Canadian Labour Economics Forum

WORKING PAPER SERIES

A Welfare Analysis of Universal Childcare: Lessons From a Canadian Reform

Sébastien Montpetit (Toulouse University)
Pierre-Loup Beauregard (University of British Columbia)
Luisa Carrer (Toulouse University)

CLEF WP #73

A Welfare Analysis of Universal Childcare: Lessons From a Canadian Reform*

Sébastien Montpetit† Pierre-Loup Beauregard‡ Luisa Carrer§

July 2024

Click here for the latest version.

Abstract

Leveraging the introduction of universal low-fee daycare in Québec in 1997, we assess the welfare effect of universal childcare provision. First, using novel data on local daycare coverage and a difference-in-differences design, we show that positive impacts on maternal labor supply and childcare use are greater in areas with larger daycare expansion, suggesting that childcare availability, not just affordability, drives these responses. We then estimate the policy's Marginal Value of Public Funds (MVPF), defined as the ratio of beneficiaries' utility gains to net governmental costs. Unlike the standard sufficient-statistics metric, which assumes a marginal change in fiscal policy, we quantify the beneficiaries' utility gains through a model of maternal labor supply and childcare choices. This allows us to relax the common marginal-policy assumption and to incorporate non-pecuniary benefits for parents. Our results indicate substantial welfare gains from universal policies, with approximately \$3.5 of benefits per dollar of net government spending – over twice the amount captured by the sufficient-statistics metric. Counterfactual simulations suggest that allocating more resources to increasing availability, rather than improving affordability, could yield even larger social returns.

Keywords: universal childcare, daycare coverage, social welfare, sufficient statistics

JEL Codes: H43, J13, J22

---

*We thank Matteo Bobba, Philippe Bontems, Marie Connolly, Nicole Fortin, Jonas Gathen, Christian Hellwig, Jean-William Laliberté, Pierre Lefebvre, Thomas Lemieux, Philip Merrigan, Tímea Laura Molnár, François Poinas, Mohamed Saleh, Katherina Thomas, Sarah Vincent, as well as workshop participants at TSE, UBC, ESG-UQÀM, CIREQ, Naples, and attendees at the SCSE, CEA, and SBCA annual conferences for valuable suggestions and comments. We further thank Pascal Doray-Demers, Wendy Kei, and Chunling Fu for their help with confidential data access and support in the UBC and UQÀM Canadian Research Data Centres. We are also indebted to Nicole Fortin and Marie Connolly whose invitations allowed Montpetit to work on this paper in Canadian Research Data Centres. Part of this research was conducted at the Quebec Interuniversity Centre for Social Statistics (QICSS) and at the UBC RDC, part of the Canadian Research Data Centre Network (CRDCN). This service is provided through the support of QICSS' Member Universities, the province of Quebec, the Canadian Foundation for Innovation, the Canadian Institutes of Health Research, the Social Science and Humanity Research Council, the Fonds de Recherche du Québec, and Statistics Canada. All views expressed in this work and remaining errors are our own.

†Toulouse School of Economics. Email: sebastien.montpetit@tse-fr.eu.

‡University of British Columbia. Email: pierreloup.beauregard@gmail.com.

§Toulouse School of Economics. Email: luisa.carrer@tse-fr.eu.

1 Introduction

Over the past century, the rise in female labor force participation has been one of the most significant transformations in the labor market. This "quiet revolution" has intensified challenges for parents with young children in balancing employment and family duties, as child-rearing responsibilities have traditionally fallen upon women (Goldin, 2006). In response, many governments have proposed expanding childcare programs to all, aiming to reduce the opportunity cost of employment for mothers and support their participation in the labor force.

However, little is known about the overall effects of universal childcare programs on society. While targeted early-childhood programs for disadvantaged families have demonstrated substantial benefits,[1] it is unclear if these positive outcomes will extend to universal programs (Baker, 2011; List et al., 2021; List, 2022). Evidence on universal childcare reforms is mixed, with outcomes varying from positive to negative across different contexts.[2] Moreover, universal programs require substantial public expenditures.[3] Due to the high costs and differences in estimated benefits, the social desirability of these policies remains unknown.

In this paper, we quantify the welfare impact of universal childcare provision by exploiting a major policy change in Québec in 1997, which introduced universal subsidized daycare and is often deemed the most ambitious childcare reform in North America (Baker et al., 2019). To this end, we calculate the policy's Marginal Value of Public Funds (MVPF), defined as the ratio of the beneficiaries' utility gains over the net governmental costs (Finkelstein and Hendren, 2020). This task is challenging for two main reasons. First, beyond earnings gains, an in-kind policy such as childcare provision can affect mothers' welfare through various other channels. Non-maternal care frees up time for market work or leisure but places the child in a different environment, impacting its human capital development. Early childhood development has long-term economic consequences.[4] While mothers likely enjoy time with their children, providing development-enhancing care can be exhausting (Chaparro et al., 2020). Additionally, increased availability reduces non-monetary costs of childcare, such as commuting time and the effort to find care when supply is limited (Bravo et al., 2022; De Groote and Rho, 2023). Second, universal programs represent a significant (non-marginal) change in the economic environment.

For these reasons, standard sufficient statistics for welfare analysis, which assume infinitesimal changes in fiscal policy, would be biased (Kleven, 2021). Our approach relaxes

[1]See, among others, Heckman et al. (2010); García et al. (2020); Bailey et al. (2021); García et al. (2023). A notable exception is Cascio (2023), who finds no average impact of targeted preschool enrollment on children's test scores in the United States.

[2] Policy impacts vary substantially depending on the specific context studied. For example, maternal labor supply effects depend largely on the counterfactual childcare market (Kline and Walters, 2016; Karademir et al., 2023). In contexts with a high prevalence of informal care (by grandparents, siblings, etc.), studies find little impact on parents' labor supply (Fitzpatrick, 2010; Havnes and Mogstad, 2011a,b; Kleven et al., 2024). Conversely, policies that crowd out parental care hours typically have positive impacts on maternal employment (Gelbach, 2002; Baker et al., 2008; Herbst, 2017; Hojman and Lopez Boo, 2022). Regarding child development, the impact depends primarily on program care quality and children's socio-economic status (Havnes and Mogstad, 2015; Kottelenberg and Lehrer, 2017; Felfe and Lalive, 2018; Cornelissen et al., 2018; Fort et al., 2020). See Hotz and Wiswall (2019) and Duncan et al. (2023) for reviews of the literature on subsidized childcare provision.

[3]For example, in his Covid Recovery Plan, US President Joe Biden announced $39 billion in investments specifically for childcare (The White House, 2021). Similarly, the current Canadian expansion, aiming to reduce daily daycare fees to $10 per day by 2026, has committed $30 billion in federal investments (Seward et al., 2023).

[4]See Almond et al. (2018) for a review of the literature on long-term impacts of childhood circumstances and Duncan et al. (2023) for a review of impacts of public investments in early childhood.

these assumptions, often implicitly made in studies focusing on beneficiaries' earnings gains (e.g. Haeck et al., 2018; Andresen and Havnes, 2019, for universal childcare reforms). Contrary to the benchmark sufficient-statistics metric, we compute beneficiaries' utility gains by estimating a simple model of maternal time allocation and childcare choices. Intuitively, we infer mothers' willingness-to-pay by simulating the reform into our estimated model, allowing beneficiaries to re-optimize their behavior. This method accounts for the in-kind nature of the transfer and the non-marginal nature of the policy.

In the first part of the paper, we estimate the reduced-form policy impacts on maternal labor supply and children's outcomes. Using novel data on regional daycare coverage rates within Québec we manually digitized, we estimate the policy's heterogeneous effects by local daycare supply. We employ an intent-to-treat (ITT) difference-in-differences approach, comparing mothers of young eligible children in Québec to their counterparts in the rest of Canada. We provide several pieces of evidence suggesting that the increase in local daycare coverage has the characteristics of a supply shock and is plausibly exogenous to parents' childcare demand, reinforcing our confidence in the main identification assumption of our analysis. Our findings indicate that the local expansion of daycare supply, not just the decrease in prices, is an important channel of impact on childcare use and maternal labor supply. In regions where daycare supply increased the most (defined as the top two terciles of daycare coverage expansion), maternal employment increased by 67% and childcare use by 38% more, even after controlling for regional-level covariates. These results suggest that increasing local daycare supply is key to the effectiveness of preschool policies.

Next, we estimate mothers' earnings gains to obtain a benchmark value of the policy's benefits following the standard sufficient-statistics approach. Specifically, we compute quantile treatment effects to estimate earnings gains across mothers' income distribution. Taking into account the effect of this heterogeneity on fiscal returns to the government, we estimate the implied fiscal externality (i.e., the return to the government from mothers' additional earnings) and calculate the policy's marginal value of public funds (MVPF). The MVPF is the ratio of the beneficiaries' willingness-to-pay (WTP) for a policy over its net cost to the government (net of fiscal externalities). When considering only mothers' earnings gains, we find a benchmark MVPF of 1.42, meaning that an additional dollar of net government spending generates $1.42 in mothers' earnings. This result indicates that the policy is welfare-improving. Nevertheless, this MVPF is relatively modest compared to targeted preschool interventions studied in Hendren and Sprung-Keyser (2020), which have MVPFs above 5.

Our reduced-form analysis also examines the impact of the policy on eligible children's educational attainment later in life as well as on fathers' labor supply. We find that the negative impacts on child behavior documented by Baker et al. (2008, 2019) do not translate into worse economic outcomes later in life. We additionally confirm null impacts on fathers' employment. This suggests the absence of fiscal impacts stemming from eligible children's economic outcomes in the long run and from fathers' earnings.[5]

[5]Nevertheless, behavioral problems could impact the government budget through other channels. In robustness exercises, we calculate the potential costs associated with increased youth criminal activity found in Baker et al. (2019), using recent estimates of the costs of crime that consider victimization costs and productivity losses. Given the relatively "benign" nature of typical juvenile crimes, this additional societal cost turns out to be

In the second part of the paper, we move beyond earnings gains to incorporate non-pecuniary benefits for mothers. We specify and estimate a model of maternal labor supply and childcare choices to infer mothers' WTP for the policy change. Our model integrates supply shortages into a model where a mother must meet childcare and time constraints while caring about the child's human capital accumulation, following Chaparro et al. (2020). The model captures key trade-offs families face, such as balancing employment and care, and deciding how much parenting effort to exert at home.

To assess our model's validity, we verify that it replicates the ITT impacts on maternal labor supply and childcare use. We simulate the policy's main features in our model, estimated using pre-reform data, and compare the simulated changes in mothers' choices to the reduced-form estimates. This comparison serves as an out-of-sample validation, demonstrating the model's accuracy in predicting policy impacts and supporting our structural assumptions on behavior. Additionally, our estimation algorithm leverages causal estimates from the reduced-form analysis to directly identify key model parameters. Specifically, we show how the Québec natural experiment can serve as an instrument for maternal care hours to identify a (potentially) non-linear cost of parenting effort.

We find that accounting for non-monetary benefits and the non-marginal nature of the policy yields a WTP more than twice as large as when considering earnings gains alone. Our structural estimator yields an MVPF estimate of 3.56, indicating substantial social returns from the policy. Our simulations reveal that only about 37% of utility gains are attributed to labor-market choices, implying that most of the increase in welfare stems from non-pecuniary benefits. These findings suggest that universal preschool policies can generate significant social returns. Furthermore, focusing on earnings gains alone would substantially underestimate the policy's benefits.

In the last part of the paper, we use our structural model of behavior to perform counterfactual analyses, to identify which features of the reform drive most of the welfare gains and provide insights on the optimal policy scheme. By removing each feature of the reform one by one, we evaluate which policy component yields the largest increase in mothers' WTP. Specifically, we compare the WTP for simulated counterfactual reforms where *(i)* childcare prices remain unchanged, *(ii)* there is no increase in daycare coverage, and *(iii)* the refundable childcare credit is maintained.[6] We find that maintaining the relatively higher pre-reform childcare price reduces the WTP for the reform by only 16%. However, reducing the daycare price without increasing coverage causes a substantial drop in WTP. Similarly, maintaining the refundable credit has little impact due to the small net price reduction it represents under the 5$/day regime. This suggests that most of the welfare gains are due to increased coverage.

Finally, we explore whether the government could have achieved even larger welfare gains under different policy schemes. We compare the Québec childcare reform to alternative levels of price reductions and daycare expansions. Consistent with our previous results indicating that

relatively small compared to mothers' gains in this context.

[6] To finance some of the policy's costs, the Québec government also modified other family benefits. The most notable fiscal change was the abolition of a refundable childcare credit for families using a low-fee space, which reimbursed a share of childcare costs to claimant families. See Section 2.1 and Baker et al. (2005) for additional details on these changes.

parents are willing to pay substantially more for increased coverage than for reduced prices, we find significantly higher MVPFs for reforms that invest more public funding into increasing availability rather than improving affordability. Thus, our results suggest that higher welfare gains could be achieved by allocating more resources towards creating additional daycare spots rather than lowering childcare fees.

This paper contributes to three strands of literature. First, we provide new evidence on the impact of universal childcare provision. In this rapidly growing literature, most studies have focused on measuring the causal impact of preschool enrollment on child development or maternal labor supply, yielding mixed evidence (see footnote 2). However, little is known about the overall societal implications of such policy changes. Notable exceptions, such as Guner et al. (2020), Daruich (2022), and Borowsky et al. (2022), estimate general-equilibrium models of the family to study impacts of childcare programs.

Unlike these studies, our work evaluates the welfare effect of an implemented reform rather than a hypothetical policy scheme. Additionally, we consider a broader range of potential non-pecuniary benefits for mothers, including the enjoyment parents derive from time spent with their children.[7] In the context of the Québec policy, we show that public provision of low-fee preschool can generate positive returns to society, even with a more diverse pool of beneficiaries compared to targeted programs. Moreover, we demonstrate that negative short-run impacts on non-cognitive outcomes (Baker et al., 2008, 2019; Haeck et al., 2015) do not necessarily translate into depressed economic outcomes later in life. This paper also highlights the role of local daycare supply in shaping the impacts of universal programs, consistent with Yamaguchi et al. (2018) for Japan and Cornelissen et al. (2018) for Germany. Lastly, we estimate mothers' WTP for the policy, including non-pecuniary gains, a dimension not addressed in previous cost-benefit analyses of universal childcare programs (e.g., Fortin et al. (2013); Haeck et al. (2018) for the same policy or Andresen and Havnes (2019) for Norway).

Second, we link sufficient statistics and structural approaches to empirically evaluate the extent of bias in sufficient-statistic estimators. Seminal papers in this literature show that, under standard assumptions, monetary gains are a sufficient statistic for beneficiaries' WTP in the case of sufficiently small fiscal policy changes (Chetty, 2009; Hendren, 2016).[8] More recently, a body of work has theoretically identified conditions under which transparent sufficient statistics for non-infinitesimal policy changes can be derived (Finkelstein and Hendren, 2020; Kleven, 2021). However, as Kleven (2021) argues, these conditions are often beyond empirical reach due to the complexity of expressing the welfare effect of large reforms as a fiscal externality. Kang and Vasserman (2022) propose welfare bounds for non-marginal reforms, but applying these bounds to policies that change many parameters in the economic environment, such as the Québec program, remains challenging.

We take an alternative approach and show how using a tractable structural model can indicate the extent of bias in sufficient-statistics methods applied to non-marginal reforms.

[7] In another related paper, Bravo et al. (2022) estimate the welfare effect of reduced distance to childcare centers induced by a national expansion in Chile, focusing on a marginal policy change. In contrast, we are interested in the overall effect of a large-scale reform.

[8] Hendren (2016) recognized that this result applies to small changes in *fiscal* policy but that pecuniary benefits are no longer sufficient if the policy in question changes the state of public good provision.

Our analysis indicates that for policies with substantial costs and significant non-pecuniary benefits, such as the Québec childcare reform, sufficient-statistic estimators can substantially underestimate welfare gains. This result has implications beyond childcare policy. Applying the sufficient-statistics framework to non-marginal reforms – a common practice as documented in our survey of MVPF estimates in Hendren et al.'s (2023) Policy Impacts Library – might significantly compromise welfare conclusions.[9] While it is well-established that the sufficient-statistic approach is biased for large policy changes, our results provide empirical insights into the magnitude of this bias. In our setting, this approach limits mothers' benefits to less than half of what we find using our structural estimator.

Third, we contribute to a growing literature that combines reduced-form ex post estimation of policy impacts and structural modelling.[10] While some recent studies, such as Griffen (2019) and Chaparro et al. (2020), specify structural models of the family to interpret experimental impacts of targeted childcare programs (Head Start and IHDP, respectively), we do so in the context of universal childcare provision. This paper also relates to Chan and Liu (2018) who study a different policy scheme, which provided cash transfers to stay-at-home mothers in Norway.[11] We contribute to this literature by showing that a tractable behavioral model, which incorporates non-pecuniary considerations of childcare decisions, can replicate the reduced-form impacts of the Québec program. Furthermore, we leverage the natural experiment to build a transparent identification argument for some key model parameters.

The rest of the paper is structured as follows. Section 2 describes the institutional background and the data we use in our empirical analysis. In Section 3, we present our reduced-form analysis. Section 4 presents our structural model of labor supply and childcare decisions and its estimation. Section 5 presents our estimates of the policy's MVPF as well as counterfactual simulations. Last, Section 6 concludes.

2 Background and Data

2.1 The Québec Childcare Reform

On September 1, 1997, a large-scale reform of preschool daycare was initiated by the provincial government of Québec, the second most populous province in Canada. At the time, the province was lagging behind the other Canadian provinces in terms of female labor-force participation. The major reform was thus designed to address this issue as well as to fight poverty and promote equality of opportunity for children (Japel et al., 2005). The centrepiece of the policy was the introduction of reduced-fare spaces in regulated childcare facilities at an out-of-pocket price of

[9]Our survey of MVPF estimates in the Policy Impacts Library suggests that computing the MVPF of large policy changes as if they were infinitesimal is common practice. We stress, however, that this exercise requires several judgment calls and should only be seen as suggestive that policies considered in this literature are often non-marginal. See Appendix E for detailed survey results.

[10]Scholars such as Heckman (2010), Todd and Wolpin (2023), and Buera et al. (2023) discuss the merits of this approach, which combines "the best of both worlds" in empirical research. On the one hand, structural models can help interpret the mechanisms through which a given policy change impacts relevant outcomes and allow for counterfactual policy experiments that deviate from the implemented policy. On the other hand, ex-post policy evaluation can be fruitfully used to identify and discipline behavioral models.

[11]Another relevant literature uses experimental results to estimate production functions of child cognition (e.g. Attanasio et al., 2020).

$5 per day per child (which increased to $7 in 2004).[12] Those low-fee spaces were allocated through the creation of a network of new regulated facilities named *Centres de la petite enfance* (CPEs). The reform was phased in by age of the child over a period of 4 years. Initially, only children aged 4 (as of September 30$^{\text{th}}$) were eligible. In the following years, the age requirement was gradually lowered: one year later, 3-year-olds became eligible, followed by 2-year-olds in September 1999. In September 2000, subsidized spaces became available to all children aged less than 59 months. Access to the program was universal so that there were no entry requirements such as labor-force participation. In other words, the only condition for eligibility was the age of the child.[13] Importantly, (gross) prices remained constant for parents until 2014 – with the exception of an increase to $7 per day per child from 2004 – with the provincial government subsidizing the remaining fees.

Eligibility for subsidized spaces, however, did not imply that parents would actually find a spot for their child. Indeed, there were important shortages of spaces, especially in the first years of implementation. There were long waiting lists at each regulated childcare facility. Figure 1, which shows the evolution of the daycare coverage rate in the province by administrative region, illustrates this low supply. In 2000, only 35% of children aged 1-4 had access to low-fee childcare services. The slow growth in the number of spaces at the beginning is in large part due to the government's decision to freeze the number of spaces in unregulated daycare.[14] To remain as for-profit entities, daycare providers could only sign an agreement with the government and open additional spaces at a reduced fee. The moratorium on the creation of for-profit daycares was lifted in June 2002, after which the for-profit market expanded.[15] New spaces kept being created at a fast pace over the following years, raising the share of children with access to subsidized spots to 65% in 2008 (Lefebvre et al., 2009). Despite successive governments' efforts to increase supply, shortages remain a reality nowadays.

The reform also included the abolition, for households with a subsidized space, of some universal family allowances as well as of a refundable childcare credit prior to the adoption of the policy.[16] The credit rate, shown in Appendix Figure A1, was decreasing with household income. As a consequence, the reform was most beneficial for middle- and high-income households and changed financial incentives mostly for those families, raising concerns about equity.

[12] The average daily gross price in March 1997 was approximately $21 (Office des services de garde à l'enfance, 997a). A reduced-fee space thus represents annual gross savings of more than $4,100 for a family signing a contract for the maximum number of days.

[13] To be eligible to the universal subsidy, families were required to enrol their child full-time for a maximum of 260 days per year. Families would typically sign yearly contracts to keep their space. The fees were billed monthly and had to be paid even if the child was absent from daycare. Note, nevertheless, that the for-profit market was not much more flexible, providers preferring to offer only full-time spaces because they were easier to manage (Haeck et al., 2018).

[14] Note that, due to long-run fertility trends, the daycare coverage rate (defined as the number of spaces per preschooler) did increase in the late 1990s even if the number of spaces stagnated somewhat.

[15] Until 2009, growth in the regulated network was still superior to that in the for-profit network. Data assembled by Haeck et al. (2016) which we complemented with recent years using ministerial reports for 2017 to 2019, however, shows that, from 2010, the for-profit network rapidly expanded as the regulated network stabilized. In the regulated network, the average annual number of newly created daycare spaces in the province from 2002 to 2009 was as high as 8,600 but dropped to 2,909 over the following decade. In contrast, between 2002 and 2009, 854 spaces were created annually in the for-profit network on average, but this figure dramatically increased to an annual growth of 6,322 spaces between 2010 and 2019.

[16] See Baker et al. (2005) for a detailed description of the changes to family allowances and other subsidies.

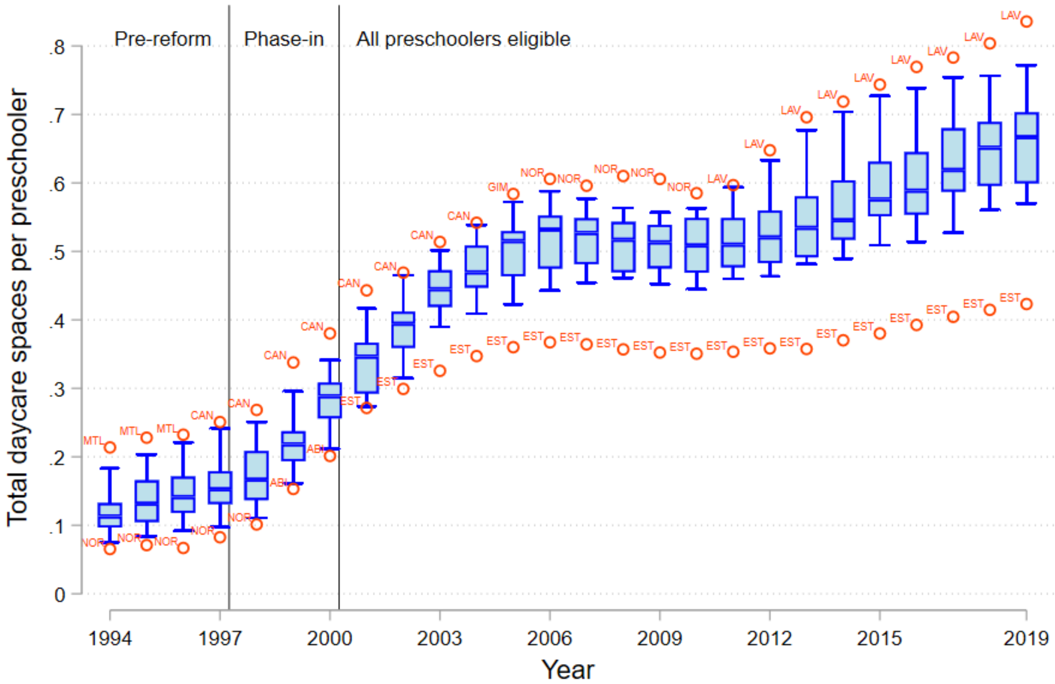


Figure 1: Evolution of the daycare coverage rate by administrative region, Québec

Note: This Figure displays the median (middle blue line), the 25th and 75th percentiles (light blue boxes), and the 10th and 90th percentiles (blue brackets), and outliers (orange circles) of the daycare coverage rate distribution at the administrative-region level in a given year. It is constructed using data from a series of ministerial reports for daycare spaces (*Ministère de la Famille*, which held various other names), from the Québec Statistical Institute for the children population, and from authors' calculations. The dependent variable is the ratio between the total number of spaces (the sum of subsidized and non-subsidized spaces) for preschool children (as of March 31st) over the number of children aged 0-4 years (as of July 1st) in a given region. Before the reform, low-income families had access to childcare subsidies and subsidies became universal in 1997. The reduced-fee program ($5/day/child) began in September 1997 only for children aged 4. All preschool children (0-4 years old) became eligible only in September 2000. In January 2004, the daily fee was raised to $7. ABI = Abitibi-Témiscamingue; CAN = Capitale-Nationale; EST = Estrie; GIM = Gaspésie-Îles-de-la-Madeleine; LAV = Laval; MTL = Montréal; NOR = Nord-du-Québec;

Quality of care. In addition to equity concerns, quality of care after the reform is one of its most controversial aspects. Given the rapid expansion of the market, maintaining sufficient quality was a challenge. Nevertheless, an audit study conducted by Japel et al. (2005) found that, despite quality being low in CPEs, they still outperformed all other childcare settings on the vast majority of the criteria they considered for quality on average. Some efforts were also made by the government to gradually increase investments in quality in the subsidized network. Appendix B provides additional details on childcare quality following the reform. [We could also mention that we care about quality as it has been shown that impacts on children's outcomes depend on care quality.]

Existing evidence on the Québec childcare reform. A substantial amount of work has been dedicated to the study of the impacts of the Québec reform on economic outcomes. Three patterns emerge from this literature. First, there are overall positive and large effects on childcare utilization and on maternal labor supply (Baker et al., 2008; Lefebvre and Merrigan, 2008).[17] Second, on average, eligible children experienced worse development outcomes (Baker et al., 2008; Kottelenberg and Lehrer, 2017; Haeck et al., 2015), with potentially long-lasting consequences (Baker et al., 2019).[18] Third, substantial heterogeneity in impacts on children were documented, most notably positive impacts among disadvantaged children (e.g. Kottelenberg and Lehrer, 2017, 2018). However, despite all the efforts invested in estimating the short- and medium-run impacts of the Québec reform on a variety of economic outcomes, we still know very little about the overall implications of the policy change. We bridge this gap by performing the first comprehensive welfare analysis of this major policy change.

2.2 Data sources

For our empirical analysis, which includes both the reduced-form ex post evaluation of the Québec reforms and the estimation of the structural model, we utilize several sources of Canadian micro-data. The main source is the National Longitudinal Survey of Canadian Youth (NLSCY), a common dataset in the literature. These data contain rich information on a representative sample of Canadian children and their parents over the period of the reform. We notably observe measures of care quality, labor-market participation of parents, and daycare expenditures, all of which are crucial to estimate parents' WTP for the policy. These repeated cross-sectional surveys covering the period 1994-1995 to 2008-2009 also contain a longitudinal component, allowing to follow a subset of children over several cycles and model human capital accumulation in early childhood.

Sample selection. As in Baker et al. (2008), we focus on two-parent families with a preschool age child because a simultaneous (but unrelated) change in Québec fiscal policy affected single parents. For the main analysis, we use survey years 1994-1995 and 1996-1997 as pre-reform data and the 2000-2001 and 2002-2003 surveys as post-reform periods. This choice is motivated by the institutional context. First, we omit the third survey wave (1998-1999) so as to focus on parents of children eligible over all of the preschool period. Doing so, we mitigate concerns over treatment effect weighting in staggered designs and anticipatory behavior (De Chaisemartin and d'Haultfoeuille, 2020).[19] Second, some features of the reform evolved a few years after its implementation. As mentioned above, daycare providers could open spaces in the unsubsidized network from June 2002 onward. Also, the daily fee was raised to $7 in 2004. Since our goal is to analyze subsidized childcare provision, we focus on the reform as it was originally conceived.[20]

[17] In a recent study, Karademir et al. (2023) find that the reform also had a small positive impact on employment of grandmothers.

[18] We note that the results of Baker et al. (2019) were challenged by Haeck et al. (2018), who found that after accounting for variation in treatment dosage, the long-term negative effects were substantially less severe than what Baker et al. (2019) estimated.

[19] Ding et al. (2020) find suggestive evidence of strategic placement of children over the implementation period, especially by families with high maternal education. High-educated mothers were significantly more likely to pay for unsubsidised spaces to guarantee a subsidized spot once such new spaces would be open. This strategic response generated a disproportionate increase in childcare use in the province among younger children not yet eligible. Karademir et al. (2023) document similar anticipatory behaviors.

[20] Relatedly, a major change in Québec's parental-leave policy occurred in 2006 and had a substantial impact on mothers' employment and earnings (see Patnaik, 2019; Karademir et al., 2023). We are therefore reticent to use all cycles of the NLSCY, especially for the evaluation of the earnings impacts, as this other policy change might introduce bias in estimates of long-term effects.

Table A2 reports summary statistics comparing Québec and the rest of Canada, our control group. In our analysis sample, we observe 34,042 children aged 0-4 and their parents.

Long-run analysis. To estimate the long-run effects of the childcare reform on eligible children, we use the Canadian Censuses of population of 2016 and 2021. These recent datasets allow us to compare individuals who are old enough to have completed their education. We relegate further details of these more standard datasets to Appendix A.1 and rather dedicate more space to describe our novel data source on daycare supply within Québec.

2.2.1 Daycare supply in Québec

While previous studies of the Québec childcare reform estimate ITT effects of the policy at the provincial level, we investigate treatment effect heterogeneity at a more granular level *within* Québec. To this end, we assemble a novel dataset of the daycare coverage rate at the administrative-region level in Québec from a series of annual management reports – as well as some reports on childcare demand for pre-reform years – of the Ministry of the Family (*Ministère de la Famille*, which also held various other names). We use this information to allow treatment dosage to vary by region of residence depending on the extent to which coverage increased over the period of analysis. Specifically, the daycare coverage rate is defined as the ratio between the number of childcare slots and the total number of preschool-age children (0-4 year olds) in the region. These reports include information on the number of daycare spaces by administrative region by type of facility (centre-based, CPE, for-profit, etc.) from 1994 to 2019. Unfortunately, prior to 2004, we cannot distinguish between regulated and for-profit spaces. This is not a major issue for our empirical analysis for two reasons. First, the share of for-profit spaces in Québec was less than 7% of total spots until 2010, thus making the for-profit market rather marginal. Second, we restrict our empirical analysis to until 2002-2003, the period over which the government froze the number of for-profit spaces.

There are 17 administrative regions in Québec, which makes them a relatively granular level given that the total provincial population was approximately 7 million inhabitants in the late 1990s. Moreover, in the Québec context, using the coverage rate at the municipal level might not be an ideal strategy since many families send their children to daycare in other cities.[21] Therefore, it might be problematic to assume that children attend facilities in their city of residence, but it appears as a reasonable assumption at the administrative-region level.

Appendix Figure A2 provides another illustration of the differential expansion across regions in Québec. Before the reform, childcare coverage was very low across the entire province, but substantial heterogeneity across regions already existed then. Coverage increased considerably from the late 1990s at different rates, with some low-coverage regions eventually catching up

[21] From other ministerial reports, we can confirm that this phenomenon is rare at the administrative-region level but is actually common at the municipal level. The share of children in daycare coming from other administrative regions is low (7.5% for Montréal and Laval and only 1.8% on average in other regions in 2001), but it is large at the municipal level. Indeed, in 2001, 25.1% and 15.8% of children in daycare in the Montréal and Laval regions were coming from another CLSC territory (a geographical unit grouping a few municipalities), respectively. A similar figure is observed in other regions: 23.6% of children in CPEs and 19% of those in other daycare came from other municipalities in 2001 (Ministère de la Famille et de l'Enfance, 2001).

with high-coverage regions, until childcare availability exceeded 0.35 spaces per preschooler in all regions in 2011.

3 Impacts of the reform

In this section, we begin our empirical analysis by estimating the impact of the Québec childcare reform on parents' behavior and exposed children's educational outcomes as they age. Using our new data on regional daycare coverage rates within Québec, we estimate heterogeneous effects of the policy on maternal employment, childcare use, and parenting practices. Next, we investigate whether the negative impacts on child health and non-cognitive outcomes documented by Baker et al. (2008, 2019) extend to economic outcomes later in life. Our empirical strategy exploits the most recent Canadian Censuses, which allow us to observe exposed children with completed education.

3.1 Impact on parents' time allocation

In this section, we estimate the impact of the Québec reform on parents' labor-market outcomes that constitute the first source of fiscal externalities. We focus on short-term impacts (i.e., until 2003) for institutional reasons mentioned in Section 2.2. For comparability reasons of the heterogeneous impacts by the local level of childcare availability, we follow the original empirical approach established in the literature on this reform.

Empirical strategy. We start by analyzing the time-allocation response of parents using the NLSCY sample following Baker et al. (2008) (henceforth BGM). We first replicate the well-established results of previous studies, notably the large impacts on maternal labor supply. The baseline specification is a standard difference-in-differences estimator, where we estimate, for mother i in province p in year t:

$$Y_{ipt} = \alpha + \beta \text{Elig}_{pt} + \gamma_p + \gamma_t + \delta X_{ipt} + \varepsilon_{ipt} \tag{1}$$

where Y_{ipt} is either a parent's labor supply (extensive and intensive margin), childcare use (intensive and extensive margin), or the frequency of reading to the child, our measure of parenting effort. The eligibility dummy Elig_{pt} takes value 1 if the household resides in Québec after the reform. γ_p and γ_t are province and survey year fixed effects. X_{ipt} is a vector of controls including age, age of the child, number of siblings, population of the area of residence, education (both parents), immigration status (both parents), and provincial unemployment rates. Standard errors are clustered at the province-by-year level.

Using our novel data on daycare coverage rates, we then investigate heterogeneity in policy impacts at the administrative-region level. Our empirical strategy employs an intent-to-treat (ITT) difference-in-differences estimator comparing two-parent families with a preschool age child in Québec to similar families in the rest of Canada. We use the same baseline set of control variables and the same sample restrictions (two-parent families) as BGM to ensure that differences in our estimates are solely due to considering local daycare supply and not to differences in design. However, to account for potential changes in composition across regions (within Québec), we also include control variables at the regional level when considering heterogeneous impacts. Our main empirical specification becomes:

$$Y_{iprt} = \alpha + \beta_1 \text{Elig}_{pt} + \beta_2 \text{Elig}_{pt} \times \text{LowExp}_r + \gamma_p + \gamma_r + \gamma_t + \sigma W_{rt} + \delta X_{ipt} + \varepsilon_{iprt} \quad (2)$$

where r indicates the administrative region of residence (within Québec only). LowExp_r equals one if region r is in the bottom tercile of the distribution of daycare expansion over the period. This variable aims at capturing regions in which daycare expanded little. The expansion level is defined as the difference between region r's daycare coverage rate in 2003 to its 1997 level. γ_r is a vector of region (within Québec only) fixed effects. Lastly, W_{rt} is a vector of regional-level control variables associated with childcare demand (shares of medium- and high-educated mothers and the number of preschoolers in the region).

Results. The main coefficients of interest are β_1 and β_2, which capture the differential effects of the policy by local daycare availability. In Tables 1 and 2, we report point estimates of the two specifications above along with results on heterogeneity by availability without regional controls. In columns (2) and (3) of Table 1, we find that the labor-supply response of mothers at the extensive margin is much stronger in regions with higher coverage. In regions where daycare expanded more, the policy boosted maternal labor-force participation 67% more on average, well above the average effect of 7.8 percentage points in the entire province. In regions in the bottom tercile of the daycare coverage rate distribution, the increase in maternal employment is substantially lower and this estimate is statistically significant. High-coverage areas thus appear to be the regions that were driving most of BGM's original result (reported in column 1).

Table 1: Heterogeneous impacts of the Québec childcare reform on mothers' employment by daycare expansion

Dep. var.:	Mother works			Mother's work hours		
	(1)	(2)	(3)	(4)	(5)	(6)
β_1 : Eligible_{pt}	0.078***	0.128***	0.156***	2.129***	3.667***	4.356***
	(0.007)	(0.018)	(0.014)	(0.298)	(1.000)	(0.654)
β_2 : Eligible_{pt}		-0.053***	-0.063***		-1.770***	-1.751***
$\times\text{LowExp}_r$		(0.007)	(0.006)		(0.598)	(0.632)
LowExp_r		0.042***			1.189**	
		(0.006)			(0.588)	
Region (r) FE			✓			✓
r-level controls			✓			✓
Mean dep. var.	0.532			17.54		
p-value of $\beta_1 + \beta_2 = 0$		0.000	0.000		0.000	0.000
R^2	0.105	0.106	0.107	0.099	0.099	0.102
N	33,758	33,758	33,758	33,637	33,637	33,637

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), immigration status (both parents), and provincial unemployment rates. Odd columns report estimates of equation (1) while even columns are regression results of equation (2). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Heterogeneous impacts of the Québec childcare reform on childcare use by daycare expansion

Dep. var.:	Child in care			Childcare hours		
	(1)	(2)	(3)	(4)	(5)	(6)
β_1 : Eligible$_{pt}$	0.138***	0.164***	0.187***	5.736***	6.614***	7.325***
	(0.032)	(0.029)	(0.034)	(1.405)	(0.917)	(0.942)
β_2 : Eligible$_{pt}$		-0.048***	-0.051***		-2.124	-2.276
$\times$LowExp$_r$		(0.014)	(0.016)		(1.443)	(1.556)
LowExp$_r$		0.018			1.182***	
		(0.013)			(0.151)	
Region (r) FE			✓			✓
r-level controls			✓			✓
Mean dep. var.	0.418			13.07		
p-value of $\beta_1 + \beta_2 = 0$		0.001	0.001		0.017	0.012
R^2	0.116	0.116	0.118	0.110	0.110	0.113
N	33,709	33,709	33,709	30,915	30,915	30,915

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), immigration status (both parents), and provincial unemployment rates. Odd columns report estimates of equation (1) while even columns are regression results of equation (2). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In columns (4) to (6) of Table 1 and in Table 2, we examine the impact of the policy on other components of households' time allocation, namely hours worked by the mother and childcare use. We estimate that, on average, (i) eligible mothers work two additional hours per week and (ii) families with a young child use childcare for almost 6 additional hours. These estimated effects are statistically significant at the 1% confidence level.

For these outcomes as well, average effects mask substantial heterogeneity by the local daycare supply change. Mean impacts on mothers' hours worked and childcare utilization are indeed less pronounced in regions where childcare supply expanded less. Only for childcare use at the intensive margin do we find an imprecisely estimated β_2, but the coefficient is nevertheless negative as for the other outcomes.

In Table A5, we further confirm previous results in that the increase in childcare use is driven by an increase in care use in formal (institutional) settings and that the labor supply of fathers is unchanged.

Taken together, our findings suggest that the relief of capacity constraints in daycare at the local level was an important driver of the policy's impacts on mothers' time allocation. In other words, not only the price decrease but also the increase in capacity at the local level was an incentive for mothers to take up employment and use childcare.

Impact on parenting practices. The policy increased maternal labor supply and thus mechanically reduced the time mothers spend at home with their children. If maternal care time and parenting effort are substitutes, we might expect mothers to compensate for the reduced time at home by spending more quality time with the child. In Table 3, we estimate the heterogeneous impact of the reform on the weekly frequency of reading to the child, our measure of parenting effort, by coverage status.

Consistent with previous results by Molnár (2023), we find that the policy had a positive impact on reading time at the bottom of the reading distribution. Point estimates suggest that parents were 4.4 percentage points more likely to read at least once per week and 5.4 percentage points less likely to never read to the child. We detect no short-run impact at the top of the reading distribution (reading daily). As for the time-allocation outcomes, the average impacts are driven by the most treated regions. For instance, the estimated decrease in the propensity to never read is almost entirely concentrated in high-expansion regions. These results thus suggest that mothers compensated for their increased work hours by exerting more effort parenting when they are home.

Table 3: Heterogeneous impact of the Québec childcare reform on weekly frequency of reading to the child by daycare expansion

Dep. var.:	Rarely/never reads			Reads at least weekly			Reads daily		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β_1 : Eligible$_{pt}$	-0.054***	-0.112***	-0.095***	0.044***	0.094***	0.046**	-0.015	-0.013	0.020
	(0.009)	(0.006)	(0.012)	(0.017)	(0.019)	(0.022)	(0.018)	(0.036)	(0.030)
β_2 : Eligible$_{pt}$		0.083***	0.079***		-0.079***	-0.074***		0.006	-0.003
$\times$LowExp$_r$		(0.010)	(0.010)		(0.019)	(0.007)		(0.023)	(0.010)
LowExp$_r$		-0.051***			0.037***			0.027	
		(0.010)			(0.010)			(0.021)	
Region (r) FE			✓			✓			✓
r-level controls			✓			✓			✓
Mean dep. var.	0.226			0.748			0.379		
p of $\beta_1 + \beta_2 = 0$		0.084	0.107		0.678	0.410		0.752	0.520
N	33,171	33,171	33,171	33,171	33,171	33,171	33,171	33,171	33,171
R^2	0.170	0.170	0.171	0.053	0.053	0.056	0.165	0.165	0.168

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), immigration status (both parents), and provincial unemployment rates. Even columns report estimates of equation (1) and odd columns are regression results without regional-level variables (shares of medium- and high-educated mothers and the number of preschoolers in the region r). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Robustness checks. We additionally perform robustness checks on our baseline results, which we report in the Appendix. First, in Table A6, we compare our results to using only Ontario, the most similar province to Québec in terms of size and economy, as a control group. The estimates are quantitatively very similar to those of our main specification. Second, we verified that our results are robust to estimating the standard errors with the wild cluster bootstrap procedure of Cameron et al. (2008) accounting for the small number of clusters.[22]

3.1.1 Threats to identification

Absence of pre-treatment trends. As with any difference-in-differences strategy, a key concern for identifying the policy's treatment effects is the possibility of differential trends between the treatment and control groups prior to the reform. Many papers on the Québec childcare program have argued and provided robust evidence that Québec and the rest of Canada

[22]These results are available in the Research Data Centre of Statistics Canada, and upon request.

(RofC) were following similar trends on a wide variety of outcomes prior to treatment (e.g. Baker et al., 2008, 2019; Haeck et al., 2015, 2018; Molnár, 2023). However, we might be concerned that our two treatment groups within Québec (high- and low-expansion regions) were evolving differently prior to the policy. While there is no direct test of the parallel trends assumption, Figure 2 provides graphical evidence that Québec high-expansion regions, Québec low-expansion regions, and the rest of Canada were on similar trends for our outcome variables before the reform.

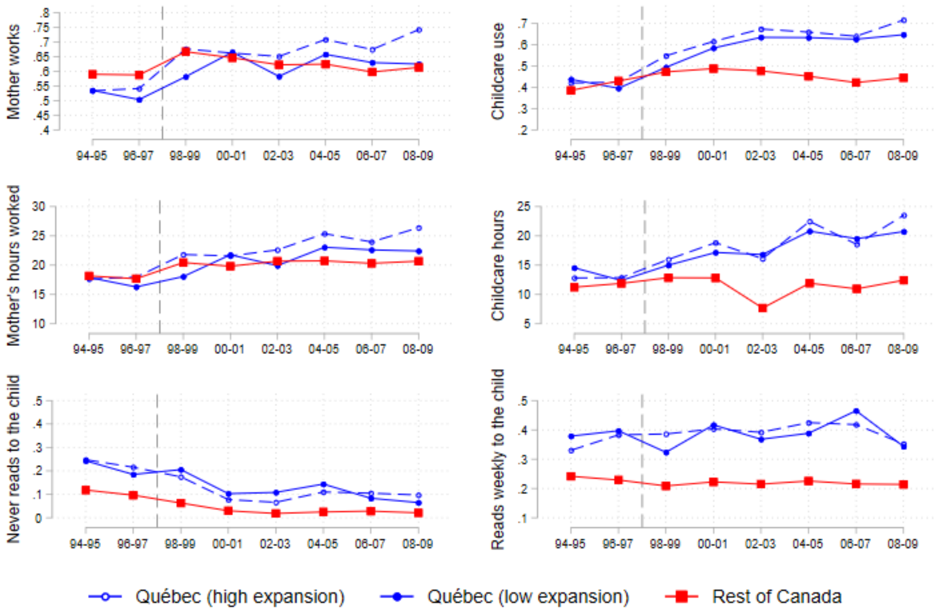


Figure 2: Mean childcare use, maternal labor supply, and reading time by daycare expansion status (in Québec) and in the rest of Canada

Note: These figures plot the means of selected outcome variables separately for three groups, namely two groups within Québec (in blue) and the rest of Canada (in red) in the NLSCY. The dotted (resp. solid) blue line represents families living in regions where the expansion of daycare coverage falls into (resp. is above) the bottom tercile in the province over our study period. The data source is the 8 waves of the NLSCY. The sample is restricted to two-parent families with a preschool-age child.

Appendix Figure A4, which reports estimated coefficients from event-study regressions, further confirms the graphical analysis. These regression results show that prior to the policy, the mean outcomes of interest were either converging or not statistically different before sharply diverging in post-policy waves.

Lastly, Table A3 presents pre-reform descriptive statistics on household characteristics and selected outcomes by expansion status. Our two treated groups are comparable prior to the policy change on all characteristics, thereby strengthening our confidence that low-expansion regions were not following differential trends.

Exogeneity of local childcare expansions. Besides comparing outcomes between Québec and the rest of Canada, our identification strategy leverages the substantial regional variation in childcare expansion within Québec to assess heterogeneity in our intention-to-treat estimates. However, for this evidence to be interpreted as causal, local childcare expansions must be plausibly exogenous to the evolution of parents' labor-market outcomes and childcare arrangements. In other words, expansions should reflect a change in daycare supply rather than being driven by increased demand. We present evidence below to support the reliability of this assumption.

First, to assess the extent to which the local daycare supply increase can be regarded as quasi-random, we explore the potential determinants of changes in regional coverage rates (see, e.g. Cornelissen et al., 2018; Yamaguchi et al., 2018). We obtain region-level information from public datasets of the Québec Statistical Institute, which we complement with other indicators from the Canadian Census of 1996. In Figure 3, we plot the correlation between the daycare expansion level and each region-level characteristic in turn. We define the expansion level as the change in the daycare coverage rate between 1997 and 2003 – namely while the gross price remained unchanged. The figure reveals that virtually all the considered characteristics are uncorrelated with expansion levels, with two exceptions: the initial coverage rate and the share of highly educated individuals in the region, which are negatively and positively correlated with the expansion level, respectively. While the negative correlation with initial coverage is mostly mechanical, we could be concerned that local daycare expansions might capture differences in average educational attainment across regions. Therefore, we control for these education shares in our regressions in the following section, along with region fixed effects, which capture variation in time-invariant regional characteristics.[23] As additional evidence to support our main identification assumption, in Appendix Table A4, we regress the expansion level on all considered characteristics. Reassuringly, we find that none of these variables can predict local childcare expansions. Indeed, we cannot reject the null hypothesis that the estimated coefficients are jointly zero.

A second potential threat to identification could arise from households endogenously sorting into different regions according to childcare availability. To the extent that such residential choice is correlated with unobservable characteristics affecting our outcomes of interest, we would be erroneously attributing the observed changes in outcomes to the increase in daycare availability. For instance, it is possible that mothers from low-coverage regions chose to move to high-coverage regions because they wished to continue working after childbirth. Such a situation would generate non-random selection into treatment and bias our estimates. To get a sense of whether this phenomenon is relevant in our setting, we study the trends in interregional migration of families with a preschool-age child in Appendix Figure A3. We find that, while there is an increase in migration flows of approximately 10% in regions which experienced the largest childcare expansion in the following years, we observe a similar trend in those where childcare provision did not expand as much. Therefore, it is not the case that families systematically moved to areas that experienced greater increases in daycare supply. This additional evidence suggests that this type of self-selection into treatment is not a major concern in our context.

Last, in Appendix Table A3, we show that pre-reform characteristics in low- versus high-expansion regions are very similar. Not only are families in the two groups of regions similar in terms of demographic characteristics (such as parents' education and the number of children in

[23] Another plausible demand-side channel could be that mothers take-up employment in public childcare services. However, data from the 1996 and 2001 Canadian Census suggests that the share of mothers' employment in our sample of interest (mothers of preschoolers in two-parent families) is very small and in fact does not increase from 1996 to 2001. Indeed, this share decreases from 3.69% in 1996 to 1.5% in 2001.

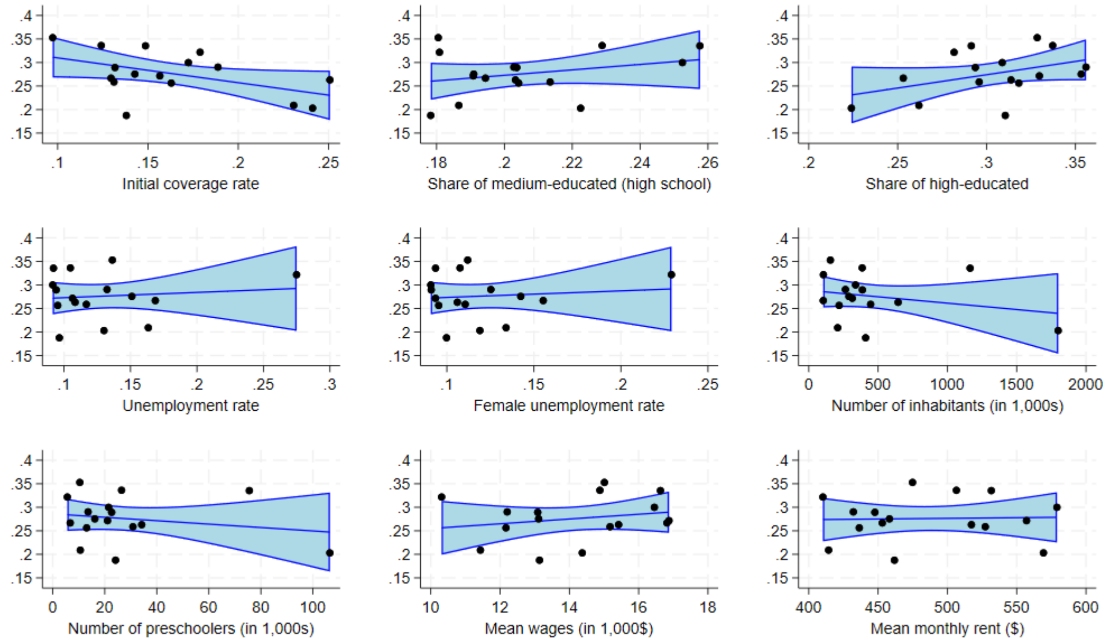


Figure 3: Regional daycare expansion and region-level characteristics

Note: This Figure illustrates the relationship between the daycare expansion level (y-axis) and region-level characteristics potentially associated with childcare demand (as indicated on the x-axis). The daycare expansion level is defined as the increase in the daycare coverage rate from 1997 to 2003. Region-level variables are calculated from 1997 data of the Québec Statistical Institute and the 1996 Canadian Census of population. Shaded areas are 95% confidence intervals.

the household), but they are also comparable along our outcomes of interest. With this evidence in hand, we now evaluate how the impacts of the price decrease interacts with changes in local supply.

3.1.2 Heterogeneity by maternal education

Last, we check whether the heterogeneity by local daycare capacity might help explain some intriguing results from previous studies. Lefebvre and Merrigan (2008) and Molnár (2023) found positive impacts of the policy on maternal employment for both the high- and low-educated mothers. These results are somewhat surprising because the financial incentives to take-up childcare were substantially stronger for better-off families. Indeed, to finance the policy, the Québec government abolished a refundable childcare credit that was rapidly decreasing with household income. For low-income families, the difference in the net price of childcare introduced by the reform was thus very small. For the poorest households, the median net price before the reform was actually approximately the same as a subsidized space under the new regime. However, even if the financial incentives were low, it is possible that low-income households responded to increased availability.

In Appendix Table A7, we use education as a proxy for income and estimate equation (1) separately for high- and low-educated mothers. Following Molnár (2023), we define high-educated mothers as those who have completed a post-secondary degree. Consistent with the literature, we find that the average employment impact is driven by high-educated mothers.

The average impact on low-educated mothers is small and insignificant. However, introducing heterogeneity by local daycare supply reveals that in higher-expansion regions, low-educated mothers do significantly increase their labor supply. This estimated impact is twice as large as that of high-educated mothers in the same regions. Moreover, we find no statistically significant difference in the impact of the policy by coverage status among high-educated mothers. These results are consistent with the financial incentives mentioned above: for high-educated mothers, our results suggest the main incentive to take-up employment was the price reduction; for low-educated mothers, access to a space was key. This also shows up in childcare take-up, where the stronger response in high-expansion regions is again driven by low-educated mothers.

3.2 Earnings gains

We now turn to analyzing the impact of the reform on earnings, which constitutes the main source of fiscal externalities. Because we find positive effects on maternal labor supply but no significant responses from fathers, we focus on mothers' labor earnings as in Lefebvre and Merrigan (2008). To be consistent with our model estimated in Section 4, we estimate equation (1) where the outcome is mothers' annual labor earnings using the income information in the NLSCY. The point estimate, along with 95% confidence intervals, is reported in Figure 4. We find that, on average, mothers of preschoolers in Québec earn an additional \$3,750 (in constant 1997 dollars) per year in the post-reform period compared to other Canadian mothers of young children. Despite using a different dataset and focusing on a different age group than Lefebvre and Merrigan (2008), we reassuringly obtain a point estimate that is quantitatively comparable to their result.[24] To assess the plausibility of the parallel-trends assumption, we also estimate an event-study regression. Appendix Figure A5 confirms that earnings of Québec mothers are not statistically different in the pre-reform waves of the NLSCY, but start diverging only in the post-policy periods.

Because earnings gains have different fiscal impacts along the income distribution, we move beyond average impacts and investigate how the policy shifts the income distribution. As we are mostly interested in impacts on income in an absolute (unconditional) sense, we use the unconditional quantile regression framework of Firpo et al. (2009). This approach estimates quantile treatment effects by comparing the cumulative earnings distributions in pre- and post-reform periods in treatment and control groups using a recentered influence function (RIF) regression. Point estimates then indicate by how much a given quantile of mothers' labor income in Québec has shifted due to the policy. The analysis reveals that there is a positive effect of about \$2,000 at the 4th and 9th deciles and a larger impact of \$4 to \$5 thousands in between. In our analysis of the fiscal externality in section 5, we take into account the impact of this heterogeneity on the government's budget.

[24] Using the Survey of Labour Income Dynamics (SLID) covering the period 1993-2002, Lefebvre and Merrigan (2008) estimate a positive impact of 2,486\$ (once adjusting to 1997 dollars) on mothers of children aged 1-5. Apart from the different sources and age ranges considered, the difference in the estimates might stem from the inclusion by the authors of 1999, where earnings gains are small and statistically insignificant, as a post-reform period. We refrain from including implementation years in our analysis for the reasons mentioned at the beginning of section 3.1.

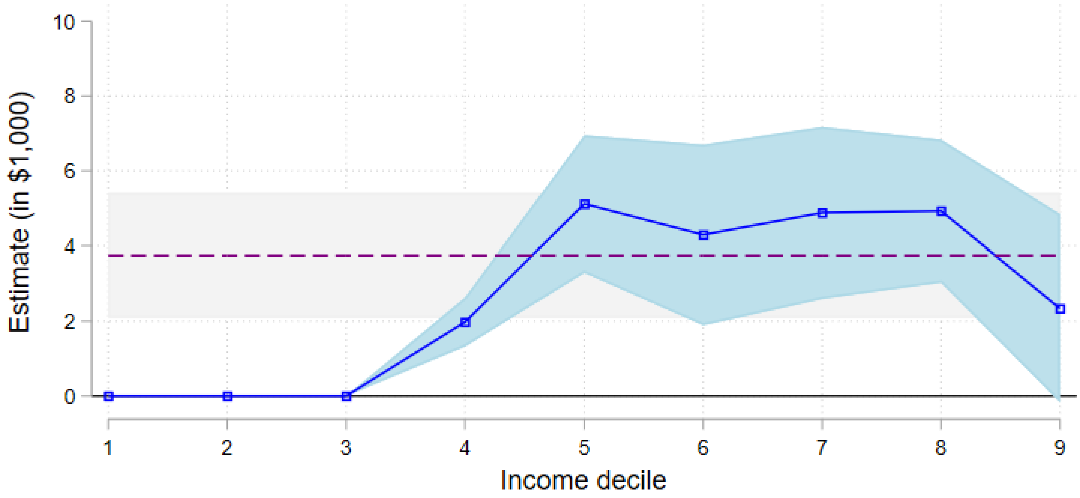


Figure 4: Distributional impact of the Québec childcare reform on mothers' labor earnings

Note: This Figure displays the mean impact (dashed purple line) of the Québec childcare reform along with point estimates from RIF unconditional quantile regressions (blue line) of the policy impact on mothers' annual labor earnings (in constant 1997 thousand dollars). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Shaded areas are 95 percent confidence intervals around each estimate.

3.3 Long-run impact on eligible children

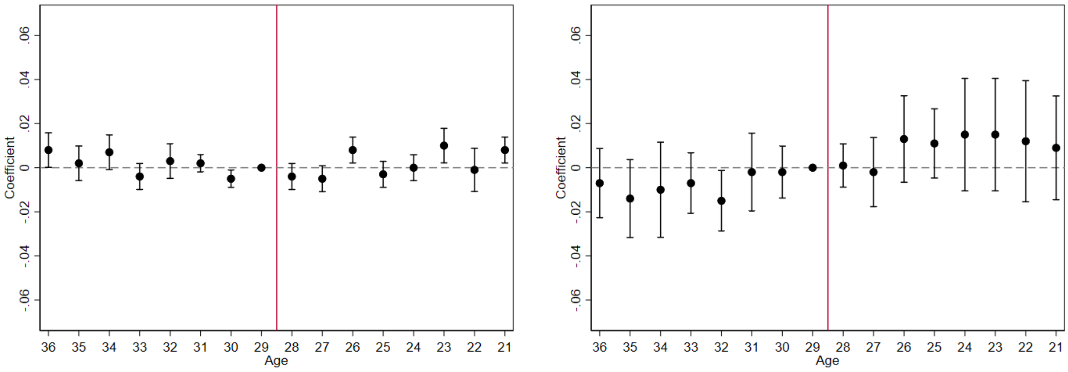
Having established that the policy has significant impacts on mothers' labor-market behavior, we now end our reduced-form analysis by investigating long-run effects on eligible children as they age. As mentioned in Section 2.1, previous evidence on the Québec childcare reform documented average negative impacts on children's non-cognitive development in the short run, but evidence is more mixed in the long run (Baker et al., 2008, 2019; Haeck et al., 2015, 2018). In this section, we assess whether those impacts have long-run implications on economic outcomes as eligible children age. Experimental evidence from targeted programs indeed suggests that boosting non-cognitive skills at a young age causes long-term improvements in economic success (Heckman et al., 2013; Algan et al., 2022). It is therefore possible that the short-run negative effects on behavior and health have translated into worse economic outcomes later in life.

To investigate this possibility, we estimate the long-run impact of the policy on eligible children's educational attainment. Using the Canadian Censuses of 2016 and 2021, we implement a triple-difference estimator, which compares same-age individuals who vary in eligibility status based on the census year and their province of birth. For individual i of age a born in province p observed in census year t, we estimate the following model:

$$Y_{iapt} = \alpha_a + \alpha_p + \theta_1 C_t + \theta_2 \, Q_i \times C_t \tag{3}$$

$$+ \sum_{a=21}^{36} \{\theta_{3,a} \, Age_a \times C_t + \theta_{4,a} \, Age_a \times Q_i + \beta_a \, Age_a \times Q_i \times C_t\} + X'_{iapt}\delta + \varepsilon_{iapt} \tag{4}$$

where Y_{iapt} is educational attainment (completion of a given degree), $Q_i = 1$ is a dummy equal to 1 if the individual is born in Québec. C_t is an indicator of whether the individual is observed in the 2021 Census ($= 0$ if observed in the 2016 Census). α_a and α_p are age and province fixed effects, respectively. X_{iapt} is a vector of controls (gender, marital status, number of children). Parameters of interest is the vector β_a, which capture the intent-to-treat policy impact.



(a) High-school degree

(b) University degree

Figure 5: Long-term effect of the Québec childcare reform on children's educational attainment

Note: These figures plot the regression coefficients on the triple interactions (β_a) from equation (3) using the 2016 and 2021 Canadian Census of population. The horizontal axis represents the individual's age. 95 percent confidence intervals shown in brackets.

The results are reported in Figure 5. We find no evidence of negative effects on educational attainment of eligible children in the long-run. This pattern holds for each educational level, namely university, high school, and college/CEGEP completion. The results for the latter, although not fully comparable across Canadian provinces, are reported in Appendix Figure A6. The results suggest a positive but statistically insignificant impact on completion of a university degree, the most comparable outcome across provinces, and no impact at lower levels. This null result is consistent with the long-run trends in educational attainment, which have been surprisingly parallel across Canadian provinces over decades (see Appendix Figure A7).

In another specification, we assess potential heterogeneity in long-run impacts at the regional level. We estimate equation (3) separately for our two treatment groups depending on the level of local childcare expansions over the period of the policy. We use individuals' place of residence 5 years before the census year as a proxy for the place of birth, which is not available in the census. Results are reported in Appendix Figure A8. As in the baseline model, there is no discernible difference in educational attainment by treatment intensity.

In light of the body of evidence documenting long-run effects of early-childhood circumstances on lifetime success, the absence of long-run impacts here might seem puzzling at first glance. However, several reasons could explain these findings. First, while Baker et al. (2019) find negative impacts on health and behavior, they obtain mixed evidence on cognitive skills. For example, they obtain insignificant impacts on some test score measures, but a positive impact on the mathematics component of PISA tests. Second, there is mixed evidence on the persistence of the short-term negative impacts on non-cognitive outcomes. While Baker et al.'s (2019) results suggest such persistence, Haeck et al. (2015, 2018, 2022) find that most negative impacts on children and parental behavior eventually fade away. Third, it is possible that compensating behavior of parents in their children's education (see Molnár, 2023) might have compensated for the impacts of daycare enrolment.

For the remainder of the paper, we treat these results (and findings in the literature) as evidence of no long-run fiscal externality from children's economic outcomes as they age. We nevertheless consider, in a robustness exercise, the long-run fiscal externality stemming from increased youth criminal activity documented by Baker et al. (2019) in our welfare analysis. Before moving to estimating the MVPF of the Québec reform, we describe the economic model used to infer mothers' willingness-to-pay.

4 Model

For our preferred MVPF estimator, we use a structural model of behavior to account for parents' behavioral responses and non-pecuniary gains of the policy. We consider a model of the family which follows Chaparro et al. (2020) (henceforth CSW), and that we adapt to our context. Our main departure from CSW is that we introduce rationing in the childcare market. Motivated by the evidence of shortages in daycare presented in Section 2, we refrain from assuming that the market is complete.[25] We further leverage our data on regional supply to explicitly introduce local coverage into the household's decision problem.

After describing the model, we briefly explain the numerical procedure to solve it. We then discuss identification of the model, where we exploit the natural experiment generated by the Québec reform to identify some of its key parameters in greater detail.

4.1 Setup

The model is that of a time-allocation problem of a mother (a unitary household) with a young child (after the parental leave period) that has to meet her child's care needs. The mother weighs the consequences of her choices on the child's development, the family budget, and her own preferences, thus providing a framework that highlights the key trade-offs families face (Becker, 1965, 1991; Del Boca et al., 2014; Chaparro et al., 2020; Berlinski et al., 2024).

4.1.1 Time constraints

A unitary household decision maker, which we refer to as the mother, with (at least) one preschool-age child makes a static decision on how to allocate her time T between market work L, child care at home T_m, and leisure ℓ. The mother's weekly time budget is given by:

$$T = L + \ell + T_m \tag{5}$$

Taking as given her child's baseline skills, denoted h_0, as well as household characteristics X, she jointly decides how to meet a child's care time constraint. While the child is awake (T_c hours), he must be cared for by the mother or in non-maternal care and thus we have:

$$T_c = T_m + T_d \tag{6}$$

[25] As documented in Section 2.2.1, childcare markets in Québec have been characterized by important shortages for several decades. Thus, assuming that childcare is available at any quality (and associated price) to every household as in CSW appears unreasonable. Moreover, in the Canadian context, it is not the case that daycare prices are a strong predictor of quality (Seward et al., 2023).

where T_d is hours of non-maternal-care.[26]

4.1.2 Budget constraint

Each hour devoted to labor-market work is remunerated at a wage rate w. To finance non-maternal care expenses and household consumption C (there is no savings), she can use her own labor income as well as non-labor income I, which includes labor income from a spouse. The mother's budget constraint writes:

$$C + (1 - \tau_d(wL + I))\, p\, T_d = wL + I \tag{7}$$

where p is the hourly price of non-maternal care. The household receives refundable provincial childcare credits that account for a share τ_d of childcare expenses. The function τ_d is decreasing in household income, from 0.75 for the poorest households to 0.26 for households with income above \$48,000 (see Appendix Figure A1 for a graph of this function).

4.1.3 Child skill production technology

In addition to choosing a time allocation (L, T_m, ℓ, T_d), the mother chooses *how* to parent. She cares about her child's achievements, which are determined through a child-development production function H taking as inputs the child's initial skills h_0 as well as time and quality of care in each care mode. We also allow the child's skills at the end of the preschool period h_1 (measured at ages 4-6) to depend on a vector of household characteristics X such as the number and ages of other children in the household and the parents' education.

$$h_1 = H(T_m, T_d, q, e; h_0, X) \tag{8}$$

where q is the quality of non-maternal care, which varies across households, and $e > 0$ is the effort devoted by the mother into adopting parenting practices that stimulate the child's development, for which we use reading time to the child as a proxy.[27] To measure non-maternal care quality, we use parents' satisfaction with the interactions the caregiver has with the child, how the caregiver praises the child, and the activities that stimulate learning as a proxy. As for daycare expenses, those variables are only measured for individuals using childcare and are only available in post-reform waves of the NLSCY. Therefore, for each household, we impute non-maternal care quality and the hourly daycare price using characteristics common to all waves (see Molnár, 2023). The variables used for prediction are the age and number of siblings in the household, parents' age, education and immigration status, and the size of the area of residence. Details of these predictions are presented in Appendix A.2.

[26] Non-maternal care includes hours in daycare centres, family-based daycare, and care by relatives. We therefore assume that care by relatives is not free. This assumption is often made in the literature because it is otherwise difficult to rationalize not taking-up free care by a relative (e.g. Bernal, 2008; Bernal and Keane, 2010; Apps et al., 2016; Griffen, 2019; Guner et al., 2020). Rather than assuming an *ad hoc* process for how informal care might be available to some families and not others, we follow this stream of literature.

[27] Compared to Chaparro et al. (2020), who treat effort as unobserved and measure "quality of care", we instead use our data on parenting practices as a proxy of the former. CSW use items of the Home Observation for Measuring the Environment (HOME) index measuring parental Support for Learning and Literacy as their measure of maternal care quality. We do not have such a variable in the NLSCY data, so we proxy for parenting effort using the frequency of reading to the child.

4.1.4 The decision problem

Mothers' utility depends on household consumption, time and effort parenting the child, leisure time, and the child's skill accumulation when he reaches school age (4-6 years old). Additionally, mothers' preferences include a non-monetary disutility of childcare use (when $T_d > 0$), intended to capture travel time to the childcare provider or search effort to find a spot. The mother's decision problem is to choose a time allocation $(\ell, L, T_m, \ell, T_d)$, a level of effort e, and a consumption level C to maximize her utility under the four constraints described above. Formally, the decision problem of a mother writes:

$$\underset{\Gamma}{Max} \quad U(C, \ell, h_1, T_m, T_d, e) \quad s.t.\ (5), (6), (7), (8) \tag{9}$$

where $\Gamma \equiv (C, L, \ell, T_m, T_d, e)$ is the vector of choices.

4.2 Functional forms

Preferences. The mother's (household's) utility function is given by:

$$U = \gamma_c \ln(C) + \gamma_\ell \ln(\ell) + \gamma_h \ln(h_1) + \gamma_m \ln(T_m) - \gamma_{e,1}\, e\, T_m^{\gamma_{e,2}} - \psi(T_d) + \varepsilon \tag{10}$$

where ε is the unobserved component of utility and $\psi(T_d)$ is the disutility of childcare use (defined below). Maternal utility thus depends on consumption, leisure time, time and effort devoted to parenting, the child's human capital accumulation, and entry costs on the daycare market. A parameter of particular interest in this equation is $\gamma_{e,2}$, the non-linearity in the disutility of parenting effort, which in CSW's terms captures an "exhaustion effect" in maternal care. This feature of CSW's model is particularly relevant in our context since reducing exhausting parenting time represents an important source of non-pecuniary utility gain for parents.

Child skills production function. The child accumulates human capital in the preschool period based on the care received and their initial skills h_0. Child skill at age of school entry denoted h_1 is given by:

$$\ln(h_1) = \delta_0 \ln(h_0) + \delta_e \frac{T_m}{T_c} \ln(e) + \delta_d \frac{T_d}{T_c} \ln(q_d) + X'\delta_m + \eta_h \tag{11}$$

where η_h is a productivity shock and the δ parameters capture the productivity of various inputs in generating skills. In this specification, quality of care in each care mode is weighted by the share of time the child spends in it so as to ensure that a given care mode impacts the child's development only when the child is actually there.[28]

Non-monetary disutility of childcare use. We introduce a non-monetary fixed utility cost that the mother bears upon entering the childcare market (Berlinski et al., 2024). This cost,

[28] We also estimated a specification allowing the productivity of each care mode to depend on initial skills h_0, thus permitting dynamic complementaries between childcare investments and baseline skills as in Cunha et al. (2010); **?**. We, however, find little evidence for such complementarities in our context.

denoted as $\psi(T_d)$, can be represented as:

$$\psi(T_d) = \mathbb{1}[T_d > 0] \, (\gamma_{d,1} - \gamma_{d,2} \, CovRate) \tag{12}$$

where $\mathbb{1}$ is the indicator function and $CovRate$ is the coverage rate in the administrative region of residence.

This cost represents several unobserved aspects of parents' childcare costs, such as travel time to the provider or search effort, and is assumed to be fixed, meaning it does not depend on hours spent in daycare but only decreases with increased coverage.[29] It is intended to capture the fact that increased local availability might reduce costs associated with travel time or the burden of finding a spot in childcare. Indeed, Bravo et al. (2022) show that the reduced distance to the nearest daycare centre induced by a national expansion in Chile is valued by families. Similarly, De Groote and Rho (2023) find that families in Leuven, Belgium, highly value proximity to daycare providers.

4.3 Model solution

Given the potential presence (and importance) of corner solutions in the model, it has to be solved numerically. Combining the budget and time constraints (6), (5), and (7) and plugging them into the objective, we are left with three choice variables. That is, the mother chooses a time allocation (L, T_d) and a level of parenting effort e. The solution algorithm works as follows. We build a grid over the feasible time allocations (applying the time constraints) and the effort level. Then, for each combination of time and effort on the grid, we compute the utility level using equation (10) and find the vector yielding the highest utility on the grid.

4.4 Identification and estimation

We adopt a transparent multi-step identification strategy following CSW. The key advantage over a joint estimation algorithm is that it better isolates the sources of variation in the data we use for identification of key parameters of the model.

The first steps consist in identifying a set of parameters that do not require additional structure. These are parameters governing the child skills production function (11) and the exhaustion effect $\gamma_{e,2}$. Taking these productivity parameters as given, we then estimate the remaining preference parameters using a logit specification.

4.4.1 Child skill production technology

We first consider identification of the productivity parameters of the child skill technology (δ). We observe the time allocation of the child across different care modes as well as proxies for care quality as perceived by the person most knowledgeable (PMK) about the child. Her parenting practices and household characteristics are also observed. In our baseline model, we estimate equation (11) by OLS using our measures of child development. We include a set of control

[29] This modelling choice is also driven by the fact that childcare providers typically open full-time spaces only (Haeck et al., 2018).

variables that might influence child development such as the age and number of siblings in the household and parents' characteristics (age, education, and immigrant status).

In this step, we leverage the panel dimension of the NLSCY, which has rarely been used in the literature, to obtain a measure of h_1 using the children observed longitudinally. We use test scores in the Peabody Picture and Vocabulary Test (PPVT) conducted during the home interview as a measure of endline skills. Development at early ages h_0 is measured using the Motor and Social Development (MSD) score. This MSD score is constructed by Statistics Canada using a series of questions on dimensions of the motor, social, and non-cognitive development of young children. The two indices are standardized by age in months and by 2-month age groups, each with mean of 100 and standard deviation of 15.

The baseline OLS estimator might be biased since childcare choices are potentially endogenous. Several variables in the error term η_h such as the mother's innate parenting ability might be correlated with childcare decisions and child development. Quality of care in each mode might also be subject to measurement error.

To address these identification issues, we also consider an instrumental-variable approach leveraging the policy change to identify δ_e and δ_d. Let $\tilde{X}_i \equiv [\frac{T_m}{T_c}\ln(e),\ \frac{T_d}{T_c}\ln(q_d)]$ denote the row vector of endogenous variables in (11). As candidate instruments for $\tilde{X}_i$, we consider the treatment dummy Eligible_{pt} and its interaction with low-coverage status. Sufficient conditions for identification are the usual instruments' relevance and exclusion-restriction conditions. The first identification assumption is that the instruments $Z_i \equiv [\text{Eligible}_{pt},\ \text{Eligible}_{pt} \times \text{LowCov}_r]$ are correlated with $\tilde{X}_i$. The policy impacts documented in Section 3.1 lend support for this assumption: the reform induced a large increase in childcare use and maternal employment, especially so in regions with higher daycare coverage. Second, the exclusion restriction requires that Z_i is uncorrelated with the error term in equation (11). The identification assumption is thus that, conditional on initial skills and household characteristics, the policy should impact child development only through childcare choices.

4.4.2 Identifying the exhaustion effect $\gamma_{e,2}$

We next consider identification of the curvature of the (dis)utility of effort $\gamma_{e,2}$. In CSW, this exhaustion effect is identified directly from reduced-form estimates of the IHDP experiment thanks to random treatment assignment. Our natural experiment gives us similar variation to exploit for identification. Specifically, we leverage the optimality condition for the effort choice and our reduced-form estimates of the policy's impacts to identify $\gamma_{e,2}$. The optimal effort level, which must be at an interior solution (it is not defined at 0), is given by the first-order condition of the maximization problem with respect to e:

$$e^* = \frac{\gamma_{h_1}}{\gamma_{e,1}}\,\delta_e\,\frac{T_m^{1-\gamma_{e,2}}}{T_c} \tag{13}$$

Taking logs on both sides yields:

$$\ln(e) = (1 - \gamma_{e,2})\ln(T_m) + \chi \tag{14}$$

where $\chi = \ln(\gamma_{h_1}) - \ln(\gamma_{e,1}) + \ln(\delta_e(h_0)) - \ln(T_c)$. Thus, optimal (log) parenting effort is determined by maternal care time and a sum of productivity and preference parameters. As CSW note, a simple OLS estimator of $(1 - \gamma_{e,2})$ would be biased because T_m is likely correlated with preference and productivity parameters in χ. The model, however, assumes that parameters in χ are time-invariant. We could thus identify $\gamma_{e,2}$ with the simple differences in mean effort and maternal care time. Nevertheless, we refrain from using this direct approach because other macro shocks could have affected the productivity of parenting effort δ_e. We therefore consider using the quasi-experimental variation to identify the exhaustion effect. We use the more conservative assumption that potential average *changes* in χ conditional on individual characteristics X are the same in Québec and the rest of Canada.

Under this assumption, we can leverage our difference-in-differences estimates to identify $\gamma_{e,2}$. Given the evidence supporting the parallel-trends assumption (conditional on household characteristics X) for parenting effort and maternal-care time, our DiD estimates identify the Intent-to-Treat effects on those outcomes. Thus, conditioning equation (14) on X as well as province ($Q = 1$ for Québec) and a post-treatment dummy and then taking the double-difference yields:

$$\begin{aligned} &\mathrm{ITT}[\ln(e)] = (1 - \gamma_{e,2})\,\mathrm{ITT}[\ln(T_m)] \\ &+ \mathbb{E}[\chi|Q = 1, post, X] \; - \; \mathbb{E}[\chi|Q = 1, pre, X] - (\mathbb{E}[\chi|Q = 0, post, X] \; - \; \mathbb{E}[\chi|Q = 0, pre, X]) \end{aligned} \tag{15}$$

where $\mathrm{ITT}[A]$ is the intent-to-treat impact on variable A. Therefore, assuming that the evolution in χ (conditional on individual characteristics) is the same in Québec and the rest of Canada, the second line in equation (15) is null and the ratio of treatment effects on parenting practices and time identifies $\gamma_{e,2}$. This assumption is plausible in our context given the evidence in the literature of parallel trends between the two groups over a wide range of outcomes.

To lend some additional support for this identification assumption, in Appendix Table A8 we report estimates of the child skill production technology in different Canadian provinces. Reassuringly, we cannot reject the hypothesis that the productivity of parenting δ_e in Ontario, the Western provinces, and Atlantic provinces is the same as that in Québec. Parenting productivity thus should not had evolved differently in Québec compared to our control group, the rest of the country.

4.4.3 Preferences

Taking as given the primitives estimated in the previous steps, we estimate preference parameters on pre-reform data using the Québec sample only. We assume that the unobserved component of utility ε follows an i.i.d. Gumbel distribution, which yields a standard logit model for preferences (McFadden, 1974). This distribution for the unobserved component of utility has the well-known advantage of yielding a closed-form expression for choice probabilities. We estimate the preference vector $\gamma \equiv (\gamma_C, \gamma_\ell, \gamma_{h_1}, \gamma_{T_m}, \gamma_{e,1}, \gamma_{d,1}, \gamma_{d,2})$ by maximum likelihood.

4.5 Estimation results

Before computing our MVPF estimates using the model in the next section, we begin by presenting the estimates of the model's main components. We then discuss the model fit and contrast the policy impacts estimated in the reduced form to the predictions of the model.

4.5.1 Model parameters

Table A9 reports model parameters estimated in the first two steps, namely the exhaustion effect ($\gamma_{e,2}$) and productivity parameters (δ). For both specifications, we compare OLS and instrumental-variable models leveraging variation from the policy change.

Child human capital. In Panel A, we report the point estimates for the key inputs of the child human capital production function. These are the productivity of initial skills (δ_0), maternal care (δ_e), and non-maternal care (δ_d). We find that those three inputs are indeed productive of child human capital. Consistent with previous literature on child development, we find that early-age skills are highly predictive of future skills (see Cunha and Heckman, 2007). The OLS estimate (column 1) suggest that a 10% increase in the Motor-Social Development Score translates into a 2.05% higher PPVT score at ages 4-6.

Childcare time and quality are also positively associated with endline child skills. In the linear regression (column 1), we find that a 1% increase in parenting quality per hour in maternal care is associated with an increase in the child's PPVT score of 0.025% Similarly, a 1% higher quality per hour in non-maternal care increases the child's endline skills by 0.034%. Those parameters, however, should be interpreted with caution because movements in one input involve manipulating several endogenous variables: an increase in maternal-care time (T_m) implies a reduction of non-maternal care (T_d). Additionally, reading time (e) might depend on T_m through the exhaustion effect.

These baseline OLS estimates might suffer from omitted-variable bias through, for example, some unobserved innate parenting ability, which is correlated with childcare choices and reading time. In column (2), we thus report estimates of the instrumental-variable model presented in section 4.4.1. As in the linear regression, we find that both initial skills and quality of care are associated with increased child development. The IV estimates suggest, however, a larger role for both care modes in producing child human capital.

Exhaustion effect. In Panel B of Table A9, we display the estimation results of the convexity of the parenting-effort cost $\gamma_{e,2}$. Columns (4) to (6) contain the results of the IV-type estimator using the policy change discussed in Section 4.4.2. As derived earlier, the exhaustion-effect parameter is given by $\gamma_{e,2} = 1 - \frac{\text{ITT}(\ln(e))}{\text{ITT}(\ln(T_m))}$. ITT estimates of the policy's impact on log reading time and log maternal-care time are reported in columns (5) and (6), respectively. We find that the reform led mothers to increase reading time by 0.08 log points and to reduce parenting time by 0.09 log points. These results suggest substitution between parenting time and effort, in line with CSW. They also imply a convexity in the cost of parenting effort (column 4) of $\gamma_{e,2} = 1.885$. Given that providing high-quality care is increasingly costly for parents, using childcare can provide some relief. Such reduction in the cost of parenting effort is potentially a key source of non-pecuniary gains for mothers.

Preferences. The final set of parameters is the preference vector γ. Table A10 shows the estimation results of the discrete-choice model (10). As expected, parents derive some utility from leisure time and time spent with the child. Moreover, our parameter estimates suggest that parents assign a significant value to their child's achievements but that parenting (effort) is costly.

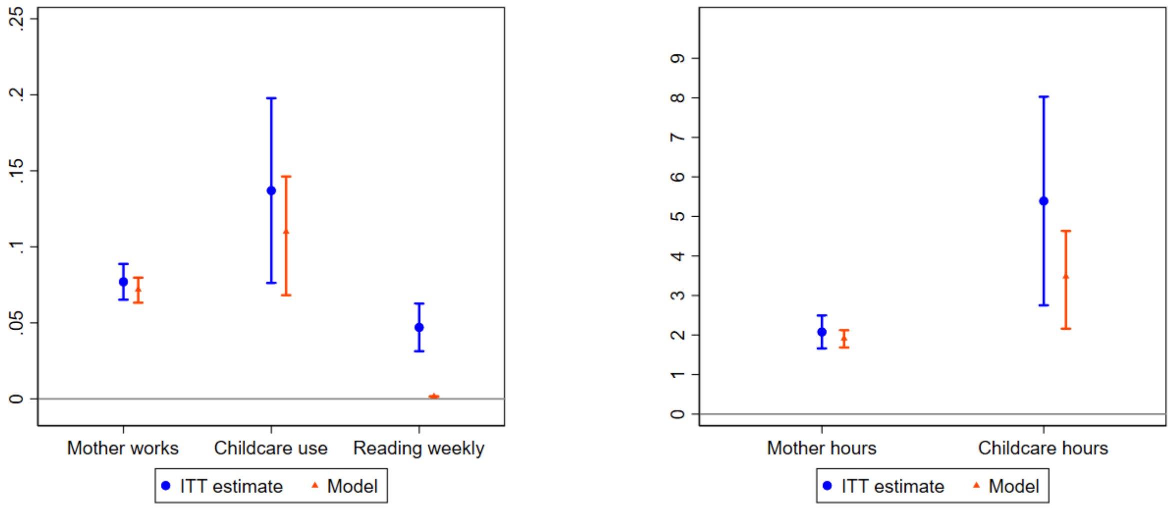
Of particular interest from the perspective of the documented heterogeneity in policy impacts by local daycare supply, our estimates reveal that increased daycare coverage substantially reduces fixed costs on the childcare market. In the last two rows of Table A10, the parameter values indicate that in a hypothetical region with complete coverage (one space per preschool-age child) the entry cost would become negligible. These results again suggest a significant role for local daycare coverage in shaping parents' time-allocation choices. We now turn to a discussion of the model's fit before using our estimated model to compute the MVPF.

4.5.2 Model fit

To assess how accurate the model predictions are compared to the actual data, we use two approaches to test the validity of the model.

In-sample fit. First, in Table A11, we assess the model in-sample fit by comparing the time-allocation choices predicted by the model to observed parents' behavior in the pre-reform data. Using our parameter estimates of the three steps along with 200 draws of the extreme-value type-1 distribution for each household in the pre-reform Québec data, we create 200 datasets of predicted choices. We then compare key market-share summary statistics from the pre-reform data (column 1) to predicted statistics from our simulated dataset (column 2). In the first three rows, we examine the performance of our simulations at predicting extensive-margin choices. We find that the model is doing a decent job for maternal employment and the share of households reading daily to the child, but over-predicts childcare use. At the intensive margin (last three rows), we find that the model cannot capture the difference between the hours worked by the mother and childcare utilization that is observable in the data. This is likely due to the strong incentives in the model to take-up childcare when the mother works. When the mother works full-time, we assume the child must attend childcare at least part-time. Moreover, when the mother works part-time, she has to sacrifice hours of leisure if the household does not use childcare. The results for the in-sample fit of the model are thus mixed. Nevertheless, as Kaboski and Townsend (2011) argue, the model's ability to reproduce the reduced-form impacts of an intervention is arguably a stronger basis for evaluating a model's usefulness.

Out-of-sample validation. Thus, second, we perform an out-of-sample validation test by verifying whether the model predicts well the ITT estimates on maternal labor supply, childcare use, and time reading to the child (Tables 1 and 3). This validation exercise is similar in spirit



Extensive margin

Intensive margin

Figure 6: Out-of-sample validation

Notes: These figures display the results of our out-of-sample validation exercise, which compares the intent-to-treat estimates of the impact of the Québec reform (results in Tables 1, 2, and 3) with predictions from the policy simulation in the model. Standard errors on model predictions are computed using the simulation procedure of Krinsky and Robb (1986).

to Chan and Liu (2018), who study a cash-for-care reform in Norway.[30]

We consider the three main aspects of the policy in our simulations: the price reduction in subsidized spaces, the local increase in supply, and the abolition of the refundable childcare credit. We map these features into the model by *(i)* reducing the hourly price of non-maternal care p_d to \$0.625 (assuming 8 hours of childcare per day), *(ii)* setting the credit rate τ_d to 0 for all households, and *(iii)* increasing the regional coverage rates $CovRate$ to their 2003 levels (second year of Cycle 5 data collection). Figure 6 summarizes the results of this exercise by contrasting the predicted behavioral responses to this policy experiment in our simulation sample to the ITT estimates from Tables 1, 2, and 3.

For inference, since the model predictions are complex non-linear functions of preference parameters, we compute the standard errors using the simulation procedure of Krinsky and Robb (1986). We draw 1,000 parameter vectors from a multinomial normal distribution and predict behavioral responses for each draw. Confidence intervals for predictions are then obtained from quantiles of the simulated distribution of labor supply and childcare choices.

We find that the model closely replicates the labor-supply response of mothers on both margins and also does a fairly good job for childcare use. Indeed, our simulation of the policy predicts a 7.19 percentage points increase in maternal employment, which is very close to the reduced-form estimate of 7.7. Similarly, the model predicts an increase of 1.91 hours at the extensive margin, in line with the positive ITT estimate of 2.08 hours. For childcare use, the model predicts a very similar impact on take-up. On the intensive margin, hours, the model

[30]This type of out-of-sample validation is also conducted by Todd and Wolpin (2006), Kaboski and Townsend (2011), and Chaparro et al. (2020), among others. See Todd and Wolpin (2023) for a review of empirical papers combining program evaluation with structural modelling.

underpredicts the use of childcare, but predictions still lie within the confidence intervals of ITT estimates. The model predicts no response of time reading to the child, in contrast with the positive impact found in the reduced-form analysis. Nevertheless, the good fit of mothers' labor supply and childcare take-up suggests the model is useful to explain key non-marginal responses.

5 Welfare analysis

In this section, we turn to the main contribution of the paper, that is, estimating mothers' welfare gains, inclusive of non-pecuniary gains. We compare a benchmark estimator, using earnings gains as a sufficient statistic, to our structural estimator, which accounts for the fact that the policy change is non-marginal. Before delving into the calculations, we first present the welfare framework we consider and how we compute the two estimators in practice. We discuss the estimators via a brief theoretical exposition and through the lens of our model (a more general theoretical analysis is provided in Appendix C.2).

5.1 Welfare framework

We build on the approach described in Hendren (2016) and Hendren and Sprung-Keyser (2020) where the social welfare impact of a policy change can be measured by its marginal value of public funds (MVPF), which is defined as the ratio of the policy's benefit to its beneficiaries (measured as their willingness to pay for that policy) to the policy's net cost to the government. That is:

$$\text{MVPF} = \frac{\text{Beneficiaries' Willingness to Pay}}{\text{Net Cost to Government}} \tag{16}$$

The net cost to the government is given by the difference between the upfront government expenditure on the policy and fiscal externalities (i.e. indirect impacts on the government budget from changes in behavior). In our context, the main fiscal externality is the return to the provincial budget due to mothers' increased labor supply, which takes the form of increased taxes collected and reduced transfers and benefits.

The higher the MVPF is, the larger the welfare gains to the beneficiaries per net dollar spent. A ratio higher than 1 indicates that the marginal benefit of the policy exceeds its marginal costs to taxpayers, suggesting that the policy is welfare-improving. This approach has several desirable features. In particular, this metric can be used to make comparisons of welfare estimates across policy domains, thus permitting to study government policy from a broader perspective (Hendren and Sprung-Keyser, 2020). Additionally, compared to other standard metrics such as the cost-benefit ratio, the MVPF has two important additional advantages. First, while the standard marginal deadweight loss of public expenditure assumes an arbitrary linear income tax rate, the MVPF framework does not make any peculiar assumption on how the government finances the policy. Second, this framework quickly identifies "Pareto-improvements" from net costs.[31]

[31] There are recent debates in Economics on the use of the MVPF as a welfare criterion to evaluate social programs. In particular, García and Heckman (2022a,b) criticize the use of this metric and suggest the use of an alternative criterion, namely the net social benefit (NSB). In robustness checks, in Appendix D, we compare our

5.2 Willingness to pay

We start by the estimation of the numerator of the MVPF, a key contribution of this paper. As already mentioned, to compute the WTP, it is crucial to distinguish two types of reforms: infinitesimal versus discrete policy changes. For sufficiently small policy changes, it can be shown that, under standard assumptions, the WTP boils down to the treatment effect on beneficiaries' earnings (Hendren, 2016). We illustrate this result below in the context of our model outlined in Section 4 and of our policy of interest. A more general exposition is provided in Appendix C.2.

Environment. Consider the model outlined in Section 4 in which the government chooses a childcare-provision policy characterized by a vector $\theta = (p, \tau_d, CovRate)$, where p is the hourly price of childcare and t_d is a childcare credit function. As in Hendren (2016) we assume that the labor and consumption-good markets are competitive so that the policy state θ has no impact on prices in those markets.[32] The decision problem thus depends on the policy state θ, which the mother takes as given. Let $\Gamma^*(\theta)$ be the vector of optimal choices under policy state θ.

Substituting the time constraints into the budget, we can rewrite (7) as:

$$C^*(\theta) + p(T_c - T_m^*(\theta)) = w(T - T_m^*(\theta) - \ell^*(\theta)) + I(\theta) \tag{17}$$

Substituting the child time constraint into the child skill technology, we can rewrite (8) as a function of maternal-care time, parenting effort, and initial skills only: $h_1^*(\theta) = \tilde{H}(T_m^*(\theta), e^*(\theta); h_0)$. We are thus left with those two constraints. Let $V(\theta) = U(C^*(\theta), \ell^*(\theta), h_1^*(\theta), e^*(\theta), T_m^*(\theta), T_d^*(\theta); \theta)$ be the agent's indirect utility under θ.

The government now implements a policy change. The reform moves the policy state θ from the status-quo policy θ_0 to some new policy state θ_1. The agent's WTP for this policy change can be measured by the standard equivalent variation (E.V.). That is, the WTP is the variation in income under θ_0 that would make the agent indifferent between the status quo and the new policy state:

$$E.V. = \frac{V(\theta_1) - V(\theta_0)}{\lambda(\theta_0)} \tag{18}$$

where λ is the mother's marginal utility of income.

WTP for a small policy change. Let us consider first, as is the case with the sufficient-statistics approach, that the policy change is infinitesimal. For an infinitesimal (marginal) policy change (in θ), at interior solutions, the numerator in (18), the difference in indirect utilities, is

MVPF estimates to calculations of the NSB and the standard cost-benefit ratio. We find that, if anything, using one of these alternative criteria reinforces our main conclusion that omitting non-monetary gains for mothers substantially affects the social desirability of the policy.

[32] In reality, we might suspect a price response of daycare providers in the private (non-CPE) network in the first years of implementation. For instance, to become more competitive with CPEs, we could expect unsubsidized daycare providers to lower their profit margin to retain some children in the private network. However, data we assembled from ministerial reports show that such pricing behavior is unlikely, at least over the time period considered in our analysis. As Appendix Figure A9 shows, average real daycare prices in the for-profit network remained relatively constant from 1994 to 1999.

the total derivative of $V(\theta_0)$ with respect to θ. This derivative yields:

$$\frac{dV(\theta_0)}{d\theta} = \gamma_{d,2} + \lambda(\theta_0)\,\frac{dI^*(\theta_0)}{d\theta} + \mu(\theta_0)\,\left[\frac{\partial \tilde{H}(\theta_0)}{\partial T_m}\frac{\partial Tm^*(\theta_0)}{\partial \theta} + \frac{\partial \tilde{H}(\theta_0)}{\partial e}\frac{\partial e^*(\theta_0)}{\partial \theta}\right] \tag{19}$$

where μ is the Lagrange multiplier on the child development constraint.
Proof. See Appendix C.1.

Therefore, the numerator of the WTP is the sum of three terms: the *direct* utility gain from the change in coverage, the *pecuniary* benefits, and the *non-pecuniary* gain stemming from the impact on child human capital. The intuition behind this result is the logic of the envelope theorem, which implies that, at the margin, behavioral responses do not have a direct effect on utility. Thus, if one additionally assumes that the utility gain from the change in coverage and child human capital gains are negligible, which may be reasonable for marginal family reforms, the difference in utilities $(V(\theta_1) - V(\theta_0))$ boils down to the policy's impact on the beneficiaries' budget constraint. The WTP is then simply given by the causal effect of the policy on earnings $(\frac{dI(\theta)}{d\theta})$. This result is powerful because it implies that the treatment effect on beneficiaries' earnings is a sufficient statistic for the numerator of the MVPF and can thus be used to make transparent welfare statements (Hendren, 2016; Hendren and Sprung-Keyser, 2020).

WTP for a non-marginal policy. Consider now a discrete (large or non-infinitesimal) policy change. We refer here to a policy change that has first-order impacts on beneficiaries' utility. In this case, such as with the Québec childcare reform, the previous result no longer holds since envelope conditions only apply to marginal reforms. In particular, behavioral responses, both for labor-market behavior and time-allocation choices, now have direct impacts on utility because the agent re-optimizes behavior. Moreover, for large reforms, non-pecuniary gains (such as child development gains) are likely important. Thus, the treatment effect on earnings of beneficiaries is a biased estimate of the WTP.

Large-policy bias. The first bias, which we label the *large-policy bias*, stems from re-optimization behavior of beneficiaries. In the model above, it is equal to the policy's direct impact on utility through consumption and leisure time choices. Since agents make non-marginal changes in budgetary choices, these no longer have a null direct impact on the difference in utilities $(V(\theta_1) - V(\theta_0))$. This bias is potentially large in our context, given that the literature has documented major impacts of the reform on economic behavior. The large changes in maternal labor supply and child care use have direct impacts on utility through changes in mothers' time allocation, which are not captured by the treatment effect on earnings.

Non-pecuniary gains. Using the treatment effect on beneficiaries' earnings as an estimator of the WTP is subject to a second bias, namely the omission of non-pecuniary benefits of the policy. In this simple model, it is equal to the policy's impact on utility through coverage and child development gains, which is captured by the first and third terms in equation (19). This bias, in fact, also applies to small reforms, and Hendren and Sprung-Keyser (2020) themselves acknowledge that it may be important in some cases.[33] We argue that non-pecuniary gains (or losses) are likely to be large in the case of child care policies, perhaps even for small-scale programs. Indeed, preschool reforms may have substantial impacts on parenting time and practices and in turn on child development, which are all valued by parents.

Social willingness-to-pay. To derive the society's WTP, one has to aggregate individual preferences taking into account preferences of the overall society. We focus on the case of a utilitarian planner who sets equal social weights to every agent.

5.2.1 Benchmark estimator

As a benchmark, we consider an estimator of the MVPF that assumes the Québec reform is infinitesimal. In the small-policy scenario, in the absence of non-pecuniary gains, as shown above, the WTP is simply the treatment effect on beneficiaries' earnings. To obtain this benchmark estimate, we use our estimates of the pecuniary impacts on mothers and assume that these are sufficient to obtain the WTP. For better comparability with our structural estimator, we focus on short-term gains for mothers.

Willingness-to-pay. We first calculate the numerator of the MVPF under the benchmark estimator, which is the treatment effect of the policy on after-tax income. We thus use our results on the short-run impact on earnings from Section 3.2. To obtain the total WTP, we multiply the quantile treatment effects by the number of mothers in each particular quantile, which yields total earnings gains of $2.469 billion.

Fiscal externality from mothers' short-term earnings gains. The second object we have to calculate is the return to the government stemming from behavioral changes. There is a first fiscal benefit due to the increased labor supply of mothers with young children. At the extensive margin, entry of mothers into the labor market expands the tax base, thus increasing tax revenues for the government. Similarly, at the intensive margin, the government collects tax revenues on additional labor income. Moreover, a second fiscal benefit for the government comes in the form of reduced tax credits and transfer payments to families, since a higher household income decreases eligibility for tax credits.

To compute the net fiscal impact of mothers' responses on the provincial budget, we use the Canadian Tax and Credit Simulator (CtaCS) developed by Milligan (2019). The CTaCS is a comprehensive software that simulates the net fiscal position (at both provincial and federal levels of government) of an individual from a set of raw inputs (e.g. province, year, raw wage income, number of young children). We calculate the fiscal return for the government using our estimates of the impact of the reform on earnings. Given that an increase in earnings has a differential effect on additional taxes paid and reduced benefits along the income distribution,

[33] For example, in their estimation of the MVPF of admission to Florida International University, changes in effort at school or other forms of utility gains during college time are assumed away (Hendren and Sprung-Keyser, 2020, p. 1230). We discuss many other reforms for which non-pecuniary gains might be important in our survey of MVPF estimates in Appendix E.

we rely on our quantile regression analysis to get a better sense of which mothers entered the labor market.

We perform the simulation in three steps. First, for each decile of the mothers' income distribution in our sample, we compute the net fiscal position of the average mother in that quantile. Then, we take the average earnings gain in a given quantile and assign it to families in that quantile. We then simulate the net additional taxes (of transfers and benefits) paid by the mother under the post-reform (1998) tax parameters. This procedure yields an estimate of the fiscal externality of $971 million.

For the purpose of our comparison between the benchmark and the structural estimators, we focus on mothers' short-run responses. For completeness of the social-welfare analysis, we further consider two other sources of fiscal externalities which have been identified in the literature.

Dynamic impact on maternal labor supply. First, Lefebvre et al. (2009) find evidence that mothers' earnings gains were lasting in the medium-run. They estimate a positive effect on earnings of mothers of older children whose child was eligible when younger of $1,995 over the period 1999-2004. Such as in our reduced-form analysis, we restrict dynamics in earnings impacts to that period to avoid capturing confounding effects of the Québec parental-leave reform of 2006, which had negative impacts on young mothers' earnings (Karademir et al., 2023). Thus multiplying these average earnings impacts by the number of mothers in two-parent families with children in those age ranges over that period, we obtain total earnings gains of about $1.102 billion for mothers of older children.[34]

Youth crime. Second, Baker et al. (2019) find that children exposed to the reform at a young age experienced long-lasting negative consequences on behavior and non-cognitive outcomes. In particular, their results suggest that the policy increased youth crime at ages 12-20 among exposed cohorts as they aged.[35] As a robustness check, we monetize these additional societal costs to verify the sensitivity of our results. For the sake of space, we report the details of the calculations in Appendix A.3. Our back-of-the-envelope calculations using estimates of costs of juvenile crime of Cohen (2020) yields an estimate of the WTP to avoid these juvenile crimes of $20.16 million (in 1997 dollars) and a negative fiscal externality of similar magnitude. Thus, these costs are somewhat negligible compared to mothers' earnings gains which amount to billions of dollars and we omit them in the remainder of the analysis. Appendix Table A12 reports the estimates of the MVPF inclusive of these additional societal costs.

5.2.2 Accounting for re-optimization behavior and non-pecuniary gains

To account for the large nature of the policy, we now refrain from assuming that envelope conditions hold. This poses a key challenge in that one can no longer express the WTP as a single treatment effect parameter. Kleven (2021) shows that practitioners would need to

[34]The data source for the number of mothers with children in given age ranges is the estimates from T1 Family Files of Statistics Canada (2023).

[35]To be sure, prevalent youth crimes are rather "benign" offences such as thefts of small amounts, mischiefs, breaking and entering, failures to appear in court, and cannabis possession (Baker et al., 2019).

estimate both "policy elasticities" and changes in elasticities along the policy path, which is arguably beyond empirical reach. An alternative approach is to "make the policy marginal" in the spirit of Bravo et al. (2022). Those authors use variation from a national childcare expansion in Chile to evaluate the welfare effect of marginally reducing the distance to a childcare centre. We do not employ this strategy for two reasons. First, as in most of the policy-evaluation literature, we are interested in studying the effect of the reform as it was implemented. Second, and most important, estimating marginal treatment effects would not be possible in our case given that we do not have at hand an instrument (such as coverage rates) with sufficiently large support over the propensity score. In fact, even if we had such an instrument, defining a policy path for a policy changing several features of the economic environment would be difficult.

Instead, we use our estimated model to compute parents' WTP. To do so, we simulate the reform in the model and estimate the WTP by computing parents' equivalent variation as in Brink et al. (2007). The equivalent variation of a parent is given by equation (18). The marginal utility of income (λ) in our model is simply given by the inverse of the mother's total income net of childcare expenses. Denoting total income by $Y \equiv wL + I$, our structural estimator of the WTP thus writes:

$$\widehat{E.V.} = \frac{\widehat{V(\theta_1)} - \widehat{V(\theta_0)}}{1/\left[\widehat{Y(\theta_0)} - \left(1 - \tau_d(\widehat{Y(\theta_0)})\right) p_d \widehat{T_d(\theta_0)}\right]} \tag{20}$$

where $\widehat{Y(\theta_0)}$ and $\widehat{T_d(\theta_0)}$ are the predicted income and childcare hours under the status quo respectively.

To measure this quantity in the model, we perform the following steps using our simulated sample. First, we obtain the indirect utility in the status quo for each synthetic mother by finding the alternative on the grid yielding the maximum utility. Status quo income net of childcare expenses is then given by the difference between the household's total income and total childcare expenses at that point on the grid. Second, we perform the policy simulation described in Section 4.5.2 where we simulate choices under the key reform parameters (the offer of a \$5/day spot, the increase in coverage locally, and the abolition of the refundable credit). The estimated indirect utility in the post-policy state $\widehat{V(\theta_1)}$ is then given by the new maximum utility on the grid. With all the estimated components in hand, we compute each synthetic mother's WTP using equation (20). Last, we take the average WTP over the 200 simulated duplicates of each mother observed in the NLSCY. To obtain a representative sample of Canadian mothers, we use the sample weights provided by Statistics Canada. Because our simulation models the reform using 2003 coverage rates as the new policy state, to maintain comparability with the benchmark estimator that calculates average impacts over the roll-out of the policy, we sum the WTP over the pre-reform data only (two years).

Our structural estimator suggests a WTP exceeding that found using the benchmark estimator by a factor of two. We find that the total WTP amounts to more than 6 billion dollars. This result thus suggests that non-pecuniary gains are important in this context, which we further investigate through counterfactual simulations in Section 5.5.

Fiscal externality. The infinitesimal-policy assumption mostly has an important implication for the WTP. However, to obtain an internally consistent structural estimator of the MVPF, we also estimate the fiscal externality within our model. To do so, we calculate mothers' predicted income gains using our simulation of the policy. We then obtain the fiscal externality using the CTaCS calculator. To do so, we divide the sample into quartiles of predicted household income and use average household characteristics and income gains in each quartile as inputs for the calculator. The total fiscal externality is then obtained by multiplying the simulated fiscal impact for each quartile by the number of mothers in that quartile. We obtain an estimate of the fiscal externality that is lower in magnitude ($909 million) but comparable to our estimates using the benchmark causal effects. This similarity is reassuring and lends further support for the ability of our model to capture key policy impacts.

5.3 Direct cost

We now consider the upfront cost of the Québec reform, which comprises two main changes to the government budget. Those are the new subsidies allocated to the daycare market and the potential savings from the abolition of the refundable childcare credits.

The main fiscal operation is the increased subsidies allocated to the daycare market. These expenditures take various forms: start-up grants, recurring operating grants to daycare centres, special needs, and other subsidies. We first sum the total subsidies over the period covered in our empirical analysis, that is, over fiscal years 1997-1998 to 2002-2003. Figure A10 shows the evolution of total subsidies to daycare facilities along with the subsidy per space. The graph shows that the rapid expansion of daycare supply over the end of the 1990s and early 2000s induced major increases in government spending. Total subsidies increased from about $470 million in fiscal year 1998-1999 to $1.206 billion in 2002-2003. However, a share of these would have most likely been spent by the government absent the reform. As can be seen in Figure A10, the government was spending nearly $300 million in the two previous fiscal years. We assume similar subsidies would have been granted absent the policy change. We project these expenses assuming the same conservative growth rate observed from 1996 to 1997 (2.1%) and deduct these counterfactual subsidies from the observed grants. These calculations yield a total amount of new subsidies allocated of $2.617 billion over our study period.

To lower the impact of subsidized daycare on public finances, the Québec government made simultaneous changes to other family allowances. In particular, for families obtaining a low-fee space, the refundable childcare credit available before the policy was abolished. To obtain an estimate of the savings generated by this fiscal policy change, we compare government expenses in this expenditure item before and after the reform. Total childcare credits allocated to families are retrieved from the Québec Ministry of Finance's annual budget. In 1996, the fiscal year just before the reform, the credit cost $192 million. In 2001, the same amount was allocated to this program ($191 million). Given the sharp decrease in the number of unsubsidized spaces (who became subsidized), this pattern is surprising. Perhaps some families, who were not claiming the credits before the reform, suddenly started doing so. We are thus reluctant here to attempt to impute what would have been spent by the government absent the Québec reform. Those savings would most likely be small in any case compared to the size of the subsidies (Fortin et al., 2013). We thus prefer to consider an upper bound on direct costs and abstract from potential savings from this source.

The reform, as expected for universal preschool subsidies, is costly. In net, abstracting from potential savings from the abolition of the refundable childcare credit for the reasons detailed above, the Québec government spent \$2.617 billion on the policy over our study period. A careful evaluation of the benefits generated by the reform is thus crucial to assess whether the policy yielded a positive return to society.

5.4 MVPF estimates

The MVPF is the ratio of the WTP to the net cost of the reform. The net cost to society is the difference between the upfront expenditure and fiscal externalities. We use our estimates from the analyses above to calculate the MVPF of the Québec childcare reform under the benchmark estimator and the structural one. Table 4 displays the different components of our MVPF calculations, which we describe below.

Benchmark estimator. First, under the benchmark sufficient-statistic estimator, we obtain an estimate of the net cost of about \$2,617M - \$971M, which yields a net expense of \$1.646 billion. Mothers' willingness-to-pay, captured by their earnings gains under the benchmark estimator, amounts to about \$2.344 billion in after-tax income (substracting the fiscal externality above to the raw earnings gains). Given the absence of evidence of long-run impacts on children from section 3.3, the benchmark estimator of the MVPF suggests parents were willing to pay about \$1.42 per net dollar spent on the reform by the government. This is a small MVPF compared to targeted preschool programs studied in Hendren and Sprung-Keyser (2020).

Structural estimator. Second, our structural estimator suggests a similar, but slightly lower fiscal externality from treated mothers of \$909 million. The willingness-to-pay, however, differs sharply. Including non-pecuniary gains for mothers more than doubles the willingness-to-pay. As a consequence, our estimate of the MVPF more than doubles as well and reaches 3.56. This estimate is much closer to MVPF estimates for targeted preschool interventions appearing in the Policy Impacts Library of Hendren et al. (2023).

5.5 Policy counterfactuals

Thus far, we have focused on estimating the MVPF of the adopted reform. Our model, however, can also be informative about *(i)* the main mechanism driving mothers' labor-supply response and *(ii)* whether the government could have obtained higher welfare gains under alternative policy schemes. Using our estimated model, we perform counterfactual simulations to shed light on these two questions.

Mechanism. First, we ask which feature of the policy is responsible for the bulk of the welfare gain. We simulate counterfactual scenarios in which we remove each feature of the policy one-by-one and compute the WTP under these alternative scenarios. First, we implement only the

Table 4: Welfare estimates

MVPF components	Mean values	External sources used
Direct cost	\$2,617M	Québec Treasury Board
Benchmark estimator		
Fiscal externality		
Tax returns and reduced transfers	\$971M	CTaCS
Willingness-to-pay		
Mothers of preschoolers	\$2,213M	CTaCS
Mothers of older children	\$1,102M	Lefebvre et al. (2009)
Taxes and reduced transfers	-\$971M	CTaCS
MVPF	1.42	
Structural estimator		
Willingness-to-pay	\$6,078M	
Fiscal externality	\$909M	CTaCS
MVPF	3.56	
Counterfactual WTP		
No price change	\$5,120M	
No coverage increase	\$362M	
Credits maintained	\$6,178M	

Notes: This table outlines the components of the MVPF under the benchmark and the structural estimators. The last column reports the external sources used for the policy's cost and other sources of fiscal externalities. The acronym CTaCS refers to the Canadian Tax and Credit Simulator of Milligan (2019). The last three rows present the values of the willingness to pay (WTP) under three counterfactual scenarios: (*i*) no price change, (*ii*) no coverage increase, and (*iii*) childcare credits maintained.

price decrease (or only the increase in coverage) while maintaining the abolition of the refundable credits. Then, we simulate the actual reform, but maintaining the refundable credits.

Our results, displayed in the second part of Table 4 suggest that most of the welfare gains are due to increased coverage. In the first counterfactual, we find that the WTP for increasing coverage rates to their 2003 level without decreasing the price is 84% that of the actual reform. However, we find that the WTP for the price reduction only is very small (only \$362M) compared to that of the actual policy, suggesting that the decrease in price is not the main driver. In our model, given that coverage operates through a reduction in entry cost (on the childcare market) and thus does not set a cap on the childcare-use response, this result is not simply mechanical. Last, not abolishing the childcare credits has a negligible effect on behavior, which is not surprising given that, at the reduced fee, obtaining further discounts only slightly lowers the net price.

Our results thus suggest that increasing childcare availability is key for the effectiveness of universal preschool policies. Those results are in line with De Groote and Rho (2023), who find large welfare gains of increasing daycare capacity on a centralized Belgian platform. In particular, they show that even a small increase in daycare capacity is sufficient to compensate advantaged families for their welfare loss under affirmative-action policies.

Alternative policies. Last, we compare the MVPF of the adopted reform to changing the main features of the policy. Specifically, we ask whether the Québec government could have achieved higher welfare gains under different price-coverage combinations.

To provide some insights into this question, we simulate behavioral responses under multiple price-coverage pairs. For simplicity, we assume a uniform coverage rate throughout the province. For each pair, we proceed in three steps as follows. First, we obtain mothers' WTP by calculating the counterfactual equivalent variation (20) using our synthetic datasets generated for the simulation of the actual reform. Second, we compute the counterfactual fiscal externality using the CTaCS calculator following the same approach as for the actual reform.

Third, we compute the counterfactual direct costs. Counterfactual government subsidies are given by the difference between the counterfactual societal (total) costs and parents' payments under the counterfactual scenario. These are impacted by both the change in the price paid by families and by the number of spots that need be created to reach the counterfactual coverage rate. We assume that the subsidies paid by the government vary linearly with the expenses made by families. In other words, this means that every additional dollar paid by families reduces government expenditures by one dollar. Next, we have to take into account the fact that, in counterfactual scenarios with high coverage rates, not all spaces are filled. Thus, in such cases, parents do not pay for every existing space, but only for those that they actually use. Consistent with the typical childcare contracts in Québec in that period (see footnote 13), we further assume that daycare centres operate over the maximum number of days (260 days).[36]

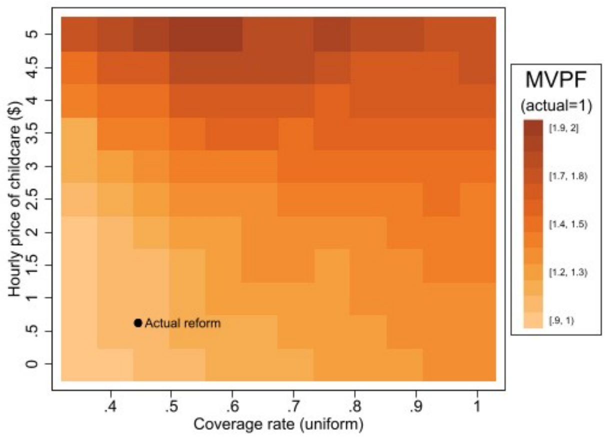


Figure 7: MVPF under counterfactual price-coverage combinations

Note: This figure plots the simulated counterfactual MVPF under different price and uniform daycare coverage combinations. The MVPF of the actual reform is normalized to 1. Darker colors represent higher values.

The results are reported in Figure 7, which shows how the simulated MVPF (where the

[36]Counterfactual government subsidies, denoted by G', are then given by the following expression:

$$\begin{aligned} G' &= \text{(Total costs)}' - \text{(Parents' payments)}' \\ &= (G + \$5 \times \text{spaces} \times 260) \times \frac{(CovRate)'}{CovRate} - p' \times 260 \times min\left\{\text{(childcare use)}', (CovRate)'\right\} \times pop \end{aligned}$$

where *pop* is the population of preschool-age children in the post-policy period and G is the total government subsidy under the actual reform. Variables with a $'$ sign refer to the values of these variables under counterfactual scenarios.

MVPF of the actual reform is normalized to 1) varies with price and coverage. We find a striking pattern: social welfare gains are generally increasing in daycare coverage but also with the fee charged to families. Together with the large WTP for increased daycare supply, this finding suggests that the government could have achieved larger welfare gains by channelling more resources towards opening spots rather than to lowering prices.

Finally, let us highlight a caveat of our analysis. Our empirical model is a partial-equilibrium framework, abstracting from general-equilibrium effects. In simulations with large coverage-rate increases, the substantial rise in maternal labor supply would likely place downward pressure on wages. Thus, we interpret the results with caution. Nevertheless, this exercise is informative about the likely direction of gains in the price-coverage space. For instance, compared to the actual reform, counterfactual estimates suggest the MVPF for a reform that doubles the price charged to families and increases the coverage rate by 5 percentage points would be 13% larger.

6 Conclusion

Childcare policies may impact social welfare through various channels. Availability of subsidized childcare can reduce the opportunity cost of employment, particularly for mothers, thereby potentially increasing the tax base. The quality and accessibility of childcare options can influence human capital development of children, potentially shaping their future labor-market outcomes. Moreover, increased availability reduces non-monetary costs associated with childcare use, such as time spent commuting to the caregiver and search effort to find a spot when supply is initially limited.

This paper incorporates these various channels into a comprehensive welfare analysis of universal preschool provision. We uncover new patterns regarding the impacts on parental behavior resulting from a universal program implemented in the late 1990s in Québec, leveraging novel data on daycare availability at the local level within the province. We show that the positive impacts on maternal labor supply and childcare use are larger in regions where daycare expanded more. These results suggest that the relief of capacity constraints at the local level, not just daycare affordability, is an important channel through which preschool reforms can boost maternal employment and childcare utilization.

Building on this insight, we estimate the value of the policy for mothers using a structural model of maternal labor supply and childcare use that incorporates the benefits of increased availability. In doing so, we explore the extent to which standard assumptions made to estimate sufficient statistics for social welfare might yield misleading results in empirical welfare analysis when applied to non-marginal reforms. Our study demonstrates how combining a reduced-form causal analysis and estimation of a tractable structural model can provide empirical insights into the magnitude of such bias. For policies with significant costs, overlooking non-pecuniary gains might compromise the conclusions about the welfare implications of a given policy. In the context of the Québec reform our estimates indicate that the benefit-to-net-cost ratio is more than twice as large when these gains are considered.

This study suggests three lessons for empirical welfare analysis of preschool reforms. First, it is the first paper to show that universal preschool reforms can yield substantial welfare gains, in particular in the form of non-pecuniary benefits for mothers. Second, it highlights the limitations of sufficient-statistic methods in welfare analysis, often implicitly used in cost-benefit analyses of large reforms. We show that, when applied to non-marginal preschool policies, this approach might omit key welfare gains that are empirically relevant. Third, it underscores the importance of local daycare supply in shaping policy impacts. Our analysis suggests that mothers have a high willingness to pay for an increase in childcare availability.

This first attempt at measuring the bias of sufficient-statistics metrics when applied to universal preschool reforms raises several questions. Studying non-marginal policies comes at the "cost" of structural assumptions on the economic problem and perhaps realism. While our model predicts the key maternal behavioral responses well, future research could extend the framework to consider labor-force dynamics or general-equilibrium effects. Other potential impacts on the Québec economy, such as gains for firms who hired mothers entering the labor force, could be explored. In addition, for cost-benefit analysis, it is essential to assess whether the sufficient-statistic approximation is reasonable in other policy domains. Given the importance of conducting appropriate cost-benefit analyses for policy, more evidence is needed to better understand which assumptions on economic behavior are reasonable in different contexts.

References

Algan, Y., Beasley, E., Côté, S., Park, J., Tremblay, R. E., and Vitaro, F. (2022). The impact of childhood social skills and self-control training on economic and noneconomic outcomes: Evidence from a randomized experiment using administrative data. *American Economic Review*, 112(8):2553–79.

Almond, D., Currie, J., and Duque, V. (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature*, 56(4):1360–1446.

Andresen, M. E. and Havnes, T. (2019). Child care, parental labor supply and tax revenue. *Labour Economics*, 61:101762.

Apps, P., Kabátek, J., Rees, R., and van Soest, A. (2016). Labor supply heterogeneity and demand for child care of mothers with young children. *Empirical Economics*, 51(4):1641–1677.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review*, 110(1):48–85.

Bailey, M. J., Sun, S., and Timpe, B. (2021). Prep school for poor kids: The long-run impacts of head start on human capital and economic self-sufficiency. *American Economic Review*, 111(12):3963–4001.

Baker, M. (2011). Innis lecture: Universal early childhood interventions: what is the evidence base? *Canadian Journal of Economics/Revue canadienne d'économique*, 44(4):1069–1105.

Baker, M., Gruber, J., and Milligan, K. (2005). Universal childcare, maternal labor supply, and family well-being. Working Paper 11832, National Bureau of Economic Research.

Baker, M., Gruber, J., and Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, 116(4):709–745.

Baker, M., Gruber, J., and Milligan, K. (2019). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy*, 11(3):1–26.

Becker, G. S. (1965). A theory of the allocation of time. *The economic journal*, 75(299):493–517.

Becker, G. S. (1991). *A treatise on the family: Enlarged edition*. Harvard university press.

Berlinski, S., Ferreyra, M. M., Flabbi, L., and Martin, J. D. (2024). Childcare markets, parental labor supply, and child development. *Journal of Political Economy*, 132(6):2113–2177.

Bernal, R. (2008). The effect of maternal employment and child care on children's cognitive development. *International Economic Review*, 49(4):1173–1209.

Bernal, R. and Keane, M. P. (2010). Quasi-structural estimation of a model of childcare choices and child cognitive ability production. *Journal of Econometrics*, 156(1):164–189.

Borowsky, J., Brown, J. H., Davis, E. E., Gibbs, C., Herbst, C. M., Sojourner, A., Tekin, E., and Wiswall, M. J. (2022). An equilibrium model of the impact of increased public investment in early childhood education. Working Paper 30140, National Bureau of Economic Research.

Bravo, D., Hojman, A., and Rodríguez (2022). Welfare effects of childcare policies. Working paper.

Brink, A., Nordblom, K., and Wahlberg, R. (2007). Maximum fee versus child benefit: a welfare analysis of swedish child-care fee reform. *International Tax and Public Finance*, 14:457–480.

Buera, F. J., Kaboski, J. P., and Townsend, R. M. (2023). From micro to macro development. *Journal of Economic Literature*, 61(2):471–503.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The review of economics and statistics*, 90(3):414–427.

Cascio, E. U. (2023). Does universal preschool hit the target? program access and preschool impacts. *Journal of Human Resources*, 58(1):1–42.

Chan, M. K. and Liu, K. (2018). Life-cycle and intergenerational effects of child care reforms. *Quantitative Economics*, 9(2):659–706.

Chaparro, J., Sojourner, A., and Wiswall, M. J. (2020). Early childhood care and cognitive development. Working Paper 26813, National Bureau of Economic Research.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics*, 1(1):451–488.

Cohen, M. A. (2020). *The costs of crime and justice*. Routledge.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who benefits from universal child care? estimating marginal returns to early child care attendance. *Journal of Political Economy*, 126(6):2356–2409.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American economic review*, 97(2):31–47.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Daruich, D. (2022). The macroeconomic consequences of early childhood development policies. Working paper.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

De Groote, O. and Rho, M. (2023). Matching disadvantaged children to daycare: Evidence from a centralized platform. Unpublished manuscript, Toulouse School of Economics.

Del Boca, D., Flinn, C., and Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies*, 81(1):137–185.

Ding, W., Kottelenberg, M. J., and Lehrer, S. F. (2020). Anticipating the (un) expected: Evidence from introducing a universal childcare policy with a shortage of spaces. Technical report, Working Paper.

Duncan, G., Kalil, A., Mogstad, M., and Rege, M. (2023). Investing in early childhood development in preschool and at home. In *Handbook of the Economics of Education, Volume 6*, pages 235–297. Elsevier.

Felfe, C. and Lalive, R. (2018). Does early child care affect children's development? *Journal of Public Economics*, 159:33–53.

Finkelstein, A. and Hendren, N. (2020). Welfare analysis meets causal inference. *Journal of Economic Perspectives*, 34(4):146–67.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Fitzpatrick, M. (2010). Preschoolers enrolled and mothers at work? the effects of universal prekindergarten. *Journal of Labor Economics*, 28(1):51–85.

Fort, M., Ichino, A., and Zanella, G. (2020). Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families. *The Journal of Political Economy*, 128(1):158–205.

Fortin, P., Godbout, L., and St-Cerny, S. (2013). L'impact des services de garde à contribution réduite du québec sur le taux d'activité féminin, le revenu intérieur et les budgets gouvernementaux. *Revue Interventions économiques. Papers in Political Economy*, 47.

García, J. L. and Heckman, J. J. (2022a). On criteria for evaluating social programs. Working Paper 30005, National Bureau of Economic Research.

García, J. L. and Heckman, J. J. (2022b). Three criteria for evaluating social programs. *Journal of Benefit-Cost Analysis*, 13(3):281–286.

García, J. L., Heckman, J. J., Leaf, D. E., and Prados, M. J. (2020). Quantifying the life-cycle benefits of an influential early-childhood program. *Journal of Political Economy*, 128(7):2502–2541.

García, J. L., Heckman, J. J., and Ronda, V. (2023). The lasting effects of early-childhood education on promoting the skills and social mobility of disadvantaged african americans and their children. *Journal of Political Economy*, 131(6):1477–1506.

Gelbach, J. B. (2002). Public schooling for young children and maternal labor supply. *American Economic Review*, 92(1):307–322.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American economic review*, 96(2):1–21.

Griffen, A. S. (2019). Evaluating the effects of childcare policies on children's cognitive development and maternal labor supply. *Journal of Human Resources*, 54(3):604–655.

Guner, N., Kaygusuz, R., and Ventura, G. (2020). Child-related transfers, household labour supply, and welfare. *The Review of Economic Studies*, 87(5):2290–2321.

Haeck, C., Lebihan, L., Lefebvre, P., and Merrigan, P. (2022). Childcare, parental behavior and well-being during childhood. *The BE Journal of Economic Analysis & Policy*, 22(2):311–360.

Haeck, C., Lebihan, L., and Merrigan, P. (2018). Universal child care and long-term effects on child well-being: Evidence from canada. *Journal of Human Capital*, 12(1):38–98.

Haeck, C., Lefebvre, P., and Merrigan, P. (2015). Canadian evidence on ten years of universal preschool policies: The good and the bad. *Labour Economics*, 36:137–157.

Haeck, C., Lefebvre, P., and Merrigan, P. (2016). La politique des services de garde à contribution réduite (sgcr) : retombées probantes, faiblesses et pistes d'améliorations. Unpublished manuscript, Groupe de recherche sur le capital humain.

Havnes, T. and Mogstad, M. (2011a). Money for nothing? universal child care and maternal employment. *Journal of Public Economics*, 95(11):1455–1465.

Havnes, T. and Mogstad, M. (2011b). No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy*, 3(2):97–129.

Havnes, T. and Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of Public Economics*, 127:100–114.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic literature*, 48(2):356–98.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010). The rate of return to the highscope perry preschool program. *Journal of Public Economics*, 94(1):114–128.

Hendren, N. (2016). The policy elasticity. *Tax Policy and the Economy*, 30(1):51–89.

Hendren, N. and Sprung-Keyser, B. (2020). A Unified Welfare Analysis of Government Policies*. *The Quarterly Journal of Economics*, 135(3):1209–1318.

Hendren, N., Sprung-Keyser, B., Stuart, E., Kelly, J., Aaronson, S., Odayappan, S., Creeron, A., Emery, J., Hutchinson, C., and Regier, J. (2023). Policy impacts library. https://policyimpacts.org/policy-impacts-library. Last accessed: 2023-01-16.

Herbst, C. M. (2017). Universal Child Care, Maternal Employment, and Children's Long-Run Outcomes: Evidence from the US Lanham Act of 1940. *Journal of Labor Economics*, 35(2):519–564.

Hojman, A. and Lopez Boo, F. (2022). Public childcare benefits children and mothers: Evidence from a nationwide experiment in a developing country. *Journal of Public Economics*, 212:104686.

Hotz, V. J. and Wiswall, M. (2019). Child care and child care policy: existing policies, their effects, and reforms. *The ANNALS of the American Academy of Political and Social Science*, 686(1):310–338.

Japel, C., Tremblay, R. E., and Côté, S. (2005). Quality counts! assessing the quality of daycare services based on the quebec longitudinal study of child development. *Choices: Investing in Our Children*, 11(5).

Kaboski, J. P. and Townsend, R. M. (2011). A structural evaluation of a large-scale quasi-experimental microfinance initiative. *Econometrica*, 79(5):1357–1406.

Kang, Z. Y. and Vasserman, S. (2022). Robust bounds for welfare analysis. Working Paper 29656, National Bureau of Economic Research.

Karademir, S., Laliberté, J.-W., and Staubli, S. (2023). The multigenerational impact of children and childcare policies. IZA Discussion Paper 15894.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2024). Do family policies reduce gender inequality? evidence from 60 years of policy experimentation. *American Economic Journal: Economic Policy*, 16(2):110–149.

Kleven, H. J. (2021). Sufficient statistics revisited. *Annual Review of Economics*, 13(1):515–538.

Kline, P. and Walters, C. R. (2016). Evaluating Public Programs with Close Substitutes: The Case of Head Start*. *The Quarterly Journal of Economics*, 131(4):1795–1848.

Kottelenberg, M. J. and Lehrer, S. F. (2017). Targeted or universal coverage? assessing heterogeneity in the effects of universal child care. *Journal of Labor Economics*, 35(3):609–653.

Kottelenberg, M. J. and Lehrer, S. F. (2018). Does quebec's subsidized child care policy give boys and girls an equal start? *Canadian Journal of Economics/Revue canadienne d'économique*, 51(2):627–659.

Krinsky, I. and Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The review of economics and statistics*, pages 715–719.

Lefebvre, P. and Merrigan, P. (2008). Child-care policy and the labor supply of mothers with young children: A natural experiment from canada. *Journal of Labor Economics*, 26(3):519–548.

Lefebvre, P., Merrigan, P., and Verstraete, M. (2009). Dynamic labour supply effects of childcare subsidies: Evidence from a canadian natural experiment on low-fee universal child care. *Labour Economics*, 16(5):490–502.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale*. Currency.

List, J. A., Suskind, D., and Supplee, L. H. (2021). *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It*. Routledge.

McFadden, D. L. (1974). Conditional logit analysis of qualitative choice behavior. In P, Z., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press.

Milligan, K. (2019). Canadian tax and credit simulator. Database, software and documentation, Version 2019-1.

Ministère de la Famille et de l'Enfance (2001). Situation des centres de la petite enfance et des garderies au québec en 2001. Technical report, Gouvernement du Québec.

Molnár, T. L. (2023). Costs of daycare, complementarities, and heterogeneous productivity of parenting time in child skill formation. Working paper.

Office des services de garde à l'enfance (1997a). Estimation régionale des besoins en services de garde régis par la loi. Technical report, Gouvernement du Québec.

Patnaik, A. (2019). Reserving time for daddy: The consequences of fathers' quotas. *Journal of Labor Economics*, 37(4):1009–1059.

Seward, B., Dhuey, E., and Pan, A. (2023). The big short: Expansion of early childhood education in post-pandemic canada. *Canadian Public Policy*, 49(3):306–329.

Statistics Canada (2023). Table 39-10-0041-01 census families with children by age of children and children by age groups. Last accessed 6 September 2023.

The White House (2021). Fact sheet: Biden-harris administration announces american rescue plan funding to rescue the child care industry so the economy can recover. https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/15/. Last accessed: 2023-04-15.

Todd, P. E. and Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American economic review*, 96(5):1384–1417.

Todd, P. E. and Wolpin, K. I. (2023). The best of both worlds: Combining randomized controlled trials with structural modeling. *Journal of Economic Literature*, 61(1):41–85.

Yamaguchi, S., Asai, Y., and Kambayashi, R. (2018). Effects of subsidized childcare on mothers' labor supply under a rationing mechanism. *Labour Economics*, 55:1–17.

A Data Appendix and additional results

A.1 Data sources

Different sources of Canadian microdata are used in this paper. We provide a brief description of these sources in this Appendix.

The Canadian Census of Population is conducted every five years since 1981 and collects information on all household members from a 20% to 25% sample representative of the Canadian population. It covers a variety of topics, of which we notably extract the province of birth, demographic characteristics as well as educational attainment. We use the 2016 and 2021 waves of the Census for the empirical analysis of children's long-run outcomes.

The NLSCY studies the development and well-being of Canadian children. Children are followed bi-annually from birth to early adulthood and the information collected includes a range of indicators of socio-emotional, cognitive, and behavioral development. The survey series began in 1994 with an initial sample of children under the age of 12, which was followed for 14 years (at two-year intervals) through 2008. In each wave of the survey, a new cohort of children under the age of two was added to focus on early childhood development. On top of including detailed information on child development outcomes, the survey also includes a questionnaire given to the parent who is deemed the most knowledgeable about the child. In that section, the parent provides information on family functioning, parental support, labor supply, and most importantly, the time allocation of care among various options (daycare center, at home by relative, at home by non-relative, and so on). Weekly expenditures on care are also reported by the parent in the last two waves and we use this data to infer the price of private-market care.

The Canadian Labor Force Surveys (LFS) are annual surveys of the working-age population – excluding some specific categories of Canadian households (individuals in Aboriginal reserves, members of the Canadian Forces, and the institutionalized) – that include labor-market and basic demographic variables. Since they include the exact date of birth, we can precisely identify eligible cohorts in the reduced-form analysis. We use the summary Table 37-10-0130-01 of Statistics Canada using the LFS for plots of long-run trends in education across Canadian provinces reported in this Appendix.

A.2 Measurement and predictions of variables

To estimate our structural model, we have to impute values for missing data on variables of interest. These include wages and non-labor income, childcare expenses, and quality of non-maternal care. We describe below how we measure these variables and

Wages and non-labor income. To estimate the model, we need to assign wage offers and to measure non-labor income for every household. This includes predicting a wage offer for non-working mothers as well as imputing the wage rate and non-labor income when income information is missing. In the NLSCY, the person most knowledgeable about the child (PMK) reports wages (for both the PMK and the partner) as well as household income. Given the absence of policy impacts on fathers' labor supply, we treat the father's income as non-labor income from the mother's point of view. We thus measure non-labor income as the difference between the reported household income and the mother's labor earnings (wages and self-employment income). We thus estimate Mincer-type models to predict real wages and income for those households. Variables used for predictions are the age and number of siblings in the household, parents' age, education and immigration status, the size of the area of residence, and a set of Census Metropolitan Area (CMA) dummies to capture local labor market variation.

Childcare expenses. The NLSCY contains measures of childcare expenses in the last two cycles. Respondents report their weekly expenses on childcare in cycles 7 and 8. We follow Molnár (2023) and measure the hourly price of childcare by dividing weekly expenses by the number of hours in institutional care. We make sure to remove households who have a subsidized space, which is observed in those waves. We then obtain predictions for childcare expenses in pre-reform data using variables common to all waves. Variables used for predictions are the age and number of siblings in the household, parents' age, education and immigration status, and the size of the area of residence.

Non-maternal care quality. We measure non-maternal care quality by constructing an index from three survey questions available in cycles 3 and 4. These questions concern parents' satisfaction with the interactions the caregiver has with the child, how the caregiver praises the child, and the activities that stimulate learning. They are phrased as follows:

> *How often would you say your caregiver praises and encourages [CHILD'S NAME], and responds promptly when he/she needs help or comforting?*
> *(1) Never (2) Rarely (3) Sometimes (4) Often*
>
> *How often does your caregiver plan activities and use toys and other materials to help [CHILD'S NAME] learn new things?*
> *(1) Never (2) Rarely (3) Sometimes (4) Often*
>
> *How often does your caregiver encourage [CHILD'S NAME]'s language development by talking to him/her and asking questions, as well as using songs and stories for this purpose?*
> *(1) Never (2) Rarely (3) Sometimes (4) Often*

Variables used for predictions are the age and number of siblings in the household, parents' age, education and immigration status, and the size of the area of residence.

A.3 Youth crime

In this section, we investigate the robustness of our main results to including long-run costs of juvenile criminal activity. Baker et al. (2019) find that children exposed to the reform at a young age experienced long-lasting negative consequences on behavior and non-cognitive outcomes. In particular, their results suggest a positive impact of the policy on youth crime at ages 12-20.

To be sure, prevalent youth crimes are rather "benign" offences such as thefts of small amounts, mischiefs, breaking and entering, failures to appear in court, and cannabis possession. Through the lens of the MVPF framework, increased criminal behavior can impact welfare

Table A1: Costs of the Québec childcare reform from increased youth criminal activity

Type of crime	Victim costs	CJS costs	Offender productivity	Impact (BGM)	WTP	Fiscal externality
Persons (assaults)	\$9,145	\$3,594	\$524	167 [59]	\$17.71M	\$6.58M
Property (theft)	\$251	\$1,922	\$89	342 [93]	\$1.28M	\$7.21M
Drugs[1]	0	\$4,523	\$786	99 [29]	\$0.85M	\$4.91M
Other[2]	0	\$176	\$86	239 [54]	\$0.32M	\$0.65M
Total					**\$20.16M**	**\$19.36M**

Notes: Costs of crime estimates are taken from Cohen (2020) and are converted in 1997 Canadian dollars using the average exchange rate in 1997 (1.3252CAD/1USD). For each crime category, we use the crime most often committed by Canadian youth (in parentheses) as reported in Baker et al. (2019) (BGM). These cost estimates include crimes committed by adults, which are more costly on average, and should thus be interpreted as upper bounds. The WTP column should be interpreted as the WTP for avoiding the committed crimes and is the sum of the victim and offender productivity costs. The estimated policy impacts are taken from Table 5, column 3 of Baker et al. (2019). Standard errors are reported in brackets.
[1] The most prevalent drug crime is cannabis possession, but the data does not allow us to distinguish between drug possession and sale. These estimates are thus likely to be upper bounds.
[2] The most prevalent crime in the "other" category is failure to appear in court, but the data does not distinguish between types of "other non-traffic violations".

through two channels: additional costs to victims and productivity losses for offenders, which reduces the WTP for the policy, and additional costs on the police and criminal justice systems, which is a negative fiscal externality. To take into account these costs to society, we perform a back-of-the-envelope calculation using estimates of costs of crime reported in Cohen (2020). Since these costs appear many years after the enaction of the policy, we apply a discount factor of 3% following Hendren and Sprung-Keyser (2020). However, results are qualitatively robust if we do not discount.[37]

We focus on the estimates from the richest specification (Table 5, column 3) in Baker et al. (2019). They find an average increase in yearly youth criminal activity of 212 crimes per 100,000 inhabitants. Given that crime rates in Québec are very low, this figure represents a rise of 22% of the mean. The authors further break down the crimes into four categories: against persons (rise of 167 crimes), against property (rise of 342), drugs (rise of 99), and other convictions (rise of 239). Since cannabis possession is likely not very costly and now legal in Canada, we focus on the other three categories. For each category, we consider the costs of the most common crime, which are non-aggravated assaults, theft of less than \$5,000, and failure to appear in court.

We first multiply the crime rates impacts by the population of exposed youth in each post-reform cohort considered in the original study to obtain the total yearly impact.[38] Second,

[37] We obtain a WTP to avoid the committed crimes of \$32.11 million and a fiscal externality of \$30.83 million when we do not discount. Results are available upon request.

[38] The population of Québec residents aged 12 to 20 years old was approximately 850,000 over the years

we multiply this number by the victimization and offender productivity costs, which enter the WTP, and the government services costs (on the criminal justice system), which imply a fiscal externality. Third, for each post-reform cohort, we discount future costs to obtain the actualized value of increased youth crime. The results are reported in Table A1. We obtain that these costs, both on the WTP and the fiscal externality, are about 20 million dollars. They are thus negligible compared to benefits stemming from mothers' earnings gains.

A.4 Appendix Figures

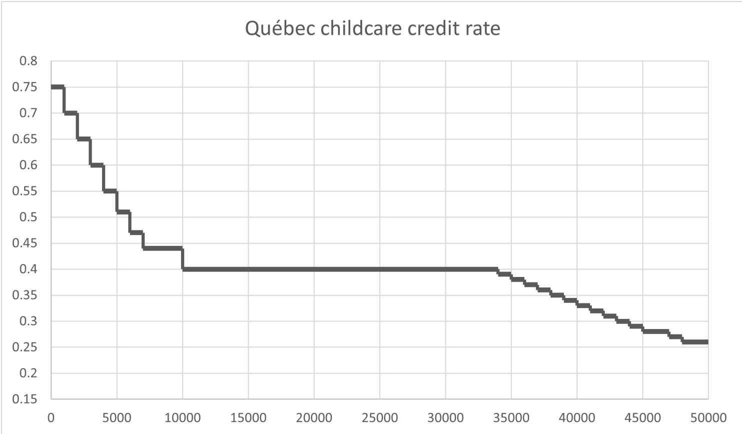
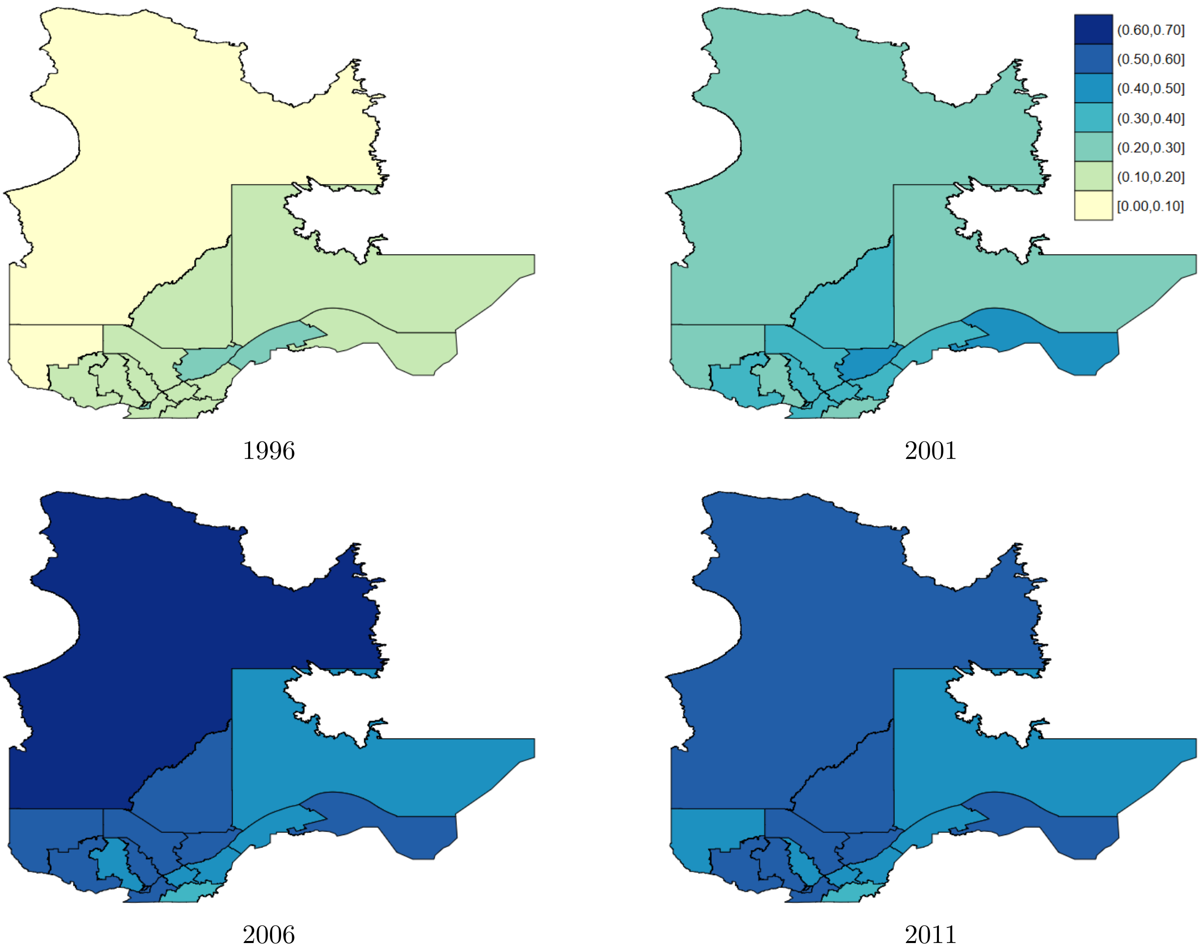


Figure A1: Refundable childcare credit rate by family income in Québec

Note: This Figure plots the refundable childcare credit rate as a function of family income. The refundable childcare credit was available to all families who used paid childcare before the 1997 daycare reform. After 1997, families using subsidized childcare were no longer eligible.

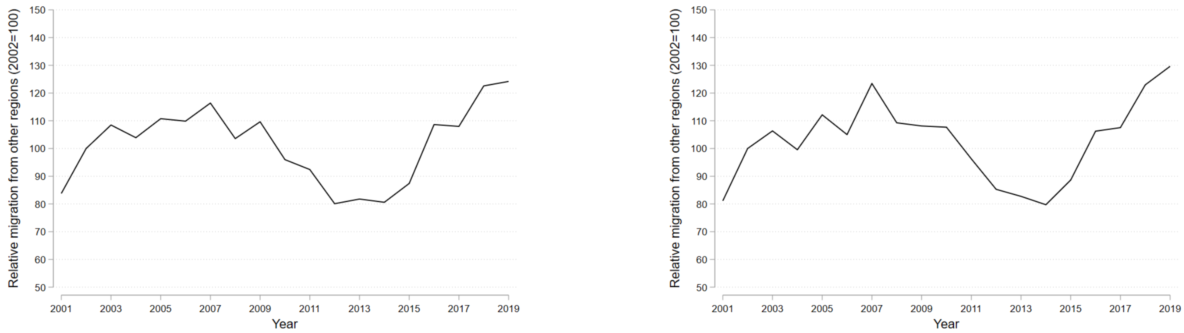
considered for this analysis (?).

Figure A2: Evolution of the total number of daycare spaces per children aged 0-4 years by administrative region, Québec



Data sources: Ministry of the Family for number of spaces and Institut de la Statistique du Québec for population of preschoolers

Figure A3: Evolution of inter-regional migration by childcare expansion status



Top tercile

Bottom tercile

Data sources: Ministry of the Family for number of daycare spaces and Institut de la Statistique du Québec for population of preschoolers and net inter-regional migration.
Notes: These figures display the evolution of the net inter-regional migration flows of preschoolers (0-4 year olds) in two groups of regions relative to 2002 (normalized to 100). The left panel shows the changes in migration to regions which are part of the top third of regions who experienced the largest childcare expansion (measured as the increase in their coverage rate from 1997 to 2003). The right panel shows the equivalent time series for the bottom third.

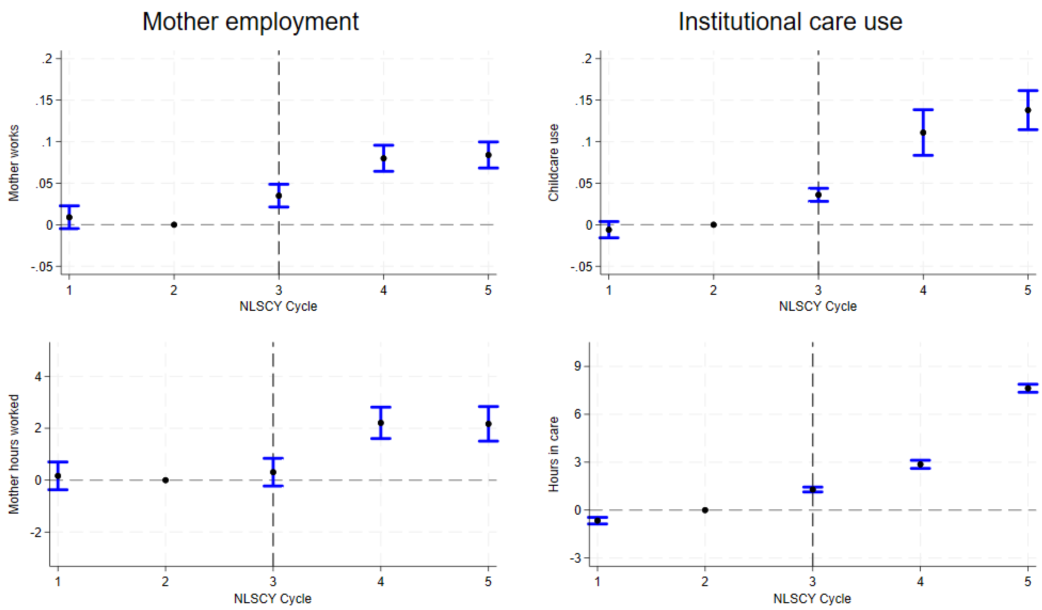


Figure A4: Dynamic impact of the Québec childcare reform on maternal supply and institutional care use

Note: These figures plot the coefficients of event-study regressions along with 95% confidence intervals. The data source is the first 5 waves of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), and immigration status (both parents). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates.

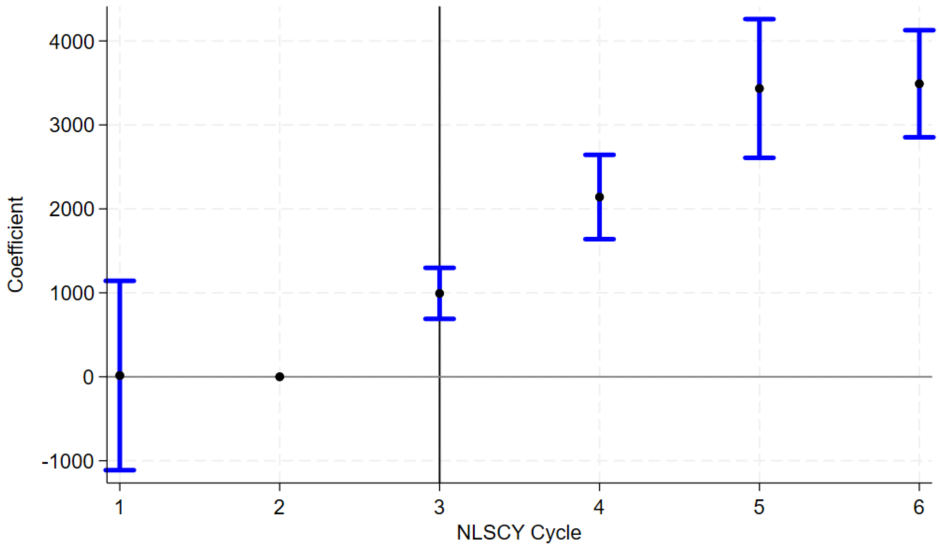
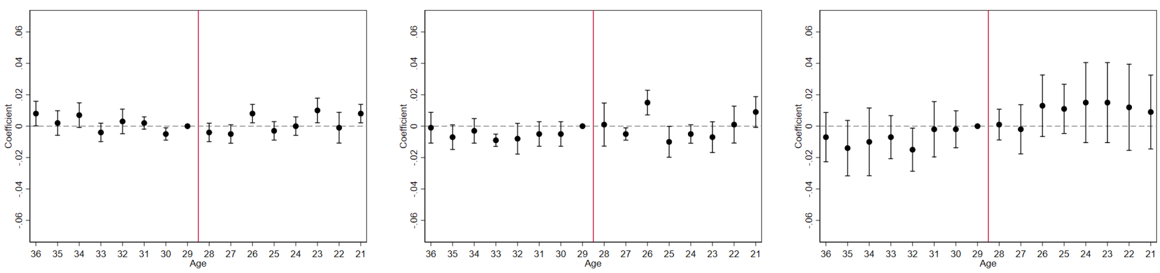


Figure A5: Dynamic impact of the Québec childcare reform on mothers' labor earnings

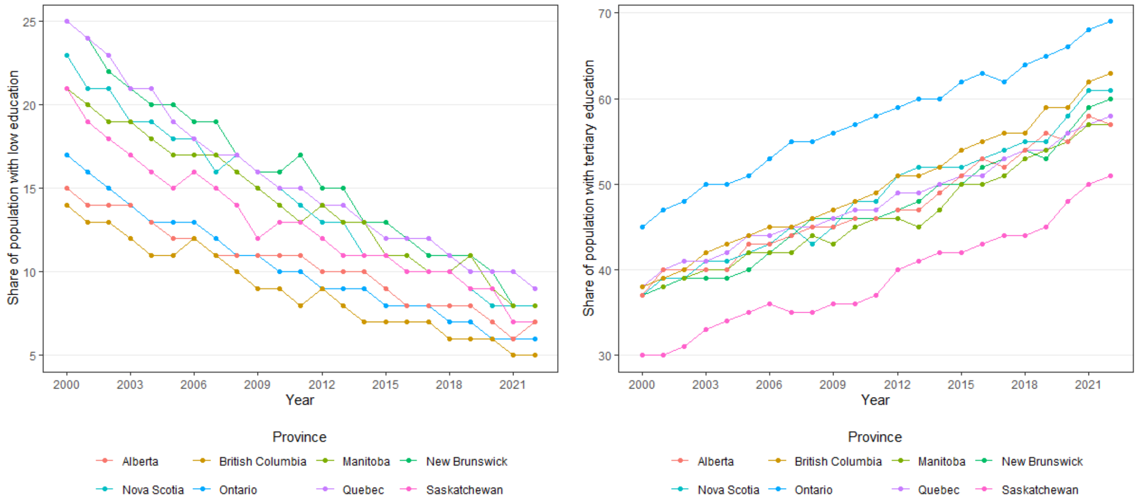
Note: These figures plot the coefficients of event-study regressions along with 95% confidence intervals. The data source is the first 5 waves of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), and immigration status (both parents). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates.



(a) High-school degree (b) College degree (c) University degree

Figure A6: Long-term effect of the Québec childcare reform on children's educational attainment

Note: These figures plot the regression coefficients on the triple interactions (β_a) from equation (3) using the 2016 and 2021 Canadian Census of population. The horizontal axis represents the individual's age. 95% confidence intervals are reported in brackets.



(a) No education

(b) Tertiary education

Figure A7: Long-term trends in educational attainment across Canadian provinces

Note: These figures plot the shares of low- and high-educated in each Canadian province from 2000 to 2022. The data source is Statistics Canada Table 37-10-0130-01 from the Canadian Labour Force Surveys.

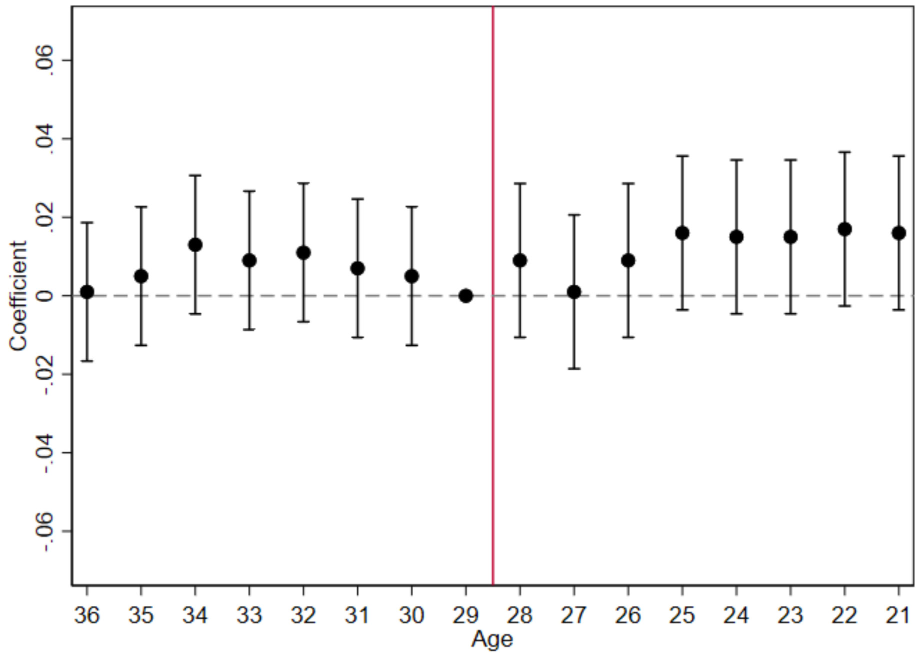
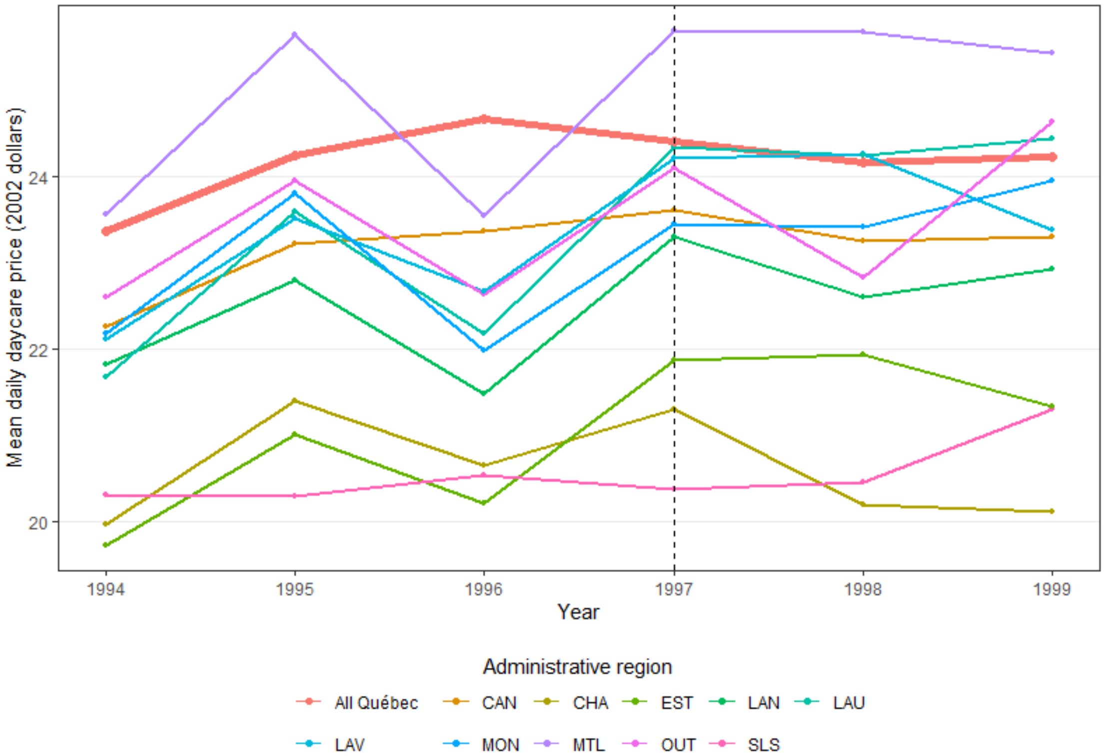


Figure A8: Long-run policy impact on children's educational attainment in low-expansion regions

Note: This figure plots the regression coefficients of event-study regressions using the 2016 and 2021 Canadian Census of population. The dependent variable is a dummy variable taking the value of one if the individual has completed university studies. The horizontal axis represents the individual's age. Robust standard errors in parentheses.

Figure A9: Evolution of average daycare prices in unregulated network by administrative region



Data sources: Ministry of the Family

Notes: This figures plots the evolution of average daily daycare prices in constant 2002 dollars in selected administrative regions in Québec. The thickest line is the average in the entire Québec province. CAN = Capitale-Nationale; CHA = Chaudière-Appalaches; EST = Estrie; LAN = Lanaudière; LAU = Laurentides; LAV = Laval; MON = Montérégie; MTL = Montréal; OUT = Outaouais; SLS = Saguenay-Lac-Saint-Jean

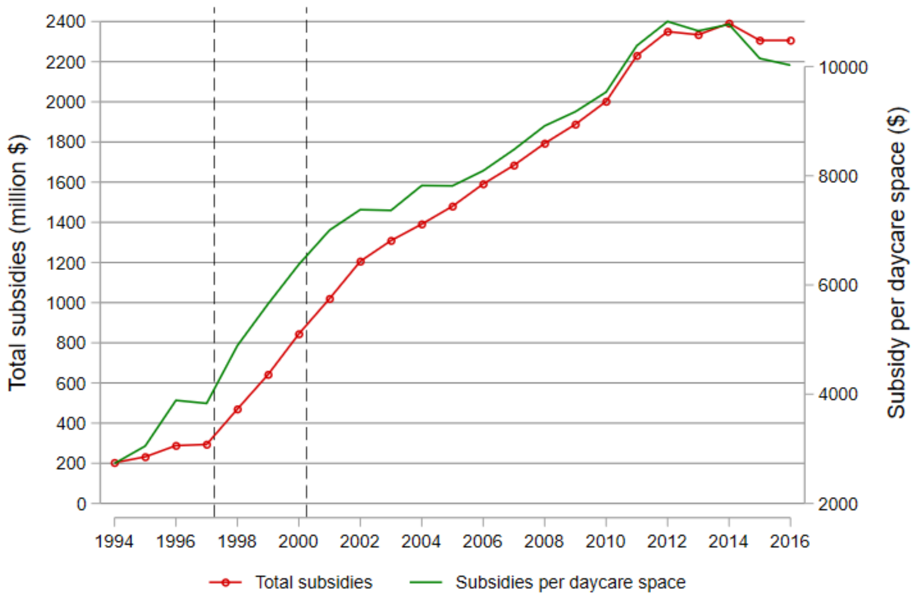


Figure A10: Evolution of daycare subsidies and subsidies per space in Québec

Note: This Figure displays the total subsidies to daycare facilities and families (red dotted line, left axis) and the subsidy per daycare space (green line, right axis). It is constructed using data from a series of budgetary reports of the Québec Treasury Board for the subsidy amounts and of the Ministry of the Family for daycare spaces. Additional costs of the program include additional administrative costs to operate the program, infrastructure subsidies to CPEs, and retirement pensions payments of daycare providers in CPEs. Those are nevertheless small in magnitude compared to direct subsidies.

A.5 Appendix Tables

Table A2: Descriptive statistics

	Québec		Rest of Canada	
	Pre	Post	Pre	Post
Panel A: household characteristics				
Age of mother	30.893	31.167	31.656	32.226
	(4.86)	(5.38)	(4.994)	(5.27)
Age of father	33.508	33.965	34.065	34.789
	(5.368)	(5.822)	(5.584)	(5.858)
Age of child	2.023	2.019	1.998	2.018
	(1.421)	(1.411)	(1.416)	(1.413)
Number of younger siblings	0.236	0.218	0.235	0.220
	(0.474)	(0.442)	(0.463)	(0.451)
Mother is immigrant	0.088	0.125	0.218	0.245
	(0.284)	(0.331)	(0.413)	(0.43)
Father is immigrant	0.096	0.126	0.209	0.239
	(0.295)	(0.332)	(0.406)	(0.427)
Number of older siblings	0.780	0.792	0.904	0.836
	(0.924)	(0.95)	(1.035)	(0.954)
Child is a girl	0.494	0.479	0.488	0.489
	(0.5)	(0.5)	(0.5)	(0.5)
Mother college degree	0.202	0.270	0.204	0.273
	(0.402)	(0.444)	(0.403)	(0.446)
Father college degree	0.195	0.239	0.215	0.262
	(0.396)	(0.427)	(0.411)	(0.44)
Mother no education	0.133	0.122	0.108	0.093
	(0.34)	(0.328)	(0.311)	(0.291)
Father no education	0.167	0.157	0.138	0.111
	(0.373)	(0.364)	(0.345)	(0.314)
Household lives in rural area	0.153	0.150	0.154	0.105
	(0.36)	(0.357)	(0.361)	(0.307)
Panel B: selected outcomes				
Child in care	0.418	0.630	0.407	0.482
	(0.493)	(0.483)	(0.491)	(0.5)
Hours in care	13.071	17.425	11.571	10.606
	(17.983)	(19.237)	(16.983)	(16.361)
Mother works	0.532	0.644	0.590	0.636
	(0.499)	(0.479)	(0.492)	(0.481)
Mother hours worked	17.541	21.575	17.929	20.250
	(18.176)	(17.982)	(17.84)	(18.448)
Father works	0.867	0.921	0.909	0.954
	(0.339)	(0.27)	(0.288)	(0.21)
Father hours worked	36.374	39.628	39.483	42.264
	(16.264)	(13.854)	(15.119)	(12.766)
Rarely/never reads	0.226	0.086	0.107	0.024
	(0.418)	(0.281)	(0.31)	(0.153)
Reads weekly	0.369	0.395	0.235	0.219
	(0.483)	(0.489)	(0.424)	(0.413)
Reads daily	0.379	0.462	0.645	0.740
	(0.485)	(0.499)	(0.479)	(0.439)
Child PPVT score	98.408	100.462	100.301	102.191
	(16.097)	(15.128)	(14.945)	(15.124)
Child MSD score	99.300	98.418	100.462	101.106
	(15.028)	(14.674)	(15.254)	(14.344)

Note: Pre-reform data is the first two waves (1994-1995 and 1996-1997) of the National Longitudinal Survey of Children and Youth (NLSCY). Post-reform data are waves 4 and 5 of the NLSCY (2000-2001 and 2002-2003). The sample contains 34,042 children aged 0-4 and their parents. Standard deviations are reported in parentheses.

Table A3: Pre-reform descriptive statistics by childcare expansion status

	Low exp.	High exp.
Panel A: Household characteristics		
Age of mother	30.625	31.104
	(4.55)	(5.062)
Age of father	33.889	33.261
	(5.815)	(5.022)
Age of child	1.981	2.060
	(1.42)	(1.424)
Number of younger siblings	0.242	0.233
	(0.484)	(0.467)
Number of older siblings	0.794	0.772
	(0.946)	(0.902)
Child is a girl	0.512	0.483
	(0.5)	(0.5)
Mother is immigrant	0.151	0.044
	(0.358)	(0.206)
Father is immigrant	0.162	0.050
	(0.369)	(0.219)
Mother college degree	0.192	0.212
	(0.394)	(0.409)
Father college degree	0.210	0.186
	(0.407)	(0.39)
Mother no education	0.174	0.104
	(0.38)	(0.305)
Father no education	0.192	0.149
	(0.394)	(0.356)
Household lives in rural area	0.131	0.165
	(0.337)	(0.371)
Panel B: Selected outcomes		
Child in care	0.417	0.422
	(0.493)	(0.494)
Hours in care	13.501	12.852
	(18.468)	(17.666)
Mother works	0.520	0.538
	(0.5)	(0.499)
Mother hours worked	17.122	17.798
	(18.355)	(18.053)
Father works	0.867	0.867
	(0.339)	(0.339)
Father hours worked	36.444	36.321
	(16.401)	(16.145)
Rarely/never reads	0.215	0.232
	(0.411)	(0.422)
Reads weekly	0.388	0.356
	(0.487)	(0.479)
Reads daily	0.373	0.384
	(0.484)	(0.487)
Child PPVT score	96.014	100.120
	(16.887)	(15.338)
Child MSD score	98.412	99.944
	(15.344)	(14.839)

Note: Data: first two waves (1994-1995 and 1996-1997) of the National Longitudinal Survey of Children and Youth (NLSCY). Low-expansion regions are administrative regions (within Québec) in the bottom tercile of the childcare expansion distribution. The sample is restricted to two-parent families with a preschool-age child. Standard deviations are reported in parentheses.

Table A4: Determinants of local childcare expansions

	(1)	(2)
Initial coverage rate	-0.6756**	-0.2208
	(0.24)	(19.54)
Number of inhabitants (in 1,000s)		-0.0003
		(0.02)
Number of preschoolers (in 1,000s)		0.0055
		(0.39)
Share of medium-educated (high school)		0.0008
		(0.35)
Share of high-educated		0.0019
		(0.28)
Unemployment rate		1.0996
		(74.08)
Female unemployment rate		-1.0255
		(111.29)
Mean wages (in 1,000$)		0.0156
		(1.42)
Mean monthly rent ($)		-0.0005
		(0.03)
Constant	0.3903***	0.2479
	(0.04)	(2.12)
p-value of joint significance	0.0056	1.0000
R^2	0.352	0.515

Note: This table reports coefficients of linear regressions of the change in the childcare coverage rate (number of spaces divided by the population of preschool-age children) from 1997 to 2003 on the initial coverage rate and baseline regional characteristics. The data sources are a series of ministerial reports for daycare spaces (*Ministère de la Famille*, which held various other names) as well as the Québec Statistical Institute and the 1996 Canadian Census for the children population and the regional characteristics. The second-to-last row reports the p-value of the hypothesis that all the coefficients on baseline regional characteristics are jointly zero. Bootstrapped standard errors (1,000 replications) in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Heterogeneous impacts of the Québec childcare reform on fathers' employment and institutional care use by daycare expansion

Dep. var.:	Institutional care		Inst. care hours		Father works		Father's work hours	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : Eligible$_{pt}$	0.131***	0.168***	5.250***	4.926***	0.005	0.043***	0.159	0.614***
	(0.011)	(0.004)	(1.387)	(0.692)	(0.007)	(0.011)	(0.396)	(0.166)
β_2 : Eligible$_{pt}$		-0.073*		-2.448***		-0.048*		-1.559*
x LowExp$_r$		(0.038)		(0.497)		(0.021)		(0.798)
Region (r) FE		✓		✓		✓		✓
r-level controls		✓		✓		✓		✓
R^2	33575	33575	33320	33320	34012	34012	31497	31497
N	0.069	0.074	0.076	0.08	0.161	0.162	0.09	0.093

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), and immigration status (both parents). Odd columns report estimates of equation (1) while even columns are regression results of equation (2). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Heterogeneous impacts of the Québec childcare reform on mothers' employment and childcare use by childcare expansion status, comparison with Ontario only

Dep. var.:	Mother works		Mother's work hours		Child in care		Childcare hours	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : Eligible$_{pt}$	0.076***	0.150***	1.850***	4.065***	0.147***	0.190***	5.481***	6.577***
	(0.002)	(0.015)	(0.138)	(0.656)	(0.033)	(0.035)	(1.547)	(0.946)
β_2 : Eligible$_{pt}$		-0.060***	-1.738**	-0.046**	-1.972			
x LowExp$_r$		(0.007)	(0.602)	(0.016)	(1.671)			
Region (r) FE		✓		✓		✓		✓
r-level controls		✓		✓		✓		✓
p-value of $\beta_1 + \beta_2 = 0$		0.000		0.000		0.001		0.012
R^2	0.116	0.119	0.105	0.110	0.127	0.130	0.114	0.119
N	15739	15739	15725	15725	15735	15735	14426	14426

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), and immigration status (both parents). Odd columns report estimates of equation (1) while even columns are regression results of equation (2). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Heterogeneous impact of the Québec childcare reform on maternal employment by daycare expansion and mother's education

Dep. var.:	Mother works				Childcare use			
	Low educ		High educ		Low educ		High educ	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : Eligible$_{pt}$	0.030	0.191***	0.095***	0.084*	0.075**	0.222***	0.158***	0.157***
	(0.019)	(0.029)	(0.079)	(0.045)	(0.014)	(0.078)	(0.027)	(0.016)
β_2 : Eligible$_{pt}$		-0.099**		0.033		-0.082*		-0.01
x LowCov$_r$		(0.047)		(0.044)		(0.042)		(0.037)
Region (r) FE		✓		✓		✓		✓
r-level controls		✓		✓		✓		✓
p-value of $\beta_1 + \beta_2 = 0$		0.031		0.000		0.046		0.001
N	10070	10070	23688	23688	10048	10048	23661	23661
R^2	0.103	0.103	0.084	0.084	0.093	0.094	0.103	0.103

Note: The data source is waves 1-2-4-5 of the NLSCY. Even columns report estimates of equation (1) and odd columns are regression results without regional-level variables (shares of medium- and high-educated mothers and the number of preschoolers in the region r). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Child skill production technology parameters in different Canadian regions

Parameter	Description	Québec (1)	Ontario (2)	Atlantic (3)	West (4)
δ_0	self-productivity	0.178***	0.155***	0.136***	0.165***
		(0.015)	(0.012)	(0.013)	(0.011)
δ_e	maternal care	0.023***	0.020***	0.019***	0.010***
		(0.004)	(0.004)	(0.004)	(0.004)
δ_d	non-maternal care	0.032***	0.012***	0.002	-0.002
		(0.005)	(0.004)	(0.005)	(0.005)
p-value of $\delta_e^{QC} - \delta_e^{p} = 0$			0.765	0.624	0.171
N		3860	5994	4879	7174

Note: This Table reports estimation results for the child human capital production function (equation 11) in different Canadian regions. Bootstrapped standard errors in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Production function and exhaustion effect parameters

Panel A: Production function parameters

Parameter	Description	OLS (1)	IV (2)
δ_0	self-productivity	0.205***	0.105***
		(0.027)	(0.018)
δ_e	maternal care	0.025***	0.226***
		(0.007)	(0.02)
δ_d	non-maternal care	0.034***	0.264***
		(0.009)	(0.018)

Panel B: Exhaustion-effect estimation

$\gamma_{e,2}$ estimate		IV First-stage	
Naive OLS (3)	IV (4)	ITT($\ln(e)$) (5)	ITT($\ln(T_m)$) (6)
1.015	1.885***	0.0796***	-0.0899***
(0.022)	(0.205)	(0.028)	(0.022)

Note: This Table reports estimation results for the child human capital production function (equation 11) and the exhaustion-effect parameter (equation 15) respectively. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Preference parameters

Parameter	Description	Estimate	SE
γ_C	consumption	1	(.)
γ_ℓ	leisure	0.312***	(0.018)
γ_{T_m}	maternal care	1.872***	(0.178)
γ_{h_1}	child skills	16.262***	(2.269)
$\gamma_{\rho,1}$	cost of effort†	0.227***	(0.0361)
$\gamma_{d,1}$	childcare use	1.643***	(0.156)
$\gamma_{d,2}$	coverage	1.615*	(0.879)
N	2518		

Note: Bootstrapped standard errors (400 replications) in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
† $\gamma_{\rho,1}$ is re-scaled (multiplied) by 10,000 for comparability.

Table A11: Model in-sample fit

	Observed outcome		Simulated outcome
	Mean	SD	Mean
Extensive margin			
Maternal employment	0.532	0.499	0.557
Childcare use	0.418	0.493	0.482
Reading daily to child	0.379	0.485	0.333
Intensive margin			
Maternal employment	17.54	18.18	14.28
Childcare use	13.07	17.98	14.43
Reading hours	3.67	2.81	3.17

Table A12: Benchmark welfare estimate including costs of juvenile crime

MVPF components	Mean values	External sources used
Direct cost	\$2,617M	Québec Treasury Board
Fiscal externality		
Tax returns and reduced transfers	\$971M	CTaCS
Youth crime (long-run)	-\$19.36M	Baker et al. (2019) and Cohen (2020)
Willingness-to-pay		
Mothers of preschoolers	\$2,213M	CTaCS
Mothers of older children	\$1,102M	Lefebvre et al. (2009)
Taxes and reduced transfers	-\$971M	CTaCS
Youth crime (long-run)	-\$20.16M	Baker et al. (2019) and Cohen (2020)
MVPF	1.40	

Notes: This table outlines the components of the MVPF under the benchmark estimator including additional societal costs of increased youth criminal activity. The last column reports the external sources used for the policy's cost and other sources of fiscal externalities. The acronym CTaCS refers to the Canadian Tax and Credit Simulator of Milligan (2019).

B Quality of care in childcare facilities

Under strong public pressure to open more spots at a reduced fee, the Québec government maintained minimal educational standards for daycare workers to facilitate entry into the profession at the implementation of the reform. A report on childcare quality in regulated settings from the *Institut de la Statistique du Québec* (Québec Statistical Institute) in 2004 emphasized the need for improving quality of care in those institutions (see **?**). Moreover, an audit study conducted by Japel et al. (2005) between 2000 and 2003 revealed that the majority of childcare settings (61%) only met the basic criteria (ensuring the children's health and safety), and that their educational component was minimal. Almost one-eighth of them failed to meet the minimum standards.

However, Japel et al. (2005) also found that CPEs, on average, outperformed all other childcare settings on the vast majority of the criteria they considered for quality. For example, 26.5% of unregulated daycares (home-based or for-profit) were rated as inadequate in terms of quality, but only 6% of CPEs were rated as such. In the same vein, only 12.5% of unregulated daycares provided more than the "minimal" quality, while 33% of CPEs were deemed to provide a good service. Therefore, this evidence suggests that quality issues were actually more important in the private childcare market.[39] One part of the solution to improve average quality, some observers argued, was thus to increase the number of reduced-fee regulated spaces in CPEs.

In response to these quality issues, the provincial government, in addition to increasing the quantity of subsidized spaces, also gradually implemented some quality changes. In 2000, as documented by Molnár (2023), the educational requirements and wages of staff in regulated facilities were substantially increased over a four-year period. The average wage of child care workers was raised by 38 to 40 percent over this time span. The staff-to-child ratios remained unchanged (except for four and five year olds whose ratio increased by 25%) despite the increase in maximum capacity (Baker et al., 2005). Qualification requirements for the staff in centre-based CPEs were raised, and they were then extended to all centre-based care in 2006. In addition to political will, the increase in parents' involvement (in the board of directors, for example) is also an important factor that led to these quality changes.

[39] One potential reason for those quality differences is the greater generosity of infrastructure subsidies given by the government to daycare providers of subsidized spaces than to those operating privately.

C Mathematical Appendix

C.1 Proof of equation (19)

Substituting the time constraints into the budget, we can rewrite (17) as:

$$C(\theta) + (1 - \tau_d(w(T - T_m(\theta) - \ell(\theta)) + I(\theta))\, p(T_c - T_m(\theta)) = w(T - T_m(\theta) - \ell(\theta)) + I(\theta)$$

Isolating non-labor income yields:

$$I(\theta) = \underbrace{\left(C(\theta) + (1 - \tau_d(w(T - T_m(\theta) - \ell(\theta)))\, p(T_c - T_m(\theta)) - w(T - T_m(\theta) - \ell(\theta))\right)}_{\equiv A(\theta)} (1 + \tau_d(T_c - T_m(\theta)))^{-1}$$

We are thus left with this constraint as well as the modified child skills production technology $\tilde{H}(T_m(\theta), e(\theta); h_0)$. The Lagrangian of the problem thus becomes:

$$\begin{aligned}
\mathscr{L} = {} & U(C(\theta), \ell(\theta), e(\theta), T_m(\theta); \theta) \\
& - \lambda \left[A(\theta)\, (1 + \tau_d(T_c - T_m(\theta)))^{-1} - I(\theta) \right] \\
& - \mu \left[h_1(\theta) - \tilde{H}(T_m(\theta), e(\theta); h_0) \right]
\end{aligned}$$

where we removed h_1 from the utility function because it is not a choice variable. The first-order conditions for optimality at interior solutions are thus given by:

$$\text{FOCs:} \quad \frac{\partial U(C^*(\theta), \cdot)}{\partial C} = \lambda\, (1 + \tau_d(T_c - T_m(\theta)))^{-1} \qquad \frac{\partial U(e^*(\theta), \cdot)}{\partial e} = \mu \frac{\partial \tilde{H}(\theta)}{\partial e}$$

$$\frac{\partial U(\ell^*(\theta), \cdot)}{\partial \ell} = \lambda\, (1 + \tau_d(T_c - T_m(\theta)))^{-1}\, \frac{\partial A(\theta)}{\partial \ell}$$

$$\frac{\partial U(T_m^*(\theta), \cdot)}{\partial T_m} = \lambda \left[\frac{\partial A(\theta)}{\partial T_m} \cdot (1 + \tau_d(T_c - T_m(\theta)))^{-1} + A(\theta) \cdot \tau_d\, (1 + \tau_d(T_c - T_m(\theta)))^{-2} \right] + \mu \frac{\partial \tilde{H}(\theta)}{\partial T_m}$$

For an infinitesimal policy change, the difference in indirect utilities (the numerator in (18)) is simply the total derivative of $V(\theta_0)$ with respect to θ, which is given by:

$$\begin{aligned}
\frac{dV(\theta_0)}{d\theta} = {} & \gamma_{d,2} + \frac{\partial U(C^*(\theta_0), \cdot)}{\partial C} \cdot \frac{\partial C^*(\theta_0)}{\partial \theta} + \frac{\partial U(\ell^*(\theta_0), \cdot)}{\partial \ell} \cdot \frac{\partial \ell^*(\theta_0)}{\partial \theta} + \frac{\partial U(T_m^*(\theta_0), \cdot)}{\partial T_m} \cdot \frac{\partial Tm^*(\theta_0)}{\partial \theta} \\
& + \frac{\partial U(e^*(\theta_0), \cdot)}{\partial e} \cdot \frac{\partial e^*(\theta_0)}{\partial \theta}
\end{aligned}$$

where $\gamma_{d,2}$ is the utility gain stemming from the change in coverage. Using the first-order conditions derived above, we have:

$$\begin{aligned}\frac{dV(\theta_0)}{d\theta} &= \lambda\left((1+\tau_d(T_c-T_m(\theta_0)))^{-1}\frac{\partial C^*(\theta_0)}{\partial\theta} + (1+\tau_d(T_c-T_m(\theta_0)))^{-1}\frac{\partial A(\theta_0)}{\partial \ell}\frac{\partial \ell^*(\theta_0)}{\partial\theta}\right)\\ &+\lambda\left[\frac{\partial A(\theta)}{\partial T_m}\cdot(1+\tau_d(T_c-T_m(\theta)))^{-1}+A(\theta)\cdot\tau_d\,(1+\tau_d(T_c-T_m(\theta)))^{-2}\right]\\ &+\mu\left(\frac{\partial\tilde{H}(\theta_0)}{\partial T_m}\frac{\partial Tm^*(\theta_0)}{\partial\theta}+\frac{\partial\tilde{H}(\theta_0)}{\partial e}\frac{\partial e^*(\theta_0)}{\partial\theta}\right)+\gamma_{d,2}\end{aligned}$$

In the expression above, the first two lines are equal to the product of the marginal utility of income (λ) and the treatment effect on earnings ($I^*(\theta)$). Therefore, using the budget constraint, we can replace those lines by $\lambda I^*(\theta)$, which yields the result. □

C.2 Willingness to pay for a policy change

In this Appendix, we generalize the results highlighting the biases in the estimation of the willingness to pay for a large reform discussed in section 5.2. We use a Hendren (2016) framework slightly generalized so as to include non-pecuniary attributes. We first discuss the WTP of a single individual and then aggregation over all beneficiaries to move to social welfare.

C.2 Individual willingness-to-pay

Environment. Consider a decision maker $i \in \mathcal{I}$ facing the problem of choosing a vector of J market goods $\mathbf{x_i} = (x_i^1, ..., x_i^J)$, which can include labor-market activity, and K non-market choice variables $\mathbf{z_i} = (z_i^1, ..., z_i^K)$ to maximize a utility function $u_i(\mathbf{x_i}(\theta), \mathbf{z_i}(\theta))$ where θ is a set of policy choices of the government (the tax schedule, the level of public-good provision, the net transfers to the agent, and so on). The government's policy choice θ can potentially affect the agent's choices $\mathbf{x_i}$ and $\mathbf{z_i}$, her after-tax income y_i, and prices of goods. The agent faces a standard budget constraint as well as a set of additional constraints on the non-market choice variables. For simplicity, we suppose this set is a singleton so that the agent has to meet the budget constraint and one constraint on $\mathbf{z_i}$ (for example, a time allocation constraint). The decision problem thus writes:

$$\max_{\mathbf{x_i}(\theta),\, \mathbf{z_i}(\theta)} u_i(\mathbf{x_i}(\theta), \mathbf{z_i}(\theta)) \quad s.t. \;\; p(\theta)\cdot\mathbf{x_i}(\theta) \leq y_i(\theta)$$
$$g(\mathbf{z_i}(\theta)) = 0$$

where $p = (p_1, ..., p_J)$ is a price vector and g is differentiable in each of its arguments. Let $V_i(\theta) = U(\mathbf{x_i}^*(\theta), \mathbf{z_i}^*(\theta))$ be the agent's indirect utility under policy state θ.

The government now implements a policy change. The reform moves the policy state θ from the status-quo policy θ_0 to some new policy state θ_1. The agent's willingness-to-pay (WTP) for this policy change can be measured by the standard equivalent variation (EV), which we denote $\Delta y_i(\theta_0)$. That is, the WTP is the variation in income under θ_0 that would make the agent indifferent between the status quo and the new policy state:

$$\Delta y_i(\theta_0) = \frac{V_i(\theta_1) - V_i(\theta_0)}{\lambda_i} \tag{21}$$

where λ_i is the agent's marginal utility of income.

WTP for small policy changes. Let us consider first, as is the case with the sufficient-statistics approach, that the policy change is infinitesimal. For an infinitesimal (marginal) policy change (in θ), the numerator in (21), the difference in indirect utilities, is the total derivative of $V_i(\theta_0)$ with respect to θ. Under the additional standard assumption that prices of goods remain unchanged at the margin (i.e. assuming competitive markets for $\mathbf{x}$), we get:

$$\frac{dV_i(\theta_0)}{d\theta} = \lambda_i \, \frac{dy_i^*(\theta_0)}{d\theta} + \mu_i \, g'(\mathbf{z_i}^*(\theta_0)) \cdot \frac{d\mathbf{z_i}^*(\theta_0)}{d\theta} \tag{22}$$

where μ_i is the Lagrange multiplier on the second constraint.
Proof. The Lagrangian of the problem writes:

$$\mathscr{L} = U(\mathbf{x_i}(\theta), \mathbf{z_i}(\theta)) - \lambda_i \left[p(\theta)\mathbf{x_i}(\theta) - y_i(\theta)\right] - \mu_i g(\mathbf{z_i}(\theta)) \tag{23}$$

and thus the solution satisfies the first-order conditions:

$$\text{FOCs:} \quad \frac{\partial U_i(\mathbf{x_i}^*(\theta), \mathbf{z_i}(\theta))}{\partial \mathbf{x_i}} = \lambda_i \, p(\theta) \qquad \frac{\partial U_i(\mathbf{x_i}(\theta), \mathbf{z_i}^*(\theta))}{\partial \mathbf{z_i}} = \mu_i \, g'(\mathbf{z_i}^*(\theta)) \tag{24}$$

where λ_i is the agent's marginal utility of income and μ_i is the Lagrange multiplier on the second constraint. Let $V_i(\theta) = U_i(\mathbf{x_i}^*(\theta), \mathbf{z_i}^*(\theta))$ be the agent's indirect utility under policy state θ.

For an infinitesimal policy change, the difference in indirect utilities (the numerator in (21)) is simply the total derivative of $V_i(\theta_0)$ with respect to θ, which is given by:

$$\frac{dV_i(\theta_0)}{d\theta} = \frac{\partial U_i(\mathbf{x_i}^*(\theta_0), \mathbf{z_i}(\theta_0))}{\partial \mathbf{x_i}} \cdot \frac{\partial \mathbf{x_i}^*(\theta_0)}{\partial \theta} + \frac{\partial U_i(\mathbf{x_i}(\theta_0), \mathbf{z_i}^*(\theta_0))}{\partial \mathbf{z_i}} \cdot \frac{\partial \mathbf{z_i}^*(\theta_0)}{\partial \theta} \tag{25}$$

Using the first-order conditions (24), we have:

$$\frac{dV_i(\theta_0)}{d\theta} = \lambda_i \, p(\theta_0) \, \frac{\partial \mathbf{x_i}^*(\theta_0)}{\partial \theta} + \mu_i \, g'(\mathbf{z_i}^*(\theta)) \frac{\partial \mathbf{z_i}^*(\theta_0)}{\partial \theta} \tag{26}$$

Taking the derivative of the budget constraint with respect to θ yields:

$$\frac{\partial y_i(\theta)}{\partial \theta} = \frac{\partial p(\theta)}{\partial \theta} \, \mathbf{x_i}(\theta) + p(\theta) \, \frac{\partial \mathbf{x_i}(\theta)}{\partial \theta}$$

Therefore, assuming that prices are not impacted by the policy change at the margin (for example, assuming competitive markets for $\mathbf{x}$), the first term on the right-hand-side is null and we obtain that the impact of the policy on earnings is simply given by:

$$\frac{\partial y_i(\theta)}{\partial \theta} = p(\theta) \frac{\partial \mathbf{x_i}(\theta)}{\partial \theta} \tag{27}$$

that is, the additional spendings induced by the policy. Substituting (27) into (26) yields the result:

$$\frac{dV_i(\theta_0)}{d\theta} = \lambda_i \frac{dy_i^*(\theta_0)}{d\theta} + \mu_i \, g'(\mathbf{z_i}^*(\theta_0)) \cdot \frac{d\mathbf{z_i}^*(\theta_0)}{d\theta}$$

□

Therefore, the numerator of the WTP is the sum of two terms, the *pecuniary* benefits and the *non-pecuniary* gains stemming from the relaxation of the second constraint (e.g. the increase in available time). The intuition behind this result is the logic of the Envelope theorem, which implies that, at the margin, behavioral responses do not have a direct effect on utility (i.e. $\partial V_i(\theta)/\partial \mathbf{x_i} = \partial V_i(\theta)/\partial \mathbf{z_i} = 0$). Thus, if one additionally assumes that non-pecuniary gains are negligible, which may be reasonable for marginal reforms, the difference in utilities boils down to the policy's impact on the beneficiaries' budget constraint. The WTP is then simply given by the causal effect of the policy on earnings ($\frac{dy_i(\theta)}{d\theta}$). This result is powerful because it implies that the treatment effect on beneficiaries' earnings is a sufficient statistic for the numerator of the MVPF (Hendren, 2016; Hendren and Sprung-Keyser, 2020). The MVPF framework thus leverages the recent "credibility revolution" in the estimation of causal effects (Angrist and Pischke, 2010) to make transparent welfare statements.

Large-policy bias. Consider now a discrete (large or non-infinitesimal) policy change. In this case, such as with the Québec childcare reform, the previous result does not hold anymore since Envelope conditions only apply to marginal reforms. In particular, behavioral responses, both for market and non-market choices, now have direct impacts on utility because the agent re-optimizes behavior. Moreover, for large reforms, non-pecuniary gains may be important. Thus, the treatment effect on earnings of beneficiaries is a biased estimate of the WTP.[40] For policies with large direct costs, as is the case of the Québec reform, underestimating the WTP might seriously affect the welfare conclusions.

The first bias, which we label the *large-policy bias* (equal to the policy's impact on utility through $\mathbf{x}$), stems from re-optimization behavior of beneficiaries. Since agents make non-marginal changes in market choices, these no longer have a null direct impact on the difference in utilities ($V(\theta_1) - V(\theta_0)$). This bias is likely to be large in our context, given that the literature has documented major impacts of the reform on economic behavior. The large changes in maternal labor supply and child care use have direct impacts on utility through changes in mothers' time allocation, which are not captured by the treatment effect on earnings.

Non-pecuniary gains. Using the treatment effect on beneficiaries' earnings as an estimator of the WTP is subject to a second bias (equal to the policy's impact on utility through $\mathbf{z}$), namely the omission of non-pecuniary benefits of the policy. This bias, in fact, also applies

[40] To see this, suppose the utility function u is additively separable in $\mathbf{x_i}$ and $\mathbf{z_i}$ such that $u_i(\mathbf{x_i}, \mathbf{z_i}) = u_i^1(\mathbf{x_i}(\theta)) + u_i^2(\mathbf{z_i}(\theta))$. The difference in indirect utilities is given by:

$$V_i(\theta_1) - V_i(\theta_0) = u_i^1(\mathbf{x_i}^*(\theta_1)) - u_i^1(\mathbf{x_i}^*(\theta_0)) + u_i^2(\mathbf{z_i}^*(\theta_1)) - u_i^2(\mathbf{z_i}^*(\theta_0))$$

The first (resp. second) difference captures the overall impact of the policy on utility via behavioral changes in $\mathbf{x_i}$ (resp. $\mathbf{z_i}$). For non-marginal policies, differences in u_i^1 and u_i^2 are no longer the partial derivatives of $V_i(\theta_0)$. The *large-policy bias* is given by: $u_i^1(\mathbf{x_i}^*(\theta_1)) - u_i^1(\mathbf{x_i}^*(\theta_0)) - \lambda_i \frac{dy_i^*(\theta_0)}{d\theta}$. The bias stemming from the omission of non-pecuniary gains is simply $u_i^2(\mathbf{z_i}^*(\theta_1)) - u_i^2(\mathbf{z_i}^*(\theta_0))$ since these are ignored by assumption.

to small reforms and Hendren and Sprung-Keyser (2020) themselves acknowledge that it may be important in some cases.[41] We argue that non-pecuniary gains (or losses) are likely to be large in the case of childcare policies (even for small-scale programs) since they may have substantial impacts on (especially mothers') parenting time and practices. Moreover, early childhood programs have substantial impacts on child development, which is valued by parents.

C.2 Social welfare

We now consider aggregation of individual beneficiaries' willingness-to-pay to obtain an estimate of the society's willingness-to-pay. Assuming there exists of a set of Pareto weights ψ_i, for each beneficiary i, social welfare at a given policy state θ is given by:

$$W(\theta) \;=\; \sum_{i\in\mathcal{I}} \psi_i V_i(\theta) \tag{28}$$

where W is the social welfare function and V_i is the indirect utility function of beneficiary i. This formulation is very general and can accommodate any social welfare function. It allows, for instance, social preferences for redistribution from richer to poorer individuals (Hendren, 2016).[42]

The society's WTP, which we denote by SWTP, for a reform is then given by (the monetary value of) the difference in social welfare between the the new (θ_1) and the status-quo (θ_0) policy states. Using the equivalent variation $\Delta y_i(\theta_0)$ as a measure of beneficiary i's WTP in dollars, the society's WTP is thus given by:

$$\text{SWTP} \;=\; \sum_{i\in\mathcal{I}} \psi_i \, \frac{V_i(\theta_1) - V_i(\theta_0)}{\lambda_i} \tag{29}$$

where we used equation (21). As Hendren and Sprung-Keyser (2020) note, the ratio ψ_i/λ_i is the marginal social utility of individual i's income.

Sufficient-statistics approach As argued in the previous section, if the reform and non-pecuniary gains are sufficiently small, the equivalent variation for a beneficiary boils down to the treatment effect on her earnings. As a naive estimator of the society's willingness-to-pay for a large reform, one can use this powerful result, as if the policy change were infinitesimal. Therefore, using equation (22), this estimator can be written as:

$$\text{SWTP}_{naive} = \sum_{i\in\mathcal{I}} \psi_i \; (y_i^*(\theta_1) - y_i^*(\theta_0)).$$

Thus, if the social welfare criterion is utilitarian, this estimator is simply the sum of (weighted) pecuniary gains of all beneficiaries. In other words, the naive estimator of the SWTP is the

[41] For example, in their estimation of the MVPF of admission to Florida International University, changes in effort at school or other forms of utility gains during college time are assumed away (Hendren and Sprung-Keyser, 2020, p. 1230). We discuss many other reforms for which non-pecuniary gains might be important in our survey of MVPF estimates in Appendix E.

[42] As additional examples, a utilitarian planner sets $\psi_i = 1 \; \forall i$ while a Rawlsian planner sets $\psi_{i'} > 0$ for i' such that $V_{i'} \leq V_i \; \forall i$ and $\psi_i = 0 \; \forall i \neq i'$.

treatment effect on (weighted) aggregate earnings.

Structural approach For discrete (large) policy changes, the Envelope theorem does not apply. Since beneficiaries do not simply react at the margin, behavioral responses have first-order impacts on utility. Therefore, one has to estimate the utility gains for each individual, which may include gains (losses) from re-optimization and non-pecuniary benefits. These can be estimated using a well-specified structural model of behavior. To obtain the equivalent variation, one can use these estimates to find, for each beneficiary, the amount of additional income that would make her indifferent between the extra cash and the implemented policy. This alternative estimator takes the form:

$$\text{SWTP}_{model} \;=\; \sum_{i\in\mathcal{I}} \psi_i \, \frac{\tilde{V}_i(\theta_1) - \tilde{V}_i(\theta_0)}{\tilde{\lambda}_i} \tag{30}$$

where $\tilde{w}$ denotes that variable w is estimated from the model.

D Comparison of MVPF estimates with other criteria

There are recent debates in Economics on the use of the MVPF as a welfare criterion to evaluate social programs. In particular, García and Heckman (2022a,b) criticize the use of this metric arguing that the MVPF approach *(i)* abstracts from the deadweight loss of taxation and thus from the social marginal value of public expenditure, *(ii)* assumes a fixed government budget and is silent about policies that loosen the government's budget constraint, *(iii)* uses an arbitrary ratio, *(iv)* cannot rank all policies, and *(v)* interprets some welfare-improving policies as "money pumps". They suggest the use of an alternative criterion, namely the net social benefit (NSB). The NSB is simply the difference between the policy's benefits and the cost to society. García and Heckman (2022b) define the NSB as:

$$\text{NSB} = \text{Benefits} - \Omega(\text{Direct cost}) \tag{31}$$

where $\Omega()$ is a potentially non-linear function, which notably captures the deadweight loss of public expenditure. In practice, however, a linearity assumption on Ω is often made in the literature and we follow this approach in our comparative exercise below. We compare our MVPF estimates with the NSB and the standard cost-benefit ratio in Table D1.

Before discussing the results we note, however, that in a reply, Hendren and Sprung-Keyser (2022) show that those critics originate from a misconception about their welfare criterion and that, in several contexts, the MVPF may be preferable to the NSB. For example, one key advantage of the MVPF framework is that it does not assume how the government finances the policy while the standard deadweight loss of taxation assumes an arbitrary linear income tax rate. Hendren and Sprung-Keyser (2020)'s criterion evaluates welfare impacts of budget-neutral programs by comparing two MVPFs: the one of an expenditure policy to the one of a revenue-raising policy. On arguments *(iv)* and *(v)*, the MVPF approach identifies policies that pay for themselves and for which recipients have a positive willingness-to-pay as Pareto improvements (defined as an infinite MVPF, not as a negative one as point *(v)* states). It is thus true that one cannot rank among Pareto improvements, but the message here is that the government should implement all those policies (at no cost) so ranking them is obsolete. Last, a fair criticism of empirical welfare analysis in García and Heckman (2022b) is that, in reality, the welfare costs of raising public revenue are likely non-linear. Such non-linearities in the deadweight loss of public expenditure cannot be accounted for in the MVPF framework, but the critique also applies to other standard criteria for evaluating social programs considered by García and Heckman (2022b). Estimating non-linear welfare costs of raising public revenue is a promising avenue for future research, but is beyond the scope of this paper. We refer the reader to those papers for a more extensive discussion.

Table D1: Comparison between MVPF and alternative social welfare criteria

Criterion	Formula	Value
Benchmark estimator		
MVPF	Benefits / Net cost	1.42
NSB	Benefits - $(1+\phi)$ Cost	-\$87.1M
CBR	Benefits / $[(1+\phi)$ Cost$]$	0.97
Structural estimator		
MVPF	Benefits / Net cost	3.56
NSB	Benefits - $(1+\phi)$ Cost	\$2675.9M
CBR	Benefits / $[(1+\phi)$ Cost$]$	1.79

Note: We assume a deadweight loss of public expenditure $\phi = 1/3$ in this example, as in Hendren and Sprung-Keyser (2022). MVPF = marginal value of public funds, NSB = net social benefit, CBR = cost-benefit ratio

Our comparative exercise in Table D1 reveals that, for the benchmark estimator, the choice of criterion substantially affects social-welfare conclusions. When focusing on earnings gains only, we find that the cost-benefit ratio (CBR) is lower than one (0.97) and that the NSB is negative (-\$87.1M), suggesting that the policy is not socially desirable. However, when incorporating non-pecuniary gains, we find that all three criteria point at the same conclusion: benefits are larger than costs and the policy should be implemented under these criteria. Indeed, we obtain a cost-benefit ratio of 1.79 and a positive NSB of more than \$2 billion under our structural estimator. Therefore, we find that using these alternative criteria reinforce our conclusion that omitting non-pecuniary gains for mothers would lead to a substantial underestimation of social-welfare gains. In fact, under the CBR and NSB criteria, abstracting from mothers' non-monetary benefits would lead one to conclude that the policy should not be adopted. This reinforces our main result that non-pecuniary gains for mothers must be accounted for in welfare analysis of universal preschool policies.

E Survey of MVPF estimates

Some authors compute social-welfare impacts of large policy changes using sufficient-statistics estimators (such as the MVPF) as if the policy were infinitesimal.[43] To assess the prevalence of this practice, we conduct our own survey of MVPF estimates appearing on the Policy Impacts Library (Hendren et al., 2023). Of course, this exercise requires some judgement calls and we therefore only take the results of our survey as suggestive.

We first have to define a criterion indicating whether a policy change can be considered as infinitesimal. Importantly, envelope conditions allowing one to express the welfare effect of a policy change as a fiscal externality concern individuals' utility maximization problem. The "size" of the policy change should thus not be evaluated using the number of recipients or as a function of how local is the treatment effect estimated, but rather by the size of behavioral responses at the *individual* level.

In line with our discussion of the *large-policy bias* in section 5.2, we define a discrete (non-infinitesimal) policy as one that induces significant behavioral re-optimization by recipients. For infinitesimal changes, by the Envelope theorem, recipients do not re-optimize at the margin and only obtain utility gains from the relaxation of constraints in their maximization problem. Ignoring non-pecuniary gains of small reforms may also be reasonable. However, large behavioral responses to a policy such as entry into the labor market suggest agents face a different economic environment and revise their optimal choices, suggesting the policy change is not small. Omitting non-pecuniary gains (or losses) of abrupt changes in behavior may also lead to important biases in welfare estimates.

We have so far surveyed the first 24 papers appearing on the Policy Impacts Library webpage. For each paper, in addition to providing basic information on the reform considered, we assess whether the policy change being studied satisfies the infinitesimal-policy criterion defined above. We also list some potential non-pecuniary gains (or losses) omitted by the authors. Lastly, we briefly discuss whether the large-policy bias or the omission of non-pecuniary gains is likely to affect the authors' welfare conclusions. In particular, we check whether papers finding apparently welfare-improving (resp. welfare-decreasing) policies are omitting utility losses (resp. gains) which biases their estimates upwards (resp. downwards).[44] In our discussions, we focus on policy impacts on outcomes studied by the authors and limit mentions to other fiscal externalities.

The detailed results are reported in Table E1. Two key findings stand out of this review. First, most papers who apply the MVPF framework do so in the context of a non-infinitesimal policy change. Out of the first 24 MVPF estimates appearing in the Policy Impacts Library, we find that at least 20 cases clearly do not satisfy our criteria for the policy studied be considered

[43] We note that it is likely not in the authors' intention to make this assumption. They might simply apply the MVPF framework because it is a convenient tool to evaluate welfare impacts of policies using reduced-form causal estimates. For most papers, the welfare analysis is not the main goal and authors may only see this exercise as illustrative of the economic returns of the reform studied.

[44] One of the stated advantage of the MVPF framework is that it "can be easily compared across programs" (Kline and Walters, 2016, p. 1815). However, for such comparisons across policy domains, which require a precise ranking of policies, obtaining a robust MVPF point estimate is crucial. Therefore, even if the biases do not affect the authors' general welfare conclusions (i.e. whether the policy is welfare-improving or not), omitting potentially large utility gains (or losses) can be problematic for policy comparisons.

as infinitesimal. For example, in many cases, the policy is found to have large impacts on labor supply at the extensive margin. Such employment responses are not marginal. Therefore, unless all beneficiaries are indifferent between working and staying out of the labor force, the large-policy bias discussed in section 5.2 applies.

Second, among those 20 cases, we argue that for at least seven, abstracting from the large nature of the policy change is likely to substantially affect the authors' welfare conclusions. For example, the MVPF of the introduction of old-age pensions in the United Kingdom is estimated to be 0.8 by Giesecke and Jäger (2021). This policy is not infinitesimal as made evident by the impacts on labor-supply and retirement decisions documented by the authors. For this reform, positive non-pecuniary gains for grandchildren such as reduced time taking care of the elderly as well as reduced financial stress related to retirement for beneficiaries are omitted. If these gains are sufficiently large, the MVPF could be higher than one, suggesting that the policy is welfare-improving rather than being a net cost to society.

Table E1: Survey of MVPF estimates in Policy Impacts Library

Authors	Causal estimates	Policy considered	Recipients	MVPF	Infinitesimal?	Agents re-optimize?	Non-pecuniary gains/losses	Welfare conclusions likely affected?
Duchini and Van Effenterre (2022)	Duchini and Van Effenterre (2022)	Introduction of primary school on Wednesday morning in France	Parents (mothers) of primary-school age children in France	3.6	No	Large adjustments in week schedule	Parenting time, more regular work schedules, child cognition	No. MVPF is already high and reduced child care contraints should generate even more welfare gains.
Deshpande and Mueller-Smith (2022)	Deshpande and Mueller-Smith (2022)	Conditioning of access to the Social Security Income on re-evaluation of disability at 18 years old	Disadvantaged youth who would have received guaranteed SSI in adulthood	-16.1	No	Impacts on crime found to be concentrated in income-generating criminal offenses. This may suggest normally sub-optimal decisions to catch-up for the income loss. Consistent with this, Deshpande (2016) finds positive impacts of SSI removal on labor supply.	Psychic costs (or benefits) of criminal activity	No. MVPF is high (in abs. value) and reduced income in youth likely to imply additional negative distortions.
Hendren and Sprung-Keyser (2020)	Zimmerman (2014)	Admission to Florida International University	Prospective students at FIU	∞	No	Admission to university likely to induce significant changes in life trajectory, preferences, etc.	utility of attending school, effort costs, moving costs from other States	No. Even if benefits might be overestimated, the net costs are negative.

Authors	**Causal estimates**	**Policy considered**	**Recipients**	**MVPF**	**Infinitesimal?**	**Agents re-optimize?**	**Non-pecuniary gains/losses**	**Welfare conclusions likely affected?**
Paradisi (2021)	Brenøe et al. (2020)	Maternal leave in Denmark (effect of having female employees close to giving birth)	Coworkers and mother's firm	0.74	Arguably yes	Loss of only one employee for a short amount of time. Firms found to compensate by marginal adjustments, but no decrease in labor demand.	Maybe change in working atmosphere	No. MVPF below 1 and no omitted positive impact.
Hendren and Sprung-Keyser (2020)	Currie and Gruber (1996); Cutler and Gruber (1996); Dave et al. (2015); Miller and Wherry (2019)	Medicaid Expansions to Pregnant Women and Infants in US States	Pregnant women and children in the USA	∞	No	Large employment responses of women and probably behavioral adjustments in health habits.	Health insurance likely an important source of stress for families.	No. The policy already pays for itself and there should not be significant losses missing.
Hendren and Sprung-Keyser (2020)	Cohodes and Goodman (2014)	Massachusetts' Adams (MA) Scholarship	High-scoring high-school students	0.72	No	Change college choice and less likely to graduate	-	No. MVPF below 1 and no omitted positive impact.
Kuka and Shenhav (2020)	Kuka and Shenhav (2020)	EITC Benefits to Recent Mothers (1993)	Low-income workers (new non-married mothers)	5.6	No	Large effects on employment and wages	Improved mother's health and their children's outcomes	No. MVPF is already high.

Authors	**Causal estimates**	**Policy considered**	**Recipients**	**MVPF**	**Infinite-simal?**	**Agents re-optimize?**	**Non-pecuniary gains/losses**	**Welfare conclusions likely affected?**
Cabral and Dillender (2021)	Cabral and Dillender (2021)	Workers Compensation Benefit Generosity	Injured workers	0.46	No	Medical spending and benefit duration increase	Workers may value the increased medical spending and the consumption-smoothing benefits afforded by more generous coverage (acknowledged by the authors)	Yes. If these benefits are sufficiently high, MVPF could be higher.
Hendren and Sprung-Keyser (2020)	Cornwell et al. (2006)	Georgia HOPE Scholarship	High-school students graduating with at least a B average	4	No	Large effects on college enrollment	Utility of attending college, effort costs	No. MVPF is very high.
Bastian and Jones (2021)	Bastian and Jones (2021)	All post-1990 EITC Expansions for Women	Low-income workers (women)	3.18-4.23	No	Large effects on employment	Improved health, reduced crime, improved child outcomes. (All accounted for only as fiscal externalities.)	No. When accounting for all fiscal externalities, policy already “pays for itself”.

Authors	Causal estimates	Policy considered	Recipients	MVPF	Infinite-simal?	Agents re-optimize?	Non-pecuniary gains/losses	Welfare conclusions likely affected?
Giesecke and Jäger (2021)	Giesecke and Jäger (2021)	Introduction of Old-Age Pensions in the UK	Individuals age 70+	0.8	No	Re-optimization of work / retirement decision. Large monetary transfer (22% of average income)	Time to help taking care of grand-children. Large change in life schedule. Financial stress relief.	Yes. MVPF close to 1. Potential welfare gain for grand-children and relief stress.
Bergolo and Cruces (2021)	Bergolo and Cruces (2021)	Conditional Cash Transfers in Uruguay	Low-income households with children and/or pregnant women. Covering 42 percent of children under 18	0.61	No	Discrete change in behaviour (health and educational conditions). Change in take-up of other social programs	Parenting time, Costs or gains of complying to the program's health and educational requirements	Maybe. Depends if the "complying with conditionalities" aspect has a positive or negative value for individuals.
Wettstein (2020)	Wettstein (2020)	Introduction of Medicare Part D	Individuals age 65+	1.98	No	If full time workers are working to get the private insurance, the policy leads to a complete re-optimization	If agents are liquidity constrained, they may value drugs more than their price, which means the WTP is higher. Gain on health and released stress from not being able to buy drugs	Yes. Some positive effects are neglected. MVPF could be higher

Authors	**Causal estimates**	**Policy considered**	**Recipients**	**MVPF**	**Infinite-simal?**	**Agents re-optimize?**	**Non-pecuniary gains/losses**	**Welfare conclusions likely affected?**
Ganimian et al. (2021)	Ganimian et al. (2021)	Early-Childhood Education in India	pre-schoolers	∞	Arguably yes	Not with the pilot. If the policy is enlarged to the whole population, it might change the population of recipients (Parents of lower or higher expected-gain children now enroll because they know that it will improve their kid's outcomes)	Schooling might be more pleasant.	No. MVPF already infinite, only additional gains.
Jácome (2022)	Jácome (2022)	Medicaid Eligibility for Teenagers in South Carolina	19 year olds losing the coverage	1.77 to 14.96	No	Induce significant life decision change. If some individuals have child to get access to the program, it change significantly their behaviours.	Reduced probability of being incarcerated. Reduced opportunities of earnings via crimes. Access to better jobs (better in non-paid compensations)	

Authors	Causal estimates	Policy considered	Recipients	MVPF	Infinitesimal?	Agents re-optimize?	Non-pecuniary gains/losses	Welfare conclusions likely affected?
Deshpande et al. (2021)	Deshpande et al. (2021)	Different evaluation standards (at known age thresholds) for eligibility to the Social Security Income in the USA	3 groups defined by leniency of evaluation standards for SSI eligibility: ¿55, 50-55, ¡50 and spillovers to homeowners	1.04	No	Large impacts on tail events, events that occur infrequently and are associated with large drops in consumption (p.152) like bankruptcy and home sale	Most treated individuals are in great financial distress so potentially large relief and positive metal health impacts of becoming eligible to SSI.	Yes. MVPF around 1 and reduced financial distress and increased access to housing are likely significant gains
Bailey et al. (ming)	Bailey et al. (ming); Hoynes and Schanzenbach (2009, 2012); Hoynes et al. (2016)	Access to Food Stamps (average $4/person/day)	US families (parents and children)	56.25	Probably not	The policy reduced food insecurity, but these are necessity purchases that would happen anyway. However, the employment response of parents suggests that, without food insecurity, some parents would not be working.	Time parents can spend away from work (leisure or child care) thanks to the food vouchers	No. MVPF very high, only additional gains.
Kline and Walters (2016)	Kline and Walters (2016)	Head Start (targeted preschool program)	Disadvantaged children in the US	1.85-2.41	Yes, if focusing only on children, but parents likely to benefit as well.	Children do not choose to attend daycare so no omitted behavioral changes	The value of child development for parents, pareting time, etc.	Possible if negative impacts on parents' utility (eg: time spent with child reduced)

Authors	Causal estimates	Policy considered	Recipients	MVPF	Infinitesimal?	Agents re-optimize?	Non-pecuniary gains/losses	Welfare conclusions likely affected?
Gray et al. (ming)	Gray et al. (ming)	Removal of work requirements for food stamps in Virginia	Potential beneficiaries of food stamps	0.86-1.15	No, which the authors acknowledge. They provide an MVPF estimate including a utility cost of working.	No strong employment responses, but removal of work requirements found to have large impacts on retention.	Contrary to Bailey et al. (2022), the authors consider utility costs of labor.	Yes. The authors themselves show that including utility costs of labor moves the MVPF from below to above 1.
Hyman (2018)	Hyman (2018)	Trade Adjustment Assistance program (retraining incentives and UI) for displaced workers	Displaced workers (due to shifts in production outside the USA)	1.14 in paper; 2.7 on Policy Impacts	No	Important employment responses to the reform. The author acknowledges the caveats of assuming Envelope conditions.	Utility costs of training programs	Possible. The MVPF estimate is slightly higher than one and some utility costs might be omitted.
Deshpande (2016)	Deshpande (2016)	Conditioning of access to the Social Security Income on re-evaluation of disability at 18 years old	Disadvantaged youth who would have received guaranteed SSI in adulthood	0.9 to 1.01	No	Changes in labor supply as a response to the income loss	Income stabilization value of SSI included in second estimate. However, utility cost of working (foregone leisure) is omitted.	Yes. When authors include the insurance value of stable SSI payments, MVPF goes from below to slightly above 1. The utility cost of foregone leisure might push the MVPF back below one.

Authors	**Causal estimates**	**Policy considered**	**Recipients**	**MVPF**	**Infinite-simal?**	**Agents re-optimize?**	**Non-pecuniary gains/losses**	**Welfare conclusions likely affected?**
Finkelstein and Hendren (2020) and Baird et al. (2016)	Baird et al. (2016)	School-based deworming program (community-wide) in Kenya	Treated school-age children and spillovers to neighboring schools	Infinite	Arguably yes	Probably not. Children would probably go to school anyway if they were not sick.	-	No. MVPF is infinite and non-pecuniary losses likely small.
Cascio (2023)	Cascio (2023)	Targeted pre-K programs in the US	Preschoolers in US States	0 (no benefit)	No	Substantial evidence from the literature that pre-K programs yield positive economic returns in the long-run	The value of child development for parents, pareting time, etc.	Yes. There should be benefits for both parents and children.
Cascio (2023)	Cascio (2023)	Universal pre-K programs in the US	Preschoolers in US States	1.96-4.27	No	Substantial evidence from the literature that pre-K programs yield positive economic returns in the long-run	The value of child development for parents, pareting time, etc.	No. MVPF already high and mostly additional gains.

Appendix References

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30.

Bailey, M. J., Hoynes, H. W., Rossin-Slater, M., and Walker, R. (forthcoming). Is the social safety net a long-term investment? large-scale evidence from the food stamps program. *Review of Economic Studies*.

Baird, S., Hicks, J. H., Kremer, M., and Miguel, E. (2016). Worms at work: Long-run impacts of a child health investment. *The quarterly journal of economics*, 131(4):1637–1680.

Baker, M., Gruber, J., and Milligan, K. (2019). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy*, 11(3):1–26.

Bastian, J. E. and Jones, M. R. (2021). Do eitc expansions pay for themselves? effects on tax revenue and government transfers. *Journal of public economics*, 196:104355.

Bergolo, M. and Cruces, G. (2021). The anatomy of behavioral responses to social assistance when informal employment is high. *Journal of Public Economics*, 193:104313.

Brenøe, A. A., Canaan, S. P., Harmon, N. A., and Royer, H. N. (2020). Is parental leave costly for firms and coworkers? Working Paper 26622, National Bureau of Economic Research.

Cabral, M. and Dillender, M. (2021). The impact of benefit generosity on workers' compensation claims: Evidence and implications. Working paper.

Cascio, E. U. (2023). Does universal preschool hit the target? program access and preschool impacts. *Journal of Human Resources*, 58(1):1–42.

Cohen, M. A. (2020). *The costs of crime and justice*. Routledge.

Cohodes, S. R. and Goodman, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6(4):251–85.

Cornwell, C., Mustard, D. B., and Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from georgia's hope program. *Journal of Labor Economics*, 24(4):761–786.

Currie, J. and Gruber, J. (1996). Health Insurance Eligibility, Utilization of Medical Care, and Child Health*. *The Quarterly Journal of Economics*, 111(2):431–466.

Cutler, D. M. and Gruber, J. (1996). Does Public Insurance Crowd out Private Insurance?*. *The Quarterly Journal of Economics*, 111(2):391–430.

Dave, D., Decker, S. L., Kaestner, R., and Simon, K. I. (2015). The effect of medicaid expansions in the late 1980s and early 1990s on the labor supply of pregnant women. *American Journal of Health Economics*, 1(2):165–193.

Deshpande, M. (2016). Does welfare inhibit success? the long-term effects of removing low-income youth from the disability rolls. *American Economic Review*, 106(11):3300–3330.

Deshpande, M., Gross, T., and Su, Y. (2021). Disability and distress: The effect of disability programs on financial outcomes. *American Economic Journal: Applied Economics*, 13(2):151–78.

Deshpande, M. and Mueller-Smith, M. (2022). Does welfare prevent crime? the criminal justice outcomes of youth removed from ssi. *The Quarterly Journal of Economics*, 137(4):2263–2307.

Duchini, E. and Van Effenterre, C. (2022). School schedule and the gender pay gap. *Journal of Human Resources*, forthcoming.

Finkelstein, A. and Hendren, N. (2020). Welfare analysis meets causal inference. *Journal of Economic Perspectives*, 34(4):146–67.

Ganimian, A. J., Muralidharan, K., and Walters, C. R. (2021). Augmenting state capacity for child development: Experimental evidence from india. Working Paper 28780, National Bureau of Economic Research.

García, J. L. and Heckman, J. J. (2022a). On criteria for evaluating social programs. Working Paper 30005, National Bureau of Economic Research.

García, J. L. and Heckman, J. J. (2022b). Three criteria for evaluating social programs. *Journal of Benefit-Cost Analysis*, 13(3):281–286.

Giesecke, M. and Jäger, P. (2021). Pension incentives and labor supply: Evidence from the introduction of universal old-age assistance in the uk. *Journal of Public Economics*, 203:104516.

Gray, C., Leive, A., Prager, E., Pukelis, K. B., and Zaki, M. (forthcoming). Employed in a snap? the impact of work requirements on program participation and labor supply. Technical report.

Hendren, N. (2016). The policy elasticity. *Tax Policy and the Economy*, 30(1):51–89.

Hendren, N. and Sprung-Keyser, B. (2020). A Unified Welfare Analysis of Government Policies*. *The Quarterly Journal of Economics*, 135(3):1209–1318.

Hendren, N. and Sprung-Keyser, B. (2022). The case for using the mvpf in empirical welfare analysis. Working Paper 30029, National Bureau of Economic Research.

Hendren, N., Sprung-Keyser, B., Stuart, E., and Emery, J. (2023). Policy impacts library. https://policyimpacts.org/policy-impacts-library. Last accessed: 2023-01-16.

Hoynes, H., Schanzenbach, D. W., and Almond, D. (2016). Long-run impacts of childhood access to the safety net. *American Economic Review*, 106(4):903–34.

Hoynes, H. W. and Schanzenbach, D. W. (2009). Consumption responses to in-kind transfers: Evidence from the introduction of the food stamp program. *American Economic Journal: Applied Economics*, 1(4):109–39.

Hoynes, H. W. and Schanzenbach, D. W. (2012). Work incentives and the food stamp program. *Journal of Public Economics*, 96(1):151–162.

Hyman, B. (2018). Can displaced labor be retrained? evidence from quasi-random assignment to trade adjustment assistance. Working Paper 3155386.

Jácome, E. (2022). Mental health and criminal involvement: Evidence from losing medicaid eligibility. Working papaer.

Kline, P. and Walters, C. R. (2016). Evaluating Public Programs with Close Substitutes: The Case of Head Start*. *The Quarterly Journal of Economics*, 131(4):1795–1848.

Kuka, E. and Shenhav, N. (2020). Long-run effects of incentivizing work after childbirth. Working Paper 27444, National Bureau of Economic Research.

Lefebvre, P., Merrigan, P., and Verstraete, M. (2009). Dynamic labour supply effects of childcare subsidies: Evidence from a canadian natural experiment on low-fee universal child care. *Labour Economics*, 16(5):490–502.

Miller, S. and Wherry, L. R. (2019). The long-term effects of early life medicaid coverage. *Journal of Human Resources*, 54(3):785–824.

Ministère de la Famille (2013–2019). Rapport annuel. Technical report, Gouvernement du Québec.

Ministère de la Famille, des Aînés et de la Condition Féminine (2005–2006). Rapport annuel de gestion. Technical report, Gouvernement du Québec.

Ministère de la Famille et de l'Enfance (1997–2003). Rapport annuel de gestion. Technical report, Gouvernement du Québec.

Ministère de la Famille et de l'Enfance (1998–2001). Situation des centres de la petite enfance et des garderies au québec. Technical report, Gouvernement du Québec.

Ministère de la Famille et des Aînés (2007–2012). Rapport annuel de gestion. Technical report, Gouvernement du Québec.

Ministère de l'Emploi, de la Solidarité sociale et de la Famille (2004). Rapport annuel de gestion 2003-2004. Technical report, Gouvernement du Québec.

Molnár, T. L. (2023). Costs of daycare, complementarities, and heterogeneous productivity of parenting time in child skill formation. Working paper.

Office des services de garde à l'enfance (1994–1997a). Estimation régionale des besoins en services de garde régis par la loi. Technical report, Gouvernement du Québec.

Office des services de garde à l'enfance (1994–1997b). Situation des centres de la petite enfance et des garderies au québec. Technical report, Gouvernement du Québec.

Office des services de garde à l'enfance et Secrétariat à la famille (1997). Rapport annuel 1996-1997. Technical report, Gouvernement du Québec.

Paradisi, M. (2021). Firms and policy incidence. Working paper.

Wettstein, G. (2020). Retirement lock and prescription drug insurance: Evidence from medicare part d. *American Economic Journal: Economic Policy*, 12(1):389–417.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.