there is an increase in the estimates for convictions. In column (3) the estimate for accused falls but remains sizable and highly significant while the estimate for convictions returns to the level seen in column (1). The estimates from the richest specifications indicate sizeable effects on crime rates. For accused across all categories we estimate a rise of 353 per 100,000 children, compared to a mean of 1872 accusations. This is a rise of 19 percent. The result is slightly higher in percentage terms for convictions at 22 percent.

The remaining rows of the table show the results for each type of crime. The impact of exposure to the Quebec program is largest for property crime; the estimates from the richest specification show an increase in accusations of these crimes of 602 per 100,000 children, or 19 percent of the mean, and for convictions for these crimes of 342 per 100,000, or about 25 percent of the mean. The estimated impacts on other criminal code violations are almost as large. Slightly smaller are the estimated impacts on crimes against persons, at 16 percent of the mean for both accusations and convictions. Finally, the impact for drug crimes is 14 percent of the mean for accusations but over 23 percent of the mean for convictions.

In the appendix, we present results separately for males and females. Gender differences are of potential importance as there is recent evidence that the impacts of non parental care vary by gender, as well as growing interest in gender

differences in childhood and adult success.[27] The estimates indicate larger absolute impacts on the crime rates for boys, particularly for other criminal code violations and drugs.[28] In fact in our richest specification some of the estimates for girls are substantively smaller and lose some statistical significance. Therefore, the gender differences in the impacts of the Quebec program on crime rates line up with the gender differences in the impact of the program on non cognitive development.

Are these findings and magnitudes credible? We consolidate in this paragraph the case we have presented. We find large contemporaneous impacts of the Quebec program on measures of children's non cognitive development that persist to school ages. These impacts are larger for boys and primarily for those who already had elevated behavioral problems. The increase in criminal behavior we document arises in the years and at ages that are consistent with a dose-response relationship with eligibility for the Quebec program. Previous research has linked improvements in similar non cognitive measures at early ages to decreased criminal behavior; we find a symmetric effect for an early-life deterioration in behavior. The estimated magnitude of the increase in crime in our most rigorous specification is 19 percent of the mean for accusations and 22 percent for convictions. For context, the overall increase in aggression at ages 5-9 was estimated to be 17 percent of a standard deviation, but our analysis indicated this increase is associated with children who already had elevated levels of this trait—those who may be most at risk for future trouble. Finally, the larger increases are for offences typical for the age—property crime and behavioral issues such as

[27] See, for example, Datta Gupta and Simonsen (2010), Felfe et al. (2015), Kottelenberg and Lehrer (2018), Baker and Milligan (2016), Bertrand and Pan (2013), Cornwell et al. (2013), Fortin et al. (2015) and Jacob (2002).

[28] Baseline crime rates are also higher for boys, but what matters here is not the share of crimes committed by boys but whether there is more criminal activity when there is a reduction in population non cognitive skills.

failure to appear in court and breaches of probation. This all said, we also emphasize that our estimates rely on assumptions about common trends across provinces, years, and cohorts which we have attempted to verify visually and statistically test.

We also note that the magnitudes we report are not outside the range found in previous studies, which also suggest large impacts of policy interventions on crime. For example, Lochner and Moretti (2004) estimate that each year of schooling reduces the probability of imprisonment by 0.1 percentage points (off a base of 0.8 percent) for whites and 0.3-0.5 percentage points (off a base of 3.6 percent) for blacks. Heller et al. (2017) find that a 27-week cognitive behavioral therapy program for teens in Chicago reduced total arrests by 28 percent and violent crime arrests by 50 percent. More directly related to early childhood, Heckman et al. (2013) report a drop of 2.3 arrests by age 27 for the Perry Preschool treated group, which was a 43% drop compared to the control group.

VI. Conclusions

The rapid growth in the labor force participation of mothers of young children has led to a strong policy interest in expanding access to non parental child care, with a particular focus on "universal" child care availability. Although that term has come to take many meanings, the best example in North America is clearly the program introduced in Quebec in the late 1990s. This program made child care much cheaper for all residents and led to an enormous expansion in use of child care by the population. Previous work has shown this policy change led to a large decline in measured non cognitive skills among young children exposed to the subsidized child care. We use this initial negative shock to assess whether the negative impact persists on longer-run outcomes in a way that mirrors findings in

the literature that link positive early-life interventions to positive longer-run outcomes.

Indeed, our evidence is generally consistent with such symmetry. We find the Quebec policy had a lasting negative impact on non cognitive skills. At older ages, program exposure is associated with worsened health and life satisfaction, and increased rates of criminal activity. Increases in aggression and hyperactivity are concentrated in boys, as is the rise in the crime rates. In contrast, we find no consistent impact on their cognitive skills.

The implications of these findings for early child care policy are potentially profound. Our findings provide strong support for the argument that the early childhood development environment is a crucial determinant of the long-term success of children. This suggests that measuring the contemporaneous impact of child care programs on development indicators is important because it is predictive of later-life success.

For policy makers, an unanswered question is whether the evidence of negative impacts is particular to the Quebec program, or whether the lessons here apply more broadly. Our findings for young children clearly contrast with evaluations of targeted programs like Perry, Abecedarian, and Head Start. The evidence we cite suggests the quality in the Quebec program is comparable to international norms, but likely worse quality than model programs such as Perry. Also, the Quebec program is universal, and the more granular analysis of, for example, Kottelenberg and Lehrer (2010) suggests that it had some positive impacts for more disadvantaged children which are offset at the mean by negative impacts for more advantaged families. If a universal program can be designed to improve contemporaneous impacts on all children, our results together with evidence such as Heckman et al. (2013) suggest such a program could lead to long-run positive outcomes.

REFERENCES

Almond, Douglas, and Janet Currie. 2011. "Human Capital Development Before Age Five." In *Handbook of Labor Economics*, Volume 4 Part B, edited by David Card and Orley Ashenfelter, 1315-1486. Amsterdam: North Holland.

Almond, Douglas, Janet Currie, and Valentina Duque. Forthcoming. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature.*

Almond, Douglas, Lena Edlund, and Mårten Palme. 2009. "Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden." *The Quarterly Journal of Economics* 124(4): 1729-1772.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103(484): 1481-1495.

Baker, Michael. 2011. "Innis Lecture: Universal Early Childhood Interventions: What Is the Evidence Base?" *Canadian Journal of Economics* 44(4): 1069-105.

Baker, Michael, and Kevin Milligan. 2010. "Evidence From Maternity Leave Expansions of the Impact of Maternal Care on Early Child Development." *Journal of Human Resources* 45(1): 1-32.

Baker, Michael, and Kevin Milligan. 2016. "Boy-Girl Differences in Parental Time Investments: Evidence from Three Countries." *Journal of Human Capital* 10(4): 399-441.

Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2005. "Universal Childcare, Maternal Labor Supply, and Family Well-Being." National Bureau of Economic Research Working Paper 11832.

Baker, Michael., Jonathan Gruber, and Kevin Milligan. 2008. "Universal Child Care, Maternal Labor Supply, and Family Well-Being." *Journal of Political Economy* 116(4): 709-45.

Bala, Nicholas, Peter J. Carrington, and Julian V. Roberts. 2009. "Evaluating the Youth Criminal Justice Act after Five Years: A Qualified Success." *Canadian Journal of Criminology and Criminal Justice* 51(2): 131-167.

Belfield, Clive R., Milagros Nores, Steve Barnett, and Lawrence Schweinhart. 2006. "The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Followup." *Journal of Human Resources* 41(1): 162-190.

Bertrand, Marianne, and Jessica Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5(1): 32-64.

Bester, Alan C., Timothy G. Conley, and Christian B. Hansen. 2011. "Inference with dependent data using cluster variance estimators." *Journal of Econometrics* 165(2): 137-151.

Bitler, Marianne P., Hilary W. Hoynes, and Thurston Domina. (2014) "Experimental Evidence on Distributional Effects of Head Start," NBER Working Paper No. 20434.

Black, Sandra E., Paul J. Devereux, Katrine V. Løken, and Kjell G. Salvanes. 2014. "Care or Cash? The Effect of Child Care Subsidies on Student Performance." *The Review of Economics and Statistics* 96(5): 824-837.

Brodeur, Abel, and Marie Connolly. 2013. "Do higher child care subsidies improve parental well-being? Evidence from Quebec's family policies." *Journal of Economic Behavior & Organization* 93: 1–16.

Carneiro, Pedro, and James J. Heckman. 2003. "Human capital policy." In *Inequality in America: What Role for Human Capital Policies?*, edited by James J. Heckman and Alan B. Krueger, 77-240. Cambridge, MA: MIT Press.

Carneiro, Pedro, and Rita Ginja. 2014. "Long Term Impacts of Compensatory Pre-School on Health and Behavior: Evidence from Head Start." *American Economic Journal: Economic Policy* 6(4): 135-73.

Carrington, Peter J., and Jennifer Scholenberg. 2005. "The Impact of the Youth Criminal Justice Act on Police Charging Practices with Young Persons: A Preliminary Statistical Assessment." Report to the Department of Justice Canada, Ottawa Canada.

Cascio, Elizabeth U. 2015. "The Promises and Pitfalls of Universal Early Education." *IZA World of Labor* 116.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. Forthcoming. "Who Benefits from Universal Child Care? Estimating Marginal Returns to early Child Care Attendance." *Journal of Political Economy*.

Cornwell, Christopher, David B. Mustard, and Jessica Van Parys. 2013. "Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School." *The Journal of Human Resources* 48(1): pp. 238-266.

Datta Gupta, Nabanita and Marianne Simonsen. 2010. "Non Cognitive Child Outcomes and Universal High Quality Child Care." *Journal of Public Economics* 94(1-2): 30-43.

Felfe, Christina, Natalia Nollenberger, Núria Rodríguez-Planas. 2015. "Can't Buy Mommy's Love? Universal Childcare and Children's Long-Term Cognitive Development." *Journal of Population Economics* 28(2): 393-422.

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux. 2009. "Unconditional quantile regressions." *Econometrica* 77(3): 953-973.

Fitzpatrick, Maria D. 2008. "Starting School at Four: The Effect of Universal

Prekindergarten on Children's Academic Achievement." *B.E. Journal of Economic Analysis and Policy* 8(1): 1-40.

Fortin, Nicole M., Philip Oreopoulos, and Shelley Phipps. 2015. "Leaving Boys Behind: Gender Disparities in High Academic Achievement." *Journal of Human Resources* 50(3): 549-579.

Gibbs, Chloe, Jens Ludwig, and Douglas L. Miller. 2013. "Head Start Origins and Impacts." In *Legacies of the War on Poverty*, edited by Martha J. Bailey and Sheldon Danziger, 39-65. New York: Russell Sage Foundation.

Gormley, William T. Jr., and Ted Gayer. 2005. "Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program." *Journal Human Resources* 40(3): 533–58.

Gormley, William T. Jr, Ted Gayer, Deborah Phillips, and Brittany Dawson. 2005. "The Effects of Universal Pre-K on Cognitive Development," *Developmental Psychology*, 41(6): 872–84.

Haeck, Catherine, Laetitia Lebihan, and Philip Merrigan. 2018. "Universal Child Care and Long-Term Effects on Child Well-Being: Evidence from Canada." *Journal of Human Capital* 12(1): 38-98.

Haeck, Catherine, Pierre Lefebvre, and Philip Merrigan. 2015. "Canadian Evidence on Ten Years of Universal Preschool Policies: The Good and the Bad," *Labour Economics* 36: 137-157.

Havnes, Tarjei, and Magne Mogstad. 2011. "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes." *American Economic Journal: Economic Policy* 3(2): 97-129.

Havnes, Tarjei, and Magne Mogstad. 2014. "Is Universal Child Care Leveling the Playing Field?" *Journal of Public Economics* 127: 100-114.

Heckman, James J., and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics* 6(1): 689-733.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the High Scope Perry Preschool Program." *Journal of Public Economics* 94: 114–28.

Heckman, James J., Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103(6): 2052-2086.

Heckman, James J., Jeffrey Smith, and Nancy Clements. 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *The Review of Economic Studies* 64(4): 487-535.

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil

Mullainathan, and Harold A. Pollack. 2017. “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago.” *The Quarterly Journal of Economics* 132(1): 1-54.

Hustedt, Jason T., W. Steven Barnett, Kwanghee Jung, and Alexandra Figueras. 2008. “Impacts of New Mexico Pre-K on Children’s School Readiness at Kindergarten Entry: Results from the Second Year of a Growing Initiative.” National Institute for Early Education Research, Rutgers University.

Jacob, Brian A. 2002. “Where the Boys Aren’t: Non Cognitive Skills, Returns to School and the Gender Gap in Higher Education.” *Economics of Education Review* 21(6): 589–598.

Japel, Christa, Richard E. Tremblay, and Sylvana Cote. 2005. “Quality Counts: Assessing the Quality of Daycare Services Based on the Quebec Longitudinal Study of Child Development” *IRPP Choices* 11(5): 1-42.

Kline, Patrick, and Christopher R. Walters. 2016. “Evaluating Public Programs with Close Substitutes: The Case of Head Start,” *The Quarterly Journal of Economics* 131(4): 1795-1848.

Kottelenberg, Michael J., and Steven F. Lehrer. 2013. “New Evidence on the Impacts of Access to and Attending Universal Child-Care in Canada.” *Canadian Public Policy* 39(2): 263-285.

Kottelenberg, Michael J., and Steven F. Lehrer. 2014. “Do the Perils of Universal Child Care Depend on the Child’s Age?” *CESifo Economic Studies* 60(2): 338-365.

Kottelenberg, Michael J., and Steven F. Lehrer. 2017. "Targeted or Universal Coverage? Assessing Heterogeneity in the Effects of Universal Child Care." *Journal of Labor Economics* 35(3): 609-653.

Kottelenberg, Michael J., and Steven F. Lehrer. 2018. “Does Quebec's Subsidized Child Care Policy Give Boys and Girls an Equal Start?” *Canadian Journal of Economics* 51(2): 627-659.

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

Lefebvre, Pierre, Philip Merrigan, and Matthieu Verstraete. 2009. "Dynamic Labour Supply Effects of Childcare Subsidies: Evidence from a Canadian Natural Experiment on Universal Child Care." *Labour Economics* 16(5): 490-502.

Lochner, Lance. 2011. “Nonproduction Benefits of Education: Crime, Health, and Good Citizenship.” In *Handbook of the Economics of Education*, Vol. 4, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 183-282. Amsterdam: Elsevier Science.

Lochner, Lance, and Enrico Moretti. 2004. “The Effect of Education on Crime:

Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review* 94(1): 155-189.

Milligan, Kevin, and Mark Stabile. 2011. "Do Child Tax Benefits Affect the Wellbeing of Children? Evidence from Canadian Child Benefit Expansions." *American Economic Journal: Economic Policy* 3(3): 175-205.

Molnar, Timea. 2017. "How Do Mothers Manage? Universal Daycare, Child Skill Formation, and the Parental Time-Education Puzzle." Ph.D. Thesis, University of British Columbia.

Trepanier, Jean. 2004. "What Did Quebec Not Want: Opposition to the Adoption of the Youth Crime Justice Act in Quebec." *Canadian Journal of Criminology and Criminal Justice* 46(3): 273-299.

Vandell Deborah Lowe, Jay Belsky, Margaret Burchinal, Nathan Vandergrift, and Lawrence Steinberg. 2010. "Do Effects of Early Child Care Extend to Age 15 Years? Results from the NICHD Study of Early Child Care and Youth Development." *Child Development*, 81(3): 737-756.

Wong, Vivian C., Thomas D. Cook, W. Steven Barnett, and Kwanghee Jung. 2007. *An Effectiveness-based Evaluation of Five State Pre-Kindergarten Programs using Regression-Discontinuity*. New Brunswick, NJ: National Institute for Early Education Research.

Zhang, Haimin. 2014. "Immigration and Crime: Evidence from Canada," Canadian Labour Market and Skills Research Network Working Paper 135.

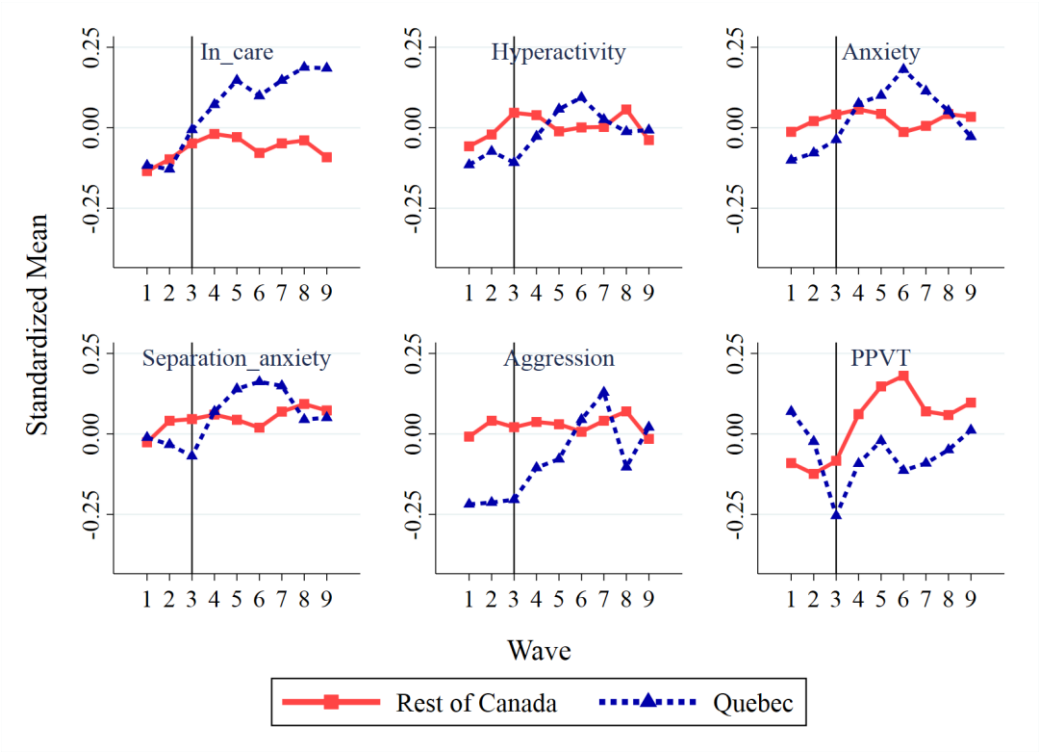

FIGURE 1. TIME TRENDS IN STANDARDIZED PRESCHOOL OUTCOMES

Notes: Authors' calculations from NLSCY/SYC data. The graph shows the mean standardized value for each of the outcomes in Table 1 across time in Quebec and the rest of Canada. The variable in care shows the deviation from the average value of the variable measured across all children age 0-4. The other variables are standardized using the mean and standard deviation.

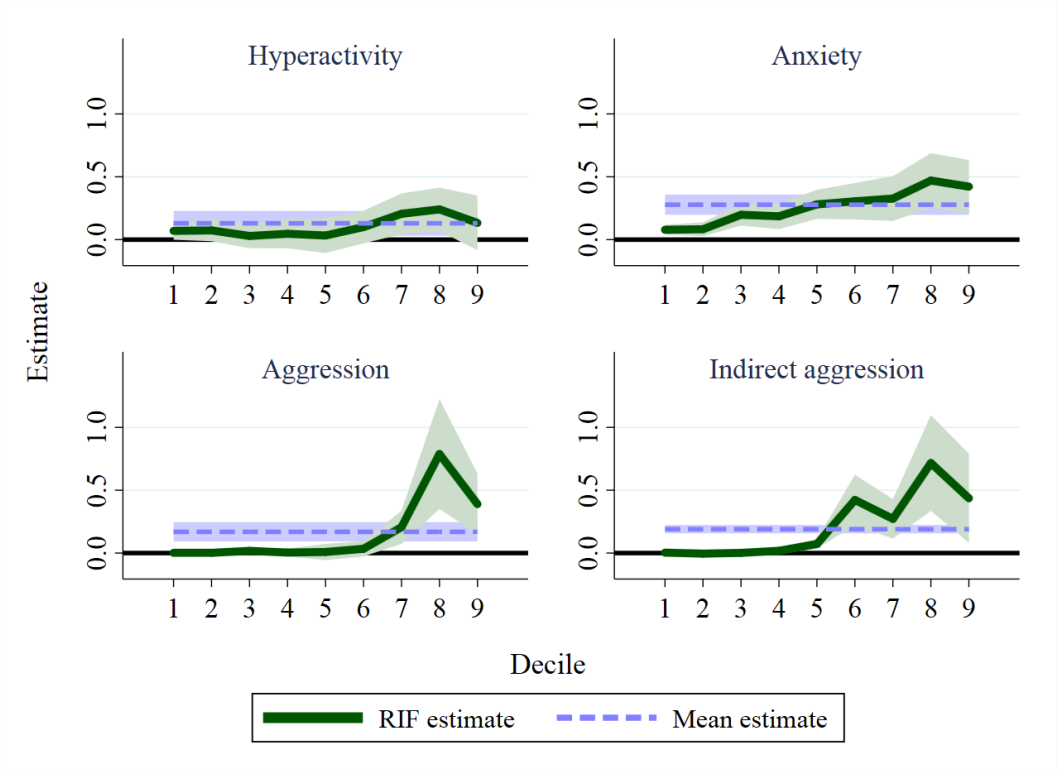

FIGURE 2. RIF REGRESSION ANALYSIS OF NON COGNITIVE SCORES, AGES 5-9

Notes: Authors' calculations from NLSCY (cycles 1, 2, 7) and the SYC (cycle 9). Displayed are the mean and RIF estimates for longer-run outcomes at ages 5-9. The dependent variable is scaled to mean zero and a unit standard deviation, so the estimates can be interpreted as fractions of a standard deviation. The mean estimates are from Table 2. The RIF estimates vary by decile cutoff and are described in the text. We show shaded 95 percent confidence intervals around each estimate.

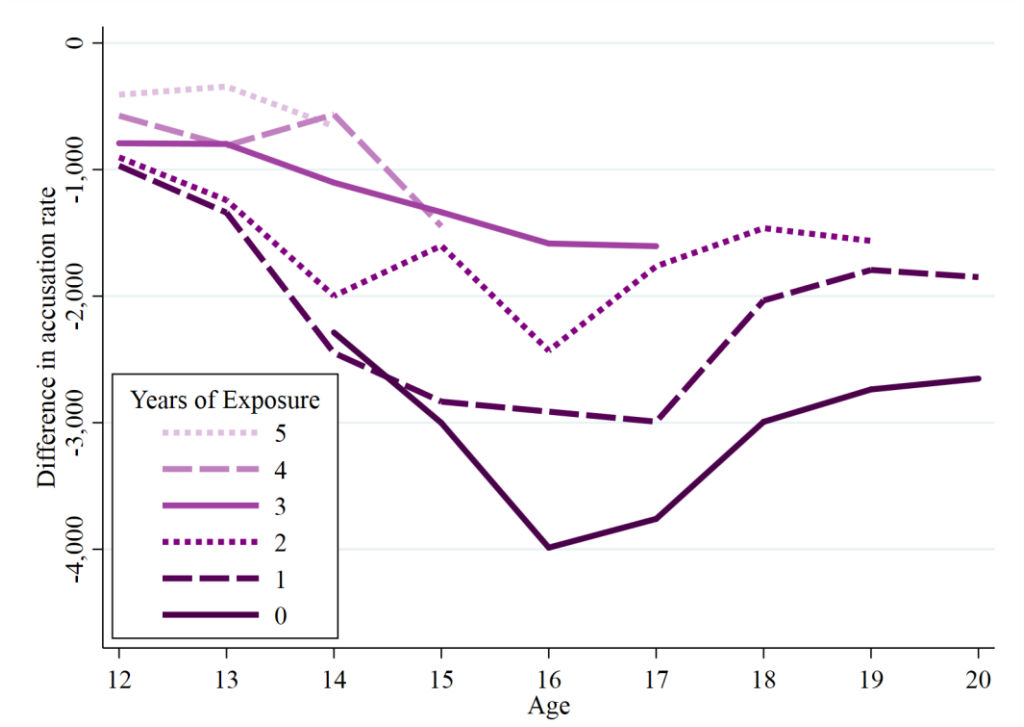

FIGURE 3. QUEBEC-REST OF CANADA DIFFERENCES IN ACCUSATION RATES BY BIRTH COHORT

Notes: Authors' calculations from UCR data. Displayed is the difference in the annual accusation rate for all crime types per 100,000 of population between Quebec and the Rest of Canada by age. Each line shows a different set of birth cohorts, arranged by years of exposure to eligibility.

TABLE 1—IMPACT OF EXPOSURE TO THE QUEBEC FAMILY PLAN ON CHILD CARE AND MEASURES OF NON COGNITIVE AND COGNITIVE OUTCOMES AT YOUNG AGES

Outcome	Mean	EXPOSURE	
Cycles	1,2,4,5	1,2,4,5	1,2,4-9
In Care	0.46	0.143	0.197
	(0.50)	[0.031]***	[0.026]***
		(0.008)***	(0.011)***
Hyperactivity	2.86	0.065	0.070
	(2.12)	[0.048]	[0.035]*
		(0.016)**	(0.023)**
Anxiety	1.25	0.115	0.119
	(1.50)	[0.027]***	[0.033]***
		(0.013)***	(0.013)***
Separation Anxiety	2.76	0.073	0.063
	(2.03)	[0.047]	[0.042]
		(0.021)***	(0.016)**
Aggression	4.98	0.117	0.172
	(2.95)	[0.040]**	[0.043]***
		(0.024)**	(0.006)***
PPVT	99.98	-0.109	-0.139
	(15.29)	[0.045]**	[0.041]***
		(0.039)*	(0.026)***

Notes: Authors' calculations from NLSCY/SYC data. Sample includes all families. The sample ages are 0-4 years for In Care; 2-3 years for (standardized) Hyperactivity, Anxiety, Separation Anxiety and Aggression; and ages 4-5 for (standardized) PPVT. Reported is the coefficient on a dummy indicating exposure to eligibility. Robust standard errors clustered on province and year of birth are in square brackets. Standard errors using the Bester et al. (2011) method are reported in round brackets. Significance is reported using p-values adjusted for multiple testing.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 2—IMPACT OF EXPOSURE TO THE QUEBEC FAMILY PLAN ON MEASURES OF NON COGNITIVE OUTCOMES, AGES 5-9

Outcome	Mean	EXPOSURE
Hyperactivity	4.09	0.130
	(3.14)	[0.050]**
		(0.008)***
Anxiety	2.49	0.278
	(2.30)	[0.041]***
		(0.022)***
Aggression	1.36	0.169
	(1.82)	[0.039]***
		(0.009)***
Indirect Aggression	1.00	0.190
	(1.55)	[0.017]***
		(0.021)***
Prosocial	13.48	0.006
	(3.90)	[0.048]
		(0.038)
Child gets along with Teacher (parent report)	0.80	-0.084
	(0.40)	[0.023]***
		(0.009)***
Hyperactivity	4.09	0.130
	(3.14)	[0.050]**
		(0.008)***

Notes: Authors' calculations from NLSCY (cycles 1, 2, 7) and the SYC (cycle 9). Sample includes all families. Reported is the coefficient on a dummy indicating exposure to eligibility. Robust standard errors clustered on province and year of birth are in square brackets. Standard errors using the Bester et al. (2011) method are reported in round brackets. Significance is reported using p-values adjusted for multiple testing.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 3—IMPACT OF EXPOSURE TO THE QUEBEC FAMILY PLAN ON STANDARDIZED TEST SCORES

	Math		Reading		Science	
	Mean	EXPOSURE	Mean	EXPOSURE	Mean	EXPOSURE
SAIP/PCAP	0.125	-0.181	0.107	-0.237	0.060	-0.062
	(0.986)	[0.118]	(1.000)	[0.181]	(0.990)	[0.059]
		(0.131)		(0.090)		(0.044)
PISA	0.119	0.312	0.144	0.106	0.122	0.043
(2009 control)	(0.998)	[0.051]***	(0.973)	[0.051]*	(0.991)	[0.083]
		(0.015)***		(0.018)**		(0.014)*
PISA	0.119	0.199	0.144	0.054	0.122	0.021
(2009 treated)	(0.998)	[0.084]*	(0.973)	[0.048]	(0.991)	[0.086]
		(0.010)***		(0.012)**		(0.021)

Notes: Authors' calculations from SAIP/PCAP and PISA test score data. Sample includes all families. Reported is the coefficient on a dummy indicating exposure to eligibility. Robust standard errors clustered on province and year of birth are in square brackets. Robust standard errors clustered on province and year of birth are in square brackets. Standard errors using the Bester et al. (2011) method are reported in round brackets. Significance is reported using p-values adjusted for multiple testing.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 4—IMPACT OF EXPOSURE TO THE QUEBEC FAMILY PLAN ON SELF-REPORTED HEALTH OUTCOMES

	Age 12-20	
	Mean	EXPOSURE
Health	2.09	0.073
	(0.85)	[0.019]***
		(0.005)***
Life Satisfaction	1.61	0.043
	(0.62)	[0.017]**
		(0.006)***
Mental Health	1.90	-0.011
	(0.88)	[0.020]
		(0.005)*

Notes: Authors' calculations from CCHS data. Sample includes all families. Reported is the coefficient on a dummy indicating exposure to eligibility. Robust standard errors clustered on province and year of birth are in square brackets. Robust standard errors clustered on province and year of birth are in square brackets. Standard errors using the Bester et al. (2011) method are reported in round brackets. Significance is reported using p-values adjusted for multiple testing.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 5—IMPACT OF EXPOSURE TO THE QUEBEC FAMILY PLAN ON CRIME RATES, AGES 12-20

	Mean	(1)	(2)	(3)
Panel A. Accused				
All	1872	514 [77]*** (181)**	590 [60]*** (71)**	353 [69]*** (35)***
Person	1839	530 [84]*** (171)*	649 [68]*** (90)**	299 [81]*** (70)**
Property	3100	566 [100]*** (190)*	932 [132]*** (193)**	602 [172]*** (11)***
Other CC	1604	639 [121]*** (396)	563 [64]*** (79)**	379 [68]*** (110)*
Drugs	945	322 [73]*** (54)**	217 [27]*** (67)**	130 [28]*** (41)*
Panel B. Convictions				
All	963	208 [37]*** (47)**	323 [39]*** (26)***	212 [44]*** (5)***
Person	1002	291 [51]*** (62)**	323 [43]*** (11)***	167 [59]*** (27)**
Property	1363	112 [61]* (30)**	455 [70]*** (41)***	342 [93]*** (31)***
Other CC	1059	289 [57]*** (117)*	311 [50]*** (100)*	239 [54]*** (15)***
Drugs	429	140 [27]*** (35)**	203 [26]*** (43)**	99 [29]*** (25)**

Notes: Authors' calculations from UCR data. Sample includes all families. In rows titled (1) are estimates from the difference in differences specification. In rows titled (2) are estimates that add all second order province, age, gender, year interactions, expect year*prov. In rows titled (3) are estimates that add province, year linear trend interactions. Reported is the coefficient on a dummy indicating exposure to eligibility. Robust standard errors clustered on province and year of birth are in square brackets. Standard errors using the Bester et al. (2011) method are reported in round brackets. Significance is reported using p-values adjusted for multiple testing.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.