Essays in Health Economics

by

Maripier Isabelle

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Economics
University of Toronto

Abstract

Essays in Health Economics

Maripier Isabelle
Doctor of Philosophy
Graduate Department of Economics
University of Toronto
2018

Webs of complex and often overlapping incentives characterize most modern healthcare systems. Some of these incentives have been purposely designed to influence the delivery of care to patients. Others emerge at the intersection of independent policies and decisions that accidentally create opportunities for arbitrage by providers of care. Using both administrative and survey data from Canada and exploiting institutional features associated with the provision of obstetric care, this thesis investigates how physicians respond to incentives embedded in their remuneration schemes, and explores longer-term consequences of such behaviour changes. The first chapter looks at primary care physicians' responses to a targeted bonus payment introduced to encourage the provision of low-volume intrapartum care in the province of Ontario. The results suggest that, while physicians do alter their behaviour in response to the introduction of explicit incentive payments, they may not do so by providing more of the services associated with the bonus. Instead, income effects may lead to a reduction in the provision of targeted services by eligible physicians, and influence their activities in other areas of their practice. The second chapter studies physicians' responses to subtler financial incentives: higher remuneration for procedures known to have close substitutes. Exploiting exogenous changes in the relative fee-for-service payment received by Canadian doctors for C-sections compared to vaginal deliveries, the findings from this chapter suggest that physicians do respond to such implicit incentives. The results also provide evidence that physicians' choice of birth delivery method might however be more strongly affected by spillovers from the publication of high-profile studies and trials results. The third chapter finally explores the potential long-term implications of physicians' responses to incentives in terms of patients' health outcomes. Investigating the impact of C-section birth on children's health outcomes later in life, and on their future consumption of care, it presents suggestive evidence that the association between the mode of delivery and some health conditions cannot be accounted for by a rich set of environmental and socioeconomic confounders. Overall, these three chapters suggest that financial incentives can shape physicians' behaviour, sometimes in unexpected ways, with both short- and long-term consequences.

Acknowledgements

My first words of gratitude go to Professor Michael Baker and Professor Mark Stabile, who taught me so much these past five years, and whose dedication, support, patience, understanding and precious advice helped and inspired me every step of the way. They are the best mentors I could ever have hoped for, and it is an honour and a privilege to work with them.

I am extremely grateful to Professor Shari Eli for her invaluable comments, guidance and generosity. Her contagious enthusiasm and capacity to approach each challenge with a smile has been and remains a wonderful source of motivation. I would also like to thank the faculty members I had the privilege to learn from, engage with and be encouraged by, as well as the staff and all my colleagues at the department of Economics for their help, kindness and friendship. In particular, I would like to thank Shannon Elliott for always making time to help me power through the more challenging days, and my officemate Genet, with whom sharing discoveries was a true privilege that made my time in the PhD program so much brighter.

I would also like to express my sincere gratitude to John Fraser and Hugh Segal, successively Masters of Massey College, who have always been there to encourage me to push my boundaries. I also want to thank the amazing friends I met at Massey College. It is a privilege to have been a part of such an uplifting, encouraging and welcoming community.

My last words of thanks go to my family and precious friends –the amazing people who were there for me from the beginning, and those who more recently let me enter their lives with open arms. They never sighed at the tears, have always been there to share the laughs and successes of the past five years, and to support me when I felt weaker. Their love is what brings meaning to my life and to my work. Finally, to my husband Taylor: Thank you for teaching me to never give up, for challenging me to keep asking questions, and for being the most supportive, loving and inspiring teammate and partner I could wish for. This dissertation could never have been without you.

Contents

Introduction

Across jurisdictions and nations, policy plays a consequential role in mediating interactions between the supply of and the demand for healthcare. Government interventions potentially impacting physicians, for example, can range from implementing regulations that determine who can provide care and in which environments, to setting the payments received for services covered by public insurance. In many modern healthcare systems, physicians' labour supply has become the target of policy makers in order to address issues such as waitlists, unmet demand for care or cost management. Indeed, financial incentives are often used as a tool to influence doctors' specialization decisions, the location of their practice, or even the scope of the services they provide. However, theoretical work suggests that predicting how physicians will respond to such incentives is often not an easy task, and empirical studies investigating this question provides mixed evidence on their capacity to elicit the intended behavioural changes.[1]

This dissertation first seeks to provide new evidence on the shorter and longer run consequences of physician financial incentives in the context of obstetric care. From a policy perspective, understanding the impact of incentives, either implicit or explicit, on this specific area of care matters since child birth has a non-trivial impact on the allocation and use of resources within healthcare systems. Indeed, birth deliveries are the most common reason for inpatient stays in Canadian and American hospitals [CIHI, 2017, McDermott et al., 2017], and cesarean sections are the most common inpatient surgery performed in hospitals both in Canada and in the United States [CIHI, 2017, HCUP, 2017]. Understanding if and how various incentives in the healthcare system can influence the allocation and the utilization of resources in obstetrics therefore amounts to mapping the consequences of policy interventions that have important potential impacts in terms of the volume of patients as well as the quantity of resources affected. From a conceptual standpoint, focusing on obstetric care, and more specifically on intrapartum care, offers an interesting opportunity to explore a context where the aggregate demand for services (birth deliveries) cannot be influenced by providers, but where incentives can generate adjustments in the intensity of the care provided or lead to changes in the mix of physicians, specialists or primary care, involved.

Another common element to the three chapters presented in this dissertation is the focus on the Canadian context, which provides an interesting setting to understand and interpret the observed consequences of changes in or features of physician incentives. First, the single-payer and universal-coverage nature of the Canadian healthcare system facilitates population-level analysis, both for patients and for physicians, therefore offering a comprehensive view of the impacts of policy changes in the healthcare system. Indeed, most government interventions affecting physicians' compensation or their work environment will affect all providers in the relevant jurisdiction and, since all Canadians are publicly insured for an extensive set of core services, they will apply to the care they provide to the full population of patients unless conditions explicitly indicate otherwise. Second, the nature of Canadian medicare facilitates the interpretation of the estimated responses to physician financial incentives as being driven by supply-side behaviour, given that changes in the payments received by providers do not translate into changes in direct costs faced by patients to access their services. Third, universal public insurance coverage also

[1] Johnson [2014], Chandra et al. [2011] and Nicholson and Propper [2011], among others, provide recent and insightful reviews of the literature.

provides an opportunity to study an environment where physicians generally cannot respond to changes in compensation or incentives by changing the mix of patients they care for based on their insurance types (for example, private or public), or by directly sorting across employers and remuneration packages. Finally, access to care for patients shouldn't be as highly correlated with socioeconomic status as in jurisdictions where private health insurance plays a more prominent role. This limits the extent to which evidence of a relationship (or of the absence thereof) between the care received and the health outcomes observed could be driven by confounders.

Capitalizing on these institutional features, and exploiting rich administrative data on physician remuneration, the first chapter of this dissertation investigates primary care physicians' responses to targeted incentives that are explicitly introduced in their remuneration schedule. While these incentives are a common tool employed by governments to try to influence the delivery of health care, especially primary care, the nature and range of their effects are complex. Both theory and existing empirical evidence suggest that increasing the payment for a medical act does not necessarily lead physicians to increase their provision of the targeted procedure. Moreover, given the relatively broad scope of primary care physicians' practice, their responses to specific bonuses or premiums may also include changes to their activities in areas of care that are not directly targeted by the incentives. Exploiting the introduction of a premium that increased the remuneration for obstetric care in Ontario, the evidence presented highlights that primary care physicians did not increase their provision of the services targeted by the premium following its introduction. On the contrary, doctors who were initially providing higher volumes of those services adjusted their provision downwards in response to the incentive. The results also suggest that the incentive might have negatively affected the provision of services in other areas of care by physicians receiving the premium. These changes in practice style are in line with the predictions of a labour supply model in which income effects are relatively strong. At a time when healthcare budgets are growing at a pace that is often qualified as unsustainable, this paper contributes to understanding the potential broader impacts of targeted financial incentives on the delivery of care, and their alignment with governments' objectives.

Instead of focusing on policies specifically designed to influence physicians' actions and labour supply, the second chapter turns to implicit incentives, which can result from the historical evolution of contracts or workplace organization, or emerge at the intersection of policies and regulations that provide opportunities for arbitrage from a provider's perspective. More specifically, it investigates the role of incentives taking the form of differential payments for substitutable services, C-section and vaginal birth deliveries, on physicians' choice of birth delivery methods. The estimated response obtained from administrative data suggests that doubling the relative compensation of C-sections over vaginal deliveries increases the probability of a C-section by at most four percentage points, thus only explaining a small fraction of the high cesarean rate observed in Canada in the last decades. While it suggests that increasing fees for certain procedures may lead to statistically significant but economically modest changes in physicians' practice style, the evidence presented points to a potentially more important role for other factors, such as the publication of trial results and research spillovers, in influencing the delivery of care.

Finally, while the two first chapters focus on immediate responses to both explicit and implicit financial incentives faced by physicians, the third chapter explores the potential long-run consequences such incentives might have on patients' health outcomes when they cause changes in physicians' practice style, and in their choice of procedures or treatments. Still in the context of obstetric care, it investigates the impact of C-section birth on the development of various health conditions later in childhood. While this relationship has attracted the attention of the media and the scientific community, the various potential confounding factors and the likely endogeneity of a child's birth delivery method have stood in the way of the identification of a causal impact. The third chapter uses rich individual-level data on Canadian infants followed from pregnancy through childhood to investigate this relationship while accounting for an extended set of variables and risk factors that could not be simultaneously accounted for in previous work. Exploiting physicians' response to financial incentives in fee-for-service environments, it also directly address the potential endogeneity of C-section birth in an instrumental variable framework. The resulting local average treatment effect estimated identifies the long-term health impact of unnecessary C-sections, performed in the presence of financial incentives. From a public policy perspective, this effect is particularly interesting, since the long-term consequences of unnecessary C-sections are weighted against weaker and uncertain short-term benefits. Overall, the results suggest a potential weak association between cesarean birth and health care utilization later in life. If they do not allow definitive conclusions to be drawn about the magnitude of the relationship between C-section birth and health in childhood, they nevertheless suggest that a causal impact may be at play, at least when health is proxied by the regular intake of prescription medication.

Overall, the evidence presented in the following chapters suggest that despite their ubiquity, financial incentives do not always translate into the behavioural changes expected. On the one hand, this may result in policies not influencing the delivery of care in the way imagined or hoped for by policy makers. On the other hand, in contexts where policy decisions may involuntarily create arbitrage opportunities for providers, mitigated responses by physicians may limit the negative consequences implicit incentives could have on both the allocation of resources within the healthcare system, and on patients' health. Adding to the growing body of evidence on physicians' labour supply, this dissertation points to the fact financial incentives should be used with caution and their impacts should be evaluated rigorously, given that the responses to such tools are often sensitive to the institutional context, the type of providers targeted and the nature of the care considered.

Chapter 1

Meeting the target? The impact of targeted financial incentives on primary care physicians' labour supply

Parts of this chapter are based on data and information provided by the Canadian Institute for Health Information. However, the analyses, conclusions, opinions and statements expressed herein are those of the author, and not necessarily those of the Canadian Institute for Health Information.

1.1 Introduction

Targeted financial incentives have been used extensively by governments in recent decades in order to influence the delivery of health care. Bonuses and premiums have been introduced to encourage the provision of certain services, to facilitate access to care for underserved or vulnerable populations, or to incentivize physicians to offer comprehensive and continuous care, among other objectives. Such incentives have played an important role in recent primary care reforms, and generally rely on three assumptions. First, that physicians respond to changes in the absolute and relative prices of the services they provide. Second, that substitution effects will dominate as prices change, such that increasing the financial payoff associated with a procedure will increase the propensity of physicians to provide it. Finally, that increasing the remuneration associated with a specific set of services will not substantially impact the balance of physicians' practice (and create the need for new incentive payments to correct unintended consequences). Given the scope of primary care physicians' practice, it is important to understand the wide range of consequences that could result from a simple policy intervention such as a the implementation of a targeted incentive payment. The literature, however, provides mixed evidence on most of the assumptions listed above, and existing work rarely investigate them all at once.

While some studies present empirical evidence of income effects generated in the context of wage and fee increases for physicians, many have found substitution effects to dominate (e.g. Rizzo and Blumenthal [1994], Thornton and Eakin [1997], Thornton [1998], Clemens and Gottlieb [2014]), although not for all services (e.g. Kantarevic et al. [2008]). Other work looking at physicians' responses to changes in their general compensation however suggests that income effects can be strong enough to generate a backward-bending labour supply curve, as well as behaviours that are consistent with income targeting (e.g. Rice and Labelle [1989], Contandriopoulos and Perroux [2013]).[1] Further work has also shown that when facing different fees for substitutable procedures, physicians respond by intensifying their recourse to the one that is more generously remunerated (e.g. Gruber et al. [1999], Alexander [2017]). These studies, however, generally focus on the impact of those implicit financial incentives on the provision of a narrow set of medical acts, and not on the overall mix or volume of care provided. Looking more specifically at primary care physicians, Lavergne et al. [2014] provide evidence that targeted incentives had little impact on providers' decisions to expand the scope of their practice. Their analysis however mostly focuses on the extensive margin responses to the incentives. Finally, the literature provides evidence that the consequences of changes in financial incentives reach beyond the provision of services directly affected by price changes. Clemens and Gottlieb [2014] show that higher Medicare fees contributed to a more rapid adoption of technology. Other studies have shown that, in addition to impacting the volume of services provided, movements towards alternatives to fee-for-service remuneration have impacted physicians' selection of patients as well as referral rates (e.g. Devlin and Sarma [2008], Kantarevic et al. [2011], Liddy et al. [2014], Rudoler et al. [2015]). It is therefore not clear how financial incentives might affect physicians' provision of the services they are associated with, at the extensive and at the intensive margins, or if they have broader consequences for the provision of other services.

To answer these questions, this paper exploits the introduction of an incentive payment meant to encourage the provision of obstetric care by physicians in Ontario, Canada, in 2000. Mainly designed for low-volume providers such as general practitioners and family physicians (hereafter referred to as GPs), the *Sole delivery premium* increased by 50% the payment received by a physician for the first 25 birth deliveries billed over a twelve-month period, effectively increasing the relative wage associated with this limited set of eligible acts compared to other procedures and introducing a non-linearity in physicians' budget constraint. Taking advantage of these features of the incentive payment, I use administrative data on physician billings to estimate the average change in GPs' provision of intrapartum care in cases where the premium should only generate an income effect (GPs initially providing more services than the maximum eligible for the premium), and on instances where it should foster both an income effect and a substitution effect (GPs initially providing fewer services than the maximum eligible for the premium). In addition to exploiting the panel nature of the administrative data to estimate the heterogenous impacts of the policy for GPs who were initially higher– and lower-volume providers of intrapartum care, I take advantage of the fact that physician remuneration parameters are set at the sub-national level in Canada and present difference-in-differences estimates of GPs' response to the premium using physician billing data from British Columbia, a Canadian province unaffected by the introduction of the premium in 2000.

[1] The existing evidence supporting strong income effects mostly comes from the analysis of responses to changes in physicians' general compensation, rather than from responses to targeted incentive payments.

The specific context of obstetric care offers an interesting opportunity to investigate GPs' responses to targeted financial incentives. First, although medical training for primary care physicians in Canada includes low-risk labour and delivery, GPs can choose to exclude these services from their practice. The share that have chosen to do so has been increasing in the decades leading to the introduction of the premium. Indeed, Godwin et al. [2002] report a drop from a 68.4% share in 1983 to just a 20% share in 1997 in the province of Ontario. Similarly, a survey of Canadian family physicians conducted in 2001 suggests that 15% of Canadian family physicians provided on-call obstetric services at the turn of the millennium, compared to 72.7% who provided general on-call services [CFCP, 2002]. Reasons often invoked to explain this decline include the personal life disruptions and higher malpractice risks associated with birth deliveries. Second, GPs who do choose to provide obstetrical care have a certain level of control over the volume of births they deliver; while they are generally responsible for the first 32 weeks of prenatal care for their (low-risk) patients, they can refer their patients to an obstetrician/gynaecologist (OB/GYN) or a colleague GP for the delivery [CIHI, 2004].[2] By the same token, GPs can accept referrals from colleagues, thereby increasing their own provision of obstetric care. Moreover, and unlike most other areas of care, observed responses to incentives in obstetrics cannot be attributed to demand inducement [Dranove, 1994]. This is interesting since most work on income effects and physician behaviour focuses on testing if physicians use their agency power to induce demand for care (quantity or intensity of services) in order to increase their earnings. Instead, this paper looks at the decision to provide certain services for which the relative price increases, in lieu of referring their patients needing these services to colleagues. A final advantage of looking at obstetric care in the Canadian context is that physician services related to childbirth are universally covered by Medicare. Changes in physician fees, which are set by the single payer, should therefore not affect the demand for these medical services, and observed changes in their provision can hence be interpreted as supply-side responses.

This study contributes to the literature in two ways. First, I estimate physicians' responses to a financial incentive that targets an area of services in which their provision decision is characterized by an important degree of discretion, and that raises the fees for these services, in turn creating a kink in physicians' budget constraint. This last feature of the incentive allows me to gain better insights on the income effect associated with such fee increases. While an important portion of the literature studying income effects among physician populations focuses on demand inducement following fee cuts or negative income shocks [Johnson, 2014], this paper focuses on a policy that increases fees. This also helps to shed some light on the symmetry of the responses observed as fees move in different directions. Second, I explore the spillover effects of an asymmetric increase in physicians' fees on the full range of services they provide.

Overall, my results indicate that it may be hard to convince GPs to broaden the scope of their practice with targeted financial incentives, at least for non-routine services such as intrapartum care: I find evidence of an income effect responsible for a little less than a 8% reduction in the volume of services targeted by the incentive billed by higher-volume providers. My results suggest that an apparent absence of response to the Sole delivery premium when looking at the full population of primary care physicians hides very different responses from GPs located on either side of the policy threshold, the incentive

[2] According to the National Family Physician Workforce Survey, only 4.2% of Canadian GPs found access to referrals to OB/GYNs to be poor or problematic in 2001 [CFCP, 2002].

resulting in a substitution effect only for lower-volume providers. I also find suggestive evidence that the introduction of targeted bonus payments may generate income effects that spill over to non-targeted areas of care. Indeed, the introduction of the Sole delivery premium is associated with a reduction in physicians' income claimed for non-obstetric services, a pattern that cannot be explained by a substitution towards targeted intrapartum services. Finally, although the patterns I observe are imprecise, the absence of change in GPs' total incomes following the introduction of the sole delivery premium is at odds with pure income-maximizing behaviour, and could be consistent with predictions from models of behaviour involving reference or target incomes.

The remainder of this paper is divided as follows. Section 1.2 describes the sole delivery premium, explains the opportunities it offers to investigate income effects through the non-linearities it introduces in physicians' budget constraint, and formulates testable hypotheses on responses for physicians on each segment of the constraint. Section 1.3 presents the data used for the empirical analysis and section 1.4 discusses the main estimating methods. Section 1.5 presents the main results for GPs' responses to the sole delivery premium for targeted services and section 1.6 explores the consequences for the provision of non-targeted services. Section 1.7 offers some concluding thoughts.

1.2 The sole delivery premium

The sole delivery premium was introduced in July 2000, following the adoption of the 2000-2004 Physician Services Agreement by the Ontario Ministry of Health and Long Term Care and the Ontario Medical Association. At the time, the premium consisted of a targeted bonus payment, billed as a distinct fee code (E411), which increased by 50% the baseline fee paid to physicians for birth deliveries, attendance at labour and/or attendance at delivery.[3] To qualify for the premium, a targeted service had to meet two conditions. First, it had to be the only one among all eligible fee codes to be billed by the physician on a given calendar day. In other words, if a physician delivered one baby and assisted at delivery for another patient in the same day, none of these procedures would qualify for the premium. However, if the same physician only performed one of these acts on a given day, then the act would qualify for the premium. Second, the premium could not be claimed for more than 25 eligible services over a twelve month period from the moment the first eligible service was billed.[4] This last condition creates an interesting non-linearity in intrapartum care providers' budget constraint: while their first 25 eligible services are remunerated at 150% of their base value, the 26^{th} and all following services are remunerated at the base fee.[5]

[3]Eligible services were initially composed of fee codes P006 (Vaginal delivery), P009 (Attendance at labour and delivery), P018 (Cesarean-section), P020 (Operative delivery), and P038 (Attendance at delivery).The premium was later extended to include fee codes P041 (Cesarean-section with tubal interruption) and E414 (High risk obstetrical premium, which does not count as one of the 25 potential premiums to be claimed by a physician since it is already billed in addition to a delivery code).

[4]To investigate changes in physicians' provision of eligible services, I define each twelve months period starting in July 2000, the moment the sole delivery premium was introduced. This period does not perfectly coincide with the twelve months over which eligible services were cumulated by each physician in the sample (for each physician, the start of a twelve-month period coincides with the date of the first eligible service provided). Nevertheless, most of the analysis focuses on general increases in the average volume of birth deliveries each year, such that the definition of the twelve-month period adopted should not have major impacts on the estimation.

[5]The 26^{th} delivery billed by a physician in a given year could qualify for the sole delivery premium if some of the first 25 births deliveries were provided on the same day. The data described in the next section does not allow me to verify if

Although available to all physicians providing the targeted services, the sole delivery premium was designed with primary care physicians in mind. The first description of the premium in the 2000-2004 Physician Services Agreement was prefaced by a clause stipulating that "it is important to maintain family physician involvement in obstetrical services" [OMA and MOHLTC, 2000]. GPs' participation in obstetric care was indeed steadily declining in the 1980s and 1990s, measured either by the proportion of primary care physicians including this area of care to their practice or by the births they delivered as a group [Chan, 2002]. Furthermore, the premium was based on a recommendation formulated by the Ontario Medical Association's Central Tariff Committee in 1999, with the stated objective of "assist[ing] low-volume obstetric practitioners without preferentially assisting those who do practice a higher volume" [Ontario Medical Association, 1999]. This objective is reflected in the conditions attached to the sole delivery premium, which are less likely to be met by obstetrician/gynaecologists. This push towards a stronger involvement of GPs in obstetric care was in line with evidence available at the time of the introduction of the premium, showing lower average intervention rates and better outcomes for low-risk obstetrics patients receiving their intrapartum care from family physicians.[6]

In October 2005, the sole delivery premium was increased to reach 100% of the fee for eligible services, as part of the 2004 Physician Services Agreement. However, this modification in the value of the incentive coincided with other substantial changes in the mode of remuneration for many GPs in Ontario (for example, changes in work organization and remuneration parameters in the context of the progressive introduction of capitation payments, blended remuneration schemes, etc.). The 2004 Physician Services Agreement also introduced various incentives for preventive care and special targeted services, including obstetric care for physicians working within Primary Care Reform (PCR) practices. Given the potential confounding effects of these changes, the main analysis in this paper focuses on the introduction of the premium in 2000 and on its impact up until 2003, a period during which GPs' remuneration was undergoing fewer changes.[7] This choice of sample period should limit the influence of other changes in GPs remuneration parameters on the analysis, and is discussed further in section 1.3.

The impact of the sole delivery premium on physicians' labour supply can be conceptualized using a simple and stylized static labour supply framework. Physicians can provide two types of services (which can be thought of as two vectors of services, given the range of activities undertaken by GPs, associated with different fees); those targeted by the incentive (s_1) and all others (s_2). GPs choose their annual provision of each type of service in order to maximize their utility, a function of consumption c and leisure l.[8] The utility function can also include certain services as direct arguments, if their mere provision is associated with utility costs (or benefits). Following Godwin et al. [2002], who suggest that providing obstetric care might be costly for certain physicians (given the disruption in their personal

this is the case. However, for low-volume intrapartum care providers such as GPs, the threshold likely rests at 25 eligible services per year.

[6]Relevant studies are listed in [Godwin et al., 2002]

[7]A few other incentives were introduced with the 2000 Physician Services Agreement, such as premium for home care applications and supervision, for complex care of the elderly, and after hour premiums (including for after hour obstetrical procedures). Most of these incentives correspond to modest amounts (between \$16 and \$10 for the first three mentioned) compared to the sole delivery premium.

[8]Utility rather than profit maximization is assumed. Given the fact that Canadian physicians are price-takers, fees for their services being regulated at the province level, only quantities enter as choice variables.

lives associated with on-call schedules, the fear of malpractice suits, etc.), I include obstetric care s_1 in physicians' utility function, given by equation (1.1).

$$U = U(c, l, s_1) \tag{1.1}$$
$$U_c > 0 \quad U_l > 0 \quad U_{s_1} \leq 0 \quad U_{cc} < 0 \quad U_{l,l} < 0 \quad U_{s_1 s_1} \leq 0$$

GPs maximize their utility function under the traditional time and budget constraints. The time constraint, given by equation (1.2) simply suggests that physicians divide their total time T between leisure, l, and the provision of both types of services, h_1 and h_2, with $s_1 = s_1(h_1)$ and $s_2 = s_2(h_2)$.

$$T = l + h_1 + h_2 \tag{1.2}$$
$$s_1 = s_1(h_1), s_2 = s_2(h_2)$$

Before the introduction of the sole delivery premium, physicians' budget constraint would simply correspond to the product of quantities and prices, summed over both types of services ($s_1 p_1 + s_2 p_2$). However, in presence of the sole delivery premium, physicians are subject to a piecewise budget constraint, summarized by equations (1.3) and (1.4). For GPs providing 25 targeted services (s_1) or less, the compensation received at the margin is $1.5 \times p_1$. For those whose provision of targeted services exceeds 25, however, each additional service s_1 provided is paid exactly at p_1, while total income derived from the provision of the first 25 eligible services is $[1.5 \times p_1]25$.

$$c = [1.5 \times p_1]s_1 + p_2 s_2 \quad \text{if } s_1 \leq 25 \tag{1.3}$$
$$c = [1.5 \times p_1]25 + p_1[s_1 - 25] + p_2 s_2 \quad \text{if } s_1 > 25 \tag{1.4}$$
$$\tag{1.5}$$

While the premium was intended to incentivize GPs' involvement in the provision of labour and delivery services, s_1, a range of responses could be expected, depending on physicians' initial level of activity in intrapartum care. For lower-volume providers, billing on average fewer than 25 eligible services per year before 2000, the premium should generate both a substitution effect and an income effect. Depending on the relative strength of each of these effects, the provision of s_1 could increase, decrease, or remain unchanged following the introduction of the premium. However, for higher-volume providers with a budget constraint given by (1.4), the premium mainly generates an income effect which should result in a reduction in the volume of s_1 provided. The potential responses associated with these cases are portrayed in figure 1.A.1 in the appendix. Higher volumes of deliveries could be achieved in aggregate if, for instance, income effects were very small or if substitution effects among low-volume providers were strong enough to compensate for potential reductions in the intrapartum care provided by high-volume providers.

The fact that obstetric care only represents a fraction of all activities undertaken by primary care physicians also means that the sole delivery premium could impact other parts of physicians' practice beyond the services it directly targets. Indeed, cross-price effects and income effects could influence the provision of non-targeted services. The nature of the data used, described in section 1.3, limits the extent to which I can investigate the direct impact of the sole delivery premium on specific non-targeted services. However, the availability of information on physicians' total income makes it possible to indirectly explore potential spillovers from the sole delivery premium on GPs' provision of care outside of obstetrics. Most importantly, and especially for high-volume providers, it provides an opportunity to test if potential income effects are strong enough to hint at income-targeting behaviour.

It should also be noted that the sole delivery premium could have an impact on GPs' decision to include or exclude intrapartum care to the range of services they provide. For example, GPs who were not active in intrapartum care prior to 2000 and whose "reservation payment" to perform birth deliveries lies somewhere between 100% and 150% of the baseline fee may, after the introduction of the incentive, decide to expand the scope of their practice to include obstetrics. Likewise, the incentive could encourage GPs who would otherwise have stopped providing intrapartum care to keep delivering births.[9] However, the data used in this paper does not provide information on GPs whose range of services does not cover obstetric care, limiting the capacity to investigate extensive margin responses in detail. Nevertheless, figure 1.1 suggests that the downward trend in the proportion of GPs including intrapartum obstetrics to their practice was not altered by the introduction of the sole delivery premium in 2000.[10]

1.3 Data

The main data for this analysis consists of physician billing records for intrapartum care extracted from the National Physician Database (NPDB). The NPDB is an administrative dataset held by the Canadian Institute for Health Information (CIHI) and formed of all billing records submitted to provincial authorities by physicians practicing in Canada. Given the single-payer nature of the Canadian healthcare system, the NPDB is a comprehensive source of information on all fee-for-service activities conducted by physicians across the country. Of course, the full range of services provided by physicians may exceed the scope of fee-for-service billings, but the vast majority of intrapartum care provided in the country is subject to piece-rate remuneration, and is therefore captured by the NPDB. The data provides longitudinal information on the volume and total value (in Canadian dollars) of the payments claimed by individual physicians for each fee code and in each three-month period. These billing records are further

[9]For example, a shock such as the birth of a GP's own child could increase the disutility cost of providing intrapartum care given the personal life disruption often associated with the provision of labour and delivery services. While the baseline fee could be insufficient to keep the benefit of providing some intrapartum care above this higher cost, the sole delivery premium could reverse this inequality and raise the benefit above the cost. An occurrence of adverse event following a birth delivery by a colleague could likewise increase the perceived malpractice risk of providing intrapartum care, and raise a GP's likelihood of opting out in the absence of sufficiently high payments. Other factors leading physicians to opt out of the provision of obstetric care, but that could potentially be counteracted by higher fees, are presented in Chan [2002], Godwin et al. [2002]

[10]The rates mapped in figure 1.1 are constructed using the number of GPs billing services eligible for the sole delivery premium observed in the NPDB as the numerator, and the total number of primary care physician in Ontario obtained from the Scott's Medical Database CIHI [2016a] as the denominator. Data from CIHI [2000-2005] on the number of GPs per age group in Ontario between 1999 and 2004 suggests that the trends remained unchanged after 2000 for physicians younger than 40 years old or older than 60 (see figure 1.A.2 in the appendix).

merged with two additional data files containing time-invariant and time-varying physician characteristics: gender, year of graduation from medical school, medical school of graduation (de-identified), age at the time of billing (in five year bins), main specialty reported in a given year for billing purposes, main geographic area of practice in a given year (statistical area classification codes defined by Statistics Canada), and total fee-for-service amount billed in a given year. I aggregate each physician's record into twelve-month periods, so that each observation corresponds to a physician-year combination.

1.3.1 Main estimating samples

The main estimating sample is formed of all family physicians or general practitioners who submitted fee-for-service billings for intrapartum care in Ontario between July 1996 and June 2004 inclusively, covering the four-year periods preceding and following the introduction of the sole delivery premium. Defining the period considered for the analysis such that it ends in June 2004 limits the risk that GPs' practice style and labour supply be affected by the introduction of new primary care models in Ontario in the early 2000s. Indeed, the mix of incentives associated with these new primary care models could have indirectly affected participating physicians' decisions with respect to the provision of intrapartum care, even if intrapartum care remained almost exclusively remunerated through fee-for-service payments for GPs in all models. It should be noted that the introduction of important primary care models took place prior to 2004.[11],but that they mostly started gaining popularity among practitioners after 2004. Sweetman and Buckley [2014] report that in 2003, approximatively 4 out of 5 primary care physicians in Ontario received their full income in the form of fee-for-service payments. Accordingly, Henry et al. [2012] report that in fiscal year 2004, fee-for-service payments represented nearly 95% of all primary care physicians' income, a share that declined sharply in the following years. The year 2005 also marked other changes in the remuneration of primary care physicians, including the elimination of a threshold system for annual earnings (another change in physicians' budget set which is discussed more in details in section 1.5). Finally, numerous financial incentives and targeted incentive payments were also introduced in Ontario as part of the 2004-2008 Physician Service Agreement.

The administrative nature of the data used makes it a rich and reliable source of information, which is nonetheless characterized by certain limitations. First, information on the precise timing at which each service billed was provided by a physician is not available. As such, it is not possible to identify if two or more of the targeted services billed by a GP were provided in the same calendar day, making them ineligible for the sole delivery premium. This is less likely to happen for GPs, who are lower-volume and lower-frequency intrapartum care providers. I therefore assume that all first 25 deliveries billed by a physician in each twelve-month period are each billed on a different calendar day.[12] Second, the data

[11] For example Primary care Networks (capitation-based) were created in the late 1990s, Family Health Networks and Family Health Groups and Comprehensive Care Models (blended fee-for-service) were introduced in 2002-04.

[12] For observations corresponding to fewer than 25 eligible services per year between 2000 and 2003, an average difference of 1.15 is observed between the number of eligible services and the number of times the sole delivery premium was billed. Meanwhile, observations corresponding to more than 25 eligible services per year were associated with an average E411 fee codes (the billing code for the sole delivery premium) of 26.15, suggesting that the calendar day requirement is likely met among GPs, and that the first 25 services billed are effectively paid at 1.5 times the baseline rate. The small differences can be caused by errors in physicians submissions, which are usually corrected in subsequent billing records, and of course by certain otherwise eligible services being provided on the same day. The discrepancies might also come from the fact that the

is based on billings submitted by physicians, which may suffer from periodic inaccuracies. Physicians involved have up to two years to correct any mistakes in submitted billings. For example, services billed but not rendered in a trimester could be subtracted from the volume billed in a subsequent period, and services rendered but not billed in a past trimester could be added to the billings of a subsequent one. Collapsing physician billings by twelve-month periods is likely to attenuate some of the measurement error coming from such discrepancies.[13] Third, the data contains no information on individual deliveries, on the characteristics of the pregnancy and/or the mother, nor on the outcomes of the delivery for her and her newborn.[14] However, certain fee codes billed by physicians are indicative of complications associated with the pregnancy and delivery, and permit identification of more involved deliveries from uncomplicated ones. This information can be used to investigate if the premium had an impact on the share of complex deliveries undertaken by GPs. Finally, the data only provides information on GPs for the years during which they provide intrapartum care. Consequently, physicians who never did during the sample period remain unobserved, posing a substantial challenge to the study of extensive margin responses to the sole delivery premium. The dataset can therefore be described as a form of unbalanced panel, information on the service provision and the time-varying characteristics of GPs (including his annual fee-for-service income and main geographic area of practice) being available only in the years when they provide intrapartum care.

Given this last feature of the data, I work with two alternative samples. The first one comprises all GPs who billed targeted services at any point between 1996 and 2003, and who are therefore observed at least once during that period. From this sample, I exclude physicians who do not consistently identify as primary care physicians. I also drop any physician who either practiced in more than one province and/or who moved from one province to another during the sample period, since the movement between provinces could affect the scope and intensity of their practice in general, and more specifically in obstetric care, and I exclude any physician with missing information on gender, main geographic area of practice, or for whom information on the year of graduation is missing. For the main analysis, I also exclude observations corresponding to the top 1% of intrapartum care billings. These observations likely represent GPs who heavily specialize in the provision of intrapartum care (and were not necessarily the ones for whom the incentive was designed), or may be an artifact of potential (rare) inaccuracies in the data. Finally, I exclude GPs who either graduated or likely retired between 1996 and 2003.[15] These two last groups are excluded from the main analysis in order to avoid interpreting respectively a retirement decision or a graduation as an extensive margin response to the premium.[16] The final

twelve-month periods defined for the purposes of this analysis (July-June) do not perfectly overlap with the twelve-month periods over which GPs cumulate eligible services.

[13] To avoid having negative values for the volume of deliveries billed in a given year, I manually subtract these corresponding values from the previous period, and adjust current billings to zero. These represent only a few occurrences in the sample.

[14] Patient information is available from other data sources, but the physician identifiers in the NPDB cannot be linked to any other administrative records which would contain information on patient and pregnancy characteristics.

[15] The data does not allow to unambiguously separate GPs who decided to exclude intrapartum care from their activities from those who retired. To identify the latter group, I assume that a GP retires if I observe him permanently stopping to provide targeted services when he is older than 65. This procedure might overestimates the number of retirees given that physicians are also likely to reduce the scope of their practice (focusing on core services) as they age [Pong, 2011]. I note that the results presented in section 1.5 are robust to including all potentially retired physicians.

[16] The top 1% intrapartum care billings and physicians who either graduated or retired between 1996 and 2003 represent the largest volume of sample exclusions (421). The full set of results including these three groups point in the same direction as the results reported in section 1.5. 24 physicians are excluded from the sample because they sometimes identify as specialists, 89 are excluded for being observed in more than one province during the sample period, 8 are

sample is composed of 1,960 Ontario primary care physicians observed between 1996 and 2003, or 78% of all GPs observed in the data. From this larger group, I define a second sample corresponding to the balanced panel of 498 GPs who billed targeted services in each of the eight sample years. This panel allows to exclusively focus on physicians' response to the premium at the intensive margin, net of any changes in the composition of physicians in the sample and of extensive margin responses.[17]

1.3.2 Summary Statistics

Descriptive statistics are presented table 1.1, and summarize the mean characteristics for all GPs in the data. Physicians are on average observed in half the years covered by the analysis. Men represent 68% of physicians in the sample, and 89% of all GPs observed graduated from a Canadian medical school. On average, physicians have 16.7 years of experience at the time of billing, and 23% of them conducted most of their activities in a metropolitan-influenced zone (which will be referred to as rural practice from here on) rather than in a a census metropolitan area or census agglomeration. On average, a GP billed the equivalent of 14.1 services targeted by the sole delivery premium each year between 1996 and 2003 (conditional on billing for targeted services), this average being slightly higher in the four years before the introduction of the premium than in the four years following it.[18] Only 5% of the deliveries performed by these physicians correspond to more involved procedures such as C-sections, breech or instrumental deliveries. Finally, GPs in the full sample derived an annual average income of $4,977 (2002 dollars) from intrapartum care between 1996 and 2003, corresponding to a little less than 3% of their total average annual fee-for-service income.

Most of these averages are characterized by substantial heterogeneity across providers. Table 1.1 also presents mean characteristics separately for GPs in the balanced panel. These physicians are slightly more experienced at the time of billing (18.9 years), are slightly more likely to practice outside of a census metropolitan area or census agglomeration (28%), and are by definition observed in all 8 years between 1996 and 2003. In addition to providing intrapartum care in a more consistent fashion, they are also on average higher volume providers of targeted services. Overall, they provided an annual average volume of 28 targeted services between 1996 and 2003, this average being slightly higher before 2000 than after. As a result, they derive a much higher income from these services, a yearly average of $10,408, just shy of 5% of their average total fee-for-service income.

The group of GPs who are not part of the balanced panel can be separated into three categories: those who only provided targeted services before the introduction of the sole delivery premium (785 GPs ap-

excluded on the basis of missing information on gender, main geographic area of practice, age or year of graduation.

[17]A larger *balanced* panel can also be created from the full set of observations by creating new entries corresponding to all missing years for the GPs excluded from the balanced sample. A null volume of targeted services is assigned to the observations created, with a few exceptions. First, all observations preceding a physician's graduation year are excluded. Second, potential retirements are identified using the procedure described above and post-retirement years are also excluded for each physicians. Based on the data available, it is however impossible to confirm if a physician who is less than 65 and who permanently becomes inactive in the intrapartum care market (i) maintained his general practice, (ii) retired, or (iii) moved out of the country. Administrative data on the migration of Ontario GPs suggests that a very small number of them left the country during the sample period [CIHI, 2000-2005]. The third explanation should therefore not apply to many observations in the sample, and not be an important source of bias when considering this constructed sample.

[18]The average taken over the smaller set of physicians observed in the sub-period 1996-1999 period is 16.6 services.

pearing on average for 1.9 years in the sample), those who only provided targeted services after 2000 (253 GPs appearing on average for 1.3 years in the sample), and those who appear both before and after 2000, but not consistently between 1996 and 2003 (424 GPs appearing on average for 5 years in the sample).[19] The size of each category is in line with the decline in the practice of obstetric care by GPs reported throughout the 1980s and 1990s [Chan, 2002, Godwin et al., 2002]. GPs who appear in the data only in certain years between 1996 and 2003 billed on average fewer targeted services each year. While physicians in the balanced sample represent 29% of all physicians observed between 1996 and 1999, they form 59% of the GPs who billed during that same period an annual average number of deliveries beyond the threshold of 25 set by the sole delivery premium.[20] Even within the set of GPs who are not included in the balanced panel, those who billed targeted services before and after the introduction of the sole delivery premium were on average higher volume providers than those who only delivered births before or after 2000 (13.86 versus 9.26 and 2.19 per year, respectively). Targeted services hence accounted for a higher proportion of the total fee-for-service income of GP observed only periodically both before and after 2000 (2.4%) than for those observed only before (1.4%) or only after (less than 1%) 2000. For these three groups combined, the average volume of targeted services provided annually is much lower after 2000 than before. Finally, the demographic characteristics of GPs who are not in the balanced panel are not uniform. Those only appearing in the data prior to the introduction of the sole delivery premium tend to be less experienced than those only appearing after its introduction (15.2 versus 17.9 years), are less likely to be men (68% versus 72%), and are less likely to practice in areas that are neither a census metropolitan area or a census agglomeration (18% versus 22%). GPs appearing in *some* years before and after the introduction of the sole delivery premium are the most similar to those in the balanced panel in terms of demographic characteristics, although they are on average less experienced at the time of billing.

The balanced panel of GPs is the preferred sample for the analysis for two main reasons. First, given that intrapartum care represents a more important part of their activities, the sole delivery premium is likely to be more salient to these GPs, which should facilitate the interpretation of the estimated response to its introduction. Second, the stable sample composition allows me to get at a response that can be thought of as an intensive margin one. Alternatively, any analysis performed on the full sample of GPs will generate results driven by three phenomenas: responses at the intensive margins, responses at the extensive margins, and changes in the sample composition. Given that data on GPs who never provided targeted services is not observed, separating these three responses is not possible. Hence, although some results presented in section 1.5 are presented for the full sample of GPs, most conclusions presented will speak to the responses estimated in the balanced panel of GPs exclusively.

[19] These groups all exclude physicians who graduated between 1996 and 2003, and those who likely retired or changed provinces during that period. As explained, entries and exits in the sample are therefore likely to reflect a decision to include/exclude intrapartum care to the services offered.

[20] See table 1.A.1 in the appendix for the complete distribution of GPs according to the years they appear in the sample and their initial average volume of targeted services before 2000.

1.4 Empirical strategy

To estimate the impact of the introduction of the sole delivery premium on GPs' provision of targeted services, I specify a simple empirical model presented in equation (1.6). The volume of targeted services provided in year t by physician i practicing in geographic zone c is a function of the main variable of interest, an indicator variable $Post2000$ equal to 1 for years following the introduction of the sole delivery premium, and 0 otherwise.[21]

$$Target\ services_{ict} = \boldsymbol{X}^{'}_{ict}\boldsymbol{\beta} + \delta Post2000_t + \boldsymbol{Z}^{'}_{ct}\boldsymbol{\eta} + \psi Trend_t + \phi_i + \upsilon_{ict} \tag{1.6}$$

I control for individual characteristics, X, that likely influence obstetrical practice. First, I include a set of dummy variables for 5-year age groups. Results from the 2007 National Physician Survey suggest that Canadian physicians aged 55 and above are more likely to reduce the scope of their practice, or to plan doing so in the near future [Pong, 2011]. Billing data from a cross section of physicians observed in 2006, also presented in Pong [2011], further confirms participation in obstetric care declines with age, even for physicians who maintain their activities in core areas of primary care. When estimating across-physician responses, I also include indicators for medical school of graduation, as well as a control indicating if a physician trained in Canada. I also include in X a vector of geographical zone fixed effects, corresponding to the main statistical area code (SAC) in which a GP has been the most active in a given year. The vector Z includes controls for other characteristics of the physician's environment that likely affect the provision of obstetric care: the number of obstetricians and gynaecologists providing intrapartum care in their SAC, the number of midwives registered in the province, and the total number of newborns. The annual value of these variables is respectively obtained from the National Physician Database, the Canadian Institute for Health Information [CIHI, 2013a], the Association of Ontario Midwives and Statistics Canada [Statistics Canada, accessed: May 8, 2017]. Given the well-documented decline in GPs' involvement in obstetric care preceding the introduction of the sole delivery premium, I also include a time trend to the main specification.[22]

In most regressions, I control for rural practice, a factor identified as having an impact on GPs' involvement in intrapartum care in Ontario given the limited availability of other types of providers [Godwin et al., 2002, Hutten-Czapski et al., 2004]. In certain specifications, I also exploit the interaction between the introduction of the sole delivery premium and the rural practice indicator. One might expect GPs in rural environments to have fewer opportunities to respond to targeted incentives such as the sole delivery premium. First, the smaller population may make it harder to seek referrals from other physicians to increase the volume of births delivered. Second, the reduced availability of other providers (especially OB/GYNs and midwives) may challenge a GP's capacity to refer a pregnant patient to a colleague and to reduce the volume of intrapartum care provided. Indeed, the results in section 1.5 suggest that responses to the sole delivery premium are concentrated among GPs whose practice is situated in a more

[21] The first year for which $Post2000 = 1$ correspond to the twelve month period following the introduction of the Sole delivery premium.
[22] All results are robust to including a quadratic trend.

urban environment.

In some specifications, I also include a set of physician fixed effects ϕ to equation (1.6), since some unobserved physician characteristics may be correlated with preferences for the provision of obstetric care. When those are included, the estimated δ can therefore be interpreted as a within-physician response. I finally use information on physicians' behaviour prior to the introduction of the sole delivery premium to investigate the income effects of the incentive in the context of the non linearities the premium introduces in GPs' budget constraint. I estimate equation (1.6) separately for physicians whose average volume of eligible services was respectively below and above the threshold of 25 between 1996 and 1999. As discussed in section 1.2, GPs typically located beyond that threshold should experience an income effect from the introduction of the policy, while those below the threshold should experience both a substitution and an income effect. The results presented in section 1.5 offer a more direct test for this theoretical prediction, and provide some insights on the relative strength of income effects on GPs' provision of targeted services.[23]

1.5 Results on the provision of targeted services

A first look at the raw data suggests that the provision of targeted services by individual GPs did not increase following the introduction of the sole delivery premium. Figure 1.2 maps the probability density function of GPs against annual volumes of services eligible to the sole delivery premium between 1996 and 1999 (solid line) and between 2000 and 2003 (dotted line). The figure is based on billings from the balanced panel of GPs, such that the differences between the densities displayed are not caused by changes in the sample composition.[24]

Figure 1.2 first highlights that despite the kink it introduced in GPs' budget set, the premium did not lead to bunching in the distribution around the threshold of 25 eligible services. This observation is confirmed by regressions of the probability that a GP's provision of intrapartum care fall within a narrow window of the threshold (not shown). This absence of bunching is not surprising given the nature of the services considered; GPs can adjust the number of births they deliver by accepting new patients or referring some of their patients to colleagues, but they likely have a limited ability to exactly choose the volume of intrapartum care provided to maximize the additional income received in bonuses and stop afterwards. Figure 1.2 also maps a shift in the distribution towards lower levels of provision after 2000. To further explore if and how GPs adjusted their provision of targeted services in response to the premium, I turn to the empirical specification presented in the previous section.

[23]The standard errors for the main results presented for the sample of Ontario physicians are clustered at the physician-level, and inference is robust to defining the clusters at the statistical area code level.

[24]Volumes in figure 1.2 are truncated at 50 per year for exposition purposes, the distribution shown still representing a little more than 90% of the observations in the sample.

1.5.1 Baseline results

Table 1.2 presents the main estimates from equation (1.6). The first two columns look at the estimates obtained using all observations available between 1996 and 2003. Column 1 presents the results when no controls aside from a time trend are included. The estimated coefficient on $Post2000$ suggest a small negative response (-0.902) to the introduction of the sole delivery premium, statistically significant at the 10% level. This response, however, becomes smaller (-0.641) and loses its statistical significance as physician characteristics and province time-varying controls are added to the specification. As mentioned above, the estimates coming from all GPs in the sample likely correspond to the combination of intensive margin responses, extensive margin responses, and changes in the sample composition through time. To get more specifically at the response at the intensive margin, columns 3 to 5 present the coefficients from equation (1.6) when the estimating sample is limited to GPs providing targeted services without interruption from 1996 to 2003. Again, the coefficients on $Post2000$ point to a negative but not statistically significant response to the incentive, corresponding to an annual reduction of approximately 1 delivery following the introduction of the sole delivery premium. A comparison of the coefficients on $Post2000$ presented in columns 2 and 4 (respectively -0.641 and -1.173) suggests that changes in the composition of the full sample likely attenuate the negative response estimated across physicians. Indeed, the observations added to the balanced panel to form the estimating sample in column 2 disproportionately come from the period preceding the introduction of the sole delivery premium, and those observations correspond to lower annual volumes of delivery compared to the pre-2000 observations in the balanced panel. Despite the fact that they represent even lower volumes of deliveries, the observations from the post-2000 period included in column 2 but excluded in column 4 represent a much smaller group. Extending the sample beyond the balanced panel therefore reduces the average volume of deliveries before 2000 more than it does after 2000, mechanically contributing to the difference in the estimates presented in columns 2 and 4.[25] Finally, the across– and within-physician responses estimated in the balanced panel (columns 4 and 5 respectively) are quite similar, given the stable composition of the panel and the fact that all physicians contribute to the identification of the parameter of interest as they are observed before and after the introduction of the premium.

Other coefficients generally have the expected signs. The availability of registered midwives in Ontario (who can be thought of as imperfect substitutes to GPs in the provision of intrapartum care) has a negative impact on GPs' provision of eligible services, although this association is not statistically different than zero. The same is generally true for the number of OB/GYNs within a GP's statistical area code, while the total number of births in a given year is positively associated with the volume of deliveries billed by GPs. Physicians nearing retirement age are likely to be less active in intrapartum care, which is consistent with evidence presented by Pong [2011]. Finally, female GPs bill on average a higher volume of labour– and delivery-related activities annually, other things equal.

Table 1.3 presents estimates from a slightly altered version of equation (1.6), in which a series of year effects are included instead of a time trend and the $Post2000$ variable. Results are presented for the full

[25] It is also possible that a more muted response in the full sample is partly driven by the fact that the sole delivery premium is not as salient an incentive for GPs who only sporadically provide intrapartum care. However, this hypothesis does not seem to be supported by the raw means presented in column 4 of table 1.1.

sample of GPs observed between 1996 and 2003 (columns 1 and 2) and for the balanced panel (columns 3 to 5). Two interesting patterns are associated with the year effect estimates. First, the effects for the years 1996 through 1999 are all negative and, once controlling for physician characteristics, follow a slightly decreasing pattern, suggesting that the volumes of targeted services provided by GPs were already declining prior to the introduction of the sole delivery premium. Second, the (negative) year effects become much larger in magnitude from 2000 on. The within-physician year effects estimated for the balanced panel of GPs (column 5) are graphed in figure 1.3. Although the year effects following the introduction of the sole delivery premium are not statistically different from each other (p-value of 0.25 in the specification with physician fixed effects), I can reject that they are statistically equal to the year effect for 1999 (p-value of 0.005) and for each of the previous years.[26] These estimates are consistent with the results presented in table 1.2 and suggest a relatively rapid, although modest, decline in the volume of eligible services provided in response to the premium. The year effects also point to a lasting effect of the premium until the end of the period studied.

The results presented so far speak to average changes in the volume of targeted services billed by GPs after the sole delivery premium changed the relative compensation received for intrapartum care. However, as described in section 1.2, different adjustments should be expected depending on where physicians were initially located in relation to the threshold of 25 targeted services per year. To test the predictions derived in section 1.2 for physicians on each side of the threshold, I therefore re-estimate equation (1.6) for physicians who billed on average fewer than 25 targeted services in the four years preceding the introduction of the premium, as well as for physicians who billed more than 25. Dividing the sample between physicians on either side of the cutoff offers an opportunity to better understand the heterogenous impacts of the financial incentive given its design. The results are shown in table 1.4. Columns 1 and 2 first present the estimated responses for lower– and higher-volume providers in the full sample. For providers whose initial provision of targeted services was below the cutoff of 25, a positive response (0.890) is estimated, suggesting that the fee increase may generate a substitution effect that is weakly dominating, driving GPs' behaviour in a way that is consistent with the objective of encouraging the provision of obstetric care by low-volume providers. However, the point estimate is not statistically significant, and it is not possible to reject that the incentive had no impact on physicians' behaviour.[27] For GPs whose initial provision of targeted services reached the threshold of 25 eligible services per year, the introduction of the sole delivery premium is associated with a statistically significant reduction of 3 deliveries annually (column 2), or slightly more than 7% of their average volume of deliveries prior to 2000. Such a response is in line with the theoretical predictions derived in section 1.2: Given the structure of the sole delivery premium, the response from physicians beyond the threshold should only

[26]Similar results are obtained in regressions excluding physician fixed effects.

[27]Physicians who did not provide targeted services between 1996 and 1999 are included in column 1. Excluding them from the sample yields a much larger, positive and statistically significant response of 1.920. However, this effect is mostly driven by changes in the sample composition. The average volume of targeted services provided prior to 2000 by GPs below the threshold of 25 is twice as large for physicians in the balanced sample compared to those who only appear in the data prior to 2000. Moreover, the latter group contributes 1231 observations between 1996 and 1999, while GPs in the balanced represent 1048 observations. Including those 1231 observations therefore reduces substantially the average volume of deliveries observed before 2000, while by construction not having a similar impact after 2000. Although their pre 2000 average volume of deliveries is closer to that of GPs in the balanced panel, including GPs who are not in the balanced sample but who appear in the full sample both before and after 2000 does not completely offset this impact, most likely because they represent a smaller number of observations. However, adding GPs observed exclusively between 2000 and 2003, for whom the average volume of deliveries is substantially lower, contributes to attenuate this impact, and brings back the estimate to a smaller and statistically non significant 0.890.

be driven by an income effect, and lead to a reduction in the volume of targeted services provided.

Very similar results are obtained when looking exclusively at GPs who provided intrapartum care in each year between 1996 and 2003, although the results are easier to interpret as intensive-margin responses. Columns 3 and 4 present the across– and within-physician estimates obtained for lower-volume providers, and suggest that the premium led GPs to increase by 0.7 or 0.8 their volume of targeted services provided annually. While the point estimates are positive, they are once again not statistically significant. Meanwhile, statistically significant negative responses are estimated among higher-volume providers. The estimates obtained for this sample are only slightly larger in magnitude than those obtained in the full sample, reaching -3.3 to -3.5. Of course, higher– and lower-volume providers differ in many respects, as shown in table 1.A.2 in the appendix. GPs who were more active in intrapartum care prior to 2000 are, for example, more likely to be women, more likely to practice in an urban environment, and earn on average more than GPs who were providing on average less than 25 targeted services per year before 2000. Different responses observed among the two subgroups of physicians may therefore be driven by differences in characteristics, as opposed to being solely driven by the structure of the incentives embedded in the sole delivery premium. The fact that the across– and within-physician estimates presented so far are very close to each other nonetheless suggest that differences in the unobserved characteristics of providers are not substantially influencing the results.[28] I also estimate a single regression in which $Post2000$ is interacted with an indicator for an initial average annual volume of deliveries exceeding the premium threshold yields comparable results (not shown). The within-physician estimates from such a specification are 0.674 (s.e. 1.039) on $Post2000$ and -4.010 (s.e. 1.003) on the interaction term. The sum of both coefficients, corresponding to high-volume providers' response, is statistically significant at the 1% level.[29]

The results obtained separately for GPs on each side of the premium threshold lend important insights. They highlight that by ignoring the different incentives faced by physicians located on each segment of the budget constraint following the introduction of the sole delivery premium, one could come to the misleading conclusion that the premium had no impact on GPs' behaviour. Meanwhile, allowing for heterogenous responses between lower– and higher-volume providers yields evidence that the latter group did adjust their behaviour after 2000, potentially due to the income effects associated with the higher payments received for the first 25 targeted services provided. The absence of a detectable change for GPs initially located below the premium threshold could be interpreted in two ways. First, low-volume providers may simply be less aware of the incentive. Indeed a stronger positive response corresponding to 1.5 is estimated for all lower-volume providers when excluding physicians who billed an average of less than 5 eligible services per year between 1996 and 1999–and for whom the premium is likely less salient. This increase is not observed in the balanced panel, potentially because GPs consistently providing intrapartum care are more likely to be made aware of changes in the relevant remuneration parameters, even when they are lower-volume providers. An alternative, and perhaps more plausible explanation

[28]The results presented in table 1.4 are also robust to controlling for physicians' baseline fee-for-service income from services other than intrapartum care.

[29]Similar results are obtained when considering the full sample rather than the balanced panel. The coefficient on $Post2000$ is very close to that reported in column 1 of table 1.4, with a point estimate of 1.229 (s.e. 0.710), and the sum of the coefficients on the interaction term and $Post2000$, corresponding to -3.963, is also close to that reported in column 2 and statistically significant at the 1% level.

is that for GPs initially located below the threshold of 25 deliveries, the income and the substitution effects associated with the financial incentive offset each other. The results obtained for the subsample of GPs initially located above the threshold could indeed suggest that income effects do partly influence the behaviour of physicians. However, for GPs initially providing few targeted services, the change in income attributable to the introduction of the incentive is expected to be less important, and the income effect may therefore be less likely to dominate the for this group, for whom the premium also results in a substitution effect. The difference in the estimates presented in columns 1 and 3 could also be consistent with this hypothesis: since the average volume of targeted services billed each year by physicians initially located below the premium threshold between 1996 and 1999 is higher in the balanced panel, the income effect can be expected to dominate for that group.

1.5.2 Robustness checks

The estimated responses presented in tables 1.2 to 1.4 can be interpreted as the change in GPs' provision of targeted services, conditional on them providing such services. The results for both lower– and higher-volume providers are somewhat amplified when replacing the missing observations for GPs in the full sample by zeros (missing observations correspond to years in which GPs did not provide targeted service).[30] The estimates are shown in panel A of table 1.5 (columns 1 to 3). While not statistically significant impact can be discerned when looking at all physicians regardless of their initial level of activity in intrapartum care, opposite and statistically significant responses are estimated when looking separately at higher– and lower-volume providers. The positive response estimated for GPs initially providing fewer than 25 targeted services per year before 2000 is very similar in magnitude to that estimated in table 1.4, but is now statistically significant. This increased precision in the estimation presented in Panel A of table 1.5 may come from the significant increase in the number of observations added to the sample when considering years during which GPs did not provide targeted services. Morevoer, the similarity in the point estimates obtained with and without years in which low-volume providers were not active in intrapartum care may suggest that the sole delivery premium did not change the trend observed in GPs' participation in obstetrics at the extensive margin, an explanation in line with the pattern shown in figure 1.1. Including the years in which GPs in sample did not provide targeted services increases the magnitude of the response estimated for higher-volume providers, as shown in column 3. The point estimate reaches -4.15 and is statistically significant. This is consistent with the disproportionate volume of observations corresponding to zero deliveries added after 2000 for this subsample, but the pattern in the data mostly comes from some GPs permanently ceasing to provide intrapartum care a few years before 2000, rather than after the introduction of the sole delivery premium. Overall, including zeros in the sample does not change the nature of the results: The apparent absence of response in the full sample results in the combination of a reduction in the volume of targeted services by higher-volume providers, and a potentially small increase by GPs who were initially lower-volume providers (and for whom the price of an additional delivery is increased by the introduction of the premium).

[30] It should be noted that time-varying physician characteristics are not available for these informations. Physicians' age group is imputed based on the information available from other observations. Values for time-varying characteristics of the main location of practice are assigned on the basis of the contemporaneous values of the characteristics for the location of practice reported by the GP in the closest year for which this information is recorded. As in all previous regressions, observations corresponding to the years preceding a GP's graduation, of following a GPs' likely retirement are excluded.

The next panels in table 1.5 explore the stability of the estimated responses to changes in the sample definition. First, panel B extends the period considered from 1995 to 2004. Doing so leaves the results relatively unaltered. The sign of the estimated response for the pooled sample consisting of all observations (column 1) is positive, but remains statistically insignificant, and most other coefficients are close although slightly smaller in magnitude than those reported in tables 1.2 and 1.4. One exception is the estimate for lower-volume providers in the full sample; the estimate becomes positive, larger in magnitude (1.63) and statistically significant. Given the primary care reforms described in section 1.3 and the growing role of alternative payment plans for GPs in 2004, both of which could contaminate the results, I choose to focus on the period spanning from 1996 to 2003. However, I note that extending the period does suggest that occasional, lower-volume providers could have responded positively to the sole delivery premium by increasing the volume of targeted services provided by 1.6 each year on average. However, this response is not observed for lower-volume providers in the balanced panel (column 5). Panels C and D of table 1.5 confirm that the main results are relatively robust to excluding very high or very low volumes of intrapartum care, although the negative response from high-volume providers is smaller and no longer statistically significant in the full sample when excluding annual volumes of targeted services greater than 50.

Panel E explores the heterogeneity of responses between men and women by interacting $Post2000$ with an indicator for male GP.[31] Although they represent a smaller proportion of all samples, especially for lower-volume providers, women's responses to the incentive are consistently stronger than men's. While the coefficient on $Post2000$ is larger in magnitude in all columns, the coefficient on the interaction term $Post2000 \times male$ is systematically of the opposite sign, suggesting that men's responses are weaker than women's. Moreover, in almost all cases, the sum of the two coefficients is not statistically different than zero. In the full sample and in the balanced panel, responses are more similar across genders for higher-volume providers, yet still slightly stronger for women. One possible explanation for these patterns could be that women may be less averse to providing obstetric care in general, making them more responsive to the change in relative prices for the first 25 birth deliveries introduced by the premium. At higher volumes of provision, the income effect associated with the premium for higher-volume providers may lead them to reduce the targeted services they provide as much if not more than men if, for example, their opportunity cost of time increases faster on average. Finally, the responses estimated among a subsample of Canadian graduates are similar to the main estimates (not shown).[32]

Another dimension along which GPs' reactions to financial incentives may differ is their environment of practice. Table 1.6 further explores the difference in the GPs' response to the premium depending on the rural or urban nature of the location from which most of their activities were conducted. The results are shown for the balanced panel of GPs but are qualitatively similar for the full sample of GPs observed in the data. Panel A first presents the results from a version of equation (1.6) to which an interaction term between $Post2000$ and a dummy variable for rural practice is added. For both lower– and higher-volume providers, the coefficients associated with the variable $Post2000$ are in line with those presented in table 1.4, while being slightly larger in magnitude. However, for the full panel and

[31] The results when equation (1.6) is estimated separately by gender for each panel and subgroup tell a similar story than those presented in panel E of table 1.5. They are however less precisely estimated given the reduced sample size for some subgroups.

[32] Limiting the sample to GPs having graduated after 1990 yields point estimates of larger magnitude, but these are quite imprecisely estimated given the reduced sample size.

for higher-volume providers, the estimate associated with the interaction of variable $Post2000$ and an indicator for rural practice strongly counteracts the coefficient on $Post2000$, having almost the same magnitude but the opposite sign. The sum of both coefficients for all subgroups cannot statistically be differentiated from zero, suggesting that the responses estimated in the full sample, and presented in earlier tables, are driven by GPs practicing in urban areas. Those responses are further confirmed in panel B, which presents the results from the estimation of equation 1.6 on a sample exclusively formed of GPs practicing in urban areas. Selection could contribute to differences in the estimated sensitivity to the sole delivery premium between urban and rural GPs. It is possible that primary care physicians who choose to practice further away from city centres are also less likely to respond to financial incentives. Another plausible explanation, discussed in the section describing the empirical strategy, is that GPs working in urban environments may have more flexibility to adjust their volume of eligible services, given the density of potential patients and the network of other physicians (GPs and OB/GYNs) to whom they can either refer patients or from whom they get referrals. GPs outside of urban centres may therefore be more likely to have slimmer margins over which to adjust in response to the introduction of the sole delivery premium.

1.5.3 Difference-in-Differences estimation

The sole delivery premium became available to all primary care physicians in Ontario in 2000, such that there is no clear control group within the province that can be used to more directly assess the premium's causal impact on GPs' behaviour. The inclusion of time trends and physician fixed effects provide some support for the hypothesis that the effects documented above are indeed responses to the premium, as do the results from estimations restricted to each side of the policy threshold. Moreover, no other incentives introduced as part of the 2000-2004 Physician Services Agreement or implemented alongside the sole delivery premium are, to my knowledge, likely to drive the change in GPs' behaviour observed after 2000.

However, to further investigate if the estimates presented above are driven by some coincidental unobserved shock affecting obstetrical practice in Canada, or by some broader trend that would affect the practice of obstetrics by primary care physicians, I obtain data on obstetric care billings for GPs in other Canadian provinces.[33] Since they were left unaffected by the introduction of the sole delivery premium (or similar incentives) in 2000, the behaviour of GPs outside of Ontario can be used as a benchmark to estimate equation (1.6) in a difference-in-differences framework. However, not all provinces are adequate control groups for this exercise. Demographic and socioeconomic changes have had different impacts across the country in the late 1990s and early 2000s, and trends in population density in urban and rural areas varied across provinces. These factors likely affected GPs' working conditions differentially across provinces during the period studied. Moreover, and although physician training is subject to national standards, health policy and the administration of health care are under the jurisdictional power of provinces in Canada. Hence, physicians' practice conditions and the parameters of their compensation vary substantially from one province to another. Such policy, socioeconomic and demographic

[33] I obtain access to this data from the National Physician Database for all provinces except Quebec, which has different data release protocols for physician billings.

disparities likely explain why trends in the provision of intrapartum care by GPs before 2000 diverge between Ontario and most other Canadian provinces or regions. British Columbia, however, stands out as a credible comparison group. First, as shown in figure 1.4, the unadjusted average volume of eligible services (or their equivalent in terms of fee codes in British Columbia[34]) provided by the balanced panel of GPs in each province generally followed parallel trends between 1996 and 1999, although higher levels are observed in Ontario. Second, the average characteristics of GPs consistently providing intrapartum care between 1996 and 2003 in British Columbia are generally close to that of their counterparts in Ontario.[35] Finally, the proportion of physicians' income from alternative payment plans, as opposed to fee-for-service, was relatively close between 1996 and 2002, with differences of 1% to 2% depending on the year [CIHI, 2005].

I therefore estimate equation (1.7) using a balanced panel of Ontario and British Columbia GPs observed between 1996 and 2002. I focus on this shorter period because a set of incentive payments for general practitioners and family physicians were introduced in British Columbia in 2003 as part of the Full Service Family Practice Incentive Program (FSFPIP). Excluding the year 2003 limits the possibility that these incentives contaminate the responses estimated in a difference-in-differences framework.[36] Additional incentive payments targeted at obstetric care were also implemented in British Columbia in the following years [Hutchison et al., 2011]. The parameters are defined as before, with the addition of the $[Post_t \times Ontario_p]$ associated with λ, the parameter of interest which captures the impact of the Sole Delivery Premium on Ontario GPs' provision of care.[37]

$$E411\ services_{ipt} = \boldsymbol{X}^{'}_{ipt}\boldsymbol{\beta} + \lambda[Post_t \times Ontario_p] + \delta Post_t + \psi Trend_t + \boldsymbol{Z}^{'}_{pt}\boldsymbol{\eta} + \phi_i + \upsilon_{ipt} \tag{1.7}$$

Panel A of table 1.7 presents the main results from equation (1.7), and shows that the break in the parallel trends observed in figure 1.4 between Ontario and British Columbia around 2000 is also detected in a regression framework when controlling for a full set of control variables, with or without physician fixed effects. The responses estimated with the pooled sample of high− and low-volume providers are negative and larger in magnitude than when only Ontario GPs are considered. The difference-in-difference estimates reach -2.7 and are statistically significant (columns 1 and 2). The estimates for higher-volume providers (columns 5 and 6) also indicate a statistically significant reduction of a little more than 5 targeted services per year, which is somewhat stronger than the corresponding results obtained using data on Ontario GPs exclusively. Again, no statistically significant change is found for lower-volume providers

[34] The dependent variable is constructed as the sum of fee codes for deliveries and attendance at labour and/or delivery, corresponding to 4000, 4014, 04017, 4018, 4025, 4050, 4052, 4104, 4105, 4021, 4106, 14104, 14105, 14109, 14108

[35] Detailed summary statistics for the balanced panel of GPs in each province are presented in table 1.A.3 in the appendix. Although small in magnitude, many differences in average characteristics are however statistically significant. Most importantly, the average GP is more likely to be a man and to practice in a rural environment in the Ontario sample, while the average annual fee-for-service income is lower in the British Columbia sample. All these characteristics are however controlled for in the analysis.

[36] An analysis of the FSFPIP has shown that the take up for these incentives was very low in the first year of implementation of the program (33.6%, coming almost exclusively from incentives related to management of diabetes), and that the take up of incentive specifically targeting obstetric care was 0% in 2003 [Hollander and Tessaro, 2009]. Indeed the inclusion of 2003 only slightly increases the magnitude of most difference-in-differences estimates, but does not change the results qualitatively.

[37] Although midwifery was regulated in both provinces during the sample period, data on the number of midwives in earlier years is not available for British Columbia, and this control variable is therefore excluded from the specification.

(columns 3 and 4). Estimating a specification in which the set of interactions between $Post2000$, $Ontario$ and a dummy variable indicating if a physician billed on average more than 25 eligible services per year prior to 2000 yields similar results (not shown). Most importantly, the interaction of the three variables is associated with a coefficient of -5.045 (s.e. 1.792), and the coefficient on the interaction of $Post2000$ and $Ontario$ is -1.559 (s.e. 0.905). The difference-in-differences estimates therefore also point to a reduction in targeted services coming from high-volume providers.

Since results from the Ontario sample suggest that the estimated responses to the sole delivery premium are driven by GPs located in urban areas, I re-estimate equation (1.7) in the balanced sample excluding rural observations.[38] The results, presented in panel B of table 1.7, are in line with the estimates from panel A. Mirroring the findings from table 1.6, the estimated responses for higher-volume providers are stronger when focusing on GPs practicing in urban locations. The estimates pooling both high– and low-volume providers also increase in magnitude as rural observations are excluded, pointing to a statistically significant reduction of a little more than 4 eligible services per year in response to the introduction of the premium. Including year effects rather than a time trend reduces the estimates on $Post2000 \times Ontario$ to levels more similar to those presented in table 1.6, but the results for lower– and higher-volume providers follow the pattern described above.[39] Finally, one potential issue with the difference-in-differences estimation approach described above is that while midwifery was introduced in the Regulated Health Professions Act in Ontario starting in 1994, it was first regulated as a profession in British Columbia in 1998. Given the role played by midwives with respect to low-risk deliveries, and the potential impact of their participation in the health care system on family physicians' and general practitioners' provision of intrapartum care practice, I re-estimate equation (1.7) excluding 1996 and 1997. The responses, estimated either with a time trend or with year effects (available upon request), are in line with those obtained when considering a longer panel.

All standard errors presented in table 1.7 are clustered by statistical area code, allowing for some degree of geographic and serial correlation in the error term. One could argue that, if physicians locational choice is based on cities or communities rather than provinces, clustering the standard errors at the statistical area code is reasonable. However, SACs may not be ideal cluter as it does not allow for spatial correlation within an entire province, the true geographic unit determining "treatment" (the introduction of the sole delivery premium). Unfortunately, the estimating sample consists only of two provinces and seven years of data, limiting the possibility to recourse to most of the typical remedies suggested for the estimation of difference-in-differences specifications with a small number of clusters [Bertrand et al., 2004]. An alternative option is to define clusters as province-year combinations, which allows for within-province correlation of the error terms, but ignores serial correlation within each province. When clustering the standard errors at this level, the estimates in columns 1 and 2 of panel A are only statistically significant only at the 10% level, and those in columns 5 and 6 are significant at the 5% level, but do not meet the 1% bar. In panel B, estimates in columns 1-2-5-6 become statistically significant at the 5% level (instead of 1%). I also note that when clustering the standard errors at the physician level (thus ignoring the possibility of geographical correlation in the error terms), the estimates in columns

[38] Pre-trends for this subsample are presented in figure 1.A.3.
[39] With year effects, results from panel A however lose their statistical significance.

1-2-5 of panel A become statistically significant at the 1% level (rather than at the 5% level), and the inference in panel B remains unchanged. While they do not change the inference substantially, none of these alternative cluster definitions are ideal.

1.5.4 Addressing potential confounding policy changes

Although the most important changes in Ontario GPs' practice environment were felt after the end of the sample period, other ancillary reforms could contaminate the results presented above. Importantly, GPs in Canada need to insure themselves against malpractice risk. Such protection is offered through membership at the Canadian Medical Protective Association (CMPA), which provides professional liability protection to physicians. It should be noted that CMPA membership fees for Ontario physicians increased by 45% over three years in 2001 as a result of a CMPA decision to vary its charges by region. Membership fees for GPs providing obstetric care were and remained higher than those for GPs choosing to exclude labour and delivery to their professional activities. The 2001 increase in the CMPA's fees would have therefore widened the gap in absolute terms between the two premiums available to GPs (although both increased proportionally), potentially discouraging some primary care physicians from providing any obstetric care (and targeted services) at all. However, the Ontario government (through its Medical Liability Reimbursement Program) absorbed most of the increase in physicians' insurance premiums, and out-of-pocket costs to physicians remained relatively constant during the period considered for this analysis [Sullivan, 2000]. Nevertheless, premiums can be reimbursed months after having been paid upfront by physicians. It is therefore possible that the 2001 increase in the CMPA premiums contributed to the reduction in the fraction of GPs choosing to provide intrapartum care, shown in figure 1.1. However, this should not drive the results presented for the balanced panel of GPs. Moreover, for physicians who kept providing intrapartum care after 2001, the CMPA's fee increase should act as a counteracting force to the effects presented in tables 1.2 to 1.6. Indeed, the increased fixed cost associated with obstetrical practice should have, if anything, led them to increase their volume of eligible services, rather than to reduce it as observed.

Another policy change which could interfere with the estimation of responses to the sole delivery premium is the gradual increase in the Ontario physician income threshold. In place until the mid 2000s across the province, the threshold set a level of annual income beyond which additional services billed were remunerated only at a fraction of the base fees. While the annual threshold for general practitioners increased slightly from $300,000 in 1997 to $330,000 in 2000, it rose substantially to $455,000 in 2002 before being eliminated in 2005 [Kantarevic et al., 2008]. This relaxation of the threshold in the early 2000s should, if anything, work against the estimates presented in tables 1.2 to 1.5, especially for higher-volume providers. Indeed, the 2000 and 2002 increases in the income threshold made it less likely that a higher remuneration for intrapartum care would push a GP's income beyond the income threshold, imposing a tax on all the services they would have to keep providing throughout the year.[40] Columns 1 to 3 of table 1.8 present alternative results that focus on subsample of physicians who were

[40] Within the balanced panel of GPs, the proportion of observations (physician-year combinations) that are above the threshold drops from 11% between 1996 and 2000 to 6.5% between 2000 and 2003d.

likely not affected by this reform whose incomes were either just below the threshold (panel B) or at least 15 thousand dollars below the threshold (panel C). A reduction of a little more than 3 eligible services per year is observed among higher-volume providers, although the estimate is only statistically significant at the 10% level. The response for low-volume providers and for the full sample are similar to those obtained with the entire balanced panel. The results are generally robust to using this narrower sample, corresponding to physicians further away from the threshold.

A second change in the threshold system, this one in 1998, is likely to interact with the estimation of responses to the sole delivery premium: payments derived from labour and delivery services were exempted from the thresholds in 1996 and 1997, but this exemption was removed in 1998 [MOHLTC, 1998].[41] To avoid capturing these two different regimes in the pre-period defined in equation (1.6), I estimate the model on a sample period starting in 1998. As shown in columns 4 to 6 of table 1.8, the results are also generally robust to only considering two years of data before the introduction of the sole delivery premium in 2000. However, the estimates are no longer statistically significant for the subsample of high-volume providers whose earnings were at least 15 thousand dollars below the income threshold set by the policy.

1.6 Spillover effects on non-targeted services

The results presented so far highlight that the sole delivery premium did not incentivize GPs to be more active in intrapartum care, and may have led those who billed higher volumes of targeted services before 2000 to reduce their provision of such procedures. These findings are at odds with the hypothesis that physicians are pure income-maximizers. Indeed, the comparison of GPs located below and above the threshold of 25 eligible services is coherent with the scenario in which the premium resulted in some income effects.[42] In this context, one remaining question is how such income effects might have influenced GPs' provision of services in areas of care outside of labour and delivery. In the absence of detailed information on billings for services not targeted by the premium, I investigate this relationship by estimating a modified version of equation (1.6) in which the outcome variable is defined either as the annual income received by GPs for targeted services or as their total annual fee-for-service earnings.

Table 1.9 presents the results from these estimations on a sample of Ontario GPs who billed for eligible services in each year between 1996 and 2003, and whose activities were mainly conducted in urban areas[43]. Panel A shows the results for the pooled sample of lower− and higher-volume providers of targeted services (based on pre-2000 annual averages), while panels B and C present the results for each group separately. Column 1 first replicate the within-physician results presented in panel B of table 1.6. The introduction of the premium is associated with an average reduction of 2.3 targeted services

[41] Kantarevic et al. [2008] suggest that the substitution effect dominated following the 1998 threshold reforms for services with relatively low volumes and high prices.

[42] It may of course be the case that, if intrapartum care enters directly and negatively into a GPs' utility function, incentives such as the sole delivery premium allow them to *buy* the option of reducing their activity in this specific area of care. As discussed in section 1.2, obstetric care could indeed be associated with direct utility costs due to its disruptive effect on physicians' personal life, etc.

[43] This last criteria reflects the fact that no response to the premium was estimated for GPs practicing in rural environments

per year among GPs in the pooled sample, a response driven by higher-volume providers who billed on average 4 fewer eligible services after 2000. Column 2 analyzes more closely the type of targeted services provided. Although the estimates in all three panels are imprecise, they suggest that GPs did not attempt to deliver more complicated births after 2000, although the absolute monetary payoff from doing so increased with the premium.[44] This is not necessarily contrary to the goal of the premium, as it may be desirable for more complex cases to be handled by OB/GYNs.

Column 3 indicates that the introduction of the premium led to statistically significant increases in the total payments received by GPs for intrapartum care, even for higher-volume providers who reduced the volume they provided in response to the incentive. All else equal, income from targeted services increased in the pooled sample by $2,262 per year (in 2002 dollars), which is in line with the expected impact of a 50% higher payment for the 25 first deliveries, less the income lost from three fewer eligible services.[45] Moreover, the estimates in column 4 suggest that the premium did not materialize into higher total earnings for GPs. Indeed, no statistically significant change is detected when looking at the impact from the sole delivery premium on the total fee-for-service income received annually. Although the negative signs on the corresponding point estimates in all panels is counterintuitive, the magnitude of the coefficient is negligible, corresponding only to 1% of the average income in the estimating sample in panel A and 1.2% in panel C.[46]

The fact that GPs' total fee-for-service earnings do not increase while average income from targeted services does so suggests that the income effects generated by the sole delivery premium could have spilled over to areas of care offered by GPs but not targeted by the premium. The responses presented in columns 3 and 4 are not dissimilar to the type of behaviour that would be expected of physicians who would have reached some form of target level of income before the premium was introduced.[47] More generally, the estimates are consistent with a model in which the marginal utility of income would decrease steeply after individuals having reached some desired income level. Under such a model of behaviour, stronger income effects should be observed for GPs whose incomes are at or past their target. Meanwhile, no reduction in the provision of intrapartum care should be expected from GPs who were below their desired level of income when the sole delivery premium was introduced. For this group, the premium should offer an opportunity to close the gap between their desired and realized levels of incomes.[48] To test this prediction, I select two groups from the estimating sample used in table 1.9: GPs whose average annual income between 1996 and 1999 was beyond $200,000 (most likely to be in the neighbourhood of their reference income) and those for whom it was equal or inferior to $150,000 (most likely to be below

[44] Complex deliveries are considered to be cesarean sections, operative deliveries breech birth, and multiple births. The baseline fee for these deliveries are higher than the fees for uncomplicated vaginal delivery or an attendance at labour and/or delivery.

[45] The average volume of deliveries in that sample after 2000 is 28. The estimated impact corresponds to an average payment (excluding the premium) of $220 (2002 dollars) once the premium on the first 25 targeted services and the reduction in the volume of services provided are accounted for.

[46] The results are smaller when using income in nominal rather than in real terms.

[47] The average physician in the main sample has 18 years of experience, and may be more likely to have reached a stable income level than younger physicians who are still building their practice. Indeed, Rizzo and Blumenthal [1996] highlights that the ratio between target and realized incomes should be a decreasing function of experience and converge towards 1 over time.

[48] Rizzo and Zeckhauser [2003] find that physicians reporting earnings below their reference income tend to have a disproportionate income growth in the following years, taking actions to close the gap between their reference and realized incomes. Rizzo and Blumenthal [1996] finds a positive association between the target-to-realized income ratio and prices charged by U.S. physicians.

their reference income). For each of these groups, I then separately estimate a version of equation 1.6 with an interaction term between *Post* and a dummy variable indicating if a physician had an average volume of eligible services superior to 25 prior to 2000. Among GPs who were likely below their income target before 2000, the coefficient on the *Post* variable is 1.921 (s.e. 4.970) and the coefficient on the interaction term is -4.664 (s.e. 4.026), the sum of both estimates being statistically undistinguishable from zero (p-value of 0.64). In the group of GPs who had likely reached their income target prior to the introduction of the premium, the coefficients on $Post2000$ and its interaction with an indicator for high-volume providers are respectively 0.152 (s.e. 1.385) and -3.851 (s.e. 1.333), and the sum of both estimates is statistically significant at the 10% level. Although the classification of GPs in those two groups relies on *ad hoc* assumptions on physicians' potential income targets, these results provide some support for stronger income effects among GPs whose initial levels of earnings were higher. They are also in line with recent evidence from Contandriopoulos and Perroux [2013], who present suggestive evidence from Quebec that in response to increases in fees, physicians may choose to increase their leisure time rather than their income, a behaviour they say is consistent with (although it doesn't need to be) income targeting.

Under certain assumptions, a simple labor-leisure framework such as the one presented in section 1.2 could predict the patterns observed in columns 3 and 4 of table 1.9, even in the absence of strong income effects such as those implied by the target income hypothesis. For example, the premium could have led GPs to substitute non-targeted (and relatively less lucrative) services for intrapartum care after the introduction of the sole delivery premium. However, the sign of the estimates in column 1 of table 1.9 –and the results presented in tables 1.2 to 1.6– provide evidence against this alternative explanation. Moreover, an adaptation of the elasticities derived in McGuire and Pauly [1991] suggests this type of substitution across services is less likely to result from the introduction of the premium if the utility cost associated with the mere provision the targeted services increases rapidly with volume, if the targeted services are disproportionately labor and time intensive, or if the marginal utility of income is very low.[49] Given the nature of the task, the personal life disruption, or the risk of malpractice associated with intrapartum care, those three scenarios (or at least the first two) are quite plausible.[50]

Alternative explanations could support the absence of increase in the total income of GPs practicing obstetrics after 2000, and future research should explore other potential narratives. Gottlieb and Clemens [2017] suggest that a permanent increase in compensation (for example through increases in fees) might incentivize physicians to spend more time in unpaid activities such as training, or to spend more time on other forms of human capital investment in the short run in order to increase their capacity to maximize income from higher fees in the long-run. They argue that this type of short-term responses could be mistaken for a backward-bending labour supply. The NPDB data does not provide information on time allocated to training. However, the reduction in targeted services is mostly observed for GPs initially delivering a high volume of births each year, and who are less likely to need additional training to reap the full benefits from the premium. Moreover, GPs in the sample have on average 18 years of experience,

[49] See glossary in McGuire and Pauly [1991].

[50] I note that given the incapacity to induce demand for childbirth, reconciling the results presented with a utility maximization framework cannot be achieved by evoking inducement of physician services as a "normal bad", but can be done by assuming that delivering births come at a utility cost for physicians that is greater than the normal cost of effort and the opportunity cost in terms of leisure.

and a shorter horizon over which to benefit from further human capital accumulation. It is therefore unlikely that the reduction in targeted services provided after 2000 comes from the type of response explored by Gottlieb and Clemens [2017].[51]

Secular downward trends in physicians' working hours could also be driving part of the results presented in columns 3 and 4 of table 1.9. Data presented in Crossley et al. [2009], for example, show that on average, Canadian physicians of all age groups and cohorts reduced their working hours between 1983 and 2001. However, they highlight that the decline was more dramatic between 1982 and 1993 (before the period studied in this paper) and partly due to secular trends and changes in the gender mix of physicians, two factors addressed by the inclusion of a time trend and physician characteristics or fixed effects in equation (1.6). The difference-in-differences results (table 1.A.6 in the appendix) also provide some reassuring evidence that the effects (or absence thereof) of the sole delivery premium on total fee-for-service income are not entirely driven by a secular reduction in working hours among Canadian GPs. Data from the National Family Physician Workforce Survey suggests that across all family physicians (including those not providing obstetrical services), the relative proportion who reduced their working hours between 1999 and 2001 was greater in British Columbia (22% versus 14%) than in Ontario (20% against 17%). Relatively similar intentions in terms of planned changes in working hours were reported in both provinces in 2001 for the period 2001 to 2003 (17% planning a reduction versus 2% planning an increase in British Columbia, those proportions reaching 19% versus 3% in Ontario) [CFCP, 2002]. Looking at the differential changes in total income after the introduction of the sole delivery premium across those two provinces should therefore help to isolate responses to the premium from common trend of GPs reducing their working hours.[52]

1.7 Conclusion

Primary care physicians' responses to the introduction of a targeted financial incentive could in theory take various forms. The interaction between substitution and income effects could lead physicians to alter their provision of the services specifically targeted by a bonus or premium in unanticipated ways. Given the scope of their practice, GPs could also respond by adjusting their activities in areas of care not targeted by the incentives they are exposed to. In this paper, I find that the introduction of a bonus payment for low-volume obstetrics in Ontario did generate stronger involvement of primary care physicians in this area of care. Indeed, GPs did not increase their provision of the services targeted by the incentive, likely because of the strong income effects associated with the premium. The results point to a reduction of up to 8% in the volume of obstetric services provided by physicians initially billing each year more targeted services than the maximum number of times the incentive payment could be claimed, and for whom the introduction of the premium should have resulted in a pure income effect. Moreover, the data suggests that the incentive might also have led GPs to reduce the volume of services they provided in areas of care not directly targeted by the bonus payment. Finally, the premium also does not seem to have impacted the rate at which GPs opted out of obstetric practice in the early 2000s.

[51] Crossley et al. [2009] also provide evidence that the hours spent by Canadian GPs on professional activities aside from patient care did not increase on average between 1997 and 2003.

[52] Using cross-sectional data on GPs from 1990, Hurley and Jeon [2007] also note that reductions in physicians' working hours may partially be offset by an increase in the volume of services billed per hour.

At face value, these results question the effectiveness of targeted incentives to shape primary care physicians' involvement in an area of care such as obstetrics, and are at odds with the income-maximizing theory of physician behaviour.

An investigation of the changes in physicians' sources of income following the introduction of the premium also highlights that the share of GPs' total income coming from intrapartum care increased after 2000, as income from labour and delivery increased and income from other services decreased. Combined with the absence of a statistically significant change in physicians' income levels after 2000, these results could be consistent with the predictions of a model that is closer to income targeting, especially if most physicians initially earned incomes close to their targets. GPs' responses to the sole delivery premium can however also be reconciled with a more traditional labor-leisure framework of physicians as utility maximizers in which targeted services are associated with a utility cost for physicians, or require an important amount of time and effort to provide.

This paper adds to existing work on physicians' responses to financial incentives and more generally to the literature on physicians' labour supply in two ways. First, the non-linearity introduced in GPs' budget constraint through a maximum number of services eligible for the incentive payment studied provides an opportunity to investigate the income effects of fee increases for physicians who should not be subject to substitution effects.[53] Second, my results suggest that focused incentives can have far-reaching consequences and distort the provision of services in various areas of care beyond the ones originally targeted. This suggests that GPs adjust their labour supply and practice style over the full range of services they provide rather than compartmentalizing their work by area of care. Of course, obstetrics is not representative of all services provided by physicians and the responses estimated in the context of the sole delivery premium may not generalize to targeted financial incentives available in other areas of care. For example, birth deliveries are associated with a higher risk of malpractice suits in case of adverse events, and on-call obstetrics can cause disruption in a GPs' personal life. Services associated with such barriers and utility costs for physicians, and who may appeal to a smaller number of providers, are however likely to be the target of government initiatives such as bonuses and premiums. Understanding the impacts of targeted financial incentives on such services is therefore important.

It should be mentioned that despite the inclusion of time trends and sample restrictions, which should limit the confounding impact of contemporaneous policy changes or changes in physicians' environment, unobserved changes around the introduction of the sole delivery premium in 2000 could affect the responses I estimate. While the results are robust to using British Columbian GPs as a control group in a difference-in-differences approach, this strategy has its limitations. However, the robustness of the results across estimation methods lends some credibility to the results presented. Overall, the estimated responses highlight the inability of the sole delivery premium to generate its stated objective of encouraging the provision of obstetric care by general practitioners and family physicians.

[53] Most other estimates of income effects are identified from changes in physicians' non-labour income, for example from changes in a spouse's income.

My findings leave some questions unanswered and identify opportunities for future research. First, the main empirical findings point to a substantial degree of heterogeneity in physicians' responses to financial incentives, especially with respect to their main environment of practice. Understanding if the absence of income effects in the population of GPs located in more rural areas is due, for example, to constraints on the demand for care, limited access to referrals, or selection into locations based on unobservable preferences is important to assess the distributional impacts of the introduction of financial incentives, both for patients and for physicians. The estimated impact on non-targeted services presented in section 1.6 are also imprecise, and the nature of the data used here does not allow to either formally test the hypothesis of target income behaviour, or to identify which non-targeted areas of care may have been affected by the sole delivery premium. These are important questions that should receive attention in the future.

Table 1.1: **Summary Statistics – Ontario GPs by years in sample**

	Balanced panel	**GPs not in balanced panel** Pre 2000 only	Post 2000 only	Pre & Post 2000	**Total**
Male	0.67	0.68	0.72	0.68	0.68
Year of graduation from medical school	1981	1982	1984	1983	1982
Graduated in Canada	0.89	0.90	0.85	0.88	0.89
Experience at billing	18.94	15.16	17.88	16.27	16.71
Rural practice (MIZ)	0.28	0.18	0.22	0.27	0.23
Targeted services/year (excl. zeros)	28.02	9.26	2.19	13.86	14.11
Average targeted services 1996-1999	28.98	9.26	–	15.65	16.60
Complex deliveries	0.03	0.06	0.13	0.04	0.05
Payment for targeted services/year	10,408	2,979	721	4,833	4,976
Total fee-for-service payment/year	229,561	209,029	175,356	201,819	208,339
Number of years in sample	8.0	1.9	1.3	5.0	4.0
Physicians (N)	498	785	253	424	1960

Notes:
1- Sample is restricted to primary care physicians who billed targeted services between 1996 and 2003, excluding the top 1% intrapartum care billings
2- Unbalanced panel is formed exclusively of physicians who did not bill targeted services each year between 1996 and 2003.
3- All physicians who identified themselves as an OB/GYN at least once in the sample period are excluded
4- All physicians who changed provinces during the sample period are excluded
5- All billings and earnings are in dollars of 2002

Table 1.2: **Impact on volume of targeted services, main specification with time trend**

	All observations		**Balanced panel**		
	(1)	(2)	(3)	(4)	(5)
Post 2000	-0.902*	-0.641	-0.841	-1.173	-1.073
	(0.549)	(0.817)	(0.594)	(0.907)	(0.975)
Male		-10.575***		-10.625***	
		(1.167)		(1.888)	
Age 30-34		6.876***		5.700**	3.332
		(1.382)		(2.605)	(2.047)
Age 35-59		8.794***		7.703**	4.132*
		(1.465)		(3.244)	(2.503)
Age 40-44		8.900***		8.521**	4.705*
		(1.574)		(3.373)	(2.811)
Age 45-49		7.870***		8.204**	1.115
		(1.750)		(3.545)	(3.173)
Age 50-54		5.697***		7.565*	-0.175
		(1.864)		(3.853)	(3.548)
Age 55-59		5.493**		5.021	-1.599
		(2.564)		(4.044)	(4.083)
Age 60-64		8.988		3.175	-6.036
		(5.805)		(4.328)	(4.734)
Age 65-69		8.681		5.771	-10.170*
		(6.603)		(7.574)	(5.443)
Age 70-74		0.253		4.268	-23.667**
		(3.253)		(10.498)	(10.224)
Age 75+		6.876***		-0.977	-26.682***
		(1.382)		(9.560)	(8.583)
OB/GYNs		0.020		-0.085	-0.081
		(0.073)		(0.082)	(0.084)
Newborns (K)		0.183**		0.158*	0.180*
		(0.083)		(0.092)	(0.096)
Midwives		-0.013		-0.062	-0.065
		(0.044)		(0.045)	(0.048)
Trend	✓	✓	✓	✓	✓
Rural & Statistical area code FE		✓		✓	✓
Medical school FE		✓		✓	
Physician FE					✓
GPs	1960	1960	498	498	498
Observations	7,917	7,917	3,984	3,984	3,984
R-squared	0.001	0.173	0.002	0.264	0.815

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Estimation sample is composed of all Ontario family physicians and general practitioners who billed targeted services between 1996 and 2003, excluding the top 1% of intrapartum care billings.
3- Estimation conditional on billing targeted services in a given twelve months period and excludes physicians who graduated or retired between 1996 and 2003.
4- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
5- The balanced panel (columns 3-5) is formed of all Ontario general practitioners and family physicians who billed targeted services in each year between 1996 and 2003.

Table 1.3: **Impact on volume of targeted services, main specification with year effects**

	All observations		**Balanced panel**		
	(1)	(2)	(3)	(4)	(5)
Year 1997	-0.778*	-0.828*	-0.918*	-1.094**	-1.142**
	(0.411)	(0.427)	(0.499)	(0.540)	(0.558)
Year 1998	-0.589	-1.049**	-1.309**	-1.670**	-1.610**
	(0.486)	(0.516)	(0.621)	(0.711)	(0.686)
Year 1999	-0.505**	-1.109**	-1.384**	-1.855**	-1.685**
	(0.540)	(0.564)	(0.664)	(0.770)	(0.694)
Year 2000	-1.586***	-2.170***	-3.064***	-3.681***	-3.484***
	(0.617)	(0.655)	(0.745)	(0.885)	(0.760)
Year 2001	-1.198*	-1.899***	-2.201***	-2.797***	-2.599***
	(0.661)	(0.709)	(0.732)	(0.926)	(0.741)
Year 2002	-1.237*	-2.108***	-2.880***	-3.491***	-3.329***
	(0.724)	(0.807)	(0.803)	(1.058)	(0.816)
Year 2003	-0.329	-1.111	-3.104***	-3.669***	-3.534***
	(0.768)	(0.875)	(0.871)	(1.197)	(0.901)
Male		-10.576***		-10.626***	
		(1.167)		(1.889)	
Age 30-34		4.408***		5.645**	
		(1.181)		(2.605)	
Age 35-59		6.871***		7.622**	
		(1.384)		(3.240)	
Age 40-44		8.788***		8.463**	
		(1.465)		(3.369)	
Age 45-49		8.894***		8.142**	
		(1.573)		(3.544)	
Age 50-54		7.867***		7.516*	
		(1.748)		(3.849)	
Age 55-59		5.694***		4.941	
		(1.863)		(4.037)	
Age 60-64		5.496**		3.123	
		(2.562)		(4.326)	
Age 65-69		8.992		5.707	
		(5.806)		(7.570)	
Age 70-74		8.681		4.207	
		(6.589)		(10.529)	
Age 75+		0.190		-0.493	
		(3.239)		(9.590)	
Rural & SAC FE		✓		✓	✓
Medical school FE		✓		✓	
Physician FE					✓
GPs	1,960	1,960	498	498	498
Observations	7,917	7,917	3,984	3,984	3,984
R-squared	0.001	0.173	0.003	0.264	0.811

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Estimation sample is composed of all Ontario family physicians and general practitioners who billed targeted services between 1996 and 2003, excluding the top 1% of intrapartum care billings.
3- Estimation conditional on billing targeted services in a given twelve months period, and excluding physicians who graduated or retired between 1996 and 2003.
4- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
5- The balanced panel is formed of all Ontario general practitioners and family physicians who billed targeted services in each year between 1996 and 2003.

Table 1.4: **Impact on volume of targeted services, by average/year before 2000**

	All observations		**Balanced panel**		**Balanced panel**	
	Below 25 pre 2000	**Above 25 pre 2000**	**Below 25 pre 2000**		**Above 25 pre 2000**	
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2000	0.890	-3.053*	0.698	0.821	-3.519**	-3.289*
	(0.619)	(1.608)	(0.786)	(0.933)	(1.713)	(1.784)
Male	-2.567***	-7.136***	-2.857***		-6.400***	
	(0.595)	(1.894)	(1.086)		(2.437)	
Age 30-34	0.054	6.079**	2.537	2.823	2.410	3.430
	(0.848)	(2.452)	(2.031)	(2.586)	(3.326)	(3.456)
Age 35-59	-0.433	6.543**	3.127	5.165*	1.859	3.027
	(0.956)	(2.797)	(2.327)	(2.987)	(3.541)	(4.241)
Age 40-44	0.706	8.610***	2.051	4.024	3.493	5.589
	(0.927)	(3.010)	(2.330)	(3.380)	(3.659)	(4.680)
Age 45-49	1.025	9.650***	1.931	2.783	1.183	-0.077
	(0.951)	(3.373)	(2.395)	(3.771)	(3.882)	(5.293)
Age 50-54	-0.030	11.104***	1.249	1.198	3.339	-0.791
	(0.979)	(3.811)	(2.418)	(4.283)	(4.328)	(5.808)
Age 55-59	-0.232	7.411*	-0.515	-0.838	3.835	-1.788
	(1.065)	(4.378)	(2.549)	(4.922)	(5.123)	(6.726)
Age 60-64	-2.189*	5.673	-3.891	-4.488	-1.404	-9.059
	(1.182)	(4.086)	(2.576)	(5.696)	(4.603)	(7.934)
Age 65-69	-3.143*	8.186	-6.534**	-7.921	-1.173	-14.471
	(1.821)	(5.502)	(3.176)	(6.529)	(7.265)	(9.140)
Age 70-74	-4.067	-11.193	-11.048***	-11.512	-24.183	-38.633**
	(2.545)	(12.479)	(3.400)	(7.410)	(15.314)	(16.307)
Age 75+	-6.955***		-17.575***	-18.372**		
	(1.797)		(3.514)	(7.931)		
OB/GYNs	-0.013	0.049	0.153	0.179	-0.102	-0.120
	(0.066)	(0.115)	(0.111)	(0.117)	(0.108)	(0.111)
Newborns (K)	0.277***	0.395**	0.200**	0.251***	-0.036	0.024
	(0.062)	(0.169)	(0.078)	(0.091)	(0.175)	(0.179)
Midwives	0.002	-0.110	-0.005	-0.003	-0.130	-0.131
	(0.032)	(0.090)	(0.040)	(0.043)	(0.085)	(0.088)
Trend	✓	✓	✓	✓	✓	✓
Rural & SAC FE	✓	✓	✓	✓	✓	✓
Physician FE				✓		✓
GPs	1,559	401	262	262	236	236
Observations	5,313	2,604	2,096	2,096	1,888	1,888
R-squared	0.088	0.174	0.248	0.570	0.239	0.707

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Estimation sample is composed of all Ontario family physicians and general practitioners who billed targeted services between 1996 and 2003, excluding the top 1% of intrapartum care billings.
3- Balanced panel (columns 3-6) is formed of all primary care physicians who billed targeted services each year between 1996 and 2003 4- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
5- The average number of targeted services per physician prior to 2000 is obtained from all years when physicians submitted relevant billings.

Table 1.5: **Robustness checks: Impact on volume of targeted services**

	All observations			**Balanced panel**		
		Pre 2000 average			Pre 2000 average	
	All	Below 25	Above 25	All	Below 25	Above 25
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Including observations with no targeted services						
Post2000	0.153	0.925***	-4.154***	-1.073	0.821	-3.289**
	(0.343)	(0.234)	(1.409)	(0.975)	(0.933)	(1.784)
Observations	15,696	12,488	3,208	3,984	2,096	1,888
R-squared	0.157	0.134	0.214	0.815	0.571	0.707
Panel B: Including observations from 1995 to 2004						
Post2000	0.740	1.632***	-2.962**	-1.046	0.463	-2.667*
	(0.705)	(0.578)	(1.375)	(0.835)	(0.827)	(1.467)
Observations	9,815	6,474	3,341	4,925	2,520	2,405
R-squared	0.176	0.095	0.168	0.790	0.575	0.678
Panel C: Excluding observations corresponding to targeted services/year $>$ 50						
Post2000	0.036	0.580	-1.437	-0.750	0.249	-2.605*
	(0.692)	(0.579)	(1.387)	(0.693)	(0.730)	(1.400)
Observations	7,154	5,288	1,866	3,420	2,079	1,341
R-squared	0.130	0.081	0.155	0.747	0.549	0.504
Panel D: Excluding observations corresponding to targeted services/year $\leq$ 5						
Post2000	-0.997	1.457*	-3.164**	-1.348	0.261	-3.214*
	(0.863)	(0.709)	(1.565)	(0.959)	(0.833)	(1.726)
Observations	5,917	3,375	2,542	3,705	1,835	1,870
R-squared	0.214	0.197	0.176	0.817	0.607	0.714
Panel E: Heterogenous effects by gender						
Post2000	-1.023	2.792***	-3.339*	-1.955	1.151	-3.749*
	(1.176)	(1.016)	(1.979)	(1.362)	(1.408)	(2.183)
Post2000 $\times$ male	0.552	-2.505***	0.515	1.288	-0.425	0.814
	(1.109)	(0.975)	(1.809)	(1.222)	(1.296)	(1.820)
p-value						
(Post2000) + (Post2000 $\times$ male)	0.58	0.65	0.10	0.51	0.45	0.11
Observations	7,917	5,313	2,604	3,984	2,096	1,888
R-squared	0.173	0.091	0.174	0.815	0.571	0.707

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Main estimation sample is composed of physicians billing targeted services between 1996 and 2003, excluding the top1% of intrapartum care billings.
3- Balanced panel (columns 4-6) is formed of all Ontario general practitioners and family physicians who billed targeted services in each year between 1996 and 2003.
4- All estimates are obtained controlling for the physicians' main SAC of practice, an indicator for rural area of practice, age categories, the total number of OB/GYNs in the province, the total number of midwives, the total number of newborns in the year, and a time trend. Columns 1-3 include a gender indicator and medical school of graduation fixed effects. Columns 4-6 include physician fixed effects.
5- The average number of targeted services per physician prior to 2000 is obtained using relevant billings for 1996-1999. 6- Estimation conditional on billing targeted services in a given twelve months period, unless mentioned otherwise (panel A).

Table 1.6: **Impact on volume of targeted services, heterogenous effects by rural practice (balanced panel)**

	All		**Below 25 pre 2000**		**Above 25 pre 2000**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Urban and Rural						
Post 2000	-2.247**	-2.044**	0.549	0.596	-4.307**	-4.133**
	(0.951)	(0.989)	(0.811)	(0.887)	(1.709)	(1.789)
Post 2000 × rural	3.088**	2.796***	0.302	0.457	4.837**	5.208**
	(1.074)	(1.037)	(1.128)	(1.061)	(2.339)	(2.432)
p-value						
(Post 2000) + (Post 2000 × rural)=0	0.48	0.55	0.44	0.40	0.85	0.70
GPs	498	498	262	262	236	236
Observations	3,984	3,984	2,096	2,096	1,888	1,888
R-squared	0.265	0.815	0.249	0.571	0.241	0.709
Panel B: Urban only						
Post 2000	-2.607**	-2.342*	0.220	0.024	-4.326**	-4.046**
	(1.176)	(1.211)	(1.086)	(1.106)	(1.851)	(1.937)
GPs	365	365	157	157	208	208
Observations	2,864	2,864	1,207	1,207	1,657	1,657
R-squared	0.254	0.806	0.282	0.593	0.257	0.709
Age groups and Rural & SAC FE	✓	✓	✓	✓	✓	✓
OB/GYN and midwives and newborns	✓	✓	✓	✓	✓	✓
Trend	✓	✓	✓	✓	✓	✓
Physician FE		✓		✓		✓

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Balanced panel is composed of all Ontario family physicians and general practitioners who billed targeted services each year between 1996 and 2003, excluding the top 1% of intrapartum care billing.
3- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
4- The average number of targeted services per physician prior to 2000 is obtained using relevant billings for 1996-1999.

Table 1.7: **Impact on volume of targeted services, difference-in-differences estimation**

	All		**Below 25 pre 2000**		**Above 25 pre 2000**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Balanced panel						
Post 2000 × Ontario	-2.740**	-2.742**	-0.273	-0.507	-5.046***	-5.155**
	(1.171)	(1.229)	(0.812)	(0.879)	(1.952)	(1.500)
Post 2000	-1.039	-0.919	-0.658*	-0.160	-2.388	-2.067
	(0.946)	(1.003)	(0.394)	(0.474)	(2.104)	(1.261)
GPs	1041	1041	590	590	451	451
Observations	7,287	7,287	4,130	4,130	3,157	3,157
R-squared	0.264	0.814	0.203	0.545	0.195	0.696
Panel B: Balanced panel, urban observations						
Post 2000 × Ontario	-4.356***	-4.335***	-0.827	-1.172	-6.758***	-6.783***
	(1.035)	(1.119)	(0.889)	(0.917)	(1.691)	(1.820)
Post 2000	-1.251	-1.038	-0.451	-0.389	-2.169	-1.660
	(1.220)	(1.266)	(0.467)	(0.527)	(2.317)	(2.489)
GPs	845	845	434	434	411	411
Observations	5,830	5,830	2,972	2,972	2,858	2,858
R-squared	0.268	0.810	0.203	0.546	0.205	0.695
Age groups	✓	✓	✓	✓	✓	✓
Rural & SAC FE	✓	✓	✓	✓	✓	✓
OB/GYN & newborns	✓	✓	✓	✓	✓	✓
Trend	✓	✓	✓	✓	✓	✓
Physician FE		✓		✓		✓

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Estimation sample is composed of all Ontario and British Columbia family physicians and general practitioners who billed targeted services between 1996 and 2003, excluding the top 1% of intrapartum care billing.
3- Balanced panel is formed of all Ontario and British Columbia general practitioners and family physicians who billed targeted services each year between 1996 and 2003.
4- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
5- The average number of targeted services per physician prior to 2000 is obtained using relevant billings for 1996-1999. 6- The average number of targeted services per physician prior to 2000 is obtained from all years between 1996 and 1999 in which physicians submitted relevant billings.

Table 1.8: **Impact on volume of targeted services, alternative samples to account for the income threshold system for Ontario physicians (balanced panel)**

	1996-2003			**1998-2003**		
		Pre 2000 average			Pre 2000 average	
	All	Below 25	Above 25	All	Below 25	Above 25
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All income levels						
Post2000	-1.073	0.821	-3.289**	-2.853**	-1.161	-4.932*
	(0.975)	(0.710)	(1.784)	(1.733)	(1.484)	(3.130)
GPs	498	262	236	498	262	236
Observations	3,894	2,096	1,888	2,988	1,572	1,416
R-squared	0.814	0.568	0.707	0.843	0.665	0.761
Panel B: Income < Threshold						
Post2000	-0.851	1.023	-3.324*	-2.047	-1.119	-3.463*
	(1.095)	(1.015)	(2.050)	(1.869)	(1.564)	(3.650)
GPs	491	259	232	491	259	232
Observations	3,510	1,894	1,616	2,691	1,449	1,242
R-squared	0.813	0.577	0.702	0.844	0.673	0.764
Panel C: Income < (Threshold - $15 K)						
Post2000	-0.884	1.087	-3.599*	-1.861	-0.715	-3.746
	(1.128)	(1.041)	(2.139)	(1.906)	(1.568)	(3.736)
GPs	488	258	230	488	258	230
Observations	3,418	1,857	1,561	2,621	1,420	1,201
R-squared	0.812	0.579	0.700	0.843	0.675	0.762
Age groups	✓	✓	✓	✓	✓	✓
Rural & SAC FE	✓	✓	✓	✓	✓	✓
OB/GYNs & midwives	✓	✓	✓	✓	✓	✓
Newborns	✓	✓	✓	✓	✓	✓
Trend	✓	✓	✓	✓	✓	✓
Physician FE	✓	✓	✓	✓	✓	✓

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Main estimation sample excluding the top1% of intrapartum care billings.
3- Balanced panel is formed of all Ontario general practitioners and family physicians who billed for E411 eligible services in each year between 1996 and 2003.
4- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
5- The average number of targeted services per physician prior to 2000 is obtained using relevant billings for 1996-1999.

Table 1.9: **Impact on targeted services (volume, complexity and income), and on total fee-for-service income (balanced panel, urban)**

	Targeted services	Share of complex births	Billings for targeted services ($2002)	Total FFS billings ($2002)
	(1)	(2)	(3)	(4)
Panel A: All GPs				
Post2000	-2.342*	0.001	2,262***	-2,438
	(1.211)	(0.005)	(434)	(2,618)
GPs	365	365	365	365
Observations	2,864	2,864	2,864	2,864
R-squared	0.805	0.788	0.805	0.915
Panel B: GPs with average targeted services/year below 25 pre 2000				
Post2000	0.024	0.009	1,547***	-1,895
	(1.106)	(0.009)	(429)	(3,742)
GPs	157	157	157	157
Observations	1,207	1,207	1,207	1,207
R-squared	0.593	0.812	0.628	0.914
Panel C: GPs with average targeted services/year above 25 pre 2000				
Post2000	-4.046**	-0.004	2,705***	-3,099
	(1.937)	(0.005)	(699)	(3,628)
GPs	208	208	208	208
Observations	1,657	1,657	1,657	1,657
R-squared	0.709	0.768	0.705	0.907
Age groups	✓	✓	✓	✓
Rural & SAC FE	✓	✓	✓	✓
OB/GYN and midwives	✓	✓	✓	✓
Trend & newborns	✓	✓	✓	✓
Physician FE	✓	✓	✓	✓

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Estimation sample is composed of all Ontario family physicians and general practitioners in a census metropolitan agglomeration of census agglomeration who billed targeted services each year between 1996 and 2003, excluding the top 1% intrapartum care billings.
3- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
4- Complex deliveries include C-sections and instrumental deliveries (forceps, vacuum)
5- All payments and billings and in $2002.

Figure 1.1: Share of Ontario primary care physicians billing targeted intrapartum care services

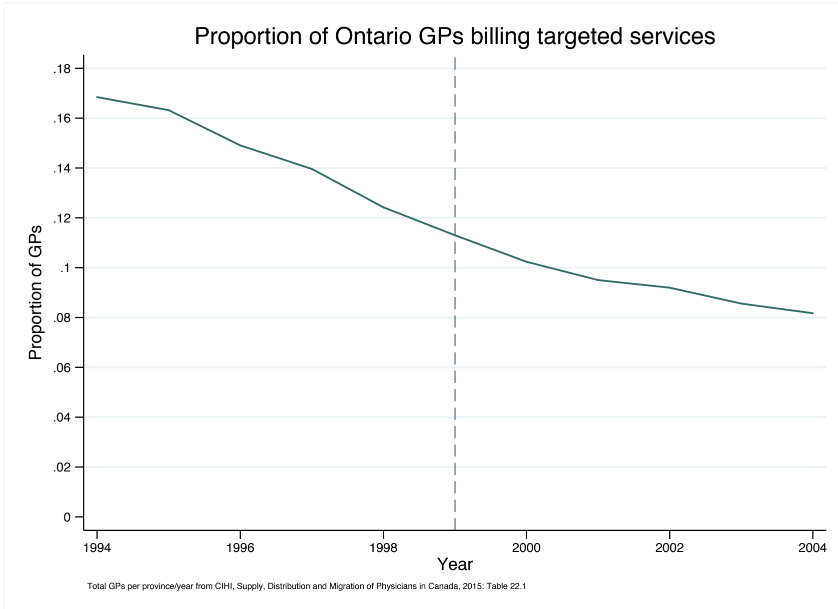


Figure 1.2: Distribution of Ontario primary care physicians against annual volume of targeted services provided

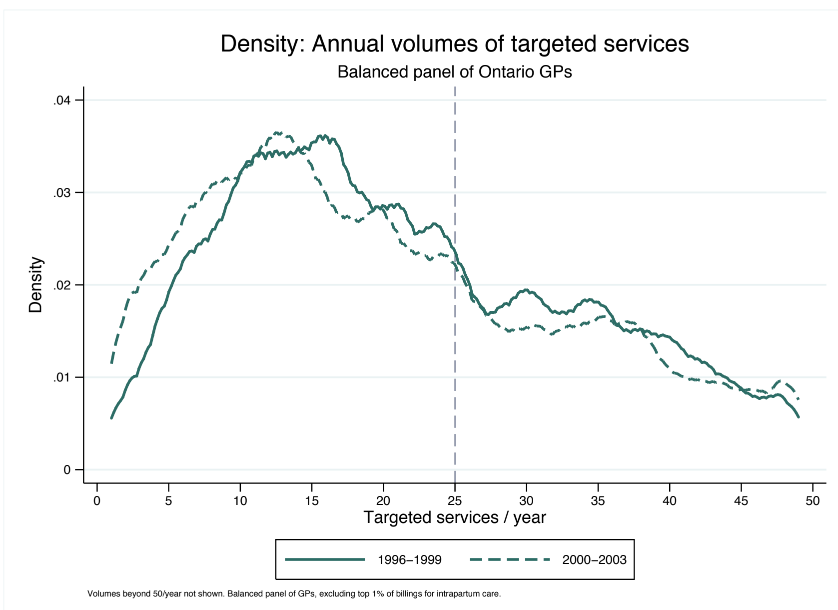


Figure 1.3: Year effects on primary care physicians' provision of targeted intrapartum care services

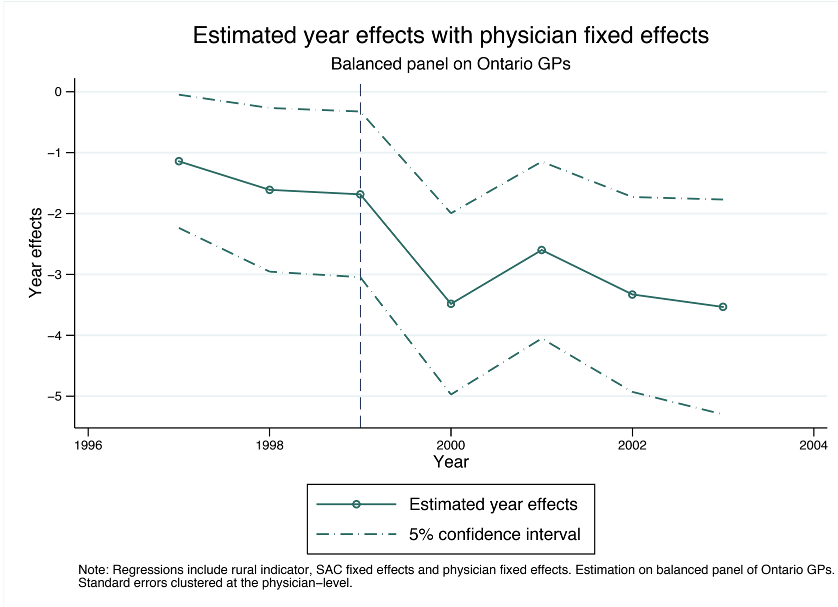
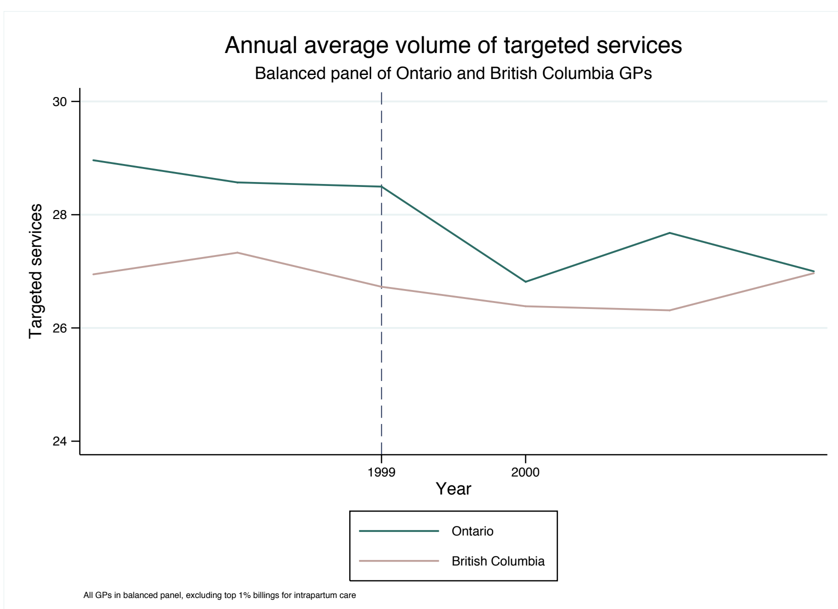


Figure 1.4: Evolution of average targeted services provided by primary care physicians in Ontario and British Columbia



Appendix

Table 1.A.1: **Distribution of Ontario GPs by years observed and average annual provision of targeted services between 1996-1999**

	Balanced Panel	GPs not in balanced panel			Total
		Pre 2000 only	Post 2000 only	Pre & Post 2000	
Below 25 pre 2000	262	700	—	344	1306
Above 25 pre 2000	236	85	—	80	401
—	—	—	253	—	253
Total	498	785	253	424	1960

Notes:
1- Sample is restricted to primary care physicians who billed for intrapartum care between 1996 and 2003, excluding the top 1% intrapartum care billings
2- All physicians who identified themselves as an OB/GYN at least once in the sample period are excluded
3- All physicians who changed provinces during the sample period are excluded
4- The balanced panel is composed of all primary care physicians who billed targeted services each year between 1996 and 2003, and who were not in the top 1% of intrapartum care billings.

Table 1.A.2: **Summary Statistics – Ontario GPs by average annual provision of targeted services between 1996-1999**

	All GPs	**Below 25 pre 2000**	**Above 25 pre 2000**
Panel A: All GPs (physicians observed between 1996 and 2003)			
Male	0.682	0.727	0.509
Year of graduation from medical school	1982	1982	1982
Graduated in Canada	0.890	0.888	0.895
Experience at billing	16.71	16.64	17.01
Rural practice (MIZ)	0.232	0.263	0.113
Eligible services/year (excl. zeros)	14.11	7.33	40.47
Average eligible services 1996-1999	16.60	8.47	43.06
Complex deliveries	0.054	0.060	0.033
Payment for eligible services/year	4,976	2,524	14,507
Total fee-for-service payment/year	208,339	201,692	234,182
Number of years in sample	4.04	3.41	6.49
Physicians (N)	1960	1,559	401
Panel B: Balanced panel (physicians observed each year between 1996 and 2003)			
Male	0.669	0.779	0.547
Year of graduation from medical school	1981	1980	1981
Graduated in Canada	0.892	0.889	0.894
Experience at billing	18.94	19.40	18.43
Rural practice (MIZ)	0.281	0.424	0.122
Eligible services/year (excl. zeros)	28.02	14.73	42.78
Average eligible services 1996-1999	28.98	14.64	44.89
Complex deliveries	0.031	0.023	0.040
Payment for eligible services/year	10,408	5,613	15,731
Total fee-for-service payment/year	229,561	216,031	244,580
Number of years in sample	8	8	8
Physicians (N)	498	262	236

Notes:
1- Sample is restricted to primary care physicians who billed for intrapartum care between 1996 and 2003, excluding the top 1% intrapartum care billings
2- All physicians who identified themselves as an OB/GYN at least once in the sample period are excluded
3- All physicians who changed provinces during the sample period are excluded
4- The balanced panel is composed of all primary care physicians who billed targeted services each year between 1996 and 2003, and who were not in the top 1% of intrapartum care billings
5- All billings and earnings are in dollars of 2002
6- Category $\leq$ 25 includes GPs who were not did not bill targeted services at all between 1996 and 1999.

Table 1.A.3: **Summary Statistics: Ontario and British Columbia GPs (balanced panel)**

	British Columbia	**Ontario**
Male	0.643	0.669
	(0.480)	(0.471)
Year of graduation from medical school	1982	1981
	(8.00)	(8.34)
Graduated in Canada	0.766	0.892
	(0.424)	(0.311)
Experience at billing	17.51	18.44
	(8.00)	(8.34)
Rural practice (MIZ)	0.124	0.282
	(0.322)	(0.448)
Rural practice (weak to no influence from CMA/CA)	0.098	0.124
	(0292)	(0.326)
Targeted services/year (excl. zeros)	26.06	28.20
	(18.17)	(19.21)
Average targeted services 1996-1999	25.69	28.98
	(17.69)	(19.71)
Average targeted services 2000-2002	26.20	27.07
	(20.63)	(19.95)
Complex deliveries	0.036	0.031
	(0.101)	(0.097)
Payment for targeted services/year	11,000	10,360
	(7,615)	(6.894)
Total fee-for-service payment/year	220,071	232,053
	(77,322)	(87.533)
Number of years in sample	7	7
	(0)	(0)
Physicians (N)	543	498

Notes:
1- Balanced panel is restricted to primary care physicians who billed for targeted services each year between 1996 and 2003
2- All physicians who identified themselves as an OB/GYN at least once in the sample period or who changed provinces are excluded
3- All physicians who changed provinces during the sample period are excluded.
4- All billings and earnings are in dollars of 2002

Table 1.A.4: **Impact on targeted services (volume, complexity and income) and on total fee-for-service income (all observations, urban)**

	Targeted services	**Share of complex births**	**Billings for targeted services ($2002)**	**Total FFS billings ($2002)**
	(1)	(2)	(3)	(4)
Panel A: All GPs				
Post2000	-1.233	-0.014	2,042***	-7,659
	(1.044)	(0.009)	(386)	(4,117)
GPs	1531	1531	1531	1531
Observations	5,880	5,880	5,880	5,880
R-squared	0.168	0.117	0.176	0.325
Panel B: GPs with average targeted services/year below 25 pre 2000				
Post2000	0.634	-0.023	1,301***	-9,982
	(0.810)	(0.015)	(311)	(6051)
GPs	1173	1173	1173	1173
Observations	3,597	3,597	3,597	3,597
R-squared	0.096	0.130	0.114	0.370
Panel C: GPs with average targeted services/year above 25 pre 2000				
Post2000	-3.274*	-0.001	3,078***	-5,357
	(1.702)	(0.004)	(606)	(3,864)
GPs	358	358	358	358
Observations	2,283	2,283	2,283	2,283
R-squared	0.188	0.490	0.184	0.455
Age groups	✓	✓	✓	✓
Rural & SAC FE	✓	✓	✓	✓
OB/GYNs and midwives	✓	✓	✓	✓
Trend & newborns	✓	✓	✓	✓
Physician FE				

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Estimation sample is composed of all Ontario family physicians and general practitioners in a census metropolitan agglomeration of census agglomeration who billed targeted services between 1996 and 2003, excluding the top 1% intrapartum care billings.
3- GPs are observed in years during which they bill targeted services.
4- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
5- The average number of targeted services per physician prior to 2000 is obtained using relevant billings for 1996-1999. 6- Complex deliveries include C-sections and instrumental deliveries (forceps, vacuum)
7- All payments and billings and in $2002.

Table 1.A.5: **Impact on targeted services (volume, complexity and income) and on total fee-for-service income (all observations, urban, with physician FE)**

	Targeted services	Share of complex births	Billings for targeted services ($2002)	Total FFS billings ($2002)
	(1)	(2)	(3)	(4)
Panel A: All GPs				
Post2000	-0.601	-0.005	2,279***	-2,818
	(0.956)	(0.005)	(349)	(2,558)
GPs	1531	1531	1531	1531
Switchers	686	686	686	686
Observations	5,880	5,880	5,880	5,880
R-squared	0.826	0.847	0.827	0.933
Panel B: GPs with average targeted services/year below 25 pre 2000				
Post2000	1.338	-0.007	1,628***	-3,896
	(0.909)	(0.008)	(355)	(3,658)
GPs	1173	1173	1173	1173
Switchers	410	410	410	410
Observations	3,597	3,597	3,597	3,597
R-squared	0.694	0.867	0.711	0.941
Panel C: GPs with average targeted services/year above 25 pre 2000				
Post2000	-3.065*	-0.001	3,052***	-1,835
	(1.742)	(0.004)	(633)	(3,461)
GPs	358	358	358	358
Switchers	276	276	276	276
Observations	2,283	2,283	2,283	2,283
R-squared	0.665	0.729	0.653	0.913
Age groups	✓	✓	✓	✓
Rural & SAC FE	✓	✓	✓	✓
OB/GYNs and midwives	✓	✓	✓	✓
Trend & newborns	✓	✓	✓	✓
Physician FE	✓	✓	✓	✓

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Estimation sample is composed of all Ontario family physicians and general practitioners in a census metropolitan agglomeration of census agglomeration who billed targeted services between 1996 and 2003, excluding the top 1% intrapartum care billings.
3- GPs are observed in years during which they bill targeted services.
4- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
5- The average number of targeted services per physician prior to 2000 is obtained using relevant billings for 1996-1999. 6- Complex deliveries include C-sections and instrumental deliveries (forceps, vacuum)
7- All payments and billings and in $2002.

Table 1.A.6: **Impact on targeted services (volume, complexity and income) and on total fee-for-service income, difference-in-differences estimation (balanced panel, urban)**

	Targeted services	**Share of complex births**	**Billings for targeted services ($2002)**	**Total FFS billings ($2002)**
	(1)	(2)	(3)	(4)
Panel A: All GPs				
Ontario × Post	-3.383***	-0.014***	1,373***	-8,078*
	(1.170)	(0.004)	(489)	(4,142)
GPs	845	845	845	845
Observations	5,830	5,830	5,830	5,830
R-squared	0.809	0.775	0.803	0.913
Panel B: GPs with average targeted services/year below 25 pre 2000				
Ontario × Post	-0.800	-0.008*	1,260***	-8,911
	(0.796)	(0.004)	(388)	(5,719)
GPs	434	434	434	434
Observations	2,972	2,972	2,972	2,972
R-squared	0.546	0.732	0.564	0.907
Panel C: GPs with average targeted services/year above 25 pre 2000				
Ontario × Post	-5.077**	-0.021***	1,393	-6,877
	(2.114)	(0.006)	(901)	(5,200)
GPs	411	411	411	411
Observations	2,858	2,858	2,858	2,858
R-squared	0.696	0.813	0.696	0.919
Age groups	✓	✓	✓	✓
Rural & SAC FE	✓	✓	✓	✓
OB/GYNs and midwives	✓	✓	✓	✓
Trend & newborns	✓	✓	✓	✓
Physician FE	✓	✓	✓	✓

Notes:
1- Statistical significance levels: * 10% ** 5% *** 1%
2- Estimation sample is composed of all Ontario family physicians and general practitioners in a census metropolitan agglomeration of census agglomeration who billed targeted services each year between 1996 and 2003.
3- Control group is formed of British Columbia family physicians and general practitioners. Treatment group is formed of Ontario family physicians and general practitioners.
4- Statistical area codes are defined by Statistics Canada using the first three digits of the postal code the most used by a physician for billing purposes in a given year.
5- The average number of targeted services per physician prior to 2000 is obtained using relevant billings for 1996-1999. 6- Complex deliveries include C-sections and instrumental deliveries (forceps, vacuum)
7- All payments and billings and in $2002.

Figure 1.A.1: Implications of the sole delivery premium on physicians' budget sets

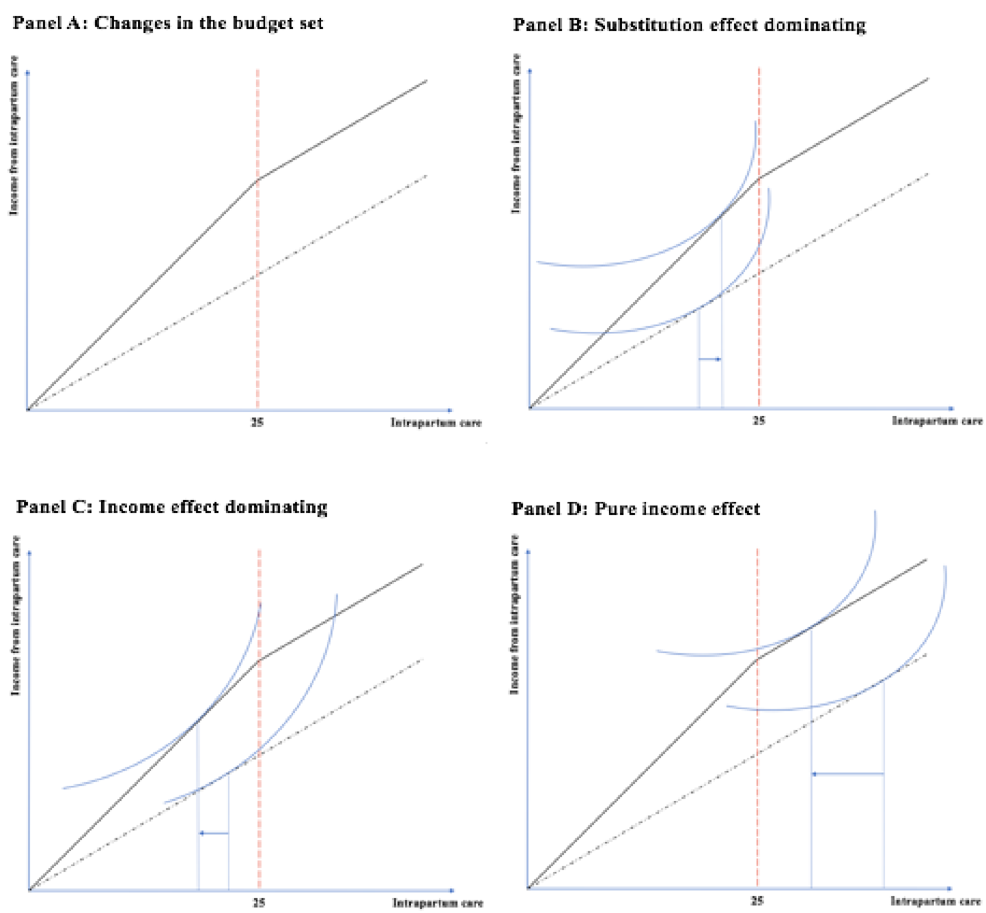


Figure 1.A.2: Share of Ontario primary care physicians billing targeted intrapartum care services, by age group

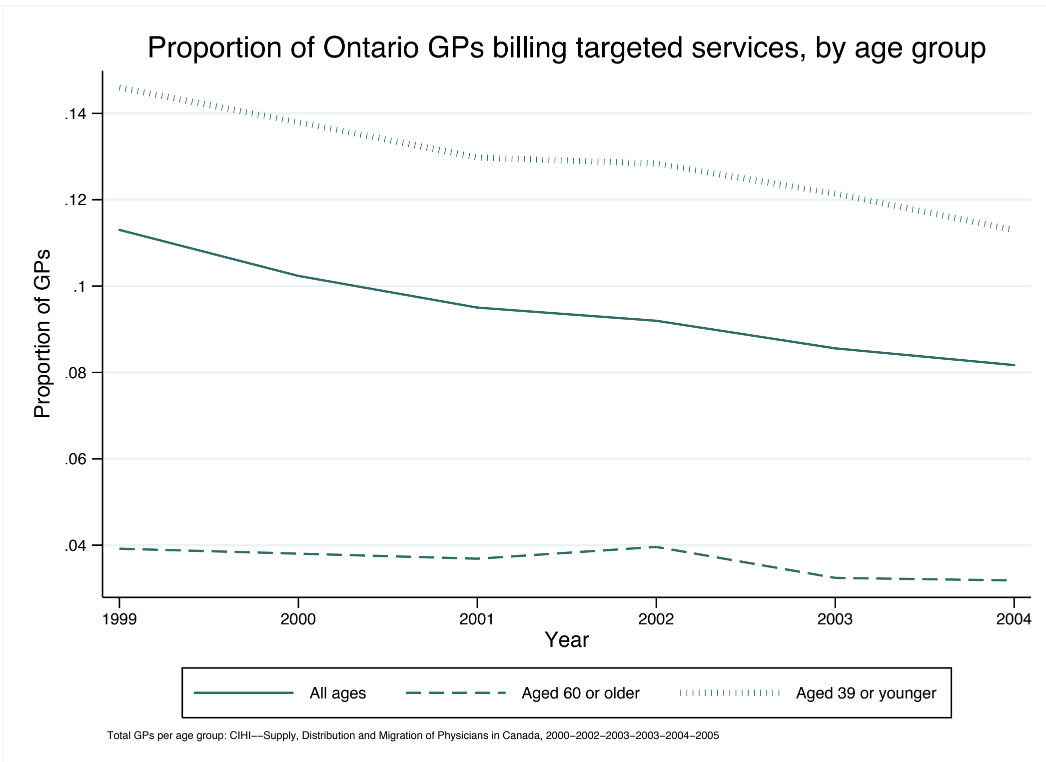
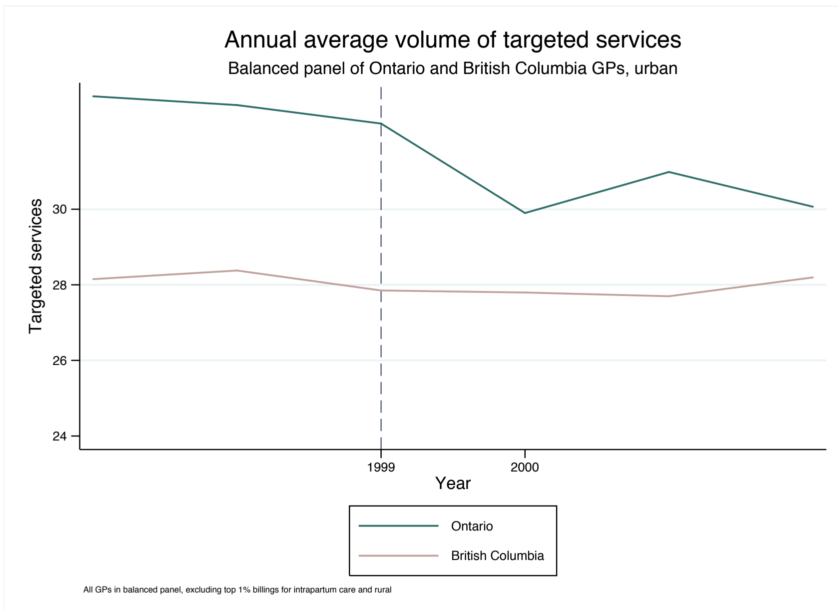


Figure 1.A.3: Evolution of average targeted services provided by primary care physicians in urban areas in Ontario and British Columbia



Chapter 2

Explaining the rise in C-sections: the contribution of physician incentives and research spillovers

This chapter is co-authored with Dr Sara Allin (Institute of Health Policy, Management and Evaluation, University of Toronto), Dr. Michael Baker (Department of Economics, University of Toronto) and Dr. Mark Stabile (INSEAD)

Parts of this paper are based on data and information provided by the Canadian Institute for Health Information. However, the analyses, conclusions, opinions and statements expressed herein are those of the authors, and not necessarily those of the Canadian Institute for Health Information.

2.1 Introduction

Why are cesarian section (C-section) births increasingly viewed as a public health care problem? First, C-section rates have been rising steadily over past decades in many developed countries. In figure 1 we graph the rates of C-section in a selection of countries in North America and Europe, as well as in Australia between 1995 and 2012. Across these countries, the rates have increased between 17 and 48 percent since 2000. Second, the World Health Organization (WHO) has identified 15 percent as the highest C-section rate justifiable on medical grounds,[1] yet Gibbons et al. [2010] report that half of a sample of countries representing over 95 percent of global births had C–section rates exceeding this level. Note

[1] The guideline emitted by the WHO in 1985 stated: "Countries with some of the lowest perinatal mortality rates in the world have cesarean section rates of less than 10%. There is no justification for any region to have a rate higher than 10-15%." [World Health Organization, 1985]. This guideline has been re-examined in 2015 and concluded that no optimal C-section rate could be identified, and that more evidence is needed to understand the health risks to mothers and babies of C-Section rates over 30% [World Health Organization, 2015].

that in figure 1, the C-section rates in all six countries are above 15 percent throughout the entire period. Third, over time C-section deliveries have consumed a rising proportion of hospitals' surgical resources. For example, in both Canada and the US, C-sections are now the most common inpatient surgery [CIHI, 2016a, Pfuntner et al., 2013].[2] Fourth, C-section rates have been found to vary across countries, within countries and across hospitals, suggesting something more than the health of the mother and child is playing a role in defining the trends. Finally, C-sections are increasingly recognized as major surgery, posing significant, and perhaps in certain cases avoidable, risks for both mother and child [SOGC, 2004a].

Among the reasons offered for rising C-sections rates, the trend to delay childbirth—the risk of complication during labor increases with maternal age—and the higher incidence of multiple births due to fertility treatments are leading medical-based accounts. In addition, defensive medicine and risk avoidance—the risk of poor or catastrophic health outcomes for the mother and child and/or the risk of lawsuits in the rare event of failed vaginal delivery—on the part of health practitioners is also thought to contribute, fanned by new technologies such as continuous electronic fetal monitoring. Finally, the convenience of C-sections is believed to appeal to both health professionals and mothers. It precludes long periods of labor and offers more predictability to doctors' and mothers' schedules.

Economic research on the rising C-section rates has mostly focused on doctors' monetary incentives to perform the procedure, as predicted by the physician-induced demand model [Evans, 1974]. While previous studies (summarized in section 2.2) suggest physicians' discretion plays some role, it is difficult to construct an overall account of the importance of this explanation. This is because much of this research focuses on a select sample of births (e.g., patients covered by Medicaid), and/or studies the practice of doctors who may have self-selected into reimbursement schemes.

In this paper, we use data from the population of births in Canada to investigate physicians' response to fees when choosing between two potentially substitutable procedures: vaginal delivery and C-section. Our focus on Canada offers important advantages. First, obstetric care in this country is remunerated according to fee-for-service agreements, so any role of relative fees should be straightforward to uncover.[3] Second, fees for these procedures vary substantially across provinces and over time, providing an attractive framework for identification. Finally, the health system in Canada is public (single payer) and universal, so we can consider the full population of hospital births within the country.

Our investigation offers two main results. First, we find that doubling the compensation for a C-section relative to a vaginal delivery increases the likelihood that a physician opts for the former by at most four percentage points, all else equal. This behavioral response is driven by obstetricians/gynaecologists (OB/GYNs). In contrast, general practitioners (GPs), whose incomes rely less heavily on activities re-

[2] C-sections totaled 100 963 surgeries in Canada in 2014-15. Knee replacements came second, with 60 607 occurrences. Over that same period, birth deliveries were responsible for nearly 370 000 inpatient stays, lasting on average 2.3 days. The second most important factor for inpatient hospitalization during that period were respiratory diseases, with nearly 88 000 occurrences [CIHI, 2016a].

[3] In 2011-2012, more than 85 percent of Canadian physicians reported being mainly compensated by fee-for-service agreements [NPS, 2013]. The Canadian Institute for Health Information estimated the proportion of physician payments made in the form of fee-for-service compensation to be around 72% [CIHI, 2016b].

lated to birth deliveries, do not exhibit sensitivity to relative fees.

Second, as in some past research, the estimates exhibit sensitivity to the specification of jurisdiction-specific trends. We find that this sensitivity is mostly resolved when we specify an interaction between a fixed effect for the announcement of the "Term Breech Trial" results [Hannah et al., 2000] and these trends. This is not intuitive, as this research changed best practice recommendations for the delivery of non-vertex deliveries, but had no explicit recommendations for the majority, vertex deliveries. To try to understand this impact we document that the release of the Trial results, and their subsequent reinterpretation, significantly impacts the growth rate of the C-section delivery of vertex births in Australia, Canada and the US. The contribution of the breech trial to the overall growth rate in C-sections appears to be much larger than the contribution of fee incentives. We are not aware that this observation has been made previously in the literature, and while we discuss some possible explanations of this effect, an investigation of its source is an important topic for future research.

2.2 Previous Research

There have been a number of empirical studies investigating the relationship between a variety of incentives and C-sections. Gruber and Owings [1996] investigate how physicians adjust the proportion of births they deliver by C-section (usually more generously remunerated than vaginal deliveries in fee-for-service contracts) in response to an income drop caused by an exogenous decline in the fertility rate. They find in a nationally representative sample of hospitals, covering the period 1970-1982, that that a 10 percent decrease in the fertility rate leads to an increase of 1 percent in the C-section rate.

Other studies look directly at the impact of changes in the relative fees of C-sections and vaginal deliveries on birth delivery outcomes. Gruber et al. [1999] provide evidence, for a sample of births covered by the US Medicaid program, that the C-section rate would rise by 3.9 percent (0.7 percentage point) in response to a $100 increase in the compensation received for a C-section, all else equal.[4] Their results also suggest that 75 percent of the difference in the C-section rates for Medicaid patients and privately insured patients between 1988-1992 in the United States could be explained by the difference in the reimbursement parameters specific to each type of insurance coverage.[5] Alexander [2015] uses data from 1990 through 2008 and exploits the different financial incentives faced by Medicaid fee-for-service and salaried American federal government employed physicians to test for physician-induced demand (PID). She finds that an increase of $100 in the fee differential between C-sections and vaginal deliveries increases the rate of primary C-sections by 4 percent within the Medicaid population in counties without a salaried hospital.[6] She finds no evidence of a relationship within counties with a salaried hospital.

[4] In a cross sectional study of birthing events in four hospitals in Cali, Columbia, Gomez and Carrasquilla [2012] also finds that privately insured patients, to whom physicians can charge more for health care, are more likely to get a primary C-section. The effect would be even stronger for repeat C-sections. They also find that 81 percent of all primary C-sections performed could find their justification on medical grounds.

[5] Grant [2009] argues that only one quarter of the effect measured by Gruber et al. [1999] remains when accounting for state-specific time trends and relaxing some sampling restrictions.

[6] Compared to Gruber et al. [1999], she examines a period during which patients had access to more information on

She also finds that the higher C-section rate prompted by financial incentives had a positive impact on infant survival rates for high-risk Medicaid pregnancies.

Other research has directly exploited the asymmetry of information between obstetricians and their patients. Johnson and Rehavi [2016] find that, in the presence of financial incentives rewarding C-section, mothers who are physicians are 7 to 12 percent less likely to have a C-section than otherwise comparable mothers without medical knowledge. In line with the PID model, they find physicians' scope to respond to financial incentives is increasing in their informational advantage over their patients.

The identification strategy in most of these studies is based in one of two characteristics of the American health system: the difference in the fee structure for the care provided to Medicaid patients and to privately insured patients, or the different remuneration agreements—fee-for-service versus fixed salary—offered across hospital types.[7] If physicians self-select on these same features of the system, then any inference must be interpreted accordingly. For example, if physicians with a higher propensity to induce demand for C-section delivery choose to work in hospitals offering fee-for-service remuneration agreements, where they will have a greater control over their income compared to at institutions offering salaried contracts, then estimates based on comparisons of salaried and fee-for-service physicians potentially overestimate the response within the population of physicians. Likewise, in a multiple-payer system, physicians with a greater propensity for inducement could also orient their practice more heavily towards privately insured patients, for whom the C-section fee premium is typically higher. Estimation on a sample of Medicaid births would therefore be less likely to capture their behavior. Finally, if it is the case that Medicaid patients are systematically different from their privately insured counterparts in ways that makes them more or less susceptible to their physician's influence, the results obtained may not apply to the population of patients as a whole.

In contrast the Canadian health care system does not allow physicians to choose between privately and publicly insured patients and to self-select into remuneration agreements for obstetric care. Furthermore, our sample covers almost the entirety of births in the ten Canadian provinces over the period studied.

One of the only papers examining physician care around birth in an environment similar to Canada's, Jensen [2014], finds that changing Danish physicians' remuneration from capitation to fee-for-service contracts, which increases incentives to provide more prenatal care, had a positive impact on infant health, especially for younger mothers. Although the institutional context she exploits limits the risk of her results being driven by physician sorting, the study cannot directly measure the impact of incentives on the quantity or quality of care provided.

birth delivery methods. This could in part explain the smaller estimates she obtains.

[7] One exception is the work of Kantarevic et al. [2008], who exploit a reform implemented in 1998 in Ontario, which imposed a ceiling on physicians' annual earnings. They investigate how physicians' choice of procedures changed as they reached the maximum amount that could be billed to the public insurance plan before receiving only partial compensation for a series of services. They find evidence of a positive and significant price elasticity of the supply of medical care for OB/GYNs. However, their study speaks to the volume of total care, rather than to the substitution between delivery methods.

2.3 Other factors influencing C-section rates

Patients' preferences have also been offered as an explanation of the high volume of C-section deliveries. One story often played up in the media is that an increasing proportion of mothers requests a C-section for aesthetic reasons, for fear of experiencing prolonged labor or for scheduling convenience. Testing this hypothesis is challenging: in Canada, for example, data on maternal requests for C-sections has not been collected consistently across provinces or over time. Despite this obstacle, a few studies argue that mothers' preferences could only play a marginal role in explaining the high rates of primary C-sections across the country. For example, using data from British Columbia (where mothers' preferences and requests for C-sections are more closely monitored), Hanley et al. [2010] provide evidence that only 2 percent of surgical deliveries are performed on mothers' requests without medical indication.[8]

Moreover, disentangling ex-ante preferences from the impact of inducement by a physician is not straightforward. In a medical note destined to Quebec physicians, Jimenez [2005] suggests that maternal preferences and apprehensions regarding delivery methods would be strongly influenced by physicians' beliefs and advice, conveyed in the prepartum phase of obstetric care. Gamble et al. [2007] also highlight the role of the health system in general and of physician advice in particular on the formation of a patient's attitude and requests towards different modes of delivery.

In parallel, the medical and legal literatures have described how tort laws, initially adopted to deter negligence, have intensified the recourse to unnecessary medical acts and contributed to a culture of defensive medicine. Using responses to the 2008 Health Tracking Physician Survey, Carrier et al. [2010] suggest that American OB/GYNs tend to overestimate the risk of a lawsuit in case of a problematic trial of labor—a risk that they would generally consider to be random—and try to avoid it by practicing defensive medicine. Interestingly, this would be observed for both fee-for-service and salaried physicians, suggesting that fear of malpractice cannot explain the different C-section rates observed between these two groups in the US. Lawthers et al. [1992] and Keeler and Brodie [1993] also report that physicians providing obstetric care may try to minimize their potential exposure to lawsuits by performing "defensive C-sections".

Exploiting variations in the true exposure to malpractice suits across geographic zones, Baicker et al. [2007] find that high malpractice costs increase the utilization of medical imaging and of other diagnostic procedures, but not the frequency of major surgeries. Using an instrumental variables framework, Kim [2007] finds no evidence that malpractice risks faced by OB/GYNs (defined as the number of malpractice claims per 1000 births or the value of malpractice claims paid per 1000 births) have an impact on physicians' choice to perform C-sections. However, Currie and MacLeod [2008] propose a theoretical framework in which increasing physician liability can reduce aggregate recourse to unnecessary C-sections by bringing their expected cost up. Their prediction hinges on the hypothesis that the probability of medical error—or bad outcome—is higher for unnecessary procedures. They find empirical support for their prediction in the context of the tort reforms implemented at the state level in the United-States

[8] The Canadian Health Services Research Foundation [2013] similarly finds that only 1 percent of C-sections in the United States would be performed on the basis of explicit mothers' preferences and would not have been performed otherwise.

between 1989 and 2001.

Another explanation for high C-section rates is technological change in health professionals' practice. Currie and MacLeod [2013] argue that the development of electronic fetal monitoring (EFM) technologies to assess mothers' and fetus' conditions during labor have had heterogeneous impacts on recourse to C-sections across the distribution of physicians with different diagnostic skills. Using electronic birth certificates for New Jersey between 1996 and 2007, they find that physicians with a limited ability to use new technology have increased their recourse to unnecessary C-sections among low-risk women following the introduction of EFM.[9]

Finally, there is very little evidence-based support in the literature for the hypothesis that physicians opt for C-section deliveries for scheduling convenience reasons [Lefevre, 2013].

2.4 The Canadian health care system

Canadian health care is a universal, single payer system that operates at the provincial level and is funded both by provincial governments and by fiscal transfers from the federal government. In the period we study, most services are remunerated according to fee-for-service agreements, in which physicians are responsible for billing the public health insurance plan for each procedure they perform at a fee that is administratively fixed by provincial authorities for a predetermined period (typically corresponding to a fiscal year). Fees do not evolve in response to the relative movements of the demand and supply for different medical services. Physicians are moreover precluded by the Canada Health Act from billing their patients in addition to the fees they receive from the public insurer.[10]

In table 2.1, we present the incidence of fee-for-service compensation for obstetric care in seven provinces, for the fiscal years 2006-2007 to 2010-2011.[11] In the less populous provinces—Prince Edward Island, Newfoundland and Labrador, New Brunswick, Saskatchewan— the rates oscillate between 40 and 80 percent. In the more populous provinces—Ontario, Quebec and Manitoba[12]— fee-for-service payments almost always represent more than 85 percent of OB/GYNs' compensation over the entire period. These statistics could be expected to be even higher for birth delivery activities specifically; the statistics presented in table 2.1 cover all care provided by OB/GYNs, including services other than birth delivery

[9] They also show that improving diagnostic skills among physicians could increase the proportion of high-risk women having a C-section and decrease the proportion of low-risk mothers giving birth by C-section. In aggregate, the impact on the latter group would outweigh the impact on the former, and a general appreciation in diagnostic skills would lead to a decrease in C-section rates.

[10] The single payer nature of provincial health care systems is directly inspired by the five principles on which rests the Canada Health Act: public administration, universality, accessibility, portability and comprehensiveness. Failure to satisfy those principles would result, for a provincial administration, in the loss of healthcare funding transfers from the federal government.

[11] The Canadian Institute for Health Information makes this information available from 2006 onwards. As of 2010-2011, the data provided by Nova Scotia, Alberta and British Columbia did not specify the proportion of total payments made to OB/GYNS under fee-for-service agreements.

[12] These three provinces account for 63 percent of all births in our sample.

that are more likely to be compensated under alternative payment schemes [CIHI, 2006]. Therefore, table 2.1 may underestimate the role of fee-for service in intrapartum care.

Changes in the fees in each province are typically induced by external scientific assessments of changes in the relative level of effort or skills required to perform different medical acts, caused for example by technological change, modifications in medical protocols or in the training for certain areas of practices. Changes in fees may also reflect political considerations as the formal negotiation process involves the provincial government and the provincial medical association (or its designated tariff committee). However, the objective of these negotiations is mostly to determine the amount of the global envelope attributed to physician remuneration, rather than to determine the relative movements in specific fees.

It might be argued that the pressure exercised by a medical association on the provincial government could nonetheless drive relative fees in a certain direction. This is an unlikely scenario in the specific case of obstetrical fees; OB/GYNs constitute a particularly small voting membership within medical associations with a limited influence in the determination of specific fees in negotiations rounds [Katz et al., 1997, Barer et al., 1996].[13] Moreover, each specialty represented in the association's tariff committee competes for its own financial interest against the other [Emery et al., 1999]. Each specialty often uses evidence or external assessment to legitimize its demands for specific fee increases, limiting the scope for endogenous fee determination.[14]

Another feature of the Canadian health care system relevant to this study is that physicians are effectively insulated from the impact of tort liability [Flood and Thomas, 2012].[15] Therefore, the risks of financial—as opposed to the reputational—costs of malpractice are unlikely to play a significant role in Canadian C-section trends.

Overall, the parameters under which physicians operate in the Canadian health system provide us with the opportunity to identify responses to exogenous changes in relative fees with few risks of contamination of our estimates through selection of physicians and patients, or other incentives from the legal environment.

[13]On the other hand, the stronger negotiating power held by general practitioners would historically have been used to increase their global remuneration by increasing the fees corresponding to medical acts for which they have a monopoly, or premiums linked to special aspects of their practice.

[14]To investigate this question further, we estimate a model in which we include leads of physician fees variables in our main estimating equation. The coefficient on the contemporaneous fee variable remains practically unchanged (from 0.039 to 0.041) and statistically significant at the 1% level, while the coefficient on the lead value of the fee variable is close to zero and not statistically significant (-0.009, with a standard error of 0.011).

[15]Almost all Canadian physicians subscribe to the Canadian Medical Protective Association (CMPA). The association negotiates with medical associations and provincial governments to establish regimes of medical liability protection for its members, along with patient compensation schemes. Although the premiums paid by physicians providing obstetric care are the highest among all specialties ($57 420 in Ontario, $20 304 in Quebec and $38 102 in the rest of Canada in 2014), they are substantially —if not entirely— reimbursed by the government in most provinces. Flood and Thomas [2012] also observe that the CMPA has a very large war chest, a reputation of defending claims very aggressively—a "scorched earth policy"— and does not experience rate premiums. Furthermore, under Canadian law unsuccessful plaintiffs may be required to pay up to two-thirds of a defendant's costs, which discourages malpractice claims.

2.5 Data

2.5.1 Individual Data on Pregnancies and Obstetric Care

Our main source of data on birth deliveries is the Hospital Morbidity Database (HMDB), administered by the Canadian Institute for Health Information (CIHI), which covers the universe of administrative discharge abstracts from acute inpatient facilities in all ten Canadian provinces.[16] Our sample covers the administrative records for all hospital births between April 1994 and March 2011 for all provinces. The only exception is Quebec, for which our data covers the period from April 2006 to March 2010.[17]

The HMDB provides a wide range of demographic characteristics for each mother, such as her birth date, her province of residence, her health risk factors, some information on her current pregnancy and previous deliveries and details on the health conditions of the infant at birth (weight, gestational age, etc.). It also contains detailed clinical information including the type of institution in which she gave birth, the date of her admission and discharge, the day of the week and time of the birth, as well as some information related to transfers to or from other institutions.[18]

Most importantly, the HMDB lists, for each mother, diagnoses made and interventions performed by physicians in acute care facilities prior to, during and after the delivery using the International Classification of Diseases codes (ICD-9, ICD-9-CM and ICD-10-CA)[19], the Canadian Classification of Health Interventions (CCI) and the Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures (CCP). The medical records also identify the factors that are medically known to be associated with higher risks of cesarean deliveries: multiple pregnancy, non-vertex presentation, etc. [ACOG, 2013, CIHI, 2006, NICE, 2012, Public Health Agency of Canada, 2008].

Finally, the administrative records identify the specialty of the physician involved in the various steps of the delivery process. When several providers are associated with a patient's episode of care, we use the type of the physician associated with the successful attempt of a delivery, but we nonetheless track any change in provider type for each patient.[20] Providers identified as family practitioners, general practitioners and community physicians are classified as GPs. Those identified as obstetrician/gynaecologists,

[16] Deliveries not happening in a hospital represent a marginal fraction of all births in Canada. For 2002-03, CIHI reports that 99% of births were delivered in acute care facilities [CIHI, 2006], with a slight decline to 98.5% by 2010 [Statistics Canada, 2016].

[17] Unlike in other provinces where acute inpatient facilities are required to provide CIHI with all discharge abstracts, the information on patients and care provided in Quebec's facilities is compiled and transferred to CIHI by the provincial Department of Health. Access to these files can be obtained only through a different data request. Our main results, however, are not sensitive to the exclusion of the four years of data from Quebec. Canadian territories are excluded from our analysis.

[18] Although almost identical from files submitted by other Canadian provinces, the data files from Quebec exclude the delivery time for all births. Moreover, certain provinces do not provide full information on mothers' previous pregnancies (characteristics and outcomes).

[19] ICD-9 and ICD-9-CM codes were gradually replaced by ICD-10-CA codes starting from 2001. From 2004 onwards, all provinces reported the diagnostics to CIHI using ICD-10-CA.

[20] If an OB/GYN and a GP are both associated with a single delivery, the delivery is classified as being performed by an OB/GYN. In most cases where this situation would arise, the GP would be paid a fixed amount for assisting the delivery. As a robustness check, we also consider the type of the first provider associated with an attempt of delivery. The direction and significance level of the results are insensitive to the definition of physician type.

surgeons, urogynaecologists, specialists in maternal-fetal medicine are classified as OB/GYNs (being more loosely defined as specialists).

We exclude from our final sample all deliveries that are not assigned to one of these two categories of physicians,[21] since other health care providers (nurse practitioners, midwives, etc.) have different remuneration agreements and compensation parameters. We also discard all records for mothers younger than 14 years old or older than 65 years old. For our main results, we also exclude multiple births, cases where the mother had a previous C-section and cases of fetal malpresentation. These cases are more likely to result in C-section delivery for medical reasons (our data indicates that cases exhibiting at least one of these characteristics have a significantly higher C-section rate, at 71.52%) and the doctors are compensated additionally for deliveries associated with such characteristics. Therefore, we restrict our main sample to the population of vertex singletons and then include the more complicated births as a robustness test.[22] Our final sample consists of 3,708,511 records. The distribution of observations across provinces is reported in table 2.2.

2.5.2 Fee Data for Obstetric Care

We obtain information on the fees paid in each fiscal year to physicians for vaginal deliveries and C-sections from provincial ministries of health and medical associations. Those amounts correspond to the information listed on the fee schedules published for physician billing purposes in the fee-for-service remuneration system. As previously mentioned, those schedules are administered at the province-level, and physician compensation for different procedures vary not only through time, but also across provinces. We obtain information on fees for the period 1994 to 2010 for most provinces, with a few exceptions: Alberta for the periods 1994-1996 and 1998-2000, Saskatchewan and Prince Edward Island for the period 1994-1997, Nova Scotia and Newfoundland and Labrador for the fiscal year 1994.

Certain provinces list slightly different fees for GPs and OB/GYNs; we therefore allocate to each birth the appropriate fee measures given the physician type observed as performing the delivery. Our results also hold when averaging the GP and OB/GYN fees for each procedure, or when using only the fees listed for OB/GYNs or only the fees listed for GPs.[23]

[21] These deliveries, when not assisted by a physician, account for 6.95 percent of all records.

[22] Johnson and Rehavi [2016] and Alexander [2015] also focus their estimation on a sample excluding cases of previous C-sections and multiple births.

[23] The payment records from the NPDB inventory also provide an alternative source of information on birth procedures performed by physicians remunerated under fee-for-service agreements. We have constructed a data set at the aggregate level (each observation corresponding to the combination of a province and a year) on the incidence of C-section and vaginal births using this source. Unlike the HMDB, NPDB payment records only include birth deliveries performed by physicians remunerated under fee-for-service agreements. Therefore, examining C-section rates in the NPDB provides a robustness check for our main results; the presence of birth deliveries not remunerated according to fee-for-service agreements in our individual-level data could bias our estimated physician response to financial incentives towards zero. Our results are however consistent across data sources.

2.5.3 Descriptive Statistics

The average C-section rate for all births in our data is 24.01 percent. Excluding all cesarean deliveries for cases of previous C-section, fetus malpresentation and multiple pregnancies, the rate in our final sample falls to 13.4 percent. Table 2.3 compares average maternal and pregnancy characteristics across delivery outcomes and provides the average of each variable in the final sample.[24] Mothers are on average 29 years of age at delivery, and mothers delivering by C-section are older than mothers delivering vaginally (aged 29.4 versus 28.5 for women having a vaginal delivery). Underlying this difference is that women older than 40 represent a greater share of the C-section group (3.6 percent versus 2.1 percent). This is likely due, in part, to the fact that pregnancy later in life is associated with specific risks and complications during labor [Joseph et al., 2005].

The proportion of deliveries performed on weekends is significantly lower for C-sections than for vaginal deliveries, which might simply be due to hospitals' constraints on the availability of operating rooms for planned surgical procedures, or might attest to the scheduling convenience of C-sections from a physician's perspective.[25] Finally, C-section deliveries are associated with longer inpatient stays and with a higher probability of exceptional postpartum complications leading to long-term hospitalization.

Of all vertex singleton births recorded in our sample, more than 73 percent were delivered by OB/GYNs, 4 percent of which we can identify as having been initially supervised by GPs. Among deliveries transferred from GPs to OB/GYNs, 30 percent ended in a C-section and 18 percent in an operative delivery. Of all mothers seen by an OB/GYN from the beginning, 17 percent delivered their child by C-section; for mothers seen by a GP from the beginning, that proportion falls to 2 percent. Overall, across the provinces, GPs' share of delivery decreased consistently from 31 percent in 1994-95 to 23 percent in 2010-11, with a low of 22 percent in 2005. Their share of all C-sections performed oscillated between 2 and 3 percent during that period.

Figure 2.2 presents the evolution of the C-section rates and of the respective fees for C-section and vaginal births by province for our sample period. We report the C-section rates for both all births (the blue lines) and for our final sample of vertex singletons whose mother did not have a previous C-section (the red line). We also show the evolution of the ratio of the fee paid for a C-section relative to the fee paid for a vaginal delivery (yellow line). Across Canadian provinces between 1994-95 and 2010-11, the average fee billed by a physician for a C-section delivery was approximately 20 percent higher than the average fee billed for a vaginal delivery, and the gap between the two fees for a delivery was on average $75. The difference in fees between the two procedures, however, exhibits considerable heterogeneity

[24] Data on mother's schooling and marital status are not available at the individual-level. Each observation is therefore matched to the corresponding province-year level proportion of married women and of women with the mentioned schooling level, as published by Statistics Canada. For those variables, the proportions reported in table 2.3 are provincial rates, weighted by the number of individual events corresponding to each delivery method for each combination of province and year. The resulting measures are relatively similar across the three groups: the average province-year rate proportion of women with some postsecondary education is around 50 percent, the proportion with less than 8 years of schooling is 2 percent and the average province-year marriage rate is 66 percent.

[25] All our results, presented in section 2.7, are robust to the inclusion of controlling for the timing (weekend) of the mother's admission to the hospital.

across provinces and through time.

The figure shows that the rates of C-section for both all births and vertex singleton births are increasing throughout the period in all provinces. This indicates that less complicated deliveries have contributed to the rise in the total C-section rates over the period. There also appears to be a positive correlation between C-section rates and movements in fees, especially in Alberta, British Columbia, New Brunswick and Prince Edward Island. Figure 2.2 also indicates that in most provinces, there is a trend break or intercept shift in C-section rates in the early 2000s (i.e., Alberta, British Columbia, Manitoba, New Brunswick, Nova Scotia, Ontario and Prince Edward Island). The obvious correlate of this development is the publication of the "Term Breech Trial" [Hannah et al., 2000] in 2000. Based on a cross-country randomized trial, this study concluded that planned C-section was safer for mothers and infants than vaginal deliveries in cases of breech presentation.[26]

Another factor influencing the C-section rates over the period is rapid improvements in electronic fetal monitoring technologies that have made it easier to detect or anticipate fetal distress during labor, and to respond preemptively by opting for a surgical delivery. Technological improvements may therefore have contributed to the rise in C-section rates in many provinces in the early 2000s [Chaillet et al., 2007],[27] however they are unlikely to be the source of the discrete change in C-section rates observed in many provinces in the early 2000s.

Finally, in figure 2.4 we report the trends in some other leading traditional medical indicators for C-sections. The increase in the rate of pregnant women with a previous history of C-section over time is a mechanical consequence of the rise in the cesarean section rate. The rate of repeat C-section is around 77 percent at the national level in our data, which can partly be explained by the fact that previous C-sections increase the risks of uterine rupture and placenta previa in subsequent pregnancies [Shearer, 1993, ACOG, 2013, NICE, 2014].[28] The trends in the proportion of multiple pregnancies are relatively flat in many of the provinces. In addition, maternal age increased steadily in all provinces between 1994-95 and 2010-11. In 2008, nearly a third of all first time mothers were older than 34 [Leon et al., 2011]. Finally, we observe a discrete decrease in the proportion of assisted deliveries in some provinces with similar timing to the increase in C-sections in figure 2.2.

[26] Daviss et al. [2010] report that while the study did have impacts on protocols for breech presentation and their rates of C-section in Canada, they were not uniform across the country.

[27] The Society of Obstetricians and Gynaecologists of Canada suggests that recourse to intensive electronic fetal monitoring may result in more instances of false positive diagnoses of fetal distress (mostly hypoxemia) leading to higher rates of cesarean sections [Chaillet et al., 2007].

[28] However, we note that various policy statements emitted by the Society of Obstetricians and Gynaecologists of Canada (SOGC) during the period covered by our study rejected the idea that pregnancies following a cesarean section should automatically lead to subsequent surgical deliveries, as described in Appendix B.

2.6 Econometric model and empirical strategy

Our main estimating equation is:

$$CSection_{ipth} = \alpha + \beta RF_{pt} + Z'pt\gamma + X'_{ipth}\phi + \delta_t + \mu_h + \epsilon_{ipth} \tag{2.1}$$

where $CSection$ is a binary variable indicating if a birth i in province p, fiscal year t and hospital h was delivered by C-section, RF is the ratio of C-section to vaginal delivery fees for the physician in charge, Z is a vector of time varying characteristics at the provincial level, X is a vector of time-varying characteristics at the birth delivery level and δ and μ are fiscal year and hospital fixed effects respectively.[29] Standard errors are clustered at the hospital level, and our results are robust to clustering at the province-fiscal year level.

Equation (2.1) includes an extensive set of maternal health conditions and characteristics related to the pregnancy, included in the set of variables X, specific to each observation. These include a set of dummy variables for maternal age, indicator for post– and pre-term delivery, complications during delivery, admission on a weekend, and other relevant health conditions of the mother. As discussed in section 2.5, some of these characteristics, in addition to having an impact on the choice of a birth delivery method, have followed specific trends throughout the period considered.

The empirical model also includes a vector of observable province time-varying characteristics, Z, including the general level of education of women in the population as well as their unemployment rate, changes in the fertility rate, the gender ratio among physicians, a measure of real GDP per capita and an indicator variable for the regulated status of midwives as health professionals in the province. These characteristics are added to capture the impact of specific features in the environment of physicians that might affect their practice conditions and their capacity to induce a demand for C-sections among their patients.[30]

2.7 Baseline Results

Estimates of equation (2.1) for our main sample of births are presented in the first four columns of table 2.4. Column 1 presents the results controlling for a limited set of observables (fixed effects for province and fiscal year of birth, and interaction terms between a dummy taking a value of one if the physician is an OB/GYN and an indicator for each province having different reimbursement rates for GPs and OB/GYNs). The estimated response to fees is positive and statistically significant. In the

[29] We also include a set of interaction terms between a dummy variable taking a value of one if the physician is an OB/GYN and a dummy variable for each province, taking a value of one if OB/GYNs and GPs are compensated at different rates. Those allow us to estimate responses within—rather than across—physician type.

[30] In certain specifications of the model, we also added variables related to immigration flows and to the amount of resources available to physicians within the health system (supply of registered nurses, percentage of hospital budgets spent on administrative expenses or capital expenditures, etc.). The results were insensitive to the addition of these variables. All province-level characteristics come from Statistics Canada, CIHI and the Canadian Association of Midwives.

second and third columns, we add additional characteristics of the birth and the province levels controls. The full set of additional controls reduces the magnitude of the estimate of β, which remains positive and statistically significant. The fourth column adds a full set of hospital fixed effects. Our estimates using the full set of controls and hospital fixed effects imply that raising the fee of a C-section by 100 percent relative to the baseline fee of a vaginal delivery would increase the probability that a physician opts for a surgical delivery by 3.9 percentage points. Given the average fee ratio of 1.20 between 1994-95 and 2010-11, our estimates of physicians' response to fee incentives explain 3.2 percent of the C-sections in our sample (24.1% of all births).

Including additional controls for the month or quarter of birth to take into account the patterns related to the seasonality of births leave our results unchanged.[31] We obtain modestly larger marginal effects when we use the full sample of births adding all cases of non-vertex presentations, previous C-sections and multiple births, cases for which medical factors could play an important role.[32] Finally, our results are also robust to measuring financial incentives as the difference (in levels) in the fees paid to physicians for each delivery method, in constant 2002 dollars, as shown in the first four columns of table 2.A.3 in the appendix.

These results are consistent with the conclusions of Clemens and Gottlieb [2014], who suggest that, when they have limited capacity to adjust the volume of care, physicians respond to fee incentives by adjusting the intensity of the care they provide. Prudence is of course required when making direct comparisons between our estimates and those from US data, given the difference in the magnitude of actual fee incentives observed in the two countries (Gruber et al. [1999] report a US$127 fee differential for Medicaid patients and a US$561 differential for privately insured patients). We nonetheless note that our results are in line with the positive and significant responses to financial incentives estimated on a Medicaid population by Gruber et al. [1999] and Alexander [2015] in the context of obstetric care, although somewhat larger. It is possible that the focus on samples of Medicaid births and of potential physician sorting across patients' insurance profiles in the American context contribute to estimating weaker responses. For example, physicians may be more likely to turn towards privately insured patients (associated with higher financial compensation for similar procedures) than to respond to financial incentives within the Medicaid population. In other words, physicians with a disposition to respond to financial incentives may self-select into a practice focusing on privately insured patients and paying more for equivalent services. Finally, Medicaid births may not be representative for other groups within the population. Meanwhile, the Canadian environment offers very little latitude for physicians to sort across remuneration schemes or patient groups, and our sample is for the full population of births.

The estimates on most of the birth characteristics have the expected associations with the probability of

[31] As an additional robustness check, we also limited our sample to births delivered between April and September of each fiscal year. Our results again remained unchanged—0.039 with a standard error of 0.013.

[32] In the specification with province controls and hospital fixed effects the estimates is 0.047 (standard error of 0.013). Controlling in this estimation for additional characteristics of the pregnancy such as malpresentation of the fetus (including breech presentation) and previous C-section, we confirm that they are positive and significant associates of C-sections, each increasing by half the chances of surgical delivery. Twinning is also a significant predictor of the delivery method, although its associated coefficient is smaller (possibly mechanically attenuated by the fact that in many cases, at least one of the fetuses is not in a vertex position, a characteristic captured by the malpresentation dummy variable).

C-section. Often a substitute to surgical delivery when complications arise in the second stage of labor, the estimated coefficient for operative delivery (the use of forceps or vacuum assisted delivery) is negative and statistically significant. The coefficient estimates associated with the full set of age dummies, not reported in the table for brevity, indicate an increasing probability of C-section with age. This is firstly the mechanical implication of maternal age on potential complications in prepartum and intrapartum care, for example in the form of dystocia [Leon et al., 2011]. Since a subset of intrapartum complications cannot be precisely accounted for as individual control variables, the age dummies could be picking up some of this impact. Secondly, the increasing coefficients associated with the age dummies could reflect the fact that physicians could have more facility to induce a demand for C-sections among older patients. This would be the case if, for example, the fear of labor complications increased with age. Knowing that they are categorized as *higher-risk* pregnancies, mothers aged 35 or more might overestimate the risks associated with a vaginal delivery and respond more positively when offered a C-section. In the same spirit, Jensen [2014] suggests that maternal age is positively correlated with patients' demands for more intense obstetric care. To investigate this potential channel, we estimated a variant of equation (2.1) that adds an interaction term between the fee variable and a dummy variable indicating if the patient is at least 35 years old. The results support the argument that a physician can use more discretion when mothers are older.[33]

We next exploit the fact that we observe physician identifiers on most discharges for the period 2001 through 2010[34] to investigate the robustness of our results to controlling for individual physician effects. These might be important if we suspect dynamic sorting of physicians practicing obstetrics with changes in the fee incentive variable. For example, physicians more sensitive to fee incentives may be responsible for a higher proportion of births when C-section fees are relatively high. As we include these fixed effects, our estimates are identified by within-physician responses to the changes in the fees. We first note that the coefficient associated with the fee incentive when our baseline specification (without physician fixed effects) is estimated on the subsample corresponding to observations for which a physician identifier is available is twice as large (0.081) and less precise than the estimate reported in column 4 of table 2.4. As we include physician fixed effects, the estimated response to fee incentives increases further, as shown in column 5 of table 2.4. This suggests that the results presented in columns 1 to 4 are not driven by more price sensitive doctors entering the market when the relative payment for C-sections increases. Moreover, the results in column 5 support the idea that the response to price incentives of existing physicians doing deliveries (within-physician variation) is stronger than the price response from physicians choosing to enter the market and delivering.

We finally amend our model to allow the effects of all covariates to vary by physician specialty. In Canada, birth deliveries can be performed by GPs, although obstetrics is only one (often marginal) com-

[33] We have also estimated equation (2.1) with a full set of controls (including province characteristics and hospital fixed effects) for subsamples defined by mothers' age. In the sample of mothers aged 35 and older the estimate for the fee ratio is 0.053 (standard error of 0.019), while in the sample of mothers less than 35 years old the estimate is 0.036 (standard error of 0.013).

[34] For the excluded years, 1994-2000, we have physician identifiers for only 65 percent of the potential sample, and lose all observations for two provinces (Manitoba and PEI). Physician identifiers on 83 percent of the potential sample for the period 2001 to 2010, although not for the data from Quebec. Also, the reporting of physician identifiers became mandatory in five provinces starting in 2001 (and in a sixth one from 2005 onwards) and so any missing identifiers are more likely to be random in later years.

ponent of their varied practice. They have the option to refer their patients to an OB/GYN, for example for higher risk pregnancies or if they decide to opt out of obstetric practice completely. In contrast, OB/GYNs' practice generally focuses more heavily on intrapartum care and deliveries. Therefore, we might expect that any discretionary contribution of the fee incentive to the choice of delivery methods would be larger for OB/GYNs, who derive a larger proportion of their incomes from birth procedures.[35] The results presented in column 6 of table 2.4 reveal that the fee responses of GPs and OB/GYNs are statistically distinct. As hypothesized, it is OB/GYNs who exhibit the greater sensitivity to the changes in the relative fee of a C-section, a result that is robust to excluding all births transferred from a GP to an OB/GYN during the delivery process. We find similar results when we estimate models separately for GP and OB/GYN deliveries (as shown in appendix table 2.A.4).

2.8 Accounting for the effects of the Hannah Term Breech Trial

We next investigate the robustness of our results to the introduction of province-specific time trends to the baseline specification. The resulting estimate for the fee variable, presented in the first row of table 2.5, is roughly one-quarter the magnitude of the estimate in column 4 of table 2.4, and no longer significantly different from zero. There is some precedence for this outcome in the US literature (e.g., Gruber et al. [1999], Grant [2009]).

Jurisdiction-specific trends are thought to control for dynamic unobserved factors that are heterogeneous at the provincial level, and that are correlated with both the dependent variable and the main explanatory variable of interest. As noted above, one easily observable manifestation of provincial level heterogeneity is the discrete trend break in the C-section rates after the release of the Term Breech Trial results. In previous specifications, the year effects accommodate any common impact of the Trial.

In the second panel of table 2.5 we more appropriately accommodate the discrete impact of this trial by interacting a dummy for the years after the Term Breech Trial report with the province specific trends. This dummy is coded for the years 2001 and later reflecting the fact that the publication of the results of the Trial was in October 2000, and in Canada a statement by the executive and council of the Society of Obstetricians and Gynaecologists of the results for breech deliveries was published in March 2001 [SOGC, 2001]. This modification has a significant impact on the estimate for the fee variable. It is now roughly three-quarters its baseline level and again statistically significant. Also, the interactions between the trends and the Term Breech Trial dummy are jointly statistically significant.[36] We observe similar changes in the estimates, and restoration of the baseline inference, when we specify the fee variable as the dollar gap in fees paid for C-section and vaginal deliveries (see columns 5 and 6 in appendix table 2.A.3).

[35] In 2002-03, the average annual payment for obstetric care was $82 000 for OB/GYNs and $12 000 for GPs [CIHI, 2006].

[36] The joint significance test is associated with a F statistic of 5.51.

While this modification of the province specific trends mostly restores our original inference, it is not obvious that it is appropriate. At least two outstanding questions are 1) why the Term Breech Trial would have an impact on the C-section rate for vertex births, and 2) whether any impact of the Trial results would differ across the provinces.

We begin our investigation of the first question in table 2.6 where we report the estimates of the year effects from our baseline specification corresponding to the columns in table 2.4. We add a column "0" which are the estimates of the year fixed effects and province fixed effects with no other added control variables. A striking pattern in these results is that in all columns, even when controlling for a full set of birth and province characteristics, and for hospital fixed effects, the year effects increase significantly starting in 2000. They keep increasing until the end of 2004, and remain relatively stable from 2005 on. We cannot reject the hypotheses that most of the estimates either pre 2000, and all the estimates post 2004 are jointly equal.[37] These results suggest that a secular increase in C-sections could be observed from 2000 to 2004 that cannot be accounted for by the changing values of the control variables.

This "unexplained" increase in the C-section rate is nominally correlated with the fallout of the Term Breech Trial. As noted above, the original research trial was published in October 2000 and the SOCG published new recommendations for breech births founded on this research in March 2001. That said, the apparent impact of this research on the C-section rate for vertex births might be a coincidence of the Canadian case.

However, evidence from other countries suggests this is not a Canadian coincidence. In figure 2.5 we report the growth rate of C-section deliveries in Australia separately for vertex and breech births between 1995 and 2005. The underlying data are population records of term singleton births taken from table 1 in Sullivan et al. [2009]. In 2001, there is a very clear impact of the Term Breech Trial on the growth of C-section deliveries for breech births, but there is also a corresponding spike in the growth rate of C-section deliveries for vertex births. The Royal Australian and New Zealand College of Obstetricians and Gynaecologists issued a statement in February of 2001 stating the higher risk in planned vaginal breech deliveries than in planned C-section deliveries for breech births as per the Trial results (Sullivan et al. [2009] p. 458).

In figure 2.6 we examine the rates of growth of C-section deliveries in the US. To construct this figure, we use data from the public-use US microdata natality files, constructed by the National Center for Health Statistics using information abstracted from all birth certificates registered in the 50 US states, the district of Columbia between 1995 and 2005. Here we see a (albeit smaller) spike in the growth of C-sections for breech births in 2001 and a corresponding spike in the growth of C-sections for vertex births between 2000 and 2003.

[37]For the period prior to 2000, we cannot reject the hypothesis that all year effects are jointly equal to zero (F statistic of 1.38 and a p-value of 0.24). We also cannot reject that the year effects for 2005 to 2010 inclusively are jointly equal (F statistic of 1.44 and p-value of 0.21).

An interesting feature of the US data is the significant deceleration in the growth rate of C-sections for breech births starting in 2005. A proximate cause of a decline in the growth rate for breech births is the publication of follow up results and commentary on the Term Breech Trial research in this period. A follow up study published in 2004 (Whyte et al. 2004) found no difference in the mortality or neurodevelopment of children at selected Trial centres by vaginal or C-section delivery. An observational study of births in France and Belgium found little difference in neonatal outcomes between breech births delivered by C-section and planned vaginal delivery [Goffinet et al., 2006]. Finally, a series of articles (e.g., Hauth and Cunningham [2002], Kotaska [2004], Glezerman [2006]) questioned the methodological soundness of the Trial and the external validity of the results. In June 2006 ACOG officially modified its recommendation for the delivery of breech presentations [ACOG, 2006].[38] The deceleration of the growth in C-sections for breech births in figure 2.6 is such that the level of these deliveries actually falls in our data by over 10 percentage points between 2004 and 2007. Note also, we observe a deceleration in the growth of C-section delivery for vertex births starting in this same year.[39]

Finally, in figure 2.7 we present C-section growth rates for Canada in comparable form to the graphs for Australia and the US. Very similar patterns are observed. Between 1999—the year before the Trial was published—and 2003, the C-section rate for breech deliveries rises from 63 percent to 77 percent. Relative to the patterns for the other two counties the response is more "spread out" possibly reflecting the sub-national organization of health care in Canada. Also, C-section rates in Australia and the US were already in excess of 80 percent before the publication of the Trial. In Canada, the rates were sub 65 percent so this research possibly led to a greater break from current practices in many health care facilities. Also importantly we observe the corresponding spikes in the growth rate of C-section for vertex births.

The second question is whether we would expect any effects of the Term Breech Trial to play out differently across provinces. Figures 2.5 through 2.7 certainly demonstrate some heterogeneity across countries. In some countries, the adoption of the Trial results was very quick. Rietberg et al. [2005] report that in the Netherlands the C-section rate for breech births rose from 50 percent to 80 percent within two months of the study's publication. In contrast, in Denmark, the rise in the rate of C-section while substantial (near 80 percent up to over 94 percent) was more gradual spreading into 2002 and 2003 [Hartnack et al., 2011].

As noted above, while there are national associations of various medical specialties in Canada, health services are organized and funded at the provincial level and doctors are governed by provincial level practice associations. There is also one small study (Daviss et al. [2010]) investigating the impact of the Term Breech Trial at Canadian health care facilities of which we are aware. The results suggest some notable heterogeneity in the responses of the maternity centres surveyed. For example, 11 of 20 centres, amongst the largest in the country, reported that it did not become required protocol to perform C-sections for breech presentations after the publication of the study. In three of the eight centers where

[38] The new recommendation deferred to the clinician's specific experience.

[39] Extending the data set until 2010, the growth rate in C-sections for breech presentations is negative until 2007. Over this same interval, the growth rate in the C-section rate for vertex births falls from 5.5 percent in 2004 to 1.5 percent in 2008.

it did become protocol, there were qualifications. Similarly, reactions to the follow up commentary on the Trial results and external validity was mixed.[40] In sum, the structure of health care in Canada and the evidence from this study certainly support our documented heterogeneous response to the Trial across provinces.

The impact of the Trial publication on the C-section rates for vertex births has not to our knowledge been previously pointed out in the literature. While beyond the scope of this study, the mechanisms that facilitated this effect are important topics for future research. Of course, there are likely common factors that have, over the years, affected the rates of C-sections for both vertex and non-vertex births. However, the spikes in the growth rate of C-sections for breech births in figures 2.5 through 2.7, so clearly associated with the Term Breech Trial results, suggest that a development related specifically to deliver of breech births had spillover effects on the delivery of vertex births.

One possible mechanism for such an effect is that an official recommendation of C-section, albeit for breech presentations, went some way to alleviate any concerns about the safety of the procedure. Of course, a recommendation of C-section delivery for breech presentations is implicitly a tradeoff of relative risks, but may have been interpreted to condone this method of delivery of other situations that pose risk to the mother or baby.

Another possibility is that the recommendation of C-section delivery for breech births necessitated the provision of staff and physical infrastructure that could not be rationalized based on the incidence of breech presentations alone. Therefore, these resources became ready and available to facilitate the advancement of C-sections for other births.

The publication and adoption of the Term Breech Trial results led to discrete shifts in the growth of C-section for both non-vertex and vertex deliveries. Accommodating this shift in our specification of province specific growth rates, as we implicitly do using year effects in our base specification, leads to estimates on the fee variables that are broadly consistent with the estimates in the base specification. A notable conclusion of our investigation of this issue is that such a relationship between the growth of C-sections for vertex and non-vertex births exists in other countries, and that the Trial could account for a large fraction of the increase in the Canadian C-section rate over the period considered.

[40] Daviss et al. [2010] report that by the Fall of 2004, 35% of the hospitals in their sample changed their internal practices (or had the intention to do so) in response to the publication of studies assessing the safety of vaginal breech birth, while 55% had no intention to make any changes. In contrast, 65% of hospitals mentioned having seen a marked increase in the rate of C-section for breech presentations in the aftermath of the publication of the Trial in 2000, 30% having seen some increase and 5% having seen no increase during the same period.

2.9 Conclusion

In this paper, we use unique features of the Canadian health care system— universal, single-payer, fee-for-service remuneration—to estimate the impact of financial incentives on physicians' choice of C-section delivery. Our empirical approach rests on an institutional context that allows us to estimate responses that should be free of the self-selection bias that may arise when physicians can sort across remuneration agreements and patient insurance types. Also, unlike past studies, the Canadian case allows us to understand how financial incentives influence physician behavior at a population level.

Using information from nearly 4 million birthing events over 17 years, we find that doubling the payment received by a physician for a C-section relative to a vaginal delivery raises the probability of surgical delivery for a birth episode by at most four percent, all else equal. This estimate is robust to numerous robustness checks, and is mainly driven by OB/GYNs, for whom the choice of birth delivery method represents an important margin along which to adjust in order to increase total remuneration.

Our results initially exhibit some sensitivity to the specification of jurisdiction specific trends. However, the sensitivity is mostly attenuated when we allow province specific responses to the publication of the Term Breech Trial. In investigating the propriety of this modification, we discover an impact of the Trial on the growth of C-section of vertex births which is also evident in Australia and the US. The contribution of the breech trial to the overall growth rate in C-sections appears to be much larger than its proportional direct impact on the delivery of non-vertex births, or the contribution of fee incentives. This point has not, to our knowledge, been made previously in the literature, and its reasons are an important topic for future research.

As C-sections represent, by far, the main cause of surgery in many countries, these findings constitute an insightful basis for policy initiatives that would seek to improve the efficiency in the allocation of resources within the health care system and the broader impacts of dissemination of new best-practices. Moreover, our results indicate that other factors such as improvement in fetal monitoring technology, delayed motherhood and preferences for delivery methods are likely to have also contributed to the increase in C-section rates over the last decades, since fee incentives can only partially explain this rise.

Table 2.1: **Share of total OB/GYN payments in fee-for-service**

	2006	**2007**	**2008**	**2009**	**2010**
Newfoundland & Labrador	63.0	62.6	60.0	61.9	56.6
Prince Edward Island	52.9	40.6	47.9	47.7	46.6
Nova Scotia	n/a	n/a	n/a	n/a	n/a
New Brunswick	79.8	81.0	77.4	79.0	69.3
Quebec	88.8	88.3	87.3	88.3	89.3
Ontario	90.5	90.7	84.8	87.7	87.9
Manitoba	89.2	88.2	88.2	87.4	87.0
Saskatchewan	75.3	75.3	76.9	76.4	72.7
Alberta	n/a	n/a	n/a	n/a	n/a
British Columbia	n/a	n/a	n/a	n/a	n/a

Source: National Physician Database, Canadian Institute for Health Information.
Notes: These payments represent all activities related to obstetric and gynaecologic care; they are not limited to birth deliveries. n/a indicates provinces and years for which the information is not available.

Table 2.2: **Total observations for fiscal years 1994 to 2010, by province**

	All births			**Vertex singletons, excluding (cases of previous C-sections)**		
	Births	Vaginal	C-section	Births	Vaginal	C-section
Newfoundland & Labrador	76 800	55 959	20 841	64 070	54 023	10 047
Prince Edward Island	17 955	12 856	5 099	14 636	12 149	2 487
Nova Scotia	145 369	110 022	35 347	118 113	102 180	15 933
New Brunswick	128 190	95 325	32 865	106 261	90 552	15 709
Quebec	329 029	252 757	76 272	272 421	240 288	32 133
Ontario	2 327 572	1 774 929	552 643	1 904 899	1 656 104	248 795
Manitoba	250 762	204 278	46 484	208 188	187 692	20 496
Saskatchewan	164 570	131 372	33 198	132 968	118 038	14 930
Alberta	387 239	289 403	97 836	312 577	267 039	45 538
British Columbia	709 564	520 066	188 898	574 378	483 498	90 880
Canada	4 537 050	3 447 567	1 089 4983	3 708 511	3 211 563	496 948

Notes: Observations from Quebec are available for the fiscal years 2006-2006 to 2009-2010. Price information for Alberta is available for the fiscal years 1997-98 and 2001-02 to 2010-2011. The sample restricted to mothers between 14 and 65 years of age, whose birth was assisted by a physician (general practitioner or specialist).

Table 2.3: **Means of main variables by delivery methods (vertex singleton births, excluding cases of previous C-section)**

	Vaginal	**C-section**	**Total**
Maternal age	28.54	29.39	28.65
Maternal age > 40 years	0.021	0.036	0.023
Maternal age > 35 years	0.145	0.191	0.151
Post-secondary education	0.506	0.514	0.507
School < 8 years	0.020	0.018	0.019
Married	0.666	0.671	0.667
Length of stay (days)	2.242	4.422	2.534
Stay > 1 month	0.000	0.004	0.001
Pre-term delivery	0.089	0.113	0.092
Post-term delivery	0.107	0.178	0.116
Operative delivery	0.141	0.009	0.124
Induced labor	0.207	0.294	0.219
Week-end delivery	0.248	0.218	0.244
Observations	3 211 563	496 948	3 708 511

Notes: All differences in means are significant at the 5% level. Marital status and highest level of schooling achieved are obtained at the province-year level from Statistics Canada. Observations from Quebec are available for the fiscal years 2006-07 to 2009-10. The years 1994-95 to 1996-97 and 1998-99 to 2000-2001 from Alberta are excluded from the analysis because of price data availability reasons. The sample is restricted to mothers between 14 and 65 years of age, whose birth was assisted by a physician (general practitioner or specialist) and who did not correspond to cases of non-vertex presentation, multiple pregnancy or previous C-section birth.

Table 2.4: **Impact of financial incentives on C-section births, main estimates**

	Baseline Specification				**Physician FE**	**Specialty**
	(1)	(2)	(3)	(4)	(5)	(6)
Fee incentive (ratio of fees CS/VD)	0.054***	0.049***	0.036***	0.039***	0.117***	0.003
	(0.008)	(0.010)	(0.010)	(0.013)	(0.022)	(0.008)
OB/GYN						0.232***
						(0.087)
Fee incentive × OB/GYN						0.040**
						(0.019)
Pre-term		0.023***	0.023***	0.023***	0.014***	
		(0.002)	(0.003)	(0.002)	(0.001)	
Post-term		0.065***	0.065***	0.068***	0.076***	
		(0.002)	(0.002)	(0.002)	(0.001)	
Operative delivery		-0.146***	-0.146***	-0.150***	-0.224***	
		(0.006)	(0.006)	(0.004)	(0.004)	
Induced labor		0.037***	0.037***	0.036***	0.010***	
		(0.002)	(0.002)	(0.002)	(0.001)	
OB/GYN × Pre-term						0.014***
						(0.002)
OB/GYN × Post-term						0.068***
						(0.002)
OB/GYN × Operative delivery						-0.179***
						(0.009)
OB/GYN × Induced labour						0.023***
						(0.002)
Province controls			✓	✓	✓	✓
Hospital fixed effects				✓		✓
Physician fixed effects					✓	
Observations	3 708 511	3 708 511	3 708 511	3 708 511	2 004 172	3 708 511

Notes: All specifications include a full set of age dummy variables, province fixed effects and year fixed effects, as well as interactions between a dummy variable for OB/GYN and an indicator for each province having different reimbursement rates for GPs and OB/GYNs. Province controls include: women's unemployment rate, proportion of women with a university degree, proportion of women without a high school degree, proportion of women physicians, population growth and a dummy variable for legislation over midwifery. Cases of non-vertex presentation, previous C-section and multiple pregnancy are excluded from the sample. Results in column 5 are obtained on a sample of observations between years 2001 to 2010 for which physician identifiers are available. Standard errors, in parenthesis, are clustered at the province-year level for columns 1, 2, 3, at the hospital level for columns 4 and 6, and at the physician level for column 5. Significance levels: *** = 1%; ** = 5% and * = 10%.

Table 2.5: **Impact of financial incentives on C-section births, including province-specific trends**

	Fee Ratio (CS/VD)
Panel A	
Province specific time trends	0.010
	(0.014)
Panel B	
Province specific time trends with interaction post	0.028**
Hannah term breech trial	(0.014)
Birth controls	✓
Province controls	✓
Hospital fixed effects	✓
Observations	3 708 511

Notes: All specifications include the full set of province fixed effects and year fixed effects, interactions between a dummy variable for OB/GYN and an indicator for each province having different reimbursement rates for GPs and OB/GYNs, as well as the birth- and province-level controls used in columns 2 to 4 of table 4. Cases of non-vertex presentation, previous C-section and multiple pregnancy are excluded from the sample. Standard errors, in parenthesis, are clustered at the hospital level. Significance levels: *** = 1%; ** = 5% and * = 10%.

Table 2.6: Impact of the Term Breech Trial on C-section births, estimates of the year effects

	(0)	(1)	(2)	(3)	(4)
Year FE: 1995	0.003	0.003**	0.001	0.003*	0.002
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)
Year FE: 1996	0.007**	0.008***	0.005**	0.009***	0.007*
	(0.003)	(0.001)	(0.002)	(0.002)	(0.003)
Year FE: 1997	0.007	0.005	0.002	0.011**	0.007*
	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)
Year FE: 1998	0.012***	0.010***	0.007*	0.015***	0.010*
	(0.003)	(0.003)	(0.004)	(0.005)	(0.006)
Year FE: 1999	0.018***	0.016***	0.010***	0.020***	0.014*
	(0.003)	(0.003)	(0.003)	(0.007)	(0.008)
Year FE: 2000	0.028***	0.028***	0.020***	0.033***	0.025***
	(0.003)	(0.002)	(0.003)	(0.007)	(0.010)
Year FE: 2001	0.033***	0.033***	0.016***	0.032***	0.023**
	(0.004)	(0.003)	(0.004)	(0.008)	(0.011)
Year FE: 2002	0.049***	0.049***	0.035***	0.051***	0.041***
	(0.003)	(0.002)	(0.002)	(0.008)	(0.011)
Year FE: 2003	0.060***	0.059***	0.045***	0.062***	0.051***
	(0.002)	(0.001)	(0.002)	(0.009)	(0.012)
Year FE: 2004	0.065***	0.065***	0.051***	0.067***	0.056***
	(0.002)	(0.002)	(0.002)	(0.009)	(0.013)
Year FE: 2005	0.070***	0.069***	0.055***	0.075***	0.063***
	(0.002)	(0.001)	(0.002)	(0.011)	(0.015)
Year FE: 2006	0.072***	0.071***	0.057***	0.079***	0.066**
	(0.002)	(0.001)	(0.002)	(0.011)	(0.016)
Year FE: 2007	0.073***	0.073***	0.058***	0.081***	0.067***
	(0.003)	(0.002)	(0.002)	(0.013)	(0.017)
Year FE: 2008	0.074***	0.073***	0.058***	0.084***	0.068***
	(0.002)	(0.002)	(0.002)	(0.013)	(0.019)
Year FE: 2009	0.075***	0.073***	0.055***	0.080***	0.065***
	(0.003)	(0.001)	(0.002)	(0.014)	(0.020)
Year FE: 2010	0.074***	0.073***	0.053***	0.080***	0.063***
	(0.004)	(0.003)	(0.003)	(0.015)	(0.021)
Fee incentive		✓	✓	✓	✓
Birth controls			✓	✓	✓
Province controls				✓	✓
Hospital fixed effects					✓
Observations	3 708 511	3 708 511	3 708 511	3 708 511	3 708 511

Notes: Columns 1 to 4 include a full set of province fixed effects and year fixed effects, as well as interactions between a dummy variable for OB/GYN and an indicator for each province having different reimbursement rates for GPs and OB/GYNs. Birth controls and province controls correspond to those reported for columns 2 to 4 in table 4. Cases of non-vertex presentation, previous C-section and multiple pregnancy are excluded from the sample. Standard errors, in parenthesis, are clustered at the province-year level for columns 0 to 3, and at the hospital level for column 4. Significance levels: *** = 1%; ** = 5% and * = 10%.

Figure 2.1: Evolution of C-section rates, selected countries for 1995-2012

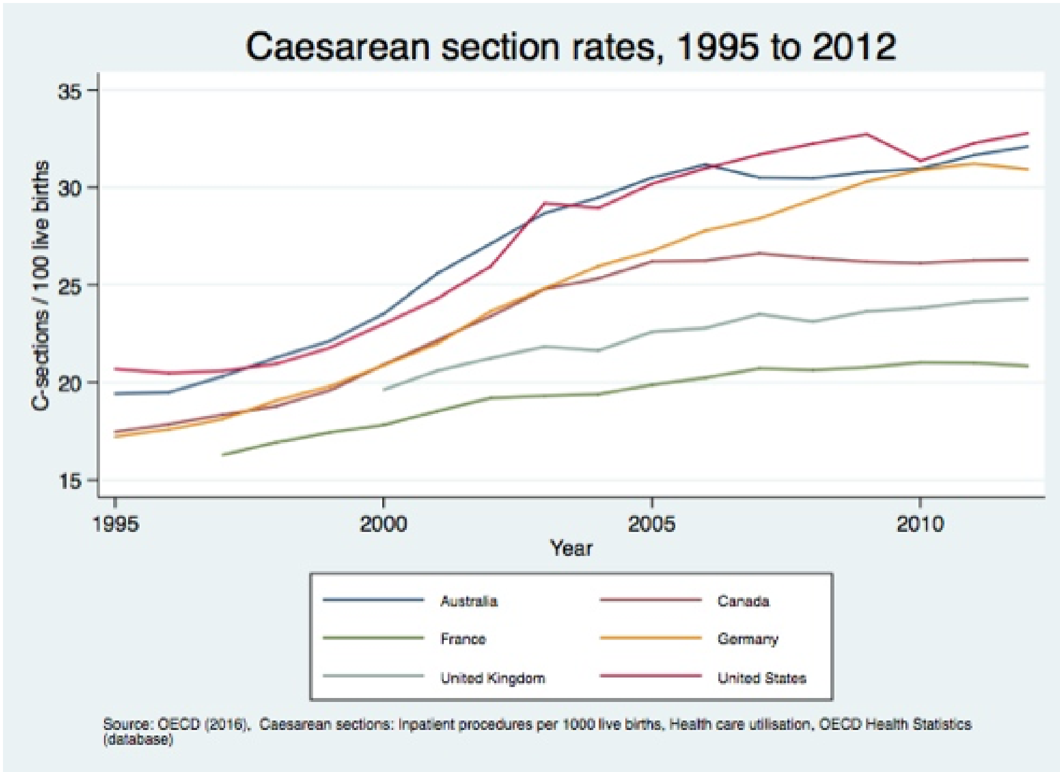


Figure 2.2: Evolution of caesarean section rates and relative fees, 1994-2010

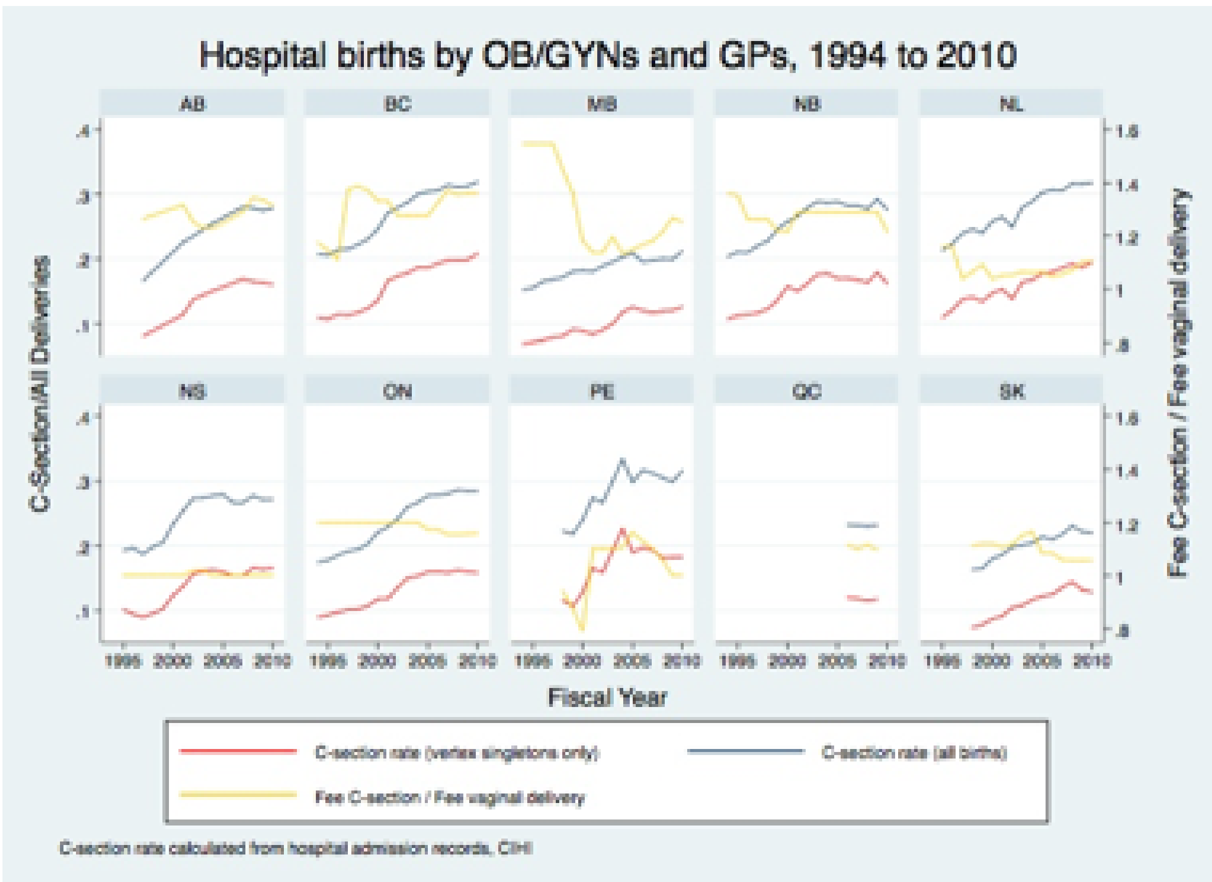
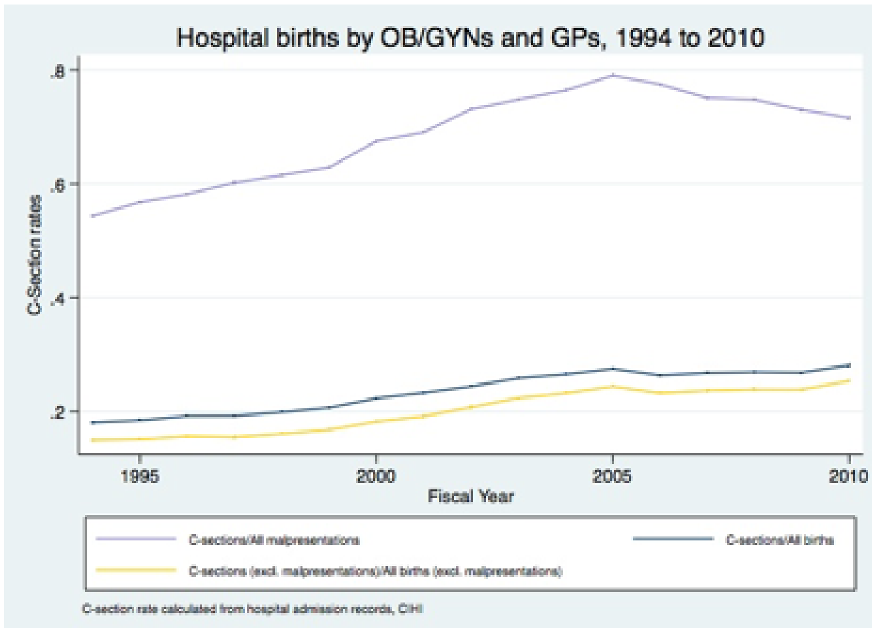


Figure 2.3: Evolution caesarean sections for cases of non-vertex births, 1994-2010

Panel A: Canada



Panel B: Provinces

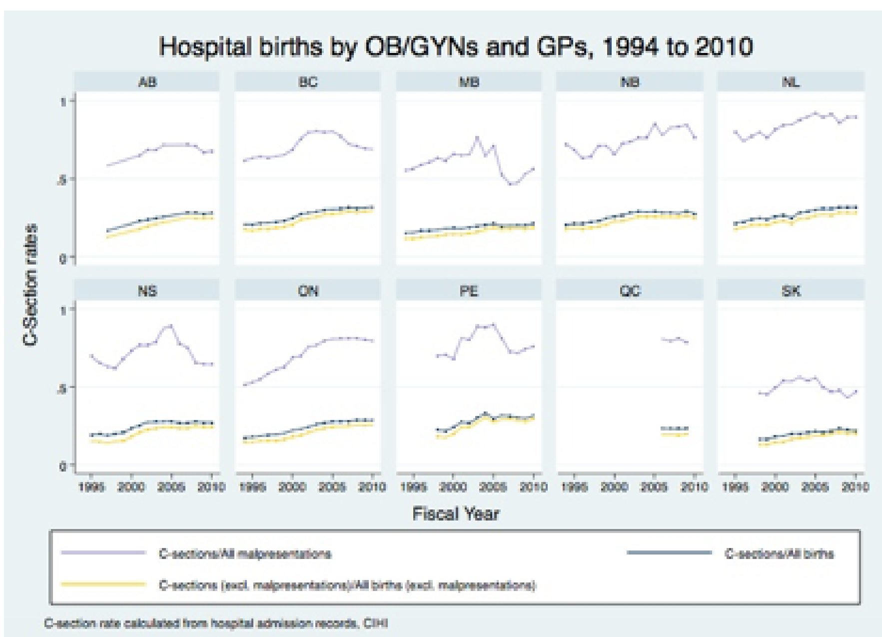


Figure 2.4: Evolution of selected mothers and pregnancy characteristics, 1994-2010

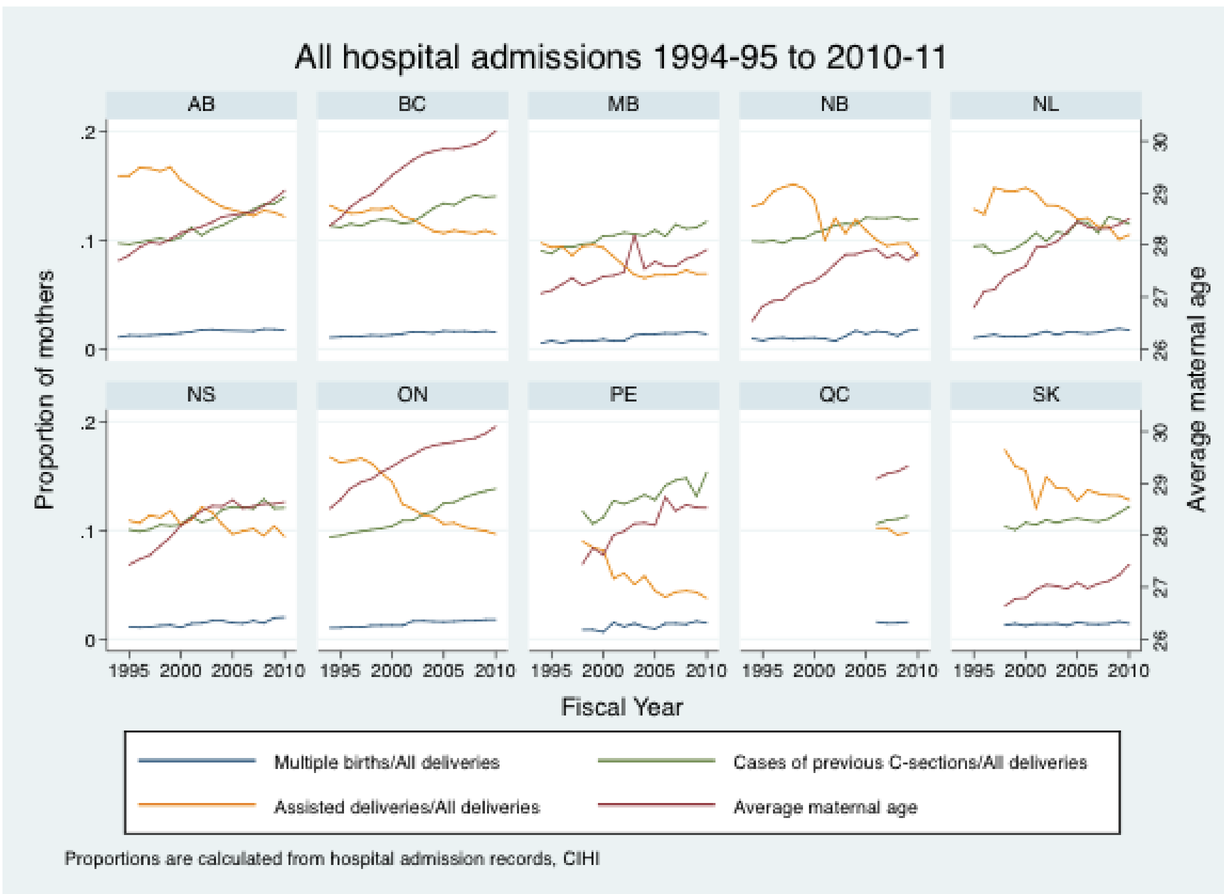


Figure 2.5: Growth in C-section rates between 1995 and 2005 in Australia, by presentation type

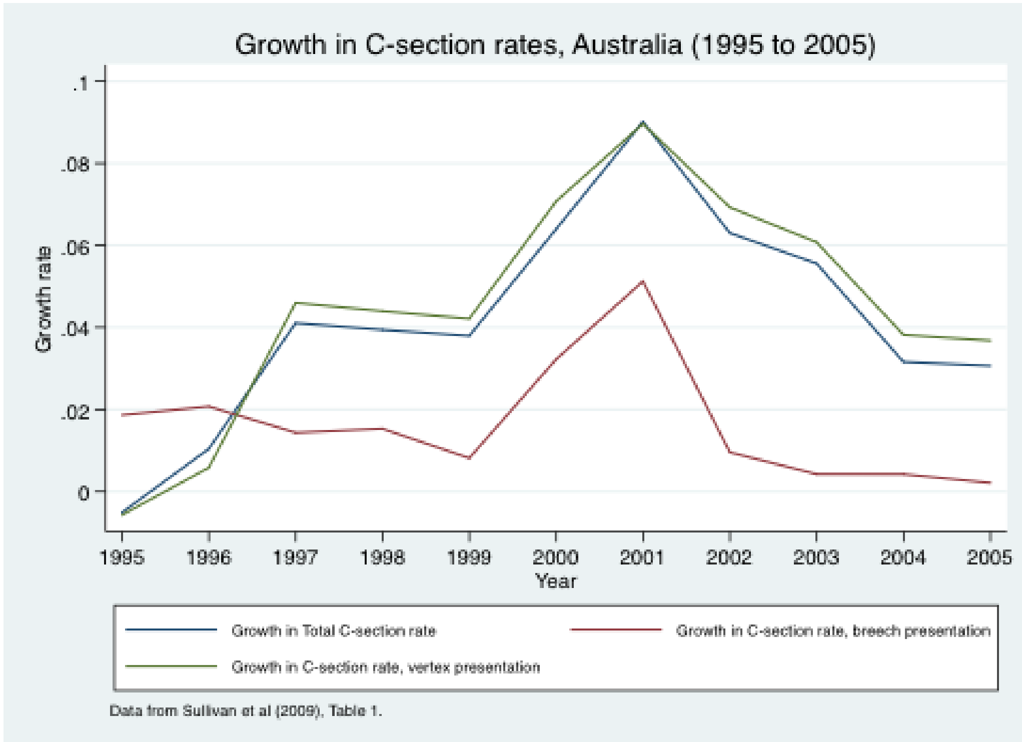


Figure 2.6: Growth in C-section rates between 1995 and 2005 in the US, by presentation type

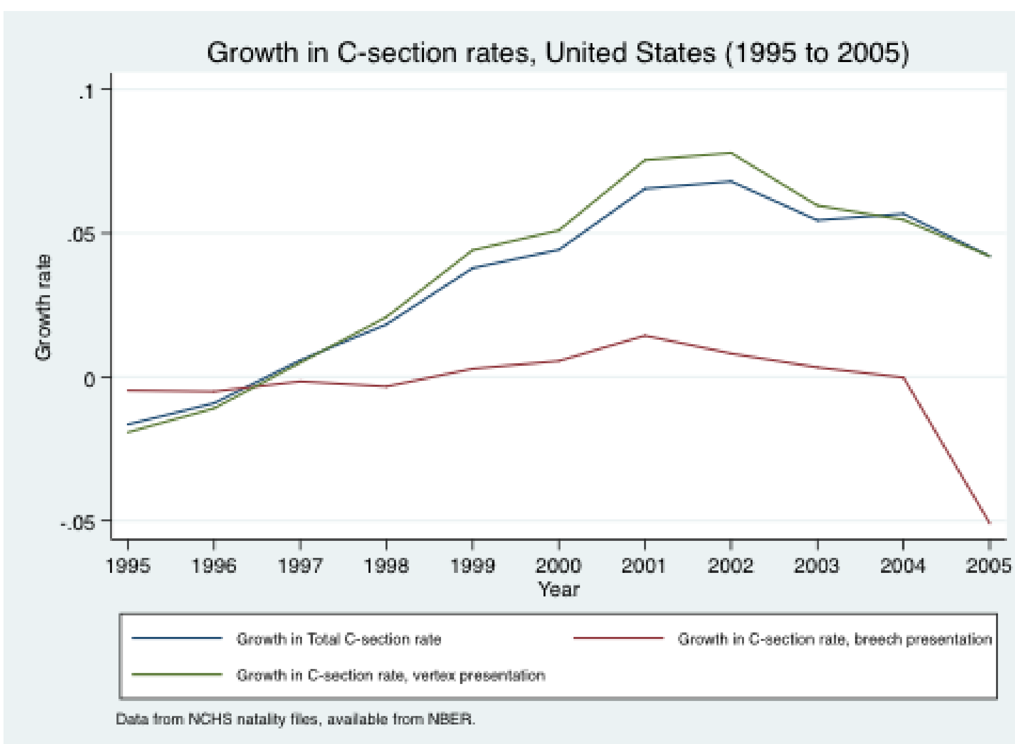
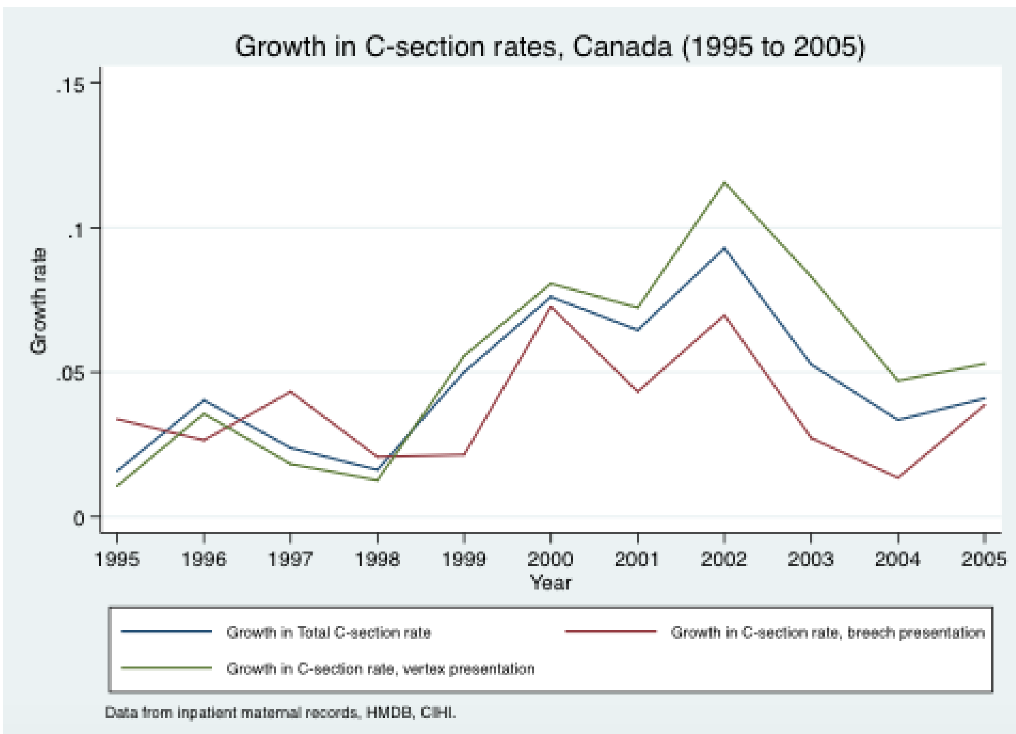


Figure 2.7: Growth in C-section rates between 1995 and 2005 in Canada, by presentation type



Appendix

Additional tables

Table 2.A.1: Annual C-section rates (%) for fiscal years 1994 to 2010, by Canadian province

	All births				**Vertex singleton births, excluding cases of previous C-section**			
	Mean	(s.d.)	Minimum	Maximum	Mean	(s.d.)	Minimum	Maximum
Newfoundland & Labrador	27.14	(3.95)	21.36	31.80	15.68	(2.64)	10.96	19.50
Prince Edward Island	28.40	(3.68)	21.85	33.31	16.99	(3.42)	10.69	22.61
Nova Scotia	24.32	(3.58)	18.73	27.96	13.49	(2.94)	9.01	16.58
New Brunswick	25.64	(25.64)	20.37	29.23	14.78	(2.56)	10.72	17.90
Quebec	23.18	(0.23)	23.04	23.29	11.80	(0.17)	11.56	12.03
Ontario	23.74	(4.18)	17.46	28.62	13.06	(2.76)	8.97	16.25
Manitoba	18.54	(1.82)	15.25	21.25	9.84	(2.06)	6.89	12.61
Saskatchewan	20.17	(2.10)	16.36	23.06	11.23	(2.20)	7.41	14.39
Alberta	25.27	(3.36)	16.65	28.04	14.57	(2.63)	8.07	16.88
British Columbia	26.62	(4.27)	20.60	31.92	15.82	(3.75)	10.72	20.84
Canada	**24.01**	**(4.23)**	**18.11**	**28.15**	**13.40**	**(2.50)**	**9.30**	**16.38**

Notes: Observations from Quebec for the fiscal years 1994-1995 to 2005-2006 and for the fiscal year 2010-11 as well as observations from Alberta for the fiscal years 2005-2006 and 2006-2007 are excluded from the analysis for data availability reasons. The sample restricted to mothers between 14 and 65 years of age, whose birth was assisted by a physician (general practitioner or specialist).

Table 2.A.2: Baseline results with additional controls for physicians' work environment

	Baseline results	**Additional controls**	
	(1)	(2)	(3)
Fee ratio (CS/VD)	0.039***	0.025**	0.030**
	(0.013)	(0.012)	(0.012)
Preterm birth	0.023***	0.023***	0.023***
	(0.002)	(0.002)	(0.002)
Post term birth	0.068***	0.065***	0.068***
	(0.002)	(0.002)	(0.002)
Operative delivery	-0.150***	-0.146***	-0.150***
	(0.004)	(0.006)	(0.004)
Induced labor	0.036***	0.037***	0.036***
	(0.002)	(0.002)	(0.002)
Registered midwives		-0.001	-0.001
		(0.005)	(0.004)
Gov expenditure in health		-0.073	-0.059
care facilities ($M)		(0.055)	(0.062)
Changes in population growth		-0.003	-0.003
		(0.002)	(0.003)
GDP per capita ($K)		0.003	0.003
		(0.002)	(0.002)
Gov investment in capital in		0.029	0.067
health care facilities ($M)		(0.024)	(0.444)
Province controls	✓	✓	✓
Hospital fixed effects	✓		✓
Observations	3 708 511	3 708 511	3 708 511

Notes: All specifications include a full set of age dummy variables, province fixed effects and year fixed effects, as well as interactions between a dummy variable for OB/GYN and an indicator for each province having different reimbursement rates for GPs and OB/GYNs. Province controls include: women's unemployment rate, proportion of women with a university degree, proportion of women without a high school degree, population growth and a dummy variable for legislation over midwifery and the proportion of women physicians. Annual capital investment expenditures per province are taken from CIHI's National Expenditures Trends 1975-2013 [CIHI, 2014]. The expenditures considered are exclusively those made by the public sector on construction, machinery, equipment, and software in hospitals. Population counts to generate expenditures per capita are from Statistics Canada's Demography Division. Cases of non-vertex, previous C-section and multiple pregnancy are excluded from the sample. Standard errors, in parenthesis, are clustered at the hospital level in columns 1 and 3, and at the province-year level in column 2. Significance levels: *** = 1%; ** = 5% and * = 10%.

Table 2.A.3: Base Estimates using fee differentials (dollars of 2002)

	Baseline specification				**Province-specific time trends**	
	(1)	(2)	(3)	(4)	(5)	(6)
Fee incentive	0.014***	0.013***	0.009***	0.010**	0.002	0.008**
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Pre-term		0.023***	0.023***	0.023***	0.022***	0.022***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Post-term		0.065***	0.065***	0.068***	0.068***	0.068***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Operative delivery		-0.146***	-0.146***	-0.150***	-0.151***	-0.151**
		(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Induced labor		0.037***	0.038***	0.036***	0.036***	0.036***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Province controls			✓	✓	✓	✓
Hospital fixed effects				✓	✓	✓
Province specific time trends					✓	
Province time trends w/ interaction post Hannah term breech trial						✓
Observations	3 708 511	3 708 511	3 708 511	3 708 511	3 708 511	3 708 511

Notes: All specifications include a full set of age dummy variables, province fixed effects and year fixed effects, as well as interactions between a dummy variable for OB/GYN and an indicator for each province having different reimbursement rates for GPs and OB/GYNs. Province controls include: women's unemployment rate, proportion of women with a university degree, proportion of women without a high school degree, proportion of women physicians, population growth and a dummy variable for legislation over midwifery. Cases of non-vertex presentation, previous C-section and multiple pregnancy are excluded from the sample. Standard errors, in parenthesis, are clustered at the province-year level for columns 1, 2, 3, 5, 6, 7 and at the hospital level for columns 4 and 8. Significance levels: *** = 1%; ** = 5% and * = 10%.

Table 2.A.4: Main estimates by physician specialy

	Births by GPs (1)	Births by OB/GYNs (2)	All births (3)
Fee ratio (CS/VD)	0.002	0.042**	0.003
	(0.007)	(0.016)	(0.008)
OB/GYN x Fee ratio			0.040**
			(0.019)
Pre-term	0.003***	0.020***	0.005***
	(0.001)	(0.002)	(0.001)
Post-term	0.011***	0.081***	0.015***
	(0.002)	(0.003)	(0.001)
Operative delivery	-0.021***	-0.203***	-0.025***
	(0.003)	(0.008)	(0.002)
Induced labor	0.006***	0.026***	0.004**
	(0.001)	(0.004)	(0.002)
OB/GYN x Pre-term			0.014***
			(0.002)
OB/GYN x Post-term			0.068***
			(0.002)
OB/GYN x Operative delivery			-0.179***
			(0.009)
OB/GYN x Induced labour			0.023***
			(0.002)
Province controls	✓	✓	✓
All interactions with OB/GYN dummy			✓
Hospital fixed effects	✓	✓	✓
Observations	980 362	2 728 149	3 708 511

Notes: All specifications include a full set of age dummy variables, province fixed effects and year fixed effects, a dummy for PB/GYN as well as interactions between a dummy variable for OB/GYN and an indicator for each province having different reimbursement rates for GPs and OB/GYNs. Province controls include: women's unemployment rate, proportion of women with a university degree, proportion of women without a high school degree, population growth and a dummy variable for legislation over midwifery and the proportion of women physicians. Results are robust to adding as controls the provincial GP to OB/GYN ratio. Column 3 includes all controls interacted with a dummy variable for OB/GYN status. Cases of non-vertex presentation, previous C-section and multiple pregnancy are excluded from the sample. Standard errors, in parenthesis, are clustered at the hospital level. Significance levels: ***= 1%; **= 5% and *= 10%.

Notes on practice guidelines, policy statements and consensus statements by the Society of Obstetricians and Gynaecologists of Canada

Concerned by the high C-section rates observed across the country in the mid 1980s, the Canadian medical community organized the *1985 National Consensus Conference on Aspects of Caesarean Birth* to propose guidelines on the safest birth delivery methods for mothers' and infants' health, while limiting recourse to unnecessary surgery [SOGC, 1986]. Despite this initiative, C-section rates in Canada remained high and started rising steadily in the mid 1990s and through the 2000s. Trying to curb this trend, the SOGC published a series of practical guidelines and policy statement to delimit the circumstances in which C-sections should be performed, most of which are described the list below.

In 2004, the SOGC reiterated that no evidence supported the idea that C-sections are less risky for mothers and newborns than vaginal deliveries. The association also highlighted that surgical deliveries are associated with risks (maternal and fetal respiratory distress, hemorrhage, lacerations, longer inpatient stay, complications such as placental insufficiency and uterine rupture in subsequent pregnancies, etc.) that should not be overlooked [SOGC, 2004a, Leon et al., 2011]. These guidelines are aligned with most of the American and European medical literature listing the risks and advantages of surgical deliveries compared to vaginal deliveries and documenting the circumstances under which C-sections should be considered.[41]

List of selected policy guidelines and statements by the SOGC, 1994-95 to 2010-11

- **Policy Statement on the Role of Obstetricians and Gynaecologists (November 1995):** Acknowledges the greater share of obstetric care undertaken by OB/GYNs, as other types of physicians gradually left that type of practice. [SOGC, 1995]

- **Policy Statement on Midwifery (February 1997):** Supports the recognition of midwifery as a licensed profession providing child delivery services for uncomplicated pregnancies. Mentions the SOGC's opposition to home births and deliveries in birthing centers, and encourages the integration of midwives in health service teams in which they would have access to health care resources and act under the direction of the physician in charge of the clinical department of obstetrics. [SOGC, 1997a]

- **Policy Statement on Vaginal Birth After Previous Caesarean Birth (December 1997):** Rejects the idea that pregnancies following a cesarean section should automatically lead to a repeat cesarean section. Suggests that all obstetric care facilities should be able to provide women with the opportunity to attempt a vaginal delivery after a cesarean birth, while recognizing the importance

[41] A few studies explore why these medical guidelines discouraging the recourse to unnecessary C-sections have had so little influence on physicians' choice of delivery method. In a randomized controlled trial led in Ontario, Lomas et al. [1991] look at the impact of medical opinion on the management of birth delivery for mothers with a previous C-section. They estimate that the production and diffusion of guidelines, audits and feedback reports are insufficient to increase the recourse to trial of labor. They however find that attempts of vaginal birth after C-section could increase by 46 to 85 percent if physicians were exposed to those guidelines in the form of an opinion from influential practitioners, rather than through written reports. Goodrick and Salancik [1996] also argue that recourse to unnecessary C-sections would increase in the absence of institutional accountability regarding the choice of medical interventions or in the presence of contradicting opinions and guidelines, which would generate grey zones facilitating the use of physicians' discretionary power.

of patient choice: "Full participation of the patient in these decisions is vital". [SOGC, 1997b]

- **Management of Twin Pregnancy (July 2000):**This consensus statement specifies the consensus on the conditions (mono amniotic twins and conjoint twins) under which a cesarean section is indicated before any trial of labor for twin pregnancies. [SOGC, 2000]

- **Guideline for Operative Vaginal Birth (August 2004):** Replaces Practice Guideline no15 by explicitly positioning c-sections and operative 52 deliveries as substitutes in case of a complicated second stage of labor. The choice between both procedures by the physician is to be guided by pregnancy-specific characteristics, without any general indication in favor of one versus the other. SOGC [2004b]

- **Joint Policy Statement on Normal Childbirth (December 2008):** Defines the concept of normal birth and mentions that a "vaginal birth following a normal pregnancy is safer for mother and child than a Caesarean section". States that "there should be a valid reason (evidence-based practice) to intervene in the natural process when labor and birth are progressing normally" and that "All pregnant and birthing women and their families should be able to make informed choices. All candidates for normal birth should be encouraged to pursue it". Discourages the practice of elective c-sections without medical indications: "A Caesarean section should not be offered to a pregnant woman when there is no obstetrical indication" SOGC et al. [2008]

- **Policy Statement on the Role of Obstetricians and Gynaecologists (July 2009):** (Update to SOGC [1995]) Acknowledges the greater share of obstetric care undertaken by OB/GYNs as other types of physicians leave that type of practice. This trend led OB/GYNs to undertake more primary care activities, because the demand for consultations by OB/GYNs is, in certain regions, too low to maintain "the minimum number of obstetricians/gynaecologists (usually three or more) required to provide 24-hour consultant access". [SOGC, 2009]

- **Policy Statement on Midwifery (July 2009):** Recommends more collaboration between care providers (including midwives) in obstetrics. Supports better access for midwives to health care facilities and resources, more information to patients as to the birth delivery options and their respective risks and benefits, as well as harmonization of obstetrical standards for care across health professionals. [SOGC and CAM, 2009]

Chapter 3

From C-section to Health Conditions: Are Children's Health Outcomes Influenced by Birth Delivery Methods?

3.1 Introduction

Across most OECD countries, C-section rates have risen substantially over the past decades. In the United States and in Canada, they have reached levels nearly twice as high as the benchmark suggested by the World Health Organization. The short-term impacts of such a trend have been heavily documented. In response to legitimate medical needs, birth by C-section can reduce risks of morbidity and even mortality for the mother, the infant or both. For example, the American College of Obstetricians and Gynecologists (ACOG) and the Society for Maternal-Fetal Medicine (SMFM) affirm that C-section is *"firmly established as the safest route of delivery"* for a series of conditions such as placenta previa and uterine rupture [Caughey et al., 2016]. The Society of Obstetricians and Gynaecologists of Canada (SOGC) and the ACOG also report that C-sections can improve the outcomes for cases of fetal macrosomia, maternal infections (such as HIV and herpes), and that they can be safer when certain complications such as abnormal or indeterminate fetal heart rate tracing, respiratory distress or arrests of labour, occur during the delivery [Caughey et al., 2016, Society of Obstetricians and Gynaecologists of Canada, 2009].[1] For most low-risk pregnancies, however, cesareans have not been shown to improve birth outcomes. In certain cases, research has rather suggested that they pose a greater risk of maternal infection, respiratory distress, haemorrhage requiring hysterectomy or transfusion, uterine rupture or complications from anaesthesia [Liu et al., 2007, Caughey et al., 2016].[2] In addition to these potential complications, unnecessary C-sections are generally associated with heftier monetary costs per patient,

[1] The relative benefits of C-section births for malpresented fetuses or multiple pregnancies are still a subject of debate in the research community and among practitioners.

[2] In its obstetric care consensus statement *Safe Prevention of the Primary Cesarean delivery*, ACOG states that "for most pregnancies, who are low-risk, cesarean delivery appears to pose greater risk of maternal morbidity and mortality than vaginal delivery." [Caughey et al., 2016]

as they often result in longer hospital stays (and generally involve longer recoveries) and require a more intensive use of resources on average than vaginal births.

While the immediate costs, risks and benefits of C-section births have been extensively documented, less is known about their consequences in the long-run. In addition to increased risks of complications for breastfeeding and in subsequent pregnancies, one of the potential impacts that has captured the attention of both the media and the scientific community is the repercussions that the procedure may have on children's propensity to develop various health conditions later in life. Interest in the association between the two phenomena has been fuelled by recent findings from clinical studies suggesting that the early development of newborns' immune system can be influenced by the choice of a birth delivery method. Previous studies have sought to investigate the persistence of these clinical findings as children age. While some have found correlations among children populations between C-section births and certain ailments such as asthma, allergies, diabetes and obesity, mixed evidence and the possibility for various confounders to contaminate the results from most observational studies have kept the literature from making direct statements of causality. Therefore, more evidence seems to be needed to develop a better understanding of the extent to which different birth delivery methods influence the risks of chronic conditions among children, independently and at a population level.

The urge to better understand the factors leading to the development of chronic conditions in childhood has been fuelled by the substantial increase in the prevalence of most of these conditions in many developed countries over the past decades. In the United States, data from the National Health Interview Survey suggests that the percentage of children suffering from asthma nearly doubled between 1980 and 1995, and has remained at historically high levels of 8% to 9% since the early 2000s [Center for Disease Control and Prevention, 2016b]. Over the past decades, U.S. obesity rates for children under 19 have also reached record levels nearing 17% between 2003 and 2011 [Odgen et al., 2014][3] and substantial increases have been observed in the rates of diabetes [Dabelea et al., 2014] and food and skin allergies [Jackson et al., 2013]. The prevalence of such conditions in the U.S. and elsewhere across the OECD[4] has exerted significant pressure on healthcare systems from a resource utilization and financing perspective. For example, the U.S. National Center for Health Statistics reports that in 2008, asthma was the third main contributing factor for hospitalizations among American children below the age of 15. Estimates of the annual medical expenditures (in terms of direct payment for care, including out-of-pocket payments, payments from a private/public insurer or from other sources) associated with a diagnosis of asthma for a child in the United States range from $531 [Wang et al., 2005] to $1377 [Miller et al., 2016].[5] Research also suggests that chronic conditions in youth come at a cost in terms of human capital accumulation. Both in Canada and in the United States, asthma has been identified as a leading cause of school absenteeism, and some empirical evidence has suggested that chronic conditions can lead to lower

[3] Although among children aged 2 to 5, this proportion has decreased during the same period.

[4] In Canada, the proportion of children aged 11 or less diagnosed with asthma is estimated to have increased from 11% in 1994-95 to 13% in 2000-2001 [Garner and Kohen, 2008]. After a period of rapid increase in the 1990s, obesity rates for children have remained relatively stable between 2004 and 2013, at approximately 13% [Rodd and Sharma, 2016] and type 1 diabetes rates for children under 14 years old are believed to have increased at an annual rate of 5 percent throughout the 1990s [Center for Chronic Disease Prevention and Control, 2011].

[5] Both studies used data from the Medical Expenditure Panel Survey. Miller et al. [2016] also estimate the additional annual costs of diabetes for a child to be $6702, and their estimate of the costs associated with a diagnosed food allergy is as high as $1044 annually, although not statistically significant at the 5% level.

academic achievement, although studies realized in the past twenty years do offer mixed results [Kohen, 2010, Center for Disease Control and Prevention, 2016a, Crump et al., 2013]. Hence, if C-section birth has a *causal* impact on the likelihood to develop a chronic condition in childhood, the long-term costs associated with that birth delivery method could extend well beyond the procedure's higher immediate price tag (both from a public health perspective and from an individual's perspective). Such a causal relationship would also suggest that reducing the rate of unnecessary C-sections may be a channel through which the monetary and non-monetary costs associated with certain health conditions could be reduced.

This paper adds to the body of literature investigating the impact of C-sections on childhood chronic conditions in two ways. First, it uses longitudinal data on 13 cohorts of Canadian children to estimate the impact of C-section birth on a series of health outcomes ranging from respiratory conditions to general healthcare utilization and obesity. It does so while controlling for confounders that have not simultaneously been accounted for in previous work, exploiting extensive information on children's socioeconomic characteristics, family background, on the characteristics related to their mother's pregnancy and on environmental triggers associated with poor health. Second, to address the potential remaining endogeneity of C-section as a birth delivery method, it suggests an instrumental variable framework building on previous research documenting physicians' response to exogenous changes in financial incentives in obstetric care [Allin et al., 2015, Gruber et al., 1999, Alexander, 2017]. More specifically, it instruments the probability that a birth is delivered by C-section using changes in the compensation received by physicians for a C-section relative to a vaginal delivery, which varies through time across and within Canadian provinces. The local average treatment effect estimated using this instrumental variable approach informs on the impact of an *unnecessary* C-section on a child's health outcomes – or to be more precise, on the long-term health impacts of C-section birth for children who were born by cesarean as a result of physicians' response to financial incentive, but who would otherwise have been delivered vaginally.

The results from both estimation strategies support, to a certain extent, the thesis of an association between C-section birth and increased health risks later in life. First, the results from a simple linear model suggest a modest yet statistically significant positive relationship between cesarean birth and the probability that a child is overweight, is diagnosed with asthma or experiences frequent episodes of wheezing. This relationship persists even when controlling for an extensive set of potential confounders. The estimated impact of C-section birth on some measure of healthcare utilization, such as the regular use of prescription medication, is also positive and statistically significant. No impact is however found when looking at ear and nose infections for children aged 0 to 3, or when considering heart conditions, kidney conditions, chronic bronchitis, and diagnoses of hyperactivity. The results also fail to reveal an association between a child's birth delivery method and a health index synthesizing a comprehensive set of health outcomes.[6] Turning to the suggested instrumental variable framework yields much noisier results. Despite this lack of precision, the two-stage least squares results and the associated reduced form estimates still point in the direction of a causal relationship between C-section birth and the regular use of prescription medication.

[6]The Health Utility Index was developed at McMaster University and synthesizes an individual's functional health based on eight dimensions.

The remainder of the paper is divided as follows. Section 3.2 presents recent studies investigating the relationship between cesarean birth and health outcomes later in childhood, and documents some of the issues limiting the ability to get at an estimated causal relationship. The main empirical frameworks are developed in section 3.3 and the data is described in section 3.4. Section 3.5 presents the main results and section 3.6 proposes a discussion of the findings and outlines potential avenues for future research. Finally, an alternative two-sample two-stage least squares estimation using more than three million birth records and clinically measured health outcomes is briefly presented in the appendix.

3.2 Literature Review

The relationship between C-sections and children's health has been the focus of an important body of research in, among other fields, epidemiology and immunology. One direction taken by clinical studies is to investigate the relationship between birth delivery methods and the diversity of infants' gut microbiota. Indeed, the composition of intestinal microbiota is believed to be of primary importance for the development of an infants' immune system, training it to recognize good bacteria from harmful ones to stop the propagation of the latter and stimulate the production of antibodies [Neu and Rushing, 2011, Maynard et al., 2012].

In that line, Azad et al. [2013] use DNA sequencing techniques to study the entire community of bacteria colonizing the gut microbiota of 24 infants', four months after their birth. They find that babies born by C-section without trial of labour had a significantly less rich and diverse microbiota compared to babies born vaginally or by emergency C-section (potentially because of a partial exposure to their mothers' vaginal bacteria during a trial of labour, or because of different antibiotic intake). Using a longitudinal sample of 198 children followed in their first year of life, Azad et al. [2015] further document that C-sections may have an impact on the composition newborns' gut microbiota in the first months of life through the administration of intrapartum antibiotic prophylaxis. Overall, differences in the initial microbiota of the newborns have been shown to persist through early infancy, potentially because of a delayed colonization of the intestine by certain bacteria, and to affect some immune and nutritional functions. For instance, Abrahamsson et al. [2012] show that the composition of gut microbiota in the month following birth was related to the likelihood of atopic eczema in children aged one. Using gene pyrosequencing techniques to analyze the composition of the bacteria of 6 infants less than 24 hours after delivery, Dominguez-Bello et al. [2010] find that the impacts of C-section birth on the composition of the microbiota extend beyond the gut. They suggest that delivery mode is *"the primary determinant of a newborn's bacterial community composition"*.

Several channels are thought to contribute to the documented association between birth delivery methods and the initial colonization of infants' microbiome [Cho and Norman, 2013, Young, 2012]. First, the most important portion of a newborn's intestinal tract colonization by bacteria is believed to happen during the labour and delivery process. Children born vaginally would be exposed to a richer variety of *pioneer* bacteria species (from the mother's vaginal canal and intestinal tract) whereas for children born by C-section, these maternal bacteria would be replaced by others that are more similar to those found on the skin or in the hospital environment. The latter would be associated with increased vulnerability to pathogens and to different responses and functions of the immune system. Reviewing the relationship between the intestinal microbiota and the development of the immune system, Maynard et al. [2012] suggest that the transmission of bacteria from mother-to-child at birth *"could be deterministic"* in terms of predisposition to asthma and allergies during childhood. Second, intake of broad spectrum antibiotics by mothers during and after the delivery process is more frequent when a C-section is performed. As these antibiotics are believed to permeate the placental barrier, they could negatively affect the diversity in the infant's intestinal microbiota and hinder the development of his immune system. Third, labour would stimulate the production of stress hormones that could prepare and stimulate the infant's immune system activity, a benefit that would not be experienced by children delivered by elective (planned) C-section. Finally, both the American College of Obstetricians and Gynaecologists and the Society of Obstetricians and Gynaecologists of Canada list breastfeeding complications as one of the risks associated with a C-section delivery. Since exclusive breastfeeding has been shown to influence the composition of newborns' microbiota, C-section births could then also indirectly influence health outcomes through the channel of infant diet.

Most clinical studies using gene-sequencing techniques, culture-based methods or comparing blood immune biomarkers provide some important insights on the mechanisms through which birth delivery methods may affect early immune system responses and influence the risk of developing health ailments. In a review of the literature, Young [2012] suggests that anomalies in the gut microbiota could be associated with a certain number of chronic conditions and atopic diseases. However, most existing work has focused on short-run outcomes, most often three to twelve months after birth, and do not generally provide much information on the persistence of the effects they measure.[7] Moreover, most of these analyses are conducted on very small samples, making it impossible to control for an extended range of confounders or to stratify results across different infant characteristics (for example, by parity).

In order to address those caveats and to better understand the long-term ramifications of the aforementioned associations, other studies have turned to observational datasets offering information on large samples of children over several years. Using information on roughly twenty-two thousand children born between 1982-1987 and 1990-1995 who were part of the GUTS cohort study, Yuan et al. [2016] find a positive association between C-section birth and the risk of obesity in childhood. They document that this relationship would persist until early adulthood and would survive as ethnicity, maternal and pregnancy characteristics, and geographic indicators are controlled for. A recent large-scale cohort study conducted by Black et al. [2015] uses Scottish administrative data on 321,287 children born between 1993 and 2007 to study the impact of planned C-section births on a series of health conditions.[8] They find a modest impact on the probability that a child develops asthma requiring hospital admissions or gets prescribed an inhaler by age 5 for respiratory problems. However, they do not identify a relationship between delivery method and obesity, inflammatory bowel disease, cancer or type 1 diabetes. Although they empirically control for maternal age, an index of deprivation, gestational age, birth weight, smoking during pregnancy and breastfeeding practices when the infant was 6 weeks old, they cannot give a causal

[7] One exception is Salminen et al. [2004] who finds effects that last up to seven years after birth delivery.
[8] They use data from children born between 2004 and 2007 to study most of the health outcomes they consider.

interpretation to their results. Sevelsted et al. [2015] follow more than two million children born between 1977 and 2012, and also find C-section birth to be associated with risks of asthma, but not with type 1 diabetes. Additionally, and unlike Black et al. [2015], their results suggest that the mode of birth delivery may be associated with inflammatory bowel disease and leukaemia later in life. They also find a relationship between C-section delivery and systemic conductive tissue disorders, juvenile arthritis, psoriasis and celiac disease. However, they empirically control for fewer (and different) potential confounders, including gender, parity and maternal diagnosis for each health outcome of interest. Their analysis therefore leaves out most parental behaviour during and after pregnancy and other environmental factors that could simultaneously affect a child's health and the likelihood that he is delivered by C-section. Conversely, Menezes et al. [2011] include variables related to parents' general health and smoking behaviour, mother's level of education at birth and household assets at birth in their analysis of the impact of C-section birth on wheezing in childhood and adolescence. Using survey data on two cohorts of Brazilian children, they cannot identify a relationship between the two phenomena. Overall, surveys of the literature report a general positive association between C-section and asthma [Cho and Norman, 2013, Blustein and Liu, 2015].[9] and mixed findings for other conditions such as allergies, type 1 diabetes, gastrointestinal infections, cancer and inflammatory bowel disease [Cho and Norman, 2013, Bager et al., 2008, Marcotte et al., 2016]. Moreover, the results obtained in cohort studies do not always survive when turning to a siblings comparison design.[10] Hence, more work is needed to assess if birth delivery methods have a causal impact on health conditions later in childhood.

3.3 Empirical framework and Idenfitication

A first strategy to estimate the impact of C-section birth on any childhood health outcome, is to specify the following model:

$$Health_{ipc} = \alpha_0 + \alpha_1 Csection_{ipc} + \mathbf{X}'_{ipc}\boldsymbol{\alpha}_2 + \lambda_p + \gamma_c + \epsilon_{ipc} \tag{3.1}$$

In equation (3.1), $Health$ is a health outcome of interest for child i born in province p and fiscal year c, and can either be a 0/1 indicator or a continuous variable. $Csection$ is a binary variable indicating if a child was born by C-section, and $\mathbf{X}$ is a vector of characteristics for the child and his environment. λ_p and γ_c are respectively province and cohort of birth fixed effects. A concern arising from previous studies, and one that has limited researchers' capacity to make causal statements about the relationship between cesarean sections and children's health outcomes, is the potential endogeneity of $Csection$ in the estimating equation above. Indeed, a series of factors that are likely to influence the likelihood of a C-section delivery are also susceptible to directly influence the probability of impairing children's health later in life. Failure to control for these in $\mathbf{X}$ would yield a biased estimate of the parameter of interest, α_1. This bias could go in either direction. For example, preferences for elective cesareans, reasons

[9] Some studies have suggested that this impact may fade with age [Maitra et al., 2004], a finding also reported in siblings studies such as Braback et al. [2013] and Almqvist et al. [2012].

[10] Blustein and Liu [2015], however, highlights that the small samples used in siblings studies can challenge identification.

related to scheduling convenience, or delayed maternal age could increase the probability of C-section birth in high income households. However, wealthier families may be able to provide better preventive care and are more likely to live in in neighbourhoods with better amenities such as air quality [Chay and Greenstone, 2005, Currie and Walker, 2011], all of which could attenuate or compensate for the negative health impacts of a C-section birth. On the reverse, omitting to control for factors such as maternal health in $\mathbf{X}$ could introduce an upward bias in the estimated α_1. Poor maternal health is positively correlated with the risk of complications in their pregnancy and of C-section birth, but those health issues may also directly affect the child's own health, for example if they have a hereditary component.

One solution to reduce the risk of omitted variable bias in the estimation of α_1 in equation (3.1) is of course to include in $\mathbf{X}$ the potential confounders that would both be associated with the probability of C-section birth and with health issues later in childhood. An alternative to directly address the potential endogeneity of the C-section variable in equation (3.1) is to turn to an instrumental variable approach. Previous work on physicians' response to financial incentives in fee-for-service remuneration contexts provide interesting insights on the choice of an instrument that would exogenously shock the probability of a C-section birth. Gruber et al. [1999] and Alexander [2017] have shown that American physicians increase their recourse to C-sections when the relative payment they receive for this procedure rises compared to the payment received for vaginal deliveries, all else equal. Looking at the full population of birth deliveries in Canada between 1994 and 2010 and using across and within province variation in relative payments for cesareans, Allin et al. [2015] find similar evidence of physician responses to price incentives. The institutional context in Canada offers an interesting set of conditions to estimate physicians' response to financial incentives in obstetric care: except for a few rare exceptions, all birth deliveries are remunerated through fee-for-service agreements, and the fees associated with birth delivery methods are set administratively by provincial health authorities. In addition to varying across provinces, these fees evolve within each province through time. Results from the aforementioned studies suggest that such changes in physicians' remuneration parameters should exogenously affect the probability that a given birth is delivered by C-section. Since the relative physician payment for a cesarean should not influence a child's later health outcomes through any other channel than the birth delivery method chosen by the attending doctor, the exclusion restriction when using relative physician fees as an instrument should also be satisfied.[11]

Using the impact of financial incentive on the likelihood of C-section birth as a first stage, the relationship between a birth delivery method and children's health can be estimated with the two-stage least squares model summarized by equations 3.2 and 3.3.

$$Health_{ipc} = \delta_0 + \delta_1 Csection_{ipc} + \mathbf{X}'_{ipc}\boldsymbol{\delta}_2 + \mu_p + \eta_c + \xi_{ipc} \tag{3.2}$$

$$Csection_{ipc} = \beta_0 + Fee_{pc}\beta_1 + \mathbf{X}'_{ipc}\boldsymbol{\beta}_2 + \omega_p + \zeta_c + \upsilon_{ipc} \tag{3.3}$$

[11] One potential problem could arise if the movements in the relative price of a C-section was correlated with the movement in other unobserved factors increasing the likelihood of developing a chronic condition. This risk can be alleviated by adding a time trend or year fixed effects to the model.

Equation (3.3) replicates the main estimating equation from Allin et al. [2015] and represents the first stage in the suggested 2SLS framework. Fee_{pc} corresponds to the ratio of the fees paid for a C-section and for a vaginal delivery ($Fee_{p,c}$=$\frac{Fee\ C-section_{p,c}}{Fee\ vaginal_{p,c}}$), and captures the financial incentive to which physicians are exposed when choosing between a C-section and a vaginal delivery in province p and fiscal year c. All the other variables in the model are as described in the context of the simple linear regression model. Equation (3.2) is the second stage, which models the main relationship of interest.[12]

Under the assumption of homogenous effects, the estimated $\hat{\delta}_1^{2SLS}$ can be interpreted as an average treatment effect. However, with heterogenous effects in this instrumental variable framework, $\hat{\delta}_1^{2SLS}$ has a local average treatment effect (LATE) interpretation: it corresponds to the impact of C-section birth on later health outcomes among children for whom physician incentives led to a C-section birth, but who would otherwise have been delivered vaginally.[13] This LATE is of particular interest from a policy perspective as it informs on the long-term consequences of cesareans that could be medically avoided. While those seem to be contributing to the rise in total C-section rates in Canada and in the United States,[14] they are the ones for which long-term harmful consequences are most likely to outweigh the potential short term benefits.

Finally, the monotonicity assumption required for the local average treatment effect to be identified demands that no physician who would have delivered a birth by C-section in the absence of financial incentives refrains from opting for a cesarean when the relative pay for doing so increases. While it seems reasonable to think that some physicians may not respond to changes in financial incentives, for example because of clinical preferences, it seems unrealistic that some doctors would be behave in a way that violates the monotonicity assumption.

[12]The framework is presented as a system of linear equations for simplicity. The first stage (and potentially the second stage, depending on the nature of the health outcome considered) is estimated as a linear probability model in a two-stage least squares approach to estimate the parameter of interest $\hat{\delta}_1^{2SLS}$. Given the binary nature of the endogenous regressor and the possibility to specify $Health$ as a binary outcome as well, the system could be thought of in terms of non-linear relationships:

$$Health_{ipc} = f(Csection_{ipc}, X'_{ipc}, epsilon_{ipc})$$
$$Csection_{ipc} = g(Fee_{ipc}, X'_{ipc}, \upsilon_{ipc})$$

However, Angrist and Pischke [2009] warn against using such a specification, highlighting that a correlation between the first stage residuals, the fitted values for the main endogenous regressors and the other variables in the model could arise if the wrong non-linear functional form is chosen in the first stage. Since the estimation of the first stage equation yields relatively similar marginal effects regardless of the model chosen (logit, probit or linear probability), the linear model seems to approximate the relationship between Fee_{ipc} and $Csection_{ipc}$ quite well. The main results are therefore reported for the 2SLS framework in which each stage is estimated using a linear probability model.

[13]Wooldrige [1997] also suggests that if the heterogenous effects take the form of a random coefficient model (if, for instance, the true $\delta_1 = \bar{\delta}_1 + \delta_{1i}$), the 2SLS estimator can consistently estimate the average treatment effect when the endogenous regressor is binary if three conditions are satisfied: (i) the instrument satisfies the exclusion restriction, (ii) the random part of the coefficient of interest is independent of the instrument and its expectation can be expressed as a linear function of the unobserved factors in the first stage, and (iii) the instrument is relevant.

[14]Trends in C-section rates in both countries depart from the evolution of the common medical indications for surgical deliveries.

3.4 Data

The main source of data used in this paper is the National Longitudinal Survey of Children and Youth (NLSCY). The survey was administered between 1994 and 2009 by Statistics Canada, and sponsored by Human Resources and Skills Development Canada[15] with the objective of documenting the physical and psychosocial development of representative cohorts of Canadians through childhood and adolescence and until the beginning of their adult lives.[16] The first wave of data was collected in the winter/spring of 1994-95 and a new cycle of the survey was administered every two years until the eighth and final cycle, in 2008-09. The original longitudinal cohort of children, surveyed at ages 0-11 in the first cycle, was followed in all subsequent cycles (until they reached 14 to 25 years of age in cycle 8). New cohorts of children aged 0 to 1 were added in each wave and longitudinally followed in subsequent cycles to monitor early childhood development. These cohorts were continuously followed until age 5 if they were initially sampled in cycle 2, until age 7 if they were sampled in cycles 4 and 5, and until cycle 8 if they were initially sampled in later cycles. The cohort of children who were first sampled at the ages of 0-1 in cycle 3 was continuously followed until in cycle 5, and reintroduced for one last time in cycle 7 at the ages of 8-9. Top-up sample of children aged 5 were also surveyed in cycles 3 and 4, as were top-up samples of children aged 2 to 5 in each of cycles 6, 7 and 8.[17] Children were selected from households sampled for the Labour Force Survey, and exceptionally from Birth Registry when made necessary by sample size requirements.[18]

In each cycle of the survey, information on children aged 9 or younger was collected via household interviews. The person most knowledgeable (PMK) about the child in the household (most often a parent and the biological mother of the child) was asked an extensive series of questions on the child's health, motor and cognitive development, skills and behaviours.[19] Most importantly for this analysis, PMKs had to answer questions regarding the child's utilization of healthcare resources and services in the year preceding the interview (such as nights spent at a hospital, prescription medications regularly taken by the child, number of visits/contacts with a physician, nurse or dentist, etc.) and health conditions (illnesses, chronic conditions, mental health indicators, weight, height, etc.). A full list of the resulting health variables relevant to this study, along with the ages at which they are documented, is presented in table 3.1. This extensive set of information allows for various health outcomes to be considered using the empirical models described in section 3.3.

An important characteristics of the NLSCY with respect to the empirical investigation of the long-term health impacts of C-section birth is the rich set of characteristics about children's family and socioeconomic environment collected as part of the interview process. Including this information in $\mathbf{X}$ in equation 3.1 allows for the analysis to go beyond previous work by controlling for a wider range of environmental and clinical confounders, thus mitigating the risk of an omitted variable bias in the estimation of α_1. More specifically, the NLSCY provides information reported by the PMK on the composition and func-

[15]Human Resources and Skills Development Canada is now known as Employment and Social Development Canada.

[16]The cohorts are representative of the population formed by the ten Canadian provinces, without the territories.

[17]The top-up samples from cycles 6 and 7 were followed in subsequent cycles as well.

[18]It should be noted that the sampling design from cycle 3 onwards excluded siblings of children in the longitudinal cohort (twins could be included in pairs in cycles 2 to 4). This features limits the scope for within-family analysis to be conducted.

[19]Direct interviews were conducted with children over the age of 9.

tioning of the household and on their home and neighbourhood environments. Extensive information is also collected in each interview on both parents' health status (including diagnoses of asthma, diabetes, chronic bronchitis, heart or liver problems and other chronic conditions), on their income and on other socio-demographic characteristics of the family. A full set of pregnancy– and birth-related characteristics is also available, including information on twinning, birthweight, gestational age and maternal age at birth, a few health conditions of the mother during pregnancy, a few pregnancy behaviour such as smoking and the reception of prenatal care, as well as breastfeeding. Tomeo et al. [1999] show that maternal recall of events relating to child birth and pregnancy is generally accurate, even decades after they happened. Comparing data collected from mothers in Providence, Rhode Island several years after childbirth, they find that the information reported with regards to smoking during pregnancy, gestational age, weight, and complications such as breech birth closely matches the information provided in medical records.[20] Recall of cesarean section was perfect in their sample, but they note that mothers did omit to report certain pregnancy complications such as high blood pressure and gestational diabetes. Accurate reporting of such characteristics and conditions is more likely to happen in the NLSCY, since the first interviews are mostly conducted a year of less after the birth of sampled children. The results presented in section 3.5 are also robust to excluding the pregnancy characteristics that have been shown to be more prone to recall bias in Tomeo et al. [1999].

In order to estimate the model using the instrumental variable framework described above, data on physicians' remuneration for C-section deliveries and for vaginal deliveries (in the form of fee-for-service payments) is collected for the fiscal years 1994-95 to 2008-09 from the historical fee schedules published on an annual basis by provincial Health Ministries and physician associations.[21] The constructed ratio of these physician fees are appended to the NLSCY records for each child on the basis of their province and full date of birth. Unfortunately, the NLSCY does not provide precise information on a child's province of birth, and only documents the province residence at the time of each interview. Each child is therefore assigned the province of residence recorded in his first interview as his province of birth.[22]

3.4.1 Main estimating sample

The main estimating sample is constructed by appending all observations from the longitudinal and early childhood development cohorts across the 8 cycles of the NLSCY. Only observations corresponding to children aged 9 or less are kept, to avoid comparability issues between the answers collected from PMKs and those collected using the self-completed questionnaires for older children. The information on

[20]They collected information from mothers on average 32 years after childbirth, which is much more than the time elapsed between a child's birth and the first interview in which information on the pregnancy and childbirth is recorded in the NLSCY.

[21]Fees are missing for a few combinations of provinces and years: Quebec, for the period 1994-97 and 2000, Alberta for the periods 1994-96 and 1998-2000, Saskatchewan and Prince-Edward Island for the period 1994-97 and Nova-Scotia and Newfoundland and Labrador for the fiscal year 1994.

[22]This imputation strategy should not be particularly problematic: a child would have to move provinces between his birth and the time of his first interview, for the most cases in his first year of life. In the NLSCY sample of children born after March 1994, just a little more than 2% of all children are observed in more than one province before they reach the age of 10, suggesting that cross-province migration should also not be a common phenomenon for very young children. Nevertheless, any wrong imputation will in the end attenuate the first stage estimates in the two-stage least squares specification.

each child in the sample is linked across survey cycles using unique child identifiers. The sample is then restricted to children for whom a complete birth date is available and who were born in Canada after March 1994. This last restriction ensures that, in the instrumental variable framework, each child can be assigned a value for the relative C-section fee paid to physicians in their province and year of birth. The combination of these selection criteria reduce the full sample from 66 800 to 39 000 children.[23] After further restricting the sample to kids for whom a birth delivery method is listed and for whom most health outcome and control variables of interest are available, the final sample consists of 21,400 children.

Table 3.2 documents the size of each birth cohort in the main estimating sample, and gives the C-section rate for each of them. The relatively large size of certain birth cohorts is driven by changes in the size of the early childhood development cohorts selected in each cycle, including the addition of top-ups in certain cycles, as described above.[24] Although unweighted, the C-section rates reported in the NLSCY are also generally consistent with the overall rates reported for the population of Canadian births during the years covered. Tables 3.3 and 3.4 present some summary statistics for the children in the main estimating sample, by birth delivery method.[25]

3.4.2 Summary statistics

The characteristics reported in table 3.3 first confirm that children delivered vaginally are on average quite different from children born by C-section. In terms of their prenatal characteristics, the latter are on average more likely to be twins and to be born prematurely, which is coherent with the identification in the medical literature of those characteristics as risk factors associated with increased recourse to surgical delivery. It should be noted that these traits might also directly influence health outcomes later in life. Children's socioeconomic backgrounds (including parental education, economic conditions and the probability of residing in a rural environment) also differ on average across birth delivery methods. For example, children from rural regions are less represented in the C-section group, a pattern that is coherent with evidence obtained looking at the universe of births in the Canadian context [CIHI, 2013b]. This last variable is an interesting example of potential confounder when investigating the impact of C-section birth on children's health: while coming from a rural environment is associated with a lower risk of C-section, it might also result in reduced exposure to certain forms of air pollution associated with respiratory problems among children [Currie and Walker, 2011, Currie, 2013].

As expected, mothers giving birth by C-section are on average older, pregnancy and delivery complications leading to C-sections being more likely to occur as maternal age increases. Cesarean babies in the sample are also more likely to have a mother with at least *some* postsecondary education, and to have a slightly higher family income relative to the low-income cutoff measure adjusted to their family composition and place of residence.[26] Children born vaginally seem more likely to have generally healthier

[23] All counts are rounded to the nearest multiple of 100 in order to protect confidentiality, by requirement of Statistics Canada's Research Data Centres

[24] For example, the number of children sampled in cycle 3 was increased compared to other cycles in order to produce provincial estimates of certain measures in young children at the demand of the federal government. This results in a substantially larger birth cohort for 1996-97.

[25] Given the sample restrictions, all the statistics and results reported in this paper are based on unweighted observations.

[26] Leeb et al. [2005] also observe a positive relationship between C-section rates and the income quintile for Canadian

parents (as measured by the absence of a chronic condition). Finally, the fact that children delivered by C-section are on average slightly younger is coherent with the increase observed in C-section rates over the period studied.

Table 3.4 presents the average health outcomes for children given their birth delivery method. A series of outcomes related to the use of medical resources offer a window into children's general health. PMKs answering for a child born by C-section in the sample were more likely to report that the child spent at least one night in a hospital in the twelve months preceding the interview (and for reasons other than the birth delivery itself in the case of children aged 0-1). Cesarean children were also more likely to have been hospitalized overnight for a respiratory or gastrointestinal ailment. However, no statistical difference is observed between both groups of children when considering hospitalizations for injuries, an outcome that can be thought of as a placebo since it is less likely to be associated with the potential consequences of C-section birth on the immune system. A larger proportion of children born by C-section takes prescription medication on a regular basis, and report using Ventolin, a puffer or an inhaler routinely. As suggested by most epidemiological studies, unadjusted means comparisons in the estimating sample suggest that children born by C-section are on average 6% (or 1.2 percentage point) more likely to suffer from at least one chronic condition. Greater prevalences of asthma (9%) and wheezing or whistling in the chest (13%) are observed in the C-section group. No statistical difference is however observed between both groups when considering the prevalence of allergies, chronic bronchitis, ear, nose and throat infections (not shown). The absence of statistical difference in the average body mass index of children across birth delivery methods masks the fact that children born by C-section are more likely to find themselves at both ends of the relevant BMI distribution. As such, the data suggest that they are 10% more likely to be in the top 5^{th} percentile of body mass index for their gender and age group, an indicator of childhood obesity used by the Centers for Disease Control and Prevention.[27] This last outcome should be considered with caution. Indeed, the overall proportion of children in the sample meeting this definition is higher than 20%. While this might reflect a very high prevalence of obesity and the fact that the rates calculated are not obtained using the sampling weight, it should be noted that the body mass index variable is also constructed using the information reported by the PMK on the child's height and weight, and that such information is likely to be rounded by the reporting adult, thus introducing noise in the body mass index calculated.

3.4.3 Relevance of the instrument in the main estimating sample

Previous research [Gruber et al., 1999, Alexander, 2017, Allin et al., 2015] has shown physician financial incentives to have a causal impact on the probability that a birth is delivered by C-section. To assess the relevance of the instrument suggested in section 3.3 in the context of this paper, the results from previous studies are replicated in the main estimating sample, using the physician fees and information

neighbourhoods. However, they find that this relationship is reversed when controlling for maternal age, which tends to be correlated with both parental income and the probability of a C-section.

[27] BMI percentiles are taken by age and gender according to the growth charts prepared by the Centers for Disease Control and Prevention.

from the NLSCY described above to estimate equation (3.3).[28]

The results from the linear probability estimation are presented in table 3.5. Column 1 describes the results obtained when the controls are limited to the set of province and fiscal year of birth fixed effects, while columns 2 and 3 report the results for specifications in which control variables are gradually introduced. The results in all three columns point to a positive and statistically significant relationship between physician financial incentives and the probability that a child is delivered by C-section. Overall, the results from the specifications including the most extensive set of control variables suggest that doubling the fee paid to physicians for a C-section relative to a vaginal delivery would increase the probability that a given birth is delivered by C-section by slightly more than 14 percentage points, all else equal. This result is significant at the 1% level and robust to limiting the sample to children observed either at ages 0-1 year or at ages 2-3, as well as to the use of alternative non-linear specifications.[29] This result is in line with those reported in Allin et al. [2015], although it is larger. The difference in the magnitude of the estimates could be due to several factors. First, the sample composition (and the combination of provinces and years considered) vary across both papers. The NLSCY sample used in this paper does not include the cohorts born in 2009 and 2010, which are included for most provinces in Allin et al. [2015]. However, the NLSCY data includes children born in Quebec between fiscal years 1998 and 2008 (with the exception of those born during the fiscal year 2000), while all births delivered in Quebec before 2006 are excluded from Allin et al. [2015]. Second, the results presented in table 3.5 are based on the unweighted NLSCY data, so the weight put on each observation differs (sometimes substantially) from the one they would receive in a study considering the full population of births. Finally, the NLSCY does not allow for several factors included in hospital records to be used as controls (most importantly maternal history of previous C-sections, which is believe to be an important predictor of C-section birth), but does allow to control for a series of socioeconomic characteristics that were not available in previous studies.

Overall, financial incentives in the form of the relative payments to physicians for C-sections compared to vaginal deliveries seem to be a relevant instrument for the potentially endogenous C-section variable in equation 3.1. The exogenous process through which these incentives are determined suggests that the instrument should also satisfy the exclusion restriction.

3.5 Results

Tables 3.6 to 3.16 present detailed estimates from equations (3.1) and (3.2) for a series of health outcomes on which C-section birth could have an impact in the long-run. The health outcomes are taken from the oldest observation available for each child in the sample. This increases the number and size of the

[28] Physicians fees are on average 14% greater for C-section than for vaginal deliveries in the sample. Graphical evolution of the raw C-section rate and physician financial incentive for each province between 1994 and 2010 is presented in figure 2.2 in Allin et al. [2015].

[29] Results using probit or logit specifications are available upon request.

birth cohorts observed in each province.[30] Age fixed effects are included in the empirical specification to account for the fact that for many health conditions, the probability of diagnosis varies with age, often increasing as children get older. The range of health outcomes considered can be classified in three broad categories. The first set consists of outcomes that speak to general patterns of healthcare utilization, such as overnight inpatient stays and regular intake of prescription medication. The second category groups outcomes that relate to respiratory conditions, which have received a lot of attention in the epidemiology literature exploring the relationship between the microbial composition of children's intestinal tract and the development of their immune system. Indicator variables for asthma, asthma impairing the child's daily activities, and wheezing and whistling in the chest are considered. A third category of variables finally tracks children's body mass index and obesity diagnosis using the age– and gender-adjusted growth charts published by the Center for Disease Control and Prevention.

For each health outcome, a table first presents results from the estimation of equation (3.1) on the full sample of children.[31] Ordinary least squares (OLS) estimates are presented for continuous dependent variables, and linear probability model (LPM) estimates are reported for binary outcomes. The first column presented for each each health outcome corresponds to a linear specification with a limited set of controls, including indicators for gender, multiple birth and gestational age (pre– or post-term birth), birthweight, area of residence (indicators for rural areas or for dense urban areas with a population of at least 500,000 inhabitants), family composition (indicator for biparental family), the ratio of the child's family income to the low-income cutoff (adjusted for family size, year and area of residence) and a series of fixed effects for, respectively, maternal age at birth and the child's age at the time of the interview. Results are then displayed for a specification that includes additional controls: indicators for the presence of a smoking parent at home[32], the presence of older siblings at home or first-born status, and a series of known health conditions of the parents (asthma, chronic bronchitis/emphysema, chronic sinusitis, high blood pressure and diabetes). The extended set of controls also identifies if the mother smoked during the pregnancy, if the child was breastfed as an infant[33], if the mother suffered from high blood pressure or diabetes during the pregnancy[34], and if she received prenatal care. In all specifications, a full set of province and fiscal-year-of-birth fixed effects are also added.[35] Finally, results from the 2SLS estimation summarized by equations (3.2) and (3.3) are displayed for each outcome, along with the first stage F-statistic and the result from a reduced-form estimation of the health outcome of interest directly on the fee ratio and on the control variables (excluding the C-section birth indicator).

[30]The results obtained when considering health outcomes at specific ages (0-1 year olds or 2-3 year olds) do not vary much qualitatively; however, looking at specific age ranges reduces the sample size (most of the observations from the non-longitudinal top-up samples are lost) and the precision with which the impact of interest is estimated. Restricting the sample to such specific age groups also invariably reduces the strength of the instrument in the 2SLS estimation, and most first-stage F statistics fall below 10.

[31]Sample sizes vary slightly across the range of outcomes considered, based on the availability of information on each health outcome in the NLSCY.

[32]The results presented in section 3.5 are robust to using the number of smokers at home.

[33]Breastfeeding status is taken as an indicators, but the results are not sensitive to using the duration of breastfeeding as an alternative control variable. Completely omitting breastfeeding as a control doesn't alter the results in a material way.

[34]Yuan et al. [2016] suggest that mother's pre-pregnancy BMI should be considered when assessing the impact of C-section birth on the risk of obesity for the child. This information is not collected as part of the NLSCY, but it is less likely to matter for most of the health outcomes looked at in tables 3.6 to 3.11.

[35]Further controlling for season of birth does not materially change the results. The results are also robust to controlling for the highest level of education achieved by the child's mother

A concern with considering the full sample of children available in the NLSCY is that the reasons for and consequences of the choice between a cesarean and a vaginal delivery differ for certain categories of births. For example, a previous C-section is one of the main predictors of cesarean for a mothers' subsequent birth deliveries (in the Canadian context as in others). Indeed, in Canada in 2014-15, the rate of repeat C-sections was slightly over 80 percent [CIHI, 2016b], despite recent guidelines by the Society of Obstetricians and Gynaecologists of Canada and the American College of Obstetricians and Gynaecologists stating that in most cases, vaginal birth after cesarean delivery (VBAC) is a safe and appropriate procedure. Twinning, or multiple births of higher order, is similarly considered as a strong predictor of C-section birth, and both types of characteristics are generally used as exclusion criteria in most studies comparing C-section deliveries with vaginal ones since the circumstances surrounding the birth and early life circumstances likely make them less representative of most births. Direct information on twinning or higher order multiple births is available from the NLSCY, and those births can therefore straightforwardly be excluded from the estimating sample. However, the data does not identify children whose mother had a history of previous C-section.[36] This may be problematic not only because of the non-representativeness of these births, but also because including them in the main sample without being able to control for maternal history of previous C-section in the main analyses could bias the estimates if this risk factor affected a child's heath outcomes independently from his own birth delivery method. In the case of the 2SLS estimation, cases of maternal history of previous C-section might increase the proportion of births in the sample that do not *respond* to a change of value in the instrument, and that therefore do not provide additional information for the identification–or that even hinder it. For each health outcome, a second table therefore presents the results obtained from a sample of singleton first-borns, the second restriction de facto excluding all cases of previous C-section.[37] The linear (panel A), non-linear (panel B) and 2SLS estimates (panel C) are presented and compared between both samples. Unfortunately, limiting the sample to singleton firstborns greatly reduces sample size; less than 50% of all observations meeting these conditions. While in many cases, focusing on the narrower sample does not substantially change the magnitude of the coefficients of interest, it does lead to a reduction in the statistical significance (below ten percent) of most estimates.[38] It should also be noted that for 2SLS estimations, excluding multiple births and children with older siblings from the sample results in the first stage F statistics falling below the value of 10 against which the weakness of an instrument is often assessed. Finally, the summary tables for each health outcome report standard errors clustered at the province-fiscal year of birth level, which restricts the possibility for errors to be correlated within provinces through time, as well as the p-values obtained when clusters are defined at the province-level and standard errors are estimated using the wild-cluster bootstrap procedure suggested by Cameron et al. [2008] to correct for the small number of clusters (ten). In most cases, the statistical significance of the estimates remains unchanged when opting for this last cluster definition.

All specifications were also estimated for a series of alternative health outcomes: the frequency of ear, nose and throat infections for children under the age of 4, hyperactivity (measured by Ritalin prescrip-

[36] This is also true for the CHMS data used in the complementary analysis described in the appendix.

[37] Breech births and fetus malpresentations would also ideally be excluded from the main estimating sample. Unfortunately, the information contained in the NLSCY does not allow to identify these births in the sample, either directly or through a proxy.

[38] One notable exception to that pattern; the impact of cesarean birth becomes 61 percent larger (and remains statistically significant at the five percent level) when looking at body mass index as an outcome and using information from singleton firstborn children only.

tion, for children aged at least 4), kidney conditions, hearth disease and general health status (from poor to excellent) as assessed by the child's parents or using an index of various health measures.[39] No statistically significant relationship between C-section birth and any of these outcomes could be found using either a linear regression framework or the 2SLS approach. This absence of relationship is not add odds with most of the literature looking at similar outcomes.

3.5.1 Healthcare utilization: Inpatient stays and use of prescription medication

The first set of results focuses on outcomes associated with children's utilization of care and on their contacts with the healthcare system. Although the outcomes considered do not provide detailed information on the specific health conditions leading to more intensive use of healthcare and on which C-section births may have a causal impact, they can inform on the long-term system costs associated with high cesarean rates.

Table 3.6 first investigates the impact of C-section birth on the probability that a child was hospitalized overnight in the year preceding the NLSCY interview (excluding hospitalization directly following birth for children aged 0 to 1). Columns 1 to 3 report the results for any type of inpatient stay (related to illnesses or injuries) as an outcome. The results from a simple linear regression with a limited set of controls (column 1) suggest that C-section birth would have a small positive impact on the probability that a child is hospitalized, corresponding to 0.7 percentage point increase in the likelihood of an overnight hospitalization (a 15% increase in the sample mean). This estimated impact cannot, however, be statistically differentiated from zero.[40] The coefficients associated with most control variables are in line with clinical findings and with evidence from research on the social determinants of health: birthweight and family income reduce the probability of poor health (as measured by the probability of overnight hospitalization), while pre-term birth increases it. One more surprising result is the negative and statistically significant estimated impact of a multiple birth. While the literature does not provide strong intuition for this result, it might be the case that controlling for birthweight, maternal age and gestational age (which are all likely to be correlated with multiple birth) captures most of the expected positive relationship between multiple birth and the likelihood of poor health in childhood.

Extending the set of controls included in the empirical specification does not substantially impact the main coefficient of interest, as reported in column 2. The coefficients associated with the additional control variables are consistent with findings from previous studies: breastfeeding and prenatal care are

[39] The health Utility Index documented in the NLSCY refers to the eight-dimensional index developed at McMaster University's Centre for Health Economics and Policy Analysis, and based on the Comprehensive Health Status Measurement System (CHSMS).

[40] When considering a sample of 0-1 year olds, the estimated impact of C-section birth on the probability of a child having been hospitalized overnight in the year preceding the interview is approximately three time larger than in the full sample (0.020) and statistically significant at the one percent level. It should be noted that the definition of the outcome variable excludes any hospitalization directly related to the birth delivery itself. An increase in magnitude is also observed in the coefficients associated with pregnancy and birth related characteristics when moving from the main estimating sample to a younger one, potentially suggesting that the health impacts of complications during the gestation period and the delivery tend to fade as a child ages.

negatively associated with the probability of being hospitalized overnight, although the association for the latter is not statistically significant. The coefficient on most complications in pregnancy, and on the presence of a smoking parent has the opposite sign, although none of these estimated impacts are statistically significant. The estimated impact of C-section birth is still not statistically significant, and is only slightly smaller. The same can be said of the results obtained when opting for a probit model (see table 3.7).

As previously discussed in section 3.3, there is still a risk that some unobserved confounding factors contaminate the estimation of the C-section coefficient reported in columns 2. Column 3 in table 3.6 therefore presents the results from a two-stage least squares estimation instrumenting the C-section birth variable with the relative fee paid to physicians for this procedure compared to a vaginal delivery. Both with the limited and extended sets of control variables, moving to the instrumental variable framework results in a considerable loss of precision. The coefficient of interest is nearly 30 times larger when estimated by 2SLS, failing to provide credible insights on the magnitude of the impact of C-section birth on healthcare utilization patterns. One potential explanation for the imprecision characterizing the 2SLS results could be a weak instrument problem. The F-statistic on the excluded instrument reported at the bottom of column 3 takes a value of 11.27, only slightly above the threshold of 10 generally suggested as a *rule of thumb* to assess the strength of an instrument. To further investigate the potential causal impact of C-section birth despite the imprecision of the two-stage least squares estimates, the coefficient from a reduced-form estimation relating the hospitalization outcome directly to the fee incentive variable (along with the set of usual control variables, but excluding the C-section indicator) is also reported at the bottom of column 3. While the estimated coefficient is positive, it is not statistically significant.[41]

Importantly given the concerns about including multiple births and cases of previous C-section in the estimating sample, the second column of table 3.7 highlights that the LPM or probit estimates of the effect of C-section birth on the probability of overnight hospitalization are reduced in half when focusing exclusively on singleton first borns, and the estimates remain non statistically significant. The 2SLS estimates obtained on this restricted sample also do not point to a relationship between the birth delivery method and the likelihood that a child spends a night at the hospital. However, the first stage F-statistic in this restricted sample, which takes a value marginally below the ten mark, points to a higher risk of weak instrument.

Of course, columns 1 to 3 investigate the relationship between C-section birth and a very loosely defined measure of healthcare utilization: staying overnight at an inpatient facility for *any* reason. This outcome may indeed capture hospitalizations for conditions that are likely unaffected by the lasting health effects of birth delivery methods. To investigate if an impact can be observed for ailments that are in line with the clinical work linking C-section birth to changes in an infant's immune system after birth, columns 4

[41] Identification of the 2SLS coefficient is compromised when looking at the smaller samples of 0-1, 2-3 or 4-5 year olds respectively. The standard errors obtained are as large as the coefficients, which are themselves also several orders of magnitude larger than the OLS estimates. The first stage F-statistics suggest that the instrument is weak for these more restrictive samples, ranging from 6.15 for the sample of 0-1 year olds to 1.5 for the sample of 4-5 year olds. The reduction in the variation observed in the values taken by the instrument when limiting the birth cohorts considered may explain the weaker first stage, and the uninformative nature of the results obtained.

to 6 look specifically at hospitalizations for gastrointestinal or respiratory problems. Contrary to initial expectations, no evidence of a causal impact of C-section birth is found, even when limiting the variables included as controls in a simple linear regression. The estimated impact is positive but small (0.004), and not statistically significant. The coefficient is also unaltered when expanding the set of controls in the estimation, as shown in column 5, and when turning to a probit estimation or to the subset of births corresponding to singleton first-borns, as shown in table 3.8. Turning to a two-stage least squares estimation again introduces a substantial level of imprecision. The coefficient associated with C-section birth becomes nearly 35 times larger than the linear probability estimates, and the associated standard error increases by a similarly large factor. The reduced-form estimation of hospitalization for gastrointestinal or respiratory problems on the relative fee paid to physicians for a C-section yields a positive coefficient (0.021), slightly smaller than that obtained when looking at any type of hospitalization. While it is statistically significant at the 10% level in the full sample, it cannot be statistically differentiated from zero in the subsample of singleton first-borns. Overall, this result does not provide firm grounds on which to conclude in a causal impact of C-section birth on hospitalizations for gastrointestinal or respiratory conditions.

Columns 7 to 9 finally present the estimation results for an outcome variable identifying if a child had been hospitalized for a severe injury in the year preceding the interview. This exercise can be seen as a placebo test. Indeed, the evidence reviewed in section 3.2 would not suggest any relationship between C-section and the occurrence of injuries, which are more likely to be the result of random accidents. The coefficients associated with C-section births are in this case negative for all specifications, and as expected, very close to zero and not statistically significant (although the 2SLS estimate remains much larger in magnitude than the estimates from linear specifications). Similar results are obtained when looking at the number of severe injuries requiring medical assistance but not necessarily resulting in a hospitalization in the year preceding the interview.

Children's utilization of healthcare resources can also be looked at through the lens of their regular use of prescription medication. Table 3.9 first considers the full sample of births to investigate the impact of C-section birth on the probability that a child takes at least one prescription medication regularly. The results from a linear specification with a limited set of controls (column 1) suggest that being born by C-section increases by 1.2 percentage point the probability that a child takes some prescription medication on a regular basis. This represents a 11% increase in the sample mean, an impact that is reduced almost by one fourth when introducing additional controls for socio-economic characteristics and family background, as shown in column 2. Moving to the instrumental variables estimation again leads to a large increase in the coefficient associated with the C-section indicator, which reaches 0.499. While the magnitude of the 2SLS estimate is too large to be credible, it is however statistically significant at the 5% level. This might suggest the existence of a causal link between C-section birth and the probability that a child needs to regularly take prescription medication. Similarly, the positive and statistically significant coefficient obtained from a reduced form estimation relating the outcome variable to the excluded instrument suggests that a causal effect may be at play. Table 3.10 presents the results for the alternative specifications and for the restricted sample. The marginal effect remains relatively unchanged when the model is estimated non-linearly. While considering singleton first-borns exclusively does not affect the magnitude of the coefficients obtained from a linear or a non-linear estimation, it does result in a loss of statistical significance. This could however be an artifact of the reduction in sample size from 21,400 to 9,000, given the fact that the reduction in the precision of the estimated effect comes from an increase in the standard error. Moreover, the 2SLS estimate is statistically different from zero (at the 10% level) even in this stricter sample, although its magnitude (0.515) is again very high. While these results are not unambiguous evidence that C-section birth has a causal impact on the probability that a child needs medication on a continuous basis, they provide some grounds on which not to completely discard this possibility. Unfortunately, the NLSCY does not provide detailed information on the type of prescription medications taken by children, or on the conditions they are associated with.

Columns 4 to 6 of table 3.9 present the results when the outcome variable is specified as the total number of prescription drugs regularly taken by a child. The coefficients obtained with a linear specification are qualitatively similar to those presented with a binary outcome, as are the estimates associated with most control variables. While the 2SLS estimate is much larger than the simple linear probability coefficients with this outcome as well, the increase in the size of the standard error is proportionally larger. The estimate obtained when instrumenting for the probability that a child is born by C-section is too noisy to be differentiated from zero, despite the fact that the first stage F statistics is again greater than 10.

3.5.2 Respiratory health conditions

Table 3.11 presents the estimated impact of C-section birth on the probability that a child suffers from asthma or from respiratory problems. This type of outcome is more closely related to clinical evidence pointing to a short-term relationship between birth delivery methods and the early development of infants' immune systems. Consistent with the evidence presented in the literature showing differences in the immune development of infants across birth delivery methods, the LPM estimate obtained when looking at the impact of C-section birth on the probability that a child is diagnosed with asthma later in life is positive and statistically significant. The main coefficient obtained with a limited set of controls is 0.012, which corresponds to a 12 percent increase in the mean for the full sample. As shown in column 2, this result is not substantially altered by the introduction of additional control variables in the regression. Even as a series of parental health conditions are added, including diagnosis of chronic respiratory conditions, the coefficient associated with C-section birth keeps a similar magnitude (0.010) and remains statistically significant at the 5% level. As expected, having a parent suffering from a respiratory condition such as asthma or chronic bronchitis is estimated to have a positive and significant impact on the probability that a child develops asthma-related symptoms or diagnosis himself, which is in line with findings suggesting the hereditary risks associated with asthma.[42] The presence of a smoking parent at home is also associated with increased risks that children in the household receive a diagnosis of asthma, and, in line with findings form the impact of air pollution on children's health (e.g. Currie [2013]), the estimated impact of the child living in a rural area is negative. More surprisingly, the coefficient associated with an indicator specifying if a child lives in a relatively larger urban zone is

[42] Liu et al. [2009] find that individuals with a family history of moderate to severe asthma are 2.4 to 4.8 times more at risk of developing asthma themselves, after adjusting for other risk factors related to behaviours and to the environment.

negative, although very small and not statistically significant. The results presented in table 3.12 again suggest that the estimated effects are robust to opting for a probit estimation.

To address the possibility that the relationship highlighted in columns 1 and 2 be spuriously caused by unobserved factors contributing simultaneously to the event of a C-section birth and to a later diagnosis of asthma, column 3 turns to the instrumental variable framework described above. Again, this estimation method leads to a substantial loss of precision, and inflates both the estimated coefficient on C-section birth and the associated standard error. The latter, however, increases some 7 times more than the coefficient, leaving the 2SLS estimates inconclusive. This may again be evidence of a weak first stage; the associated F-statistic of 11.2 is only slightly larger than the usual threshold used to determine the strength of an instrument. However, the reduced form estimates, presented at the bottom of column 3, also fail to provide suggestive evidence of a causal relationship of C-section birth on an asthma diagnosis. Looking at more serious cases of asthma impairing the participation of a child's in normal activities yields the same type of results (columns 4 to 6), as does turning to conditions not requiring a diagnosis, such as the probability that parents report episodes of wheezing or whistling in the chest for their child in the year preceding the interview, as shown in table 3.14. Similar results are also obtained when looking at children's regular use of inhalers or puffers. In line with the results from recent studies, additional results provide no conclusive evidence of a relationship between C-section birth and allergies, or chronic bronchitis.[43]

A potential explanation for the difference between the LPM and the 2SLS results presented in tables 3.11 to 3.14 is that the extensive set of controls included in the linear or probit regressions is not capturing all confounding factors simultaneously correlated with C-section birth and with the likelihood that a child develops a respiratory problem. For example, it is possible that parents who are more likely to request a C-section are also more likely to be frequent users of the healthcare system, thus increasing the probability that they consult a physician to seek a formal diagnosis of asthma if they believe that their child experiences some respiratory issues. Alternatively, some unobserved medical conditions during the pregnancy increase both the risks complications at delivery and the probability that a child develops respiratory problems later in life independently from the recourse to a specific birth delivery method. In this case, LPM results could point to a statistically significant non-causal positive relationship. This association should however not be picked up by the 2SLS estimates, which given the choice of instrument, inform on the long-term health consequences of unnecessary C-sections. Given the lack of precision associated with the two-stage least squares estimates, it is however hard to completely reject the possibility that C-section birth has a causal impact on the development of respiratory conditions later in childhood.

For all respiratory outcomes mentioned above, limiting the sample definition to first born singletons does not substantially alter the magnitude of the linear and non-linear point estimates, but it once again results in the results being no longer statistically significant, as shown in panels A and B of tables 3.12, 3.13 and 3.14. Panel C of those same tables presents the main coefficients of interest from the 2SLS estimation, along with the associated standard errors, the first stage F-statistic and the coefficient from a reduced form regression exploring the relationship between occurrences of asthma and the relative fee

[43]Those results are not displayed for brevity, but are available upon request.

received by physicians for a C-section delivery. As with the health outcomes previously considered, the first stage from the 2SLS estimation is weaker in the restricted sample than in the full one. Qualitatively, the results are nevertheless in line with those obtained in the full sample of births; a positive but very imprecisely estimated second stage fails to detect a statistical relationship between respiratory problems and cesarean birth, as does the reduced form estimation.

3.5.3 Body mass index and obesity

As discussed in section 3.2, previous work provides mixed evidence on the relationship between C-section birth and the risks of obesity among children. Table 3.15 investigates this relationship by looking at two different outcomes: a child's body mass index, and the probability that it is in the top five percentiles of the BMI distribution for children of their age and gender. Two caveats related to these outcomes should be highlighted here. First, a clinical measure of a child's *body mass index* is not readily available in the NLSCY. BMI is rather calculated using the information provided by the PMK on the child's height and weight.[44] Research has however shown that parents are likely to err or to be imprecise when reporting their child's measurements, mostly because of a tendency to round true measures to the nearest unit [Special Survey Division, 2009]. The constructed measure of a child's BMI is therefore likely plagued with at least *some* degree of measurement error. This imprecision in each child's measure of BMI will also affect where they are placed in the relevant distributions used to assess obesity.[45] A second limitation associated with outcomes relating to body mass index is that they are available for a reduced number of children (13,700 compared to more than 21,400 for most other outcomes). This is explained by two reasons: (i) non-response rates for each of the two measures used to derive the BMI variable are relatively high; and (ii) the CDC growth charts used to derive the variable for obesity are available for children aged 2 to 20 (all children observed only at 0 or 1 in the sample are therefore excluded from the analysis focusing on obesity). Given this substantive change in sample size, the estimated coefficient for the first stage of the 2SLS estimation is reported at the bottom of table 3.15.

With these limitations in mind, the results from linear regressions presented in columns 1 and 2 of table 3.15 suggest a positive impact of C-section birth on a child's probability to be fall beyond the 95^{th} percentile of the age and gender adjusted BMI distributions published by the CDC. This result is replicated using a probit specification, but does not survive when the sample is restricted to singleton first borns.[46] Interestingly, and unlike the results obtained with most respiratory conditions, adding control variables does not reduce the magnitude of the estimated coefficient associated with C-section birth (in fact it increases it marginally). The same is true when looking at a continuous measure of BMI as an outcome, as shown in columns 4 and 5, although the impact is small, C-section birth is associated with a 0.149 increase in body mass index when controlling for the most extensive set of potential confounders. This last estimate is however not statistically significant when the standard errors are clustered at the province level and a correction for the small number of clusters is applied. In addition to the issues

[44] According to the formula $BMI = \frac{weight(kg)}{height^2(m)}$

[45] In general, the constructed measure of BMI seems to overestimate the probability that children fall within either tails of the distribution for their age and gender.

[46] BMI is only available for 5,700 children when excluding children who have older siblings or who are not singletons.

already discussed with regards to this outcome variable, the sensitivity of the inference suggests that the link between C-section birth and the risks of obesity must be interpreted with caution. Moreover, as with previous outcomes, the coefficients and standard errors associated with the C-section variable are substantially inflated when turning to a 2SLS estimation. Although positive, the coefficient is no longer statistically significant, and the reduced-form estimation on the excluded instrument and controls does not provide evidence of a causal impact. Other environmental factors such as exposure to second hand smoke in the house, smoking during pregnancy or birthweight are estimated to be strongly associated with obesity and BMI. Linear, non-linear and instrumental variables estimates further suggest that C-section birth has no significant impact on the probability that a child is underweight, with a body mass index below the 5^{th} percentile for his age and gender, or is overweight but not obese, with a BMI between the 85^{th} and 95^{th} percentile for his age group and gender.[47]

3.5.4 Discussion of the two-stage least squares estimates

The 2SLS estimates presented above are characterized by an important degree of imprecision. Several factors could explain why turning to the suggested instrumental variables framework may introduce noise in the estimation. First, the choice of linear specifications for each stage might not be ideal given the binary nature of both the endogenous regressor and, in most cases, the dependent variable. However, turning to non-linear models estimated by maximum likelihood does not yield the expected improvements.[48]

A second reason for the poor performance of the 2SLS may be related to the estimation of the first stage in the NLSCY. To address this possibility, the appendix presents a two-sample two-stage least squares approach in which the first stage is estimated using data on more than 3 million births. The second stage is then estimated using the Canadian Health Measures Survey, another data set documenting the health of Canadian children (with clinically measured outcomes). Unfortunately, the results obtained are just as imprecise as the 2SLS estimates presented in tables 3.6 to 3.15. Overall, there may not be enough variation in the instrument in either samples to lend the strength that would be necessary to obtain precise estimates. Indeed, the estimated first-stage F-statistics are in most cases only slightly higher than the threshold of 10 suggested as a rule of thumb to assess the strength of an instrument [Staiger and Stock, 1997], or marginally below that value for the sample of first-born singletons. Moreover, they are only marginally higher than the critical value suggested by Stock and Yogo [2005] for a two-stage least square test with a size of 15% for a model that includes one endogenous regressor and one instrument.[49]

Finally, the large magnitude of the coefficients estimated with the 2SLS approach compared to the linear probability coefficients could partially be explained by the local average treatment effect interpretation

[47] These results are not presented, but are available upon request.

[48] The maximum likelihood estimation of an endogenous switching model in which both equations have binary outcome variables, for example, yields even less precise results than the 2SLS estimates.

[49] Attempts to estimate the impact of C-section birth on the set of health outcomes described using the special regressors approach and employing maternal age at birth or family income to low-income cutoff ratio as a special regressor also did not improve the precision of the estimates, potentially because the value of the kurtosis for these variables is not high enough in the NLSCY sample [Lewbel, 2012].

of the 2SLS estimates. Clinical studies have suggested that the impact of C-section birth on the immune system of infants would be mostly observed for babies born by elective cesarean, as opposed to emergency ones. Since elective C-sections often take place before labour begins, newborns are unlikely to be even partially exposed the to the maternal bacteria in the vaginal canal that can colonize the microbiota and reinforce the immune system of babies during labour [Black et al., 2015, Neu and Rushing, 2011]. Moreover, the intake of intrapartum antibiotics may differ depending on the elective or emergency nature of the C-section, and may impact the development of infants' gut microbiota [Azad et al., 2015]. As the NLSCY data does not provide the information needed to separate emergency C-sections from elective ones, it is possible that the impact estimated in the OLS/LPM framework is attenuated compared to the effect that would be identified if only elective C-sections could be considered. Given the instrument used in this paper, the 2SLS estimates of the long-term health impacts of C-section birth are identified from the births that are delivered by C-section because financial incentives led physicians to choose this method over a vaginal delivery. The births from which the 2SLS estimates are identified from are therefore more likely to be elective C-sections– and a higher estimated impact of cesareans on health outcomes may therefore be expected from this identification strategy.[50] For example, Yuan et al. [2016] find an association between C-section and obesity that is much stronger for individuals without known risk factors for cesarean delivery, even when adjusting for a variety of pregnancy and delivery characteristics. The LATE estimated with the 2SLS approach could convey a similar story, whereby an unnecessary C-section would have a relatively higher impact on a child's health, given his otherwise good predispositions. Altogether, it is however unlikely that the magnitude of the instrumental variable estimates could be entirely explained by this interpretation of the local average treatment effect.

3.6 Conclusion

C-section rates have reached record highs across the OECD. This paper seeks to shed some light on the long-term consequences this phenomenon might have on children's health in two ways. First, it exploits the rich information on children's socioeconomic characteristics, family health history and environment contained in the National Longitudinal Survey of Children and Youth to investigate the relationship between C-section birth and various health conditions while controlling for a more extensive set of potential confounders than have simultaneously been included in previous studies. Second, it addresses remaining endogeneity concerns associated with cesarean birth by turning to an instrumental variables approach that exploits exogenous changes in physicians' financial incentive to opt for C-section delivery as shocks on the probability that a child is born by cesarean.

An initial set of findings highlights that the positive association between C-section birth and a series of health outcomes survives the inclusion of a comprehensive set of confounders in simple linear models. More specifically, the relationship between cesarean birth and the likelihood that a child takes prescription medication on a regular basis or suffers from asthma, although sometimes attenuated, remains

[50] Black et al. [2015] however only finds a statistical difference in the health outcomes of children born by planned and unplanned C-section when looking at type 1 diabetes. They do find a difference in health outcomes of children born by C-section compared to children delivered vaginally for a wider range of health outcomes, including severe asthma requiring hospitalization.

positive and statistically significant when adding socioeconomic status, parent diagnosis for related health conditions, the presence of smokers at home, and a series of pregnancy-related characteristics to the list of controls included in the empirical specification considered. A relationship between birth delivery method and the risk of childhood obesity can also be found, even as these variables are controlled for. However, when taking into account family medical history and a wider set of socioeconomic characteristics, no robust relationship is found between cesarean birth and the probability of overnight hospitalization. Overall, these findings point to the fact that in most cases, the association between C-section birth and health outcomes later in childhood can be overestimated when the influence of various factors in a child's environment and background are not taken into account. Estimates from non-linear specifications lead to similar conclusions. While most results are not statistically significant when restricting the analysis to first born singletons, the magnitude of the linear and non-linear estimates is generally comparable across both samples.

The evidence generated using the suggested instrumental variable approach, however, is harder to interpret. In fact, the two-stage least squares estimates for most health outcomes are both too large to be credible and too imprecise to be really informative, potentially because of a lack of strength in the instrument. Nevertheless, the 2SLS estimates provide some evidence of a causal relationship between C-section birth and the probability that a child regularly takes prescription medication. This result is further supported by a positive and statistically significant reduced form estimate of this outcome on the excluded instrument. Given the choice of instrumental variable, this result can be interpreted as evidence of an impact of unnecessary cesareans on children's health as proxied by the regular intake of prescription drugs. Unfortunately the magnitude of this relationship is hard to assess given the imprecision of the 2SLS estimates.

Overall, although the estimates presented in this paper do not provide some clean measure of the increased health risks that children born by C-section may face later in childhood, they do provide some ground on which not to discard the possible association between health outcomes and the choice of birth delivery method. More specifically, the results presented suggest that some of the clinical findings linking cesarean birth to short-term negative impacts on the development of infants' immune system may persist in the long-run. At the very least, the evidence presented suggests that relationships previously documented between C-section birth and health impairments such as respiratory conditions in childhood are likely not entirely an artifact of omitted control variables relating to parental health, prenatal history and most characteristics pertaining to a child's socioeconomic background and environment, since a positive association can still be found when those factors are controlled for. Future research should nevertheless investigate further the causal nature of these associations, and seek to quantify their importance in order to better understand the long-term implications and costs of the high C-section rates observed in many countries, especially in cases where the procedure may not be necessary.

Table 3.1: List of health outcomes form the NLSCY

Variable	**Ages**	**Cohorts**	**Details**
General health (assessed by parent)	0+	1994-2008	Health status assessed by person most knowledgeable. 1= Excellent and 5=Poor. A dummy indicating if health is excellent is derived from this variable.
Body mass index	2+	1997-2007	The variable bmi is created for cycles 1 and 2 using (weight)/(height^2). In all cycles, bmi is observed for children aged 24 months or older
Body mass index (percentile)	2+	1997-2007	Percentile (3^{rd}-5^{th}-10^{th} – 15^{th} - 25^{th} -50^{th} – 75^{th} – 85^{th} – 90^{th} – 95^{th} – 97^{th}) of body mass index per month of age and gender, according to the grow charts published by the CDC. Indicators are derived from this variable to flag very low bmi (<=5^{rd} percentile) or very high bmi (>=95^{th} percentile) compared to age group/gender.
Inpatient stay (overnight)	0+	1994-2008	Indicator of the child having spent a night in a hospital in the year preceding the interview. Dummy variables are created for hospital stays overnight due to: Respiratory illness or disease, gastrointestinal ill or disease, immune related disease (respiratory or gastrointestinal), injury, other health issue.
Asthma	0+	1994-2008	Indicator of the child having been diagnosed with asthma. Although asthma may be thought of as unlikely to be diagnosed at early ages, the proportion of kids having been diagnosed with asthma is relatively similar in all age groups, and declines for older children, although the sample is much smaller for children aged 6 or more. For children diagnosed with asthma, another variable flags if the child's activity are impaired by the condition, and another indicates if he has had asthma attacks in the year preceding the interview.
Wheezing or Whistling	0+	1994-2008	Indicator of the child had had wheezing or whistling in the chest in the year preceding the interview (relatively prevalent for all age groups, but a little more for those under 5).
Nose/throat infection	0-3	1994-2008	Indicator of the child having at least rarely a nose/throat infection. An indicator variable for having frequent nose or throat infections is also derived.

Variable	Ages	Cohorts	Details
Ear infection	0-3	1994-2008	Indicator of the child having had an ear infection since birth. A variable also indicates how many times the child has had an ear infection since birth, but is capped at 4. This variable should be looked at within age groups (e.g. only for children aged 0, only for children aged between 12 and 15 months, etc.) – there are very few non responses in each age group. However, only looking at 3 year olds limits the birth cohorts to 1994-2006.
Chronic condition	0+	1994-2008	This variable is derived in the NLSCY from all answers given to the questions specifically targeting specific chronic conditions. However, the list of questions asked increases with the cycles, and it is not clear that the definition is consistent across cycles. Excludes asthma in all cycles.
Allergies	0+	1994-2008	The definition of this variable is not consistent across cycles; In cycles 1-3, only a generic question on allergies is asked whereas in cycles 4-8, specific questions are asked about food/respiratory/other forms or allergies. A general variable for any allergies is constructed in the late cycles to have ensure consistency. In all cases, the allergies need to be diagnosed by a health professional.
Bronchitis	0+	1994-2008	Indicator of the child having been diagnosed with bronchitis
Heart condition	0+	1994-2008	Indicator of the child having been diagnosed with a heart condition or disease.
Epilepsy	0+	1994-2008	Indicator of the child having been diagnosed with epilepsy.
Kidney condition	0+	1994-2008	Indicator of the child having been diagnosed with a kidney condition or disease.
Other chronic condition	0+	1994-2008	Indicator of the child having been diagnosed with another long term condition (excluding learning disabilities or nervous/psychological/emotional problems) The definition for cycles 4-8 also excludes attention deficit disorders, eating disorders (cycle 8), diabetes (cycles7-8). Asthma is excluded from the conditions covered by this variable in all cycles.
Regular intake of prescription medicine	0+	1994-2008	Variable derived from individual questions about taking the following prescription medicines on a regular basis: Ventolin/inhaler/puffer, Ritalin, Tranquilizers/nerve pills, anti-convulsants/anti-epileptics, other medication. Indicators for each medication category are available, as is a variable indicating the number of categories of medicines that a child regularly takes.

Table 3.2: Birth cohort sizes associated C-section rates, main estimating sample

Birth cohort (fiscal year)	**C-section rate** (s.d.)	**Observations** (rounded nearest 100)
Apr 1994 – Mar 1995	0.197	1,100
	(0.398)	
Apr 1995 – Mar 1996	0.185	1,000
	(0.388)	
Apr 1996 – Mar 1997	0.206	600
	(0.405)	
Apr 1997 – Mar 1998	0.207	4,300
	(0.406)	
Apr 1998 – Mar 1999	0.184	1000
	(0.425)	
Apr 1999 – Mar 2000	0.236	900
	(0.425)	
Apr 2000 – Mar 2001	0.255	900
	(0.436)	
Apr 2001 – Mar 2002	0.225	1,300
	(0.436)	
Apr 2002 – Mar 2003	0.225	1,200
	(0.418)	
Apr 2003 – Mar 2004	0.270	1,900
	(0.444)	
Apr 2004 – Mar 2005	0.258	1,500
	(0.444)	
Apr 2005 – Mar 2006	0.261	1,800
	(0.439)	
Apr 2006 – Mar 2007	0.275	1,200
	(0.447)	
Apr 2007 – Mar 2008	0.253	1,600
	(0.435)	
Apr 2008 – Mar 2009	0.274	1,000
	(0.446)	
Full sample	0.234	21,400
	(0.425)	

Notes:
1- The larger volume of observations from birth cohort for 1997-98 is partly due to oversampling of aged 1 and 5 in cycle 3. This oversampling was a response to the intention of the government of Canada to increase its capacity to monitor children's outcomes in the early years.
2- Number of observations are rounded at nearest 100 in concordance with rules by the Canadian Research Data Centre Network.

Table 3.3: Average selected characteristics, by birth delivery method

	C-section births	**Vaginal births**	**Difference (CS-VD)**	**All births**
Age	4.26	4.53	-0.28***	4.47
Multiple birth (0/1)	0.057	0.018	0.038***	0.027
Preterm birth (0/1)	0.157	0.087	0.070***	0.104
Postterm birth (0/1)	0.006	0.004	0.002*	0.004
Birthweight	3.39	3.46	-0.06***	3.44
Maternal age at birth	30.07	28.92	1.14***	29.19
Boy (0/1)	0.526	0.508	0.018**	0.512
Rural (0/1)	0.201	0.216	-0.015**	0.213
Area of residence >500K (0/1)	0.237	0.247	-0.009	0.244
Family income to LICO	2.47	2.41	0.06**	2.42
Mother's highest degree=high school (0/1)	0.179	0.182	-0.003	0.181
Mother's highest degree= some postsec (0/1)	0.467	0.450	0.017**	0.454
Mother's highest degree=university (0/1)	0.291	0.288	0.004	0.289
Father has chronic condition (0/1)	0.390	0.373	0.017**	0.377
Mother has chronic condition (0/1)	0.476	0.430	0.045***	0.441
Fee ratio (CS/VD)	1.3	1.1	0.1**	1.1
Observations	21,400			

Notes:
1- Number of observations are rounded at nearest 100 in concordance with rules by the Canadian Research Data Centre Network.

Table 3.4: Average selected health outcomes, by birth delivery method

	C-section births	**Vaginal births**	**Difference (CS-VD)**	**All births**
Overnight stay at the hospital in past year (0/1)				
For any reason	0.054	0.045	0.009**	0.047
For respiratory or gastrointestinal problem	0.026	0.019	0.006**	0.021
For injury	0.003	0.004	-0.001	0.004
Regular intake of prescription medication (0/1)	0.106	0.094	0.012**	0.097
Number of prescription medication regularly taken	0.12	0.11	0.02***	0.11
Any chronic condition (0/1)	0.203	0.191	0.012***	0.194
Allergy (0/1)	0.124	0.119	0.005	0.120
Chronic bronchitis (0/1)	0.016	0.015	0.001	0.015
Asthma diagnosis (0/1)	0.107	0.098	0.009**	0.100
Asthma hindering participation in activities (0/1)	0.016	0.013	0.003*	0.014
Regular use of Ventolin/puffers/inhalers (0/1)	0.061	0.056	0.005*	0.057
Wheezing/Whistling (0/1)	0.207	0.183	0.024***	0.188
Body mass index	17.49	17.44	0.06	17.45
Observations	21,400			

Notes:
1- Number of observations are rounded at nearest 100 in concordance with rules by the Canadian Research Data Centre Network.
2- Total number of observations for health outcomes corresponding to body mass index and obesity measure are only available for 13,100 observations, since those measures are only considered for children aged 2 and above, in concordance with the CDC guildelines.

Table 3.5: First stage – Impact of financial incentive of C-section birth

	(1)	(2)	(3)
Fee Ratio	0.149***	0.129***	0.144***
	(0.041)	(0.041)	(0.042)
Boy		0.010*	0.010*
		(0.006)	(0.006)
Multiple birth		0.214***	0.214***
		(0.027)	(0.027)
Pre-term birth		0.107***	0.089***
(<259 days)		(0.012)	(0.012)
Post-term birth		0.115**	0.098**
(>293 days)		(0.045)	(0.046)
Birthweight		0.006	0.011*
(kg)		(0.007)	(0.007)
Smoker at home			0.021***
			(0.007)
Rural area		-0.016**	-0.013*
		(0.007)	(0.007)
Urban area		-0.017**	-0.021***
(500,000+)		(0.008)	(0.008)
Both parents at home		-0.028**	-0.024**
		(0.011)	(0.011)
Family income/ LICO		0.001	-0.002
(adjusted)		(0.002)	(0.002)
Pregnancy diabetes			0.071***
			(0.014)
High blood pressure			0.076***
during pregnancy			(0.010)
Mother smoked			-0.020**
during pregnancy			(0.009)
Prenatal care			0.001
			(0.028)
Prov & Year of birth FE	✓	✓	✓
Age FE	✓	✓	✓
Maternal age FE		✓	✓
Parents' health conditions		✓	✓
Older siblings		✓	✓
Observations	21,400	21,400	21,400

Notes:
1- Significance levels: ***=0.01, **=0.05, *=0.10.
2- All standard errors are clustered at the province-birth cohort level.
3- Estimations based on unweighted sample.

Table 3.6: Impact of C-section birth on the probability that a child has been hospitalized in the year preceding the interview

	All hospital stay			**Respiratory/Gastrointestinal**			**Injury**		
	LPM	LPM	2SLS	LPM	LPM	2SLS	LPM	LPM	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
C-section birth	0.007	0.006	0.179	0.004	0.004	0.140	-0.001	-0.001	-0.040
	(0.004)	(0.004)	(0.146)	(0.003)	(0.003)	(0.090)	(0.001)	(0.001)	(0.037)
Boy	0.013***	0.013***	0.011***	0.009***	0.009***	0.007***	-0.000	-0.000	0.000
	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)
Mutiple birth	-0.019**	-0.020**	-0.058*	0.005	0.005	-0.025	-0.004***	-0.004***	0.005
	(0.009)	(0.009)	(0.034)	(0.007)	(0.007)	(0.022)	(0.001)	(0.001)	(0.008)
Pre-term birth	0.026***	0.022***	0.007	0.013***	0.011**	-0.001	-0.000	-0.001	0.003
(<259 days)	(0.006)	(0.006)	(0.014)	(0.005)	(0.005)	(0.009)	(0.001)	(0.001)	(0.004)
Post-term birth	0.019	0.016	-0.001	0.007	0.005	-0.009	0.019	0.019	0.022
(>293 days)	(0.036)	(0.036)	(0.039)	(0.016)	(0.016)	(0.019)	(0.015)	(0.015)	(0.016)
Birthweight	-0.010***	-0.010***	-0.012***	-0.004*	-0.004*	-0.006**	-0.001	-0.000	0.000
(kg)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)
Smoker at home		0.003	-0.001		0.003	0.001		0.001	0.002
		(0.003)	(0.005)		(0.002)	(0.003)		(0.001)	(0.001)
Rural area	0.006*	0.005	0.008*	0.002	0.001	0.003	0.000	-0.000	-0.001
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)
Urban area	-0.006	-0.004	0.000	-0.006**	-0.005*	-0.002	0.002	0.002*	0.001
(500,000+)	(0.004)	(0.004)	(0.005)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)
Two parents	0.000	0.002	0.006	-0.002	-0.001	0.003	-0.002	-0.002	-0.003
	(0.005)	(0.005)	(0.007)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Income/ LICO	-0.002***	-0.001**	-0.001	-0.001	-0.000	0.000	-0.000*	-0.000	-0.000
(adjusted)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Pregnancy		0.003	-0.009		0.003	-0.006		0.001	0.003
diabetes		(0.007)	(0.013)		(0.005)	(0.008)		(0.002)	(0.004)
Pregnancy		0.007	-0.006		0.001	-0.010		0.002	0.005
high blood pr.		(0.005)	(0.012)		(0.004)	(0.007)		(0.002)	(0.003)
Pregnancy		0.006	0.009		-0.001	0.002		0.001	0.001
smoking		(0.006)	(0.007)		(0.003)	(0.004)		(0.002)	(0.002)
Prenatal care		-0.013	-0.013		0.000	0.001		-0.012	-0.012
		(0.017)	(0.019)		(0.009)	(0.011)		(0.008)	(0.008)
Prov & Year									
of birth FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maternal age FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parents' health		✓	✓		✓	✓		✓	✓
Older siblings		✓	✓		✓	✓		✓	✓
First stage F			11.269			12.124			11.24
Reduced form			0.025			0.021*			-0.005
			(0.019)			(0.011)			(0.005)
Observations	21,400	21,400	21,400	20,800	20,800	20,800	21,400	21,400	21,400

Notes:
1- Maternal and paternal health conditions include indicators for asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes.
2- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10.
3- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.
4- Low-income cutoff (LICO) is adjusted for year, family size and geographic area.
5- Similar results are obtained for columns 7-9 when using the number of severe injuries requiring medical assistance in the 12 months preceding the interview (capped at 5).
5- Observations rounded at the closest multiple of 100.

Table 3.7: Impact of C-section birth on the probability that a child has been hospitalized overnight in the year preceding the interview

		All births	**Singleton first borns**
Panel A: OLS results	**C-section birth**	0.006	0.003
		(0.004)	(0.005)
	P-value		
	(cluster=province)	0.215	0.436
	Wild cluster bootstrap p-value		
	(cluster=province)	0.220	0.402
	Full controls	✓	✓
Panel B: Probit	**C-section birth**	0.004	0.002
	(marginal effect)	(0.003)	(0.004)
	Full controls	✓	✓
Panel C: 2SLS	**C-section birth**	0.179	0.141
		(0.146)	(0.161)
	First-stage F stat	11.269	9.41
	Reduced form	0.025	0.024
		(0.019)	(0.027)
	Full controls	✓	✓
	Observations	21,400	9,000

Notes:

1- Full controls include indicators for gender, multiple birth, pre-term and post-term births, birthweight, indicators for maternal age at birth, two parents at home and density of population, family income to LICO ratio, indicators for age, as well as fixed effects for province and fiscal year of birth, indicators for smoking parents, breastfeeding in infancy, a series of pregnancy complications (high blood pressure, pregnancy diabetes, maternal smoking during pregnancy, prenatal care) and a series of maternal and paternal health conditions (asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes), as well as older siblings.
2- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10. P-values obtained by implementing a wild-cluster bootstrap-t procedure (described in Carmeron, Gerbach and Miller (2008)), are also presented to allow for errors be correlated across years within province.
3- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.
4- Observations rounded at the closest multiple of 100.

Table 3.8: Impact of C-section birth on the probability that a child has been hospitalized overnight for a gastrointestinal or respiratory condition in the year preceding the interview

		All births	**Singleton first borns**
Panel A: OLS results	**C-section birth**	0.004	0.003
		(0.003)	(0.004)
	P-value		
	(cluster=province)	0.256	0.434
	Wild cluster bootstrap p-value		
	(cluster=province)	0.312	0.470
	Full controls	✓	✓
Panel B: Probit	**C-section birth**	0.003	0.002
	(marginal effect)	(0.002)	(0.002)
	Full controls	✓	✓
Panel C: 2SLS	**C-section birth**	0.140	0.068
		(0.090)	(0.098)
	First-stage F stat	12.124	11.245
	Reduced form	0.021*	0.013
		(0.011)	(0.018)
	Full controls	✓	✓
	Observations	20,800	8,800

Notes:

1- Full controls include indicators for gender, multiple birth, pre-term and post-term births, birthweight, indicators for maternal age at birth, two parents at home and density of population, family income to LICO ratio, indicators for age, as well as fixed effects for province and fiscal year of birth, indicators for smoking parents, breastfeeding in infancy, a series of pregnancy complications (high blood pressure, pregnancy diabetes, maternal smoking during pregnancy, prenatal care) and a series of maternal and paternal health conditions (asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes), as well as older siblings.

2- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10. P-values obtained by implementing a wild-cluster bootstrap-t procedure (described in Carmeron, Gerbach and Miller (2008)), are also presented to allow for errors be correlated across years within province.

3- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.

4- Observations rounded at the closest multiple of 100.

Table 3.9: Impact of C-section birth on the probability that a child regularly takes prescription medication

	Any prescription medication (0/1)			**Number of prescription medication**		
	LPM	LMP	2SLS	LPM	LPM	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
C-section birth	0.012**	0.009*	0.499**	0.016**	0.013*	0.523
	(0.005)	(0.005)	(0.230)	(0.007)	(0.007)	(0.329)
Boy	0.034***	0.034***	0.029***	0.040***	0.040***	0.035***
	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.007)
Multiple birth	-0.019	-0.017	-0.124**	-0.019	-0.016	-0.128*
	(0.016)	(0.016)	(0.054)	(0.018)	(0.018)	(0.075)
Pre-term birth	0.020**	0.014	-0.029	0.030***	0.023**	-0.022
(<259 days)	(0.009)	(0.009)	(0.023)	(0.011)	(0.011)	(0.032)
Post-term birth	0.095**	0.084*	0.035	0.137**	0.124**	0.074
(>293 days)	(0.041)	(0.043)	(0.044)	(0.061)	(0.062)	(0.063)
Birthweight	-0.014***	-0.014***	-0.020***	-0.016***	-0.015***	-0.021***
(kg)	(0.004)	(0.004)	(0.006)	(0.005)	(0.006)	(0.007)
Smoker at home		0.010*	-0.001		0.011*	0.000
		(0.005)	(0.007)		(0.006)	(0.009)
Rural area	-0.016***	-0.016***	-0.010	-0.018**	-0.017**	-0.011
	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)	(0.009)
Urban area	-0.010	-0.008	0.002	-0.011	-0.009	0.002
(500,000+)	(0.007)	(0.007)	(0.009)	(0.008)	(0.008)	(0.011)
Two parents	-0.008	-0.003	0.009	-0.006	-0.000	0.013
	(0.009)	(0.009)	(0.012)	(0.011)	(0.012)	(0.015)
Income/ LICO	-0.001	0.000	0.001	-0.001	0.000	0.001
(adjusted)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Pregnancy		0.005	-0.029		0.008	-0.028
diabetes		(0.008)	(0.018)		(0.010)	(0.025)
Pregnancy		0.002	-0.035*		0.003	-0.036
high blood pr.		(0.008)	(0.020)		(0.010)	(0.028)
Pregnancy		-0.013*	-0.003		-0.010	-0.000
smoking		(0.008)	(0.010)		(0.009)	(0.013)
Prenatal care		-0.009	-0.009		-0.008	-0.008
		(0.024)	(0.028)		(0.029)	(0.033)
Prov & Year of birth FE	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓
Maternal age FE	✓	✓	✓	✓	✓	✓
Parents' health		✓	✓		✓	✓
Older siblings	✓	✓		✓	✓	
First stage F			11.17			11.17
Reduced form			0.070***			0.073*
			(0.020)			(0.038)
Observations	21,400	21,400	21,400	21,400	21,400	21,400

Notes:
1- Maternal and paternal health conditions include indicators for asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes.
2- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10.
3- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.
4- Low-income cutoff (LICO) is adjusted for year, family size and geographic area.
5- Observations rounded at the closest multiple of 100.

Table 3.10: Impact of C-section birth on the probability that a child regularly takes prescription medication

		All births	Singleton first borns
Panel A: OLS results	**C-section birth**	0.009*	0.011
		(0.005)	(0.008)
	P-value		
	(cluster=province)	0.027	0.176
	Wild cluster bootstrap p-value		
	(cluster=province)	0.079	0.246
	Full controls	✓	✓
Panel B: Probit	**C-section birth**	0.008*	0.008
	(marginal effect)	(0.005)	(0.007)
	Full controls	✓	✓
Panel C: 2SLS	**C-section birth**	0.499**	0.515*
		(0.230)	(0.313)
	First-stage F stat	11.168	9.444
	Reduced form	0.070***	0.086**
		(0.020)	(0.042)
	Full controls	✓	✓
	Observations	21,400	9,000

Notes:

1- Full controls include indicators for gender, multiple birth, pre-term and post-term births, birthweight, indicators for maternal age at birth, two parents at home and density of population, family income to LICO ratio, indicators for age, as well as fixed effects for province and fiscal year of birth, indicators for smoking parents, breastfeeding in infancy, a series of pregnancy complications (high blood pressure, pregnancy diabetes, maternal smoking during pregnancy, prenatal care) and a series of maternal and paternal health conditions (asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes), as well as older siblings.

2- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10. P-values obtained by implementing a wild-cluster bootstrap-t procedure (described in Carmeron, Gerbach and Miller (2008)), are also presented to allow for errors be correlated across years within province.

3- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.

4- Observations rounded at the closest multiple of 100.

Table 3.11: Impact of C-section birth on the probability that a child has asthma

	Asthma diagnosis			**Asthma impairing daily activities**		
	LPM	LMP	2SLS	LPM	LPM	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
C-section birth	0.012***	0.010**	0.085	0.004**	0.004*	0.058
	(0.004)	(0.004)	(0.228)	(0.002)	(0.002)	(0.063)
Boy	0.047***	0.047***	0.046***	0.007***	0.008***	0.007***
	(0.005)	(0.005)	(0.006)	(0.002)	(0.002)	(0.002)
Multiple birth	-0.014	-0.014	-0.030	-0.012***	-0.011***	-0.023
	(0.016)	(0.016)	(0.049)	(0.004)	(0.004)	(0.015)
Pre-term birth	0.042***	0.036***	0.029	0.003	0.002	-0.003
(<259 days)	(0.008)	(0.008)	(0.022)	(0.003)	(0.003)	(0.007)
Post-term birth	0.074**	0.060*	0.053	0.039	0.037	0.031
(>293 days)	(0.036)	(0.035)	(0.038)	(0.025)	(0.025)	(0.026)
Birthweight	-0.007*	-0.008*	-0.008*	-0.003	-0.003*	-0.004*
(kg)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)
Smoker at home		0.012***	0.010*		0.002	0.001
		(0.004)	(0.005)		(0.002)	(0.002)
Rural area	-0.014**	-0.014**	-0.013**	0.001	0.001	0.002
	(0.006)	(0.006)	(0.006)	(0.002)	(0.002)	(0.002)
Urban area	-0.007	-0.004	-0.002	-0.000	0.000	0.001
(500,000+)	(0.006)	(0.005)	(0.007)	(0.002)	(0.002)	(0.002)
Two parents	-0.024***	-0.017**	-0.015	-0.003	-0.001	0.000
	(0.008)	(0.008)	(0.010)	(0.005)	(0.005)	(0.005)
Income/ LICO	-0.002**	-0.000	-0.000	-0.001**	-0.001	-0.001
(adjusted)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Pregnancy		0.017*	0.012		0.002	-0.002
diabetes		(0.010)	(0.017)		(0.004)	(0.006)
Pregnancy		0.009	0.003		0.001	-0.003
high blood pr.		(0.006)	(0.018)		(0.003)	(0.006)
Pregnancy		-0.007	-0.006		-0.003	-0.002
smoking		(0.006)	(0.008)		(0.003)	(0.003)
Prenatal care		-0.002	-0.002		-0.006	-0.006
		(0.016)	(0.016)		(0.011)	(0.011)
Prov & Year of birth FE	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓
Maternal age FE	✓	✓	✓	✓	✓	✓
Parents' health		✓	✓		✓	✓
Older siblings		✓	✓		✓	✓
First stage F			11.19		11.14	
Reduced form			0.012			0.008
			(0.031)			
Observations	21,400	21,400	21,400	21,400	21,400	21,400

Notes:
1- Information on asthma impairing daily activities (columns 4-6) is collected only from respondents with a diagnosis of asthma.
2- Maternal and paternal health conditions include indicators for asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes.
3- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10.
4- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.
5- Low-income cutoff (LICO) is adjusted for year, family size and geographic area.
6- Observations rounded at the closest multiple of 100.

Table 3.12: Impact of C-section birth on the probability that a child has a diagnosis of asthma

		All births	Singleton first borns
Panel A: OLS results	**C-section birth**	0.010**	0.009
		(0.004)	(0.007)
	P-value		
	(cluster=province)	0.011	0.081
	Wild cluster bootstrap p-value		
	(cluster=province)	0.038	0.114
	Full controls	✓	✓
Panel B: Probit	**C-section birth**	0.009**	0.008
	(marginal effect)	(0.004)	(0.006)
	Full controls	✓	✓
Panel C: 2SLS	**C-section birth**	0.085	0.323
		(0.228)	(0.321)
	First-stage F stat	11.19	9.359
	Reduced form	0.012	0.054
		(0.031)	(0.048)
	Full controls	✓	✓
	Observations	21,400	9,000

Notes:
1- Full controls include indicators for gender, multiple birth, pre-term and post-term births, birthweight, indicators for maternal age at birth, two parents at home and density of population, family income to LICO ratio, indicators for age, as well as fixed effects for province and fiscal year of birth, indicators for smoking parents, breastfeeding in infancy, a series of pregnancy complications (high blood pressure, pregnancy diabetes, maternal smoking during pregnancy, prenatal care) and a series of maternal and paternal health conditions (asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes), as well as older siblings.
2- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10. P-values obtained by implementing a wild-cluster bootstrap-t procedure (described in Carmeron, Gerbach and Miller (2008)), are also presented to allow for errors be correlated across years within province.
3- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.
4- Observations rounded at the closest multiple of 100.

Table 3.13: Impact of C-section birth on the probability that a child asthma that hinders his participation in daily activities

		All births	Singleton first borns
Panel A: OLS results	**C-section birth**	0.004***	0.003
		(0.002)	(0.003)
	P-value		
	(cluster=province)	0.002	0.097
	Wild cluster bootstrap p-value		
	(cluster=province)	0.002	0.146
	Full controls	✓	✓
Panel B: Probit	**C-section birth**	0.003**	0.002
	(marginal effect)	(0.001)	(0.002)
	Full controls	✓	✓
Panel C: 2SLS	**C-section birth**	0.058	0.148
		(0.063)	(0.112)
	First-stage F stat	11.143	9.359
	Reduced form	0.008	0.025
		(0.008)	(0.016)
	Full controls	✓	✓
	Observations	21,400	9,000

Notes:

1- Full controls include indicators for gender, multiple birth, pre-term and post-term births, birthweight, indicators for maternal age at birth, two parents at home and density of population, family income to LICO ratio, indicators for age, as well as fixed effects for province and fiscal year of birth, indicators for smoking parents, breastfeeding in infancy, a series of pregnancy complications (high blood pressure, pregnancy diabetes, maternal smoking during pregnancy, prenatal care) and a series of maternal and paternal health conditions (asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes), as well as older siblings.

2- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10. P-values obtained by implementing a wild-cluster bootstrap-t procedure (described in Carmeron, Gerbach and Miller (2008)), are also presented to allow for errors be correlated across years within province.

3- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.

4- Observations rounded at the closest multiple of 100.

Table 3.14: Impact of C-section birth on the probability that a child experienced wheezing or whistling in the chest in the year preceding the interview

		All births	Singleton first borns
Panel A: OLS results	**C-section birth**	0.013**	0.010
		(0.007)	(0.008)
	P-value		
	(cluster=province)	0.015	0.177
	Wild cluster bootstrap p-value		
	(cluster=province)	0.044	0.202
	Full controls	✓	✓
Panel B: Probit	**C-section birth**	0.013**	0.010
	(marginal effect)	(0.006)	(0.008)
	Full controls	✓	✓
Panel C: 2SLS	**C-section birth**	0.130	0.239
		(0.266)	(0.341)
	First-stage F stat	10.973	9.302
	Reduced form	0.018	0.040
		(0.036)	(0.054)
	Full controls	✓	✓
	Observations	21,400	9,000

Notes:

1- Full controls include indicators for gender, multiple birth, pre-term and post-term births, birthweight, indicators for maternal age at birth, two parents at home and density of population, family income to LICO ratio, indicators for age, as well as fixed effects for province and fiscal year of birth, indicators for smoking parents, breastfeeding in infancy, a series of pregnancy complications (high blood pressure, pregnancy diabetes, maternal smoking during pregnancy, prenatal care) and a series of maternal and paternal health conditions (asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes), as well as older siblings.

2- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10. P-values obtained by implementing a wild-cluster bootstrap-t procedure (described in Carmeron, Gerbach and Miller (2008)), are also presented to allow for errors be correlated across years within province.

3- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.

4- Observations rounded at the closest multiple of 100.

Table 3.15: Impact of C-section birth on a child's BMI

	Bmi $\geq$ 95th pctile			**Bmi**		
	(1)	(2)	(3)	(4)	(5)	(6)
C-section birth	0.017*	0.018*	0.285	0.137	0.149*	0.897
	(0.010)	(0.010)	(0.295)	(0.086)	(0.085)	(2.232)
Boy	0.048***	0.049***	0.044***	0.185**	0.187***	0.175**
	(0.008)	(0.008)	(0.009)	(0.071)	(0.070)	(0.076)
Mutiple birth	0.003	0.002	-0.059	-0.130	-0.138	-0.308
	(0.025)	(0.024)	(0.070)	(0.219)	(0.206)	(0.544)
Pre-term birth	0.017	0.011	-0.011	0.139	0.102	0.040
(<259 days)	(0.014)	(0.014)	(0.026)	(0.121)	(0.125)	(0.222)
Post-term birth	0.108	0.096	0.064	0.873	0.774	0.685
(>293 days)	(0.066)	(0.067)	(0.079)	(0.598)	(0.620)	(0.665)
Birthweight	0.046***	0.049***	0.045***	0.547***	0.591***	0.580***
(kg)	(0.009)	(0.009)	(0.009)	(0.073)	(0.074)	(0.084)
Smoker at home		0.059***	0.055***		0.478***	0.468***
		(0.008)	(0.009)		(0.082)	(0.086)
Rural area	0.022**	0.014	0.019	0.186**	0.123	0.136
	(0.010)	(0.010)	(0.012)	(0.086)	(0.083)	(0.096)
Urban area	-0.006	0.005	0.009	-0.032	0.065	0.078
(500,000+)	(0.009)	(0.009)	(0.011)	(0.075)	(0.072)	(0.089)
Two parents	-0.051***	-0.034**	-0.026	-0.632***	-0.487***	-0.464***
	(0.013)	(0.013)	(0.016)	(0.146)	(0.142)	(0.162)
Income/ LICO	-0.012***	-0.007***	-0.006**	-0.101***	-0.059***	-0.057***
(adjusted)	(0.002)	(0.002)	(0.002)	(0.020)	(0.019)	(0.020)
Pregnancy		0.040**	0.018		0.199	0.136
diabetes		(0.018)	(0.030)		(0.175)	(0.247)
Pregnancy		0.014	-0.004		0.167	0.116
high blood pr.		(0.015)	(0.025)		(0.119)	(0.204)
Pregnancy		0.052***	0.056***		0.559***	0.570***
smoking		(0.011)	(0.012)		(0.112)	(0.115)
Prenatal care		-0.029	-0.038		-0.240	-0.266
		(0.048)	(0.051)		(0.367)	(0.369)
Prov & Year of birth FE	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓
Maternal age FE	✓	✓	✓	✓	✓	✓
Parents' health		✓	✓		✓	✓
Older siblings		✓	✓		✓	✓
First stage coefficient			0.161***			0.161***
			(0.048)			(0.048)
First stage F stat			11.34			11.34
Reduced form			0.046			0.154
			(0.046)			(0.372)
Observations	13,100	13,100	13,100	13,100	13,100	13,100

Notes:
1- Bmi is estimated using reported information on children's height and weight.
2- Maternal and paternal health conditions include indicators for asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes.
3- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10.
4- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.
5- Low-income cutoff (LICO) is adjusted for year, family size and geographic area.
6- Sample is restricted to children aged 2 and older, in concordance with the CDC's BMI-for-age growth charts used to define obesity.
7- Observations rounded at the closest multiple of 100.

Table 3.16: Impact of C-section birth on a child's bmi

		All births	**Singleton first borns**
Panel A: OLS results	**C-section birth**	0.149*	0.241**
		(0.085)	(0.113)
	P-value		
	(cluster=province)	0.214	0.209
	Wild cluster bootstrap p-value		
	(cluster=province)	0.348	0.320
	Full controls	✓	✓
Panel B: 2SLS	**C-section birth**	0.897	4.113
		(2.232)	(2.884)
	First Stage		
	F stat	11.339	10.532
	Coefficient on instrument	0.161***	0.243***
		(0.048)	(0.075)
	Reduced form	0.145	1.001
		(0.372)	(0.667)
	Full controls	✓	✓
	Observations	13,100	5,700

Notes:

1- Full controls include indicators for gender, multiple birth, pre-term and post-term births, birthweight, indicators for maternal age at birth, two parents at home and density of population, family income to LICO ratio, indicators for age, as well as fixed effects for province and fiscal year of birth, indicators for smoking parents, breastfeeding in infancy, a series of pregnancy complications (high blood pressure, pregnancy diabetes, maternal smoking during pregnancy, prenatal care) and a series of maternal and paternal health conditions (asthma, chronic bronchitis, sinusitis, high blood pressure and diabetes), as well as older siblings.

2- Standard errors (in parenthesis) are clustered at the province-year level. Significance levels: ***=0.01, **=0.05, *=0.10. P-values obtained by implementing a wild-cluster bootstrap-t procedure (described in Carmeron, Gerbach and Miller (2008)), are also presented to allow for errors be correlated across years within province.

3- Sample limited to 1 observation per child (oldest), with all information related to pregnancy and birth delivery carried from the earliest observation for each child.

4- Observations rounded at the closest multiple of 100.

5- Sample is limited to children aged 2 and older, in concordance with the CDC's BMI-for-age growth charts used to derive the relevant percentiles

Appendix

Two-Sample Two-Stage Least Squares

In an attempt to improve on and complement the two-stage least squares estimates presented in the main section of this paper, the impact of C-section birth on health outcomes in childhood is also estimated using a two-sample two-stage least squares (TS2SLS) strategy [Angrist and Krueger, 1992]. This approach is implemented to (i) consider alternative health outcomes available in a dataset (the Canadian Health Measures Survey, or CHMS) that does not provide information on each child's birth delivery method, and (ii) increase the sample size in the first stage of the estimation to potentially improve the precision of the estimates. Overall, the LATE interpretation of the 2SLS estimates discussed in the main sections of the paper also applies to the estimates developed below.

The Hospital Morbidity Database (HMDB) is used to estimate the first stage of the TS2SLS model. The HMDB is held by the Canadian Institute for Health Information and contains administrative information on the population of births delivered in Canadian provinces, including information on the delivery method, on maternal health conditions and pregnancy risk factors, as well as on the medical services and care provided during the inpatient stay associated with the delivery and until the mother's discharge from the hospital. Records from the HMDB for the fiscal years 1994 to 2010 are used, with the exclusion of births in the province of Quebec, which is only included in the data set starting from 2006. These records are combined with information on physician fees to estimate the first stage of the TS2SLS. The sample overall consists of 3,980,160 births for which information on delivery, pregnancy and relative C-section fees is available[51] and that correspond to combinations of province-fiscal year of birth that are represented in the CHMS data, used in the second stage.

The second stage is estimated using the Canadian Health Measures Survey, a nationally representative survey administered by Statistics Canada in partnership with Health Canada and the Public Health Agency of Canada. The CHMS contains clinical information on various health outcomes for children, including body mass index (from which an indicator for obesity can be derived), and diabetes diagnosis.[52] The first wave of the survey was initiated between March 2007 and February 2009, and data from the third and most recent wave, was collected between January 2012 and December 2013. Designed to address the limitations of health information gathered in Canada, the data collection process behind the CHMS consists of home interviews documenting health conditions and socio-economic characteristics,[53] supplemented by a thorough analysis of medication prescribed to respondents and precise clinical measurements of various health-related variables and environment characteristics. In addition to documenting health outcomes that are not included in the NLSCY questionnaire, the CHMS also tracks the age at which certain conditions were diagnosed and provides consistently and rigorously measured

[51]Data on fees is also not available for the period covering the fiscal years 1994-96 and 1998-2000 in Alberta, for the period 1994-97 in Saskatchewan and Prince-Edward Island, and for the fiscal year 1994 in Nova Scotia and Newfoundland and Labrador.

[52]Other health conditions available in the NLSCY are also documented in the CHMS, but the TS2SLS results for these outcomes are less precise are not presented in table 3.A.2 for brevity. Other outcomes exclusively available in the CHMS data (such as cancer diagnosis) have very low prevalence rates, limiting the ability to consider them for the analysis.

[53]For children aged 3 to 11, a parent/guardian present during the interview and clinic visit provides some answers.

characteristics such as height and weight during clinic visits by the respondents (allowing for body mass index and obesity to be precisely recorded). It should be noted that given the time periods for the collection of data in each survey (1994 to 2008 for the NLSCY and 2007 to 2013 for the CHMS), the health outcomes for the cohorts born between 1994 and 2008 are generally observed at a later age in the CHMS than in the NLSCY. Hence, the results from a simple 2SLS analysis using the NLSCY and a TS2SLS using the CHMS can be thought of as providing answers to the impact of C-section birth on children's health in, respectively, the shorter– and longer-run.

Despite its strengths, the CHMS has certain limitations. First, it does not directly provide individual-level information on birth delivery methods (C-sections or vaginal deliveries). As such, the impact of C-section birth on health conditions in childhood cannot be investigated in this dataset using OLS or 2SLS estimators. However, the data does contain a series of pregnancy related variables that are also available in the HMDB, along with the province and fiscal year of birth of each respondent, making it a good candidate for the estimation of the second stage in the TS2SLS framework. A second shortcoming is that (unlike the NLSCY) the CHMS does not focus on children or youth: its target population is formed of Canadians aged 3 to 79[54], and the estimating sample from this data therefore excludes any child aged 0 to 2. Overall, when focusing only on children born after 1994 and for whom no relevant information is missing, the final CHMS sample comprises close to 3000 individuals.[55] Another limitation is that, for each wave of the CHMS, information is only collected in a limited number of provinces (between five and seven). Observations from a total of eight provinces are included in the final sample used for the TS2SLS analysis, which pools data from all available waves of the survey. The absence of children from two provinces limits the variation in the instrument for C-section birth used in the first stage of TS2SLS analysis, since physician fees are set by provincial authorities.[56]

The proposed two-sample two-stage least squares analysis can be summarized by equations (4) to (7). For simplicity, all exogenous variables and instruments are combined in $Z_k = [Fee_k, X_k']$, where $k = \{1, 2\}$ indicating the data set used for the first or second stage. The variables have the same definition as in equations (3.2) and (3.3).

$$\begin{aligned}
&\text{First stage:} && Csection_1 = \gamma_0 + Z_1'\gamma_1 + \upsilon_1 && (4)\\
&\text{First stage estimator:} && \hat{\gamma}_1 = [Z_1'Z_1]^{-1}Z_1'Csection_1 && (5)\\
&\text{Cross-sample fitted values:} && \widehat{Csection}_{21} = Z_2[Z_1'Z_1]^{-1}Z_1'Csection_1 && (6)\\
&\text{Second stage estimator:} && Health_2 = \alpha_0 + \alpha_1\widehat{Csection}_{21} + X_2'\alpha_2 + \epsilon_2 && (7)
\end{aligned}$$

A few conditions have to be satisfied for this approach to be successful.[57] First, the two samples used

[54] Cycle 1 only surveys Canadians aged 6 to 79. However, cycles 2 and 3 extend the population of study to include Canadians aged 3 to 5.

[55] Detailed information on sample composition in terms of birth cohorts is provided in table 3.A.1.

[56] Only observations for which the province and fiscal year of birth is available both in the CHMS and in the HMDB are kept for the TS2SLS estimation of the first and second stage.

[57] Inoue and Solon [2010] shows that under those conditions, the TS2SLS is more likely to yield asymptotically efficient estimates than the two-sample instrumental variable estimator.

should be representative of the same population: in this case, both the HMDB sample and the CHMS sample correspond to the population of children born in Canadian provinces between 1994 and 2010, with the aforementioned exclusion of certain province-year of birth combinations.Second, the data used in the first stage should include the instrument (relative payment for C-sections at the province-fiscal year level) and the endogenous right-hand side variable (C-section) while the data set for the second stage should include the instrument and the health outcomes of interest [Angrist and Krueger, 1992]. This condition is also satisfied using the HMDB and the CHMS since values for the instrument can be imputed in each dataset using each observation's date and province of birth. Third, the exogenous variables included in the model should also be available in both samples. This is the case for control variables related to pregnancy characteristics. Most socio-demographic characteristics available in the CHMS are not available in the HMDB, and are therefore not included in the specification.

The results from equation (4) are summarized at the bottom of table 3.A.2. The estimated effect suggests that, controlling for pre-term birth (gestational age of less than 259 days), post-term birth (gestational age of more than 293 days), multiple birth, a series of dummy variables for maternal age, as well as province and fiscal year of birth fixed effects, doubling the relative fee for a C-section compared to a vaginal delivery would increase the probability that a child is born by C-section by 5 percentage points.[58] The estimated impact is statistically significant at the 1% level. Using the full set of estimated first-stage coefficients, the predicted C-section rate in the CHMS sample for the period is 25.3%, close to the observed rate in the corresponding population of births, and to the C-section rate estimated in the NLSCY over a similar period.

Table 3.A.2 also presents the results from the second stage estimation for two outcomes: obesity and diabetes. The estimates from this second stage are subject to more problems than the 2SLS estimated presented in section 3.5. The estimated impact of C-section birth on the probability that a child is obese[59] suffers from severe imprecision, and its magnitude (1.237) is not credible. The estimated impact of C-section birth on the probability that a child suffers from diabetes is negative, and is again very imprecisely estimated. The negative coefficient is surprising, although the result is not statistically different from zero. The few occurrences of diabetic children in the sample could contribute to the imprecision with which the coefficient of interest is estimated.[60] The estimates obtained using other outcomes are also marked by an important lack of precision, even when compared to the 2SLS estimates obtained in the NLSCY. For example, the estimated probability that a child has received a diagnosis of asthma by age 5 corresponds to an increase of 42 percentage point (not shown), with a standard error twice as large as the estimated coefficient.

Overall, no gains are achieved by turning from the 2SLS estimator to this version of the TS2SLS esti-

[58]The estimated impact of relative physician fees on C-section birth is slightly larger (0.056) when the only controls included are province and fiscal year of birth fixed effects.

[59]BMI $\geq 95^{th}$ percentile of the child's age and gender group, according to the growth charts made available by the Center for Disease Control and Prevention

[60]The standard errors presented have not been corrected using the covariance matrix for the first stage disturbances, as this correction would require obtaining further authorization with respect to the movement of data and output between secure facilities. It is to be expected that making that correction would further increase the imprecision of the estimates, as suggested in Inoue and Solon [2010].

mator. The smaller sample size in the CHMS might have a role to play in generating imprecision in the estimation. However, it is likely the weakness of the instrument that is at the heart of the problem, something that cannot not resolved by increasing the size of the sample in the first stage. Finally, it should be noted that using the NLSCY to estimate the second stage also does not generate an improvement from the 2SLS results presented in the main section of this paper.

Table 3.A.1: Cohorts matching between the DAD/HMDB and he CHMS

DAD/HMDB		**CHMS**			**DAD/HMDB**		**CHMS**		
Year of birth		Cycle	Year	Age	Year of birth		Cycle	Year	Age
1994	⟶	Cycle 1	2007	13	2003	⟶	Cycle 1	2007	—
			2008	14				2008	–
			2009	15				2008	6
	⟶	Cycle 2	2009	15			Cycle 2	2009	6
			2010	16				2010	7
			2011	17				2011	8
		Cycle 3	2012	18			Cycle 3	2012	9
			2013	19				2013	10
1995	⟶	Cycle 1	2007	12	2004	⟶	Cycle 2	2009	5
			2008	13				2010	6
			2008	14				2011	7
		Cycle 2	2009	14					
			2010	15			Cycle 3	2012	8
			2011	16				2013	9
		Cycle 3	2012	17	2005	⟶	Cycle 2	2009	4
			2013	18				2010	5
.				.				2011	6
.				.			Cycle 3	2012	7
.				.				2013	8
2001	⟶	Cycle 1	2007	6	2006	⟶	Cycle 2	2009	3
			2008	7				2010	4
			2008	8				2011	5
		Cycle 2	2009	8			Cycle 3	2012	6
			2010	9				2013	7
			2011	10	2007	⟶	Cycle 2	2009	—
		Cycle 3	2012	11				2010	3
			2013	12				2011	4
2002	⟶	Cycle 1	2007	–			Cycle 3	2012	5
			2008	6				2013	6
			2008	7	2008	⟶	Cycle 2	2009	—
		Cycle 2	2009	7				2010	—
			2010	8				2011	3
			2011	9			Cycle 3	2012	4
		Cycle 3	2012	10				2013	5
			2013	11	2009	⟶	Cycle 3	2012	3
								2013	4
					2010	⟶	Cycle 3	2012	—
								2013	3

Table 3.A.2: Impact of C-section birth on children's health, TS2SLS estimates

	Obesity	**Diabetes**
C-section	1.237	-0.138
	(0.930)	(0.091)
Multiple birth	-0.042	0.042
(twins or more)	(0.029)	(0.028)
Pre-term birth	0.023	-0.006*
($<$ 259 days)	(0.029)	(0.003)
Post-term birth	0.008	-0.004
($>$293 days)	(0.057)	(0.002)
Province FE	✓	✓
Fiscal year FE	✓	✓
Maternal age	✓	✓
First stage	0.049***	0.049***
	(0.011)	(0.011)
Reduced form	0.153	-0.012
	(0.178)	(0.027)
Observations	2910	2920

Notes:

1- Significance levels: ***=0.01, **=0.05, *=0.10.

2- All standard errors are clustered at the province-year level but the variance covariance matrix has not yet been adjusted for the TS2SLS estimator [Angrist and Pischke [2009], Inoue and Solon [2010]].

3- Observations for combinations of province and fiscal-year available in the HMDB and the CHMS only.

References

T.R. Abrahamsson, H.E. Jakobsson, A.F. Andersson, B. Bjorksten, L. Engstrand, and M.C. Jenmalm. Low diversity of the gut gicrobiota in infants with atopic eczema. *Journal of Allergy and Clinical Immunology*, 129(2), February 2012.

ACOG. Committee Opinion no 340: Mode of term singleton breech delivery. *Obstetrics and Gynecology*, 108(1):235–7, 2006.

ACOG. Committee Opinion No 559: Cesarean delivery on maternal request. Technical report, American Association of Obstetricians and Gynaecologists, April 2013.

Diane Alexander. Does physician compensation impact procedure choice and patient health? Evidence from medicaid C-section use. Working paper, Princeton University, 2015.

Diane Alexander. Does physician compensation impact procedure choice and patient health? Evidence from Medicaid C-section use. Working paper 2017-07, Federal Reserve Bank of Chicago, 2017.

Sara Allin, Michael Baker, Maripier Isabelle, and Mark Stabile. Physician incentives and the rise in C-section: Evidence from Canada. *NBER Working paper*, (21022), 2015.

C. Almqvist, S. Cnattingius, P. Lichtenstein, and C. Lundholm. The impact of birth mode of delivery on childhood asthma and allergic disease - A sibling study. *Clinical and Experimental Allergy*, 42, 2012.

Joshua D. Angrist and Alan B. Krueger. The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, 87, 1992.

Joshua D. Angrist and Jorn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, 2009.

M.B. Azad, T. Konya, R.R. Persaud, D.S. Guttman, R.S. Chari, C.J. Field, M.R. Sears, P.J. Mandhane, S.E. Turvey, P. Subbarao, A.B. Becker, J.A. Scott, and A.L. Kozyrskyj. Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: A prospective cohort study. *BJOG*, 6, 2015.

Meghan B. Azad, Theodore Konya, Heather Maughan, David S. Guttman, Catherine J. Field, Radha S. Chari, Malcolm R. Sears, Allan B. Becker, James A. Scott, and Anita L. Kozyrskyj. Gut microbiota of healthy canadian infants: Profiles by mode of delivery and infant diet at 4 months. *Canadian Medical Association Journal*, 185(5), February 2013.

P Bager, J. Wohlfahrt, and T. Westergaard. Caesarean delivery and risk of atopy and allergic disease: Meta-analyses. *Clinical and Experimental Allergy*, 38, 2008.

Katherine Baicker, Elliott S. Fischer, and Amitabh Chandra. Malpractice liability costs and the practice of medicine in the medicare program. *Health Affairs*, 26(3), May 2007.

Morris L Barer, Jonathan Lomas, and Claudia Sanmartin. Re-minding our Ps and Qs: Medical cost controls in Canada. *Health Affairs*, 15(2), 1996.

Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1), 2004.

Mairead Black, Siladitya Bhattacharya, Sam Philip, Jane E. Norman, and David J. McLernon. Planned cesarean delivery at term and adverse outcomes in childhood health. *Journal of the American Medical Association*, 314(21), December 2015.

Jan Blustein and Jianmeng Liu. Time to consider the risks of caesarean delivery for long term child health. *British Medical Journal*, 350, June 2015.

L. Braback, C. Ekeus, A.J. Lowe, and A. Hjern. Confounding with familial determinants affects the association between mode of delivery and childhood asthma medication - A national cohort study. *Allergy, Asthma and Clinical Immunology*, 9, 2013.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3), August 2008.

Canadian Health Services Research Foundation. Myth: C-sections are on the rise because more mothers are asking for them. *Journal of Health Services Research and Policy*, 17(1):63–64, 2013.

Emily Carrier, James D. Reschovsky, Michelle M. Mello, Ralph C. Mayrell, and David Katz. Physicians' fears of malpractice lawsuits are not assuaged by tort reforms. *Health Affairs*, 29(9), 2010.

Aaron B Caughey, Alison G. Cahill, Jeanne-Marie Guise, and Dwight J. Rouse. Safe prevention of the primary cesarean delivery (reaffirmed). Obstetric Care Consensus Series 1, American College of Obstetricians and Gynecologists and Society for Maternal-Fetal Medicine, March 2016.

Center for Chronic Disease Prevention and Control. Diabetes in Canada: Facts and figues from a public health perspective. Technical report, Public Health Agency of Canada, 2011.

Center for Disease Control and Prevention. Asthma and schools, 2016a. URL `http://www.cdc.gov/HealthySchools/asthma/`.

Center for Disease Control and Prevention. Asthma national health interview survey (nhis) data, March 2016b. URL `http://www.cdc.gov/asthma/nhis/default.htm`.

CFCP. Updates data release of the 2001 National Family Physicians Workforce Survey. Technical report, The College of Family Physicians of Canada, 2002.

Nils Chaillet, Eric Dube, Marylene Dugas, Diane Francoeur, Johanne Dube, Sonia Gagnon, Lucie Poitras, and Alexandre Dumont. Identifying barriers and facilitators towards implementing guidelines to reduce cesarean section rates in Quebec. *Bulletin of the World Health Organization*, 85(10), October 2007.

Benjamin T.B. Chan. The declining comprehensiveness of primary care. *Canadian Medical Association Journal*, 166(4):429–434, 2002.

Amitabh Chandra, David Cutler, and Zirui Song. Who ordered that? The economics of treatment choices in medical care. In Mark V. Pauly, Thomas G. McGuire, and Pedro P. Barros, editors, *Handbook of health economics*, volume 2. Elsevier, 2011.

Kenneth Chay and Michael Greenstone. Does air quality matter? Evidence from the housing market. *Journal of Political Economy*, 113(2), 2005.

Clara E. Cho and Mikael Norman. Cesarean section and development of the immune system in the offspring. *American Journal of Obstetrics and Gynecology*, 208(4), April 2013.

CIHI. Supply, distribution and migration of Canadian physicians. Annual reports, Canadian Institute for Health Information, 2000-2005.

CIHI. Giving birth in Canada: Providers of maternity and infant care. Technical report, Canadian Institute for Health Information, 2004.

CIHI. Canada's health care providers: 2005 chartbook, 2005.

CIHI. Giving birth in Canada: The costs. Report, Canadian Institute for Health Information, 2006.

CIHI. Canada's health care providers: 1997 to 2011 - Midwives, 2013a.

CIHI. Hospital births in Canada: A focus on women living in rural and remote areas. Technical report, Canadian Institute for Health Information, July 2013b.

CIHI. National health expenditure trends, 1975 to 2013. Chapter 3: Total health expenditures by use of funds. Technical report, Canadian Institute for Health Information, 2014.

CIHI. Supply, distribution and migration of Canadian physicians, 2015: Table 22.1. Data Tables, Canadian Institute for Health Information, 2016a.

CIHI. Discharge Abstract Database: Childbirth indicators by place of residence. Technical report, Canadian Institute for Health Information, 2016b.

CIHI. Inpatient hospitalizations, surgeries and newborn indicators, 2014-2015. Data tables, Canadian Institute for Health Information, 2016a.

CIHI. Physicians in Canada, 2015. National physician database 2014-2015. Data release, Canadian Institute for Health Information, 2016b.

CIHI. Inpatient hospitalizations, surgeries and newborn indicators, 2015-16. Technical report, Canadian Institute for Health Information, 2017.

Jeffrey Clemens and Joshua D. Gottlieb. Do physicians' financial incentives affect medical treatment and patient health. *American Economic Review*, 104(4), 2014.

Damien Contandriopoulos and Melanie Perroux. Fee increases and target income hypothesis: Data from Quebec on physicians' compensation and service volumes. *Healthcare Policy*, 9(2), 2013.

Thomas F. Crossley, Jeremiah Hurley, and Sung-Hee Jeon. Physician labour supply in Canada: A cohort analysis. *Health Economics*, 18(4), 2009.

Casey Crump, Diana Rivera, Rebecca London, Melinda Landay, Bill Erlendson, and Eunice Rodriguez. Chornic health conditions and school performance among children and youth. *Annals of Epidemiology*, 23(4), April 2013.

Janet Currie. Pollution and infant health. *Child Development Perspectives*, 7(4), 2013.

Janet Currie and Bentley MacLeod. First do no harm? Tort reforms and birth outcomes. *Quarterly Journal of Economics*, 123(2), May 2008.

Janet Currie and Bentley MacLeod. Diagnosis and unnecessary procedure use: Evidence from C-section. Working paper 18977, National Bureau of Economic Research, 2013.

Janet Currie and Reed Walker. Traffic congestion and infant health: Evidence from E-ZPass. *American Economics Journal: Applied Economics*, 3(1), January 2011.

Dana Dabelea, Elizabeth J. Mayer-Davis, Sharon Saydah, Giuseppina Imperatore, Barbara Linder, Jasmin Divers, Ronny Bell, Angela Badaru, Jennifer W. Talton, Tessa Crume, Angela D. Liese, Anwar T. Merchant, Jean M. Lawrence, Kristi Reynolds, Lawrence Dolan, Lenna L. Liu, and Richard F. Hamman. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. *Journal of the American Medical Association*, 311(17), 2014.

Betty-Ann Daviss, Kenneth C. Johnson, and Andre Lalonde. Evolving evidence since the term breech trial: Canadian response, European dissent, and potential solutions. *Journal of Obstetrics and Gynaecology Canada*, 32(3), March 2010.

Rose Anne Devlin and Sisira Sarma. Do physician remuneration schemes matter? The case of Canadian family physicians. *Journal of Health Economics*, 25(7), 2008.

Maria G. Dominguez-Bello, Elizabeth K. Costello, Monica Contreras, Magda Magris, Gilda Hidalgo, Noah Fierer, and Rob Knight. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences*, 107(26), 2010.

David Dranove. Physician-induced demand for childbirths. *Journal of Health Economics*, 13(1), 1994.

J.C. Herbert Emery, Chris Auld, and Mingsham Lu. Paying for physician services in Canada: The institutional, historical and policy contexts. Working paper 99-6, Institute of Health Economics, University of Calgary, 1999.

Robert G. Evans. Physician induced demand: Some empirical evidence and implications. In editor M. Perlman, editor, *The Economics of Health and Medical Care*, pages 162–173. MacMillan, London, 1974.

Colleen Flood and Bryan Thomas. Missing the mark on patient safety. In Ken Oliphant and Richard W. Wright, editor, *Medical Malpractice and Compensation in Global Perspective*. De Gruyter Berlin/Boston, 2012.

Jenny A. Gamble, Chris Creedy, Debra K. McCourt, Jane Weaver, and Sarah Beake. A critique of the literature on women's request for cesarean section. *Birth*, 34(4), December 2007.

Rochelle Garner and Dafna Kohen. Changes in the prevalence of asthma among Canadian children. *Health Reports*, 19(2), 2008.

Luz Gibbons, Jose M. Belizan, Jeremy A. Lauer, Ana P. Betran, Mario Merialdi, and Fernando Althabe. The global numbers and costs of additionally needed and unnecessary caesarean sections performed per year: Overuse as a barrier to universal coverage. Background Paper 30, World Health Report, 2010.

Marek Glezerman. Five years to the term breech trial: The rise and fall of a randomized control trial. *American Journal of Obstetrics and Gynecology*, 194(1), January 2006.

Marshall Godwin, Geoffrey Hodgetts, Rachelle Seguin, and Susan MacDonald. The Ontario family medicine resident cohort study: factors affecting residents' decisions to practise obstetrics. *Canadian Medical Association Journal*, 166(2):179–184, 2002.

Francois Goffinet, Marion Carayol, Jean-Michel Foidard, Sophie Alexander, Serge Uzan, Damien Subtil, and Gerard Berart. Is planned vaginal delivery for breech presentation at term still an option? Results of an observational prospective survey in France and Belgium. *American Journal of Obstetrics and Gynecology*, 194(4):1002–1011, 2006.

Olga Gomez and Gabriel Carrasquilla. Factors associated with unjustified cesarean section in four hospitals in Cali, Colombia. *International Journal of Quality in Health Care*, 11(5):385–89, 2012.

Elizabeth Goodrick and Gerald R. Salancik. Organizational discretion in responding to institutional practices: Hospitals and cesarean births. *Administrative Science Quarterly*, 41(1), March 1996.

Joshua D. Gottlieb and Jeffrey Clemens. When practicing isn't enough: Can human capital theory explain physicians' on-the-job investment. 16th IZA/SOLE Transatlantic Meeting of Labor Economists, June 2017.

Darren Grant. Physician financial incentives and cesarean delivery: New conclusions from the healthcare cost and utilization project. *Journal of Health Economics*, 28(1), 2009.

Jonathan Gruber and Maria Owings. Physician financial incentives and cesarean section delivery. *The RAND Journal of Economics*, 27(1):99–123, 1996.

Jonathan Gruber, John Kim, and Dina Mayzlin. Physician fees and procedure intensity: The case of cesarean delivery. *Journal of Health Economics*, 18(4), August 1999.

Gillian E. Hanley, Patricia A. Janssen, and Devon Greyson. Regional variation in the cesarean delivery and assisted vaginal delivery rates. *Obstetrics and Gynecology*, 115(6), June 2010.

Mary E. Hannah, Walter J. Hannah, Sheila A. Hewson, Ellen D. Hodnett, Saigal Saroj, and Andrew R. Willan. Planned cesarean section versus planned vaginal birth for breech presentation at term: A randomized multicentre trial. *Lancet*, 356(9239), October 2000.

Julie E. Hartnack, Steen Rasmussen, and Lone Krebs. Consequences of the term breech trial in Denmark. *Acta Obstetricia et Gynecologica Scandinavica*, 90(7), July 2011.

John C. Hauth and Gary Cunningham. Vaginal breech delivery is still justified. *Obstetrics and Gynecology*, 99(6), June 2002.

HCUP. Most common operations during inpatient stays, 2015. web, 2017.

David A. Henry, Susan E. Schultz, Richard H. Glazier, R. Sacha Bhatia, Irfan A. Dhalia, and Andreas Laupacis. Payments to Ontario physicians from Ministry of Health and Long-Term Care sources 1992/93 to 2009/10. Ices investigative report, Institute for Clinical Evaluative Sciences, February 2012.

Marcus J. Hollander and Angela Tessaro. Evaluation of the full service family practice incentive program and the practice support program. Report, Hollander Analytical Services Ltd, June 2009.

Jeremiah Hurley and Sung-Hee Jeon. The relationship between physician labour supply. service volume and service intensity. *Canadian Public Policy*, 33, Supplement 2007.

Brian Hutchison, Jean-Frederic Levesque, Erin Strumpf, and Natalie Coyle. Primary health care in Canada: systems in motion. *Milbank quarterly*, 89(2):256–288, 2011.

Peter Hutten-Czapski, J. Roger Pitblado, and Steve Slade. Scope of family practice in rural and urban settings. *Canadian Family Physician*, 50(11), 2004.

Atsushi Inoue and Gary Solon. Two-sample instrumental variable estimator. *Review of Economics and Statistics*, 92(3), 2010.

Kristen D. Jackson, LaJeana D. Howie, and Lara J. Akinbami. Trend in allergic conditions among children: United-States, 1997-2011. Technical Report 121, National Center for Health Statistics, Centers for Disease Control and Prevention, May 2013.

Vibeke Myrup Jensen. Happy doctor makes happy baby? Incentivizing physicians improves quality of prenatal care. *Review of Economics and Statistics*, 96(5), December 2014.

Vania Jimenez. La Cesarienne sur demande: Une manifestation de la culture de la peur. *Le medecin du Quebec*, 40(7):65–69, July 2005.

Erin M. Johnson. Physician-induced demand. In *Encyclopedia of Health Economics*. Elsevier, 2014.

Erin M. Johnson and Marit Rehavi. Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, 8(1), 2016.

K. S. Joseph, Alexander C. Allen, Linda Dodds, Linda Ann Turner, Heather Scott, and Robert Liston. The perinatal effects of delayed childbearing. *Obstetrics and Gynecology*, 105(6):1410–18, 2005.

Jasmin Kantarevic, Boris Krajl, and Darrel Weinkauf. Income effect and physician labour supply: Evidence from the threshold system in Ontario. *Canadian Journal of Economics*, 41(4):1262–1284, November 2008.

Jasmin Kantarevic, Boris Krajl, and Darrel Weinkauf. Enhanced fee-for-service model and physician productivity: Evidence from Family Health Groups in Ontario. *Journal of Health Economics*, 30(1), 2011.

Steven Katz, Cathy Charles, Jonathan Lomas, and H. Gilbert Welch. Physician relations in Canada: Shooting inward as the circle closes. *Journal of Health Politics, Policy and Law*, 22(6), November 1997.

Emmett B. Keeler and Mollyann Brodie. Economic incentives in the choice between vaginal delivery and cesarean section. *The Milbank Quarterly*, 71(3):365–404, 1993.

Beomsoo Kim. The impact of malpractice risk on the use of obstetrics procedures. *The Journal of Legal Studies*, 36(S2), June 2007.

Dafna E. Kohen. Asthma and school functioning. *Health Reports*, 21(4), December 2010.

Andrew Kotaska. Inappropriate use of randomized trials to evaluate complex phenomena: Case study of vaginal breech delivery. *British Medical Journal*, 329, October 2004.

M. Ruth Lavergne, Sandra Peterson, Rachael McKendry, Saskia Sivananthan, and Kimberlyn McGrail. Full-service family practice in British-Columbia: Policy interventions and trends in practice 1991-2010. *Healthcare Policy*, 9(4):32–47, 2014.

Ann G. Lawthers, A. Russell Localio, Nan M. Laird, Stuart Lipsitz, Liesi Hebert, and Troyen A. Brennan. Physicians' perceptions of the risk of being sued. *Journal of Health Politics, Policy and Law*, 17(3), Fall 1992.

Kira Leeb, Akerke Baibergenova, Eugene Wen, Greg Webster, and Jennifer Zelmer. Are there socio-economic differences in cesarean section rates in Canada? *Healthcare Policy*, 1(1), 2005.

Melanie Lefevre. *Physician induced demand for C-sections: Does convenience incentive matter?* PhD thesis, Boston University and Universite de Liege, 2013.

Juan Andree Leon, Michael Kramer, Reg S. Sauve, and Mark Walker. In due time: Why maternal age matters. Analysis in brief, Canadian Institute for Health Information, September 2011.

Arthur Lewbel. An overview of the special regressor method. In Aman Ullah, Jeffrey Racine and Liangjun Su, editor, *Handbook of applied nonparametric and semi-parametric econometrics and statistics*. Oxford University Press, 2012.

Jinhu Li, Jeremiah Hurley, Philip DeCicca, and Gioia Buckley. Physician response to pay-for-performance: evidence from a natural experiment. *Health Economics*, 23(8):962–978, 2013.

Clare Liddy, Jatinerpreet Singh, Ryan Kelly, Simone Dahrouge, Monica Taljaard, and Jamie Younger. What is the impact of primary care model type on specialst referral rates? A cross-sectional study. *BMC Family Practice*, 15, February 2014.

S Liu, KS Joseph, M Heaman, R Sauve, and MS Kramer. Maternal mortality and severe morbidity associated with low-risk planned cesarean delivery versus planned vaginal delivery at term. *Canadian Medical Association Journal*, 176, 2007.

T. Liu, R. Valdez, PW. Yoon, D. Crocker, R. Mooneshinghe, and MJ. Khoury. The association between family history of asthma and the prevalence of asthma among U.S. adults: National Health and Nutrition Examination Survey, 1999-2004. *Genetics in Medicine*, 11(5), May 2009.

Jonathan Lomas, Murray Enkin, Geoffrey M. Anderson, Eugene Vayda, and Joel Singer. Opinion leaders vs audit and feedback to implement practice guidelines: Delivery after previous Cesarean section. *Journal of the American Medical Association*, 265(17), May 1991.

A. Maitra, A. Sherriff, D. Strachan, and J. Henderson. Mode of delivery is not associated with asthma or atopy in childhood. *Clinical and Experimental Allergy*, 34, 2004.

Erin L. Marcotte, Thomas P. Thomopoulos, Claire Infante-Rivard, Jacqueline Clavel, Eleni Th. Petridou, Joachim Schuz, Sameera Ezzat, Beth A. Mueller, Ana M. Mora, Catharina Wesseling, Alkistis Skalkidou, Wafaa M. Rashed, Stephen S. Francis, Roula Arjouche, Friederike Erdmann, Laurent Orsi, and Logan Spector. Caesarean delivery and risk of childhood leukemia: A pooled analysis from the childhood leukemia international consortium. *The Lancet - Hematology*, 3(4), April 2016.

Craig L. Maynard, Charles O. Elson, Robin D. Hatton, and Casey T. Weaver. Reciprocal interactions of the intestinal microbiota and immune system. *Nature*, 489, September 2012.

Kimberly W. McDermott, Anne Elixhauser, and Ruirui Sun. Trends in hospital inpatient stays in the United States 2005-2014. Statistical brief 225, HCUP, 2017.

Thomas G. McGuire and Mark V. Pauly. Physician response to fee changes with mutiple payers. *Journal of Health Economics*, 10(4), 1991.

A.M.B. Menezes, P.C. Hallal, A.M. Matijasevich, A.J.D. Barros, B.L. Horta, C.L.P. Araujo, D.P. Gigante, I.S. Santos, G. Minten, M.R. Domingues, S.C. Dumith, and F.C. Barros. Caesarean sections and risk of wheezing in childhood and adolescence: Data from two birth cohort studies in Brazil. *Clinical and Experimental Allergy*, 41(2), February 2011.

Gabrielle F. Miller, Edward Coffied, Zanie Leroy, and Robin Wallin. Prevalence and costs of five chornic conditions in children. *Journal of School Nursing*, April 2016.

MOHLTC. Changes to physician payments for 1998/99. Bulletin 4309, Ontario Ministry of Health and Long Term Care, 1998.

Josef Neu and Jona Rushing. Cesarean versus vaginal delivery: Long-term infant outcomes and the hygiene hypothesis. *Clinics in Perinatalogy*, 38(2), June 2011.

NICE. Caesarean cection. Clinical guideline 132. Technical report, National Institute for Health and Clinical Excellence, 2012.

NICE. Advice on offering caesarean section in specific situations. Pathway, National Institute for Health and Clinical Excellence, February 2014.

Sean Nicholson and Carol Propper. Medical workforce. In Mark V. Pauly, Thomas G. McGuire, and Pedro P. Barros, editors, *Handbook of health economics*, volume 2. Elsevier, 2011.

NPS. 2013 Canadian Survey results. Technical report, National Physician Survey, 2013.

C.L. Odgen, M.D. Carroll, B.K. Kit, and K.M. Flegal. Prevalence of childhood and adult obesity in the United-States; 2011-12. *Journal of the American Medical Association*, 311(8), 2014.

OMA and MOHLTC. Physician service agreement 2000-2004. Technical report, Ontario Medical Association and Ministry of Health and Long Term Care, 2000.

Ontario Medical Association. Central tariff committee recommendation. Communications with the OMA, 1999.

Anne Pfuntner, Lauren M. Wier, and Carol Stocks. Most frequent procedures performed in U.S. hospitals. HCUP Statistical Brief 165, Agency for Healthcare Research and Quality, 2013.

Robert W. Pong. Putting away the stethoscope for good? Toward a new perspective on physician retirement. Technical report, Canadian Institute for Health Information, 2011.

Public Health Agency of Canada. Canada perinatal health report. Technical report, Government of Canada, 2008.

Thomas H. Rice and Roberta J. Labelle. Do physicians induce demand for medical services? *Journal of Health Politics, Policy and Law*, 14(3), 1989.

Christine C. Rietberg, Patty M. Elferink-Stinkens, and H. A. Visser, Gerard. The effect of the term breech trial on medical intervention behaviour and neonatal outcome in the Netherlands: An analysis of 35,453 term breech infants. *BJOG An International Journal of Obstetrics and Gynaecology*, 112 (2), March 2005.

John A. Rizzo and David Blumenthal. Physician labor supply: Do income effects matter? *Journal of Health Economics*, 13(2), 1994.

John A. Rizzo and David Blumenthal. Is the target income hypothesis an economic heresy? *Medical Care Research and Review*, 53(3), 1996.

John A. Rizzo and Richard J. Zeckhauser. Reference incomes, loss aversion, and physician behavior. *Review of Economics and Statistics*, 85(4):909–922, 2003.

Celia Rodd and Atul K. Sharma. Recent trends in the prevalence of overweight and obesity among Canadian children. *Canadian Medical Association Journal*, May 2016.

David Rudoler, Raisa Deber, Janet Barnsley, Richard H. Glazier, Adrian Rohit Dass, and Audrey Laporte. Paying for primary care: the factors associated with physician self-selection into payment models. *Health Economics*, 24(9):1229–1242, 2015.

S. Salminen, G.R. Gibson, A.L. McCartney, and E. Isolauri. Influence of mode of delivery on gut microbiota composition in seven year old children. *Gut*, 53, 2004.

Astrid Sevelsted, Jakob Stokholm, Klaus Bonnelykke, and Hans Bisgaard. Cesarean section and chronic immune disorder. *Pediatrics*, 135(1), January 2015.

Elizabeth L. Shearer. Cesarean section: Medical benefits and costs. *Social Science and Medicine*, 37 (10):1223–1231, November 1993.

Society of Obstetricians and Gynaecologists of Canada. Caesarean section. Web, March 2009. URL `http://pregnancy.sogc.org/labour-and-childbirth/caesarean-section/`.

SOGC. Indications for cesarean section: Final statement of the panel of the national consensus conference on aspects of cesarean birth. Consensus reference report, Society of Obstetricians and Gynaecologists of Canada, June 1986.

SOGC. Policy statement on the role of obsetricians/gynaecologists in Canada. *Journal of Obstetrics and Gynaecology of Canada*, (42), November 1995.

SOGC. Policy statement on midwifery. *Journal of Obstetrics and Gynaecology of Canada*, 66, 1997a.

SOGC. Policy statement on vaginal birth after previous caesarean birth. *Journal of Obstetrics and Gynaecology of Canada*, 68, December 1997b.

SOGC. Consensus statement on the management of twin pregnancies. *Journal of Obstetrics and Gynaecology of Canada*, 91, July 2000.

SOGC. Statement on vaginal breech. Press release, March 2001.

SOGC. Cesarean section on demand - SOGC's position. Press release, Society of Obstetricians and Gynecologists of Canada, March 2004a.

SOGC. Guideline for operative vaginal birth. *Journal of Obstetrics and Gynaecology of Canada*, 148, August 2004b.

SOGC. Policy statement on the role of obstetricians and gynaecologists in Canada. *Journal of Obstetrics and Gynaecology of Canada*, 229, July 2009.

SOGC and CAM. Policy statement on midwifery. *Journal of Obstetrics and Gynaecology of Canada*, 227, July 2009.

SOGC, AWHONN, CAM, CFCP, and SRPC. Joint policy statement on normal birth. *Journal of Obstetrics and Gynaecology of Canada*, 221, December 2008.

Special Survey Division. National Longitudinal Survey of Children and Youth, cycle 8 - User guide. Technical report, Statistics Canada, 2009.

Douglas Staiger and James H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, May 1997.

Statistics Canada. Table 102-4516 - Live births and fetal deaths (stillbirths), by place of birth (hospital and non-hospital), Canada, provinces and territories, annual, 2016.

Statistics Canada. Table 102-4502 - Live births, by month, Canada, provinces and territories. Database, CANSIM, accessed: May 8, 2017.

James H. Stock and Motohiro Yogo. Testing for weak instruments in linear IV regression. In James H. Stock and Donald W. K. Andrews, editor, *Identification and inference for econometric models: Essays in honor of Thomas J. Rothenberg*. Cambridge University Press, 2005.

Elizabeth Anne Sullivan, Katinka Moran, and Michael Chapman. Term breech singletons and caesarean section: A population study, Australia 1991-2005. *Journal of Obstetrics and Gynaecology*, 49, 2009.

Patrick Sullivan. CMPA, OMA try to heal rift over regional rating. *Canadian Medical Association Journal*, 163(2), 2000.

Arthur Sweetman and Gioia Buckley. Ontario's experiment with primary care reform. *SPP Research Paper*, 7(11), 2014.

James Thornton. The labour supply behaviour of self-employed solo practice physicians. *Applied Economics*, 30(1), 1998.

James Thornton and B. Kelly Eakin. The utility-maximizing self-employed physician. *Journal of Human Resources*, 32(1), 1997.

Janet W. Tomeo, Catherine A. anand Rich-Edwards, Karin B. Michels, Catherine S. Berkey, David J. Hunter, A. Lindsay Frazier, Walter C. Willett, and Stephen L. Buka. Reproducibility and validity of maternal recall of pregnancy-related events. *Epidemiology*, 10(6), November 1999.

Yan Wang, Yuna Zhong, and Lani Wheeler. Direct and indirect costs of asthma in school-age children. *Preventing Chronic Disease*, 2(1), January 2005.

Jeffrey M. Wooldrige. On the two-stage least squares estimation of the average treatment effect in a random coefficient model. *Economic Letters*, 56, 1997.

World Health Organization. Appropriate technology for birth. *Lancet*, 2, August 1985.

World Health Organization. WHO statement on caesarean section rates, April 2015.

V.B. Young. Intestinal microbiota in health and disease. *Current Opinion in GastroEnterology*, 28(1), January 2012.

Changzhen Yuan, J. Gaskins, Audrey, Arianna I. Blaine, Cuilin Zhang, Matthew W. Gillman, Stacey A. Missmer, Alison E. Field, and Jorge E. Chavarro. Association between cesarean birth and risk of obesity in offspring in childhood, adolescence and early adulthood. *Journal of the American Medical Association: Pediatrics*, September 2016.