



HAL
open science

Essays in health economics

Alex Proshin

► **To cite this version:**

Alex Proshin. Essays in health economics. Economics and Finance. Université Panthéon-Sorbonne - Paris I, 2021. English. NNT : 2021PA01E020 . tel-03483880

HAL Id: tel-03483880

<https://theses.hal.science/tel-03483880>

Submitted on 16 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE
PARIS SCHOOL OF ECONOMICS

T H È S E

pour l'obtention du titre de Docteur en Sciences Économiques

Présentée et soutenue publiquement le 9 février 2021 par

Alex PROSHIN

Essays in health economics

Préparée sous la direction de Lise ROCHAIX et Audrey LAPORTE

Composition du jury :

Brigitte DORMONT	Professeure, Université Paris Dauphine	
Randall ELLIS	Professeur, Boston University	Rapporteur
Mark STABILE	Professeur, INSEAD	Rapporteur
Hélène HUBER-YAHI	Maître de conférences, PSE - Université Paris 1-Panthéon Sorbonne	Suffragante
Grégoire MERCIER	Médecin, CHU Montpellier	Suffragant
Lise ROCHAIX	Professeure, PSE - Université Paris 1-Panthéon Sorbonne	Directrice
Audrey LAPORTE	Professeure, University of Toronto	Co-directrice

L'Université Paris 1 Panthéon Sorbonne n'entend donner aucune approbation, ni improbation aux opinions émises dans cette thèse ; elles doivent être considérées comme propres à leur auteur.



Thanks

Saying that my PhD adventure has been quite a ride would clearly be a huge understatement. It has been a truly wild journey, to the point that if someone told me back in 2016 where it all would actually take me to, I would have reckoned this person to be totally out of their mind... During my years as a PhD candidate I have learned more than thought I could possibly do. The knowledge I acquired and the skills I learned not only fulfilled me both professionally and personally, but also, perhaps even more importantly, affected in innumerable ways my outlook on the world. And need I say that these 5 years condensed perhaps more events than quite a few people would live in their entire lives. And all these experiences were as unexpected and diverse as they came.

Despite all those twists and turns, one thing has always remained constant. Namely, support from my PhD supervisors – Lise and Audrey. I would love to thank both of them for their unwavering belief in my potential and success, on good days and bad ones. Without all this, I would not have become who I currently am both professionally and personally.

I express my gratitude to Jean-Claude, for his immense patience and devotedly hard work, which made our days at Hospinnomics ever more smooth and comfortable. I thank H el ene Huber-Yahi - my second thesis committee member – for her intellectually deep insights which always kept my work improving. I thank Christine Meyer, for her detailed guidance, expertise, and the chance to learn the meaning of 'la f erule fran aise'. I express my gratitude to Alexandre Cazenave-Lacroutz, Adrian Rohit-Dass, Marie-Caroline Cl ement and Zeynep Or for their fruitful collaborations, constructive guidance and pieces of advice. I also extend my gratitude to the members of my defense committee – Brigitte Dormont, Gr egoire Mercier, Mark Stabile and Randall Ellis – for their challenging discussions and very constructive feedback from which this dissertation benefited enormously.

Needless to say, this journey would not have been as joyful and complete without my coworkers, who I thank all deeply. I would like to thank Setti for her assistance with countless practical issues, and in particular the unforgettable 2-days impromptu preparation for a language test, which risks never fading in my living memory. I thank Ivan,

who during his Masters and PhD probably became more Russian than my Russian friends. Thelma, for her precious advice on legal questions and Lucie for her warm encouragements. Antoine, for his responsiveness and remarkable tactfulness. Rolando, for his honesty and inspiring determination. Lorène and Estelle for helping me make good language choices in emails I was too scared to send... And, of course, Maria, Thomas, Emilie, Sameh, Fernando, Caroline, Léa, Marc-Antoine, and all Hospinnomics team for fascinating discussions in the 'salle d'attente' and mutual assistance when it was needed.

This journey – from the school desk up until today – would certainly not have been possible without Olga, to whom I owe everything and who has largely defined who I am and am proud to be. I bow to Anya, whose wisdom and resilience is monumental. Vova and Lena, whose optimism, benevolence, warmth and stunning perspicacity can convert even a hardened pessimist into a raging optimist. Thanks Miriam, my roommate, for being my supporter, and putting up with my late night test corrections and class preparations. Tuyana, for the fond memory of you sitting in the last row and cheerleading me when I was nervously teaching my first class in a cold Paris 1 classroom. I keep it dearly in my heart. Kolya for his zen mindfulness that I aspire to. Ilya, for your precious and very knowledgeable advice. Sergi, Tyler, Anna, Dasha, Jesus (no, not the one from the Bible), Bogdan, Polina, Lucine, Zhenya, Olya for your discussions that made my world so much brighter a place.

And, my parents, who have done such a lot that perhaps several theses will not be enough to describe it. I owe you a lot. Spasibo.

Table of Contents

General Introduction	9
1 Impact of Tariff Refinement on the Choice Between Scheduled Cesarean Section and Normal Delivery: Recent Evidence from France	17
1 Introduction and literature review	18
2 Obstetric care in France: institutional context	22
3 Data	28
4 Empirical strategy	31
5 Main results	39
6 Robustness checks	43
7 Discussion and policy implications	47
8 Conclusion	50
9 Appendixes	52
2 Impact of Quality-based Procedures on orthopedic care quantity and quality in Ontario Hospitals	69
1 Introduction and literature review	70
2 Pay-for-performance (P4P) and Prospective Payment Systems (PPS): re- view of existing research	72
3 Canadian healthcare context	79
4 Research questions	88
5 Data	89
6 Models and main results	92
7 Robustness checks	100
8 Discussion and conclusion	110
9 Appendixes	112

3	Impact of Child Subsidies on Child Health, Well-being and Investment in Child Human Capital: Evidence from Russian Longitudinal Monitoring Survey 2011-2017	155
1	Introduction and literature review	156
2	Maternity Capital institutional context	161
3	Data and descriptive statistics	165
4	Empirical strategy and main results	169
5	Robustness checks	180
6	Discussion and conclusion	188
7	Appendixes	191
	General conclusion	207
	Bibliography	213

General Introduction

A crisis is an opportunity riding the dangerous
wind

Chinese Proverb

In any moment of decision, the best thing you
can do is the right thing, the next best thing is
the wrong thing, and the worst thing you can do
is nothing.

Theodore Roosevelt, 26th U.S.

President

This PhD dissertation attempts to shed light on the challenges different policy-makers are facing while implementing public policies, looking at distributional and welfare effects of policy interventions. More specifically, what is the price an individual or society is willing to pay for individual health and well-being, and how to allocate these funds in a way that is conducive to achieving public policy goals? The goal of this introduction is to provide a summary of the thesis main results, and to set them in this broader economic perspective.

The three chapters of this PhD dissertation consider very different, yet in many ways comparable, public interventions. Chapter 1 investigates the effect of an increased level of tariff differentiation – called DRG refinement – on the rate of scheduled C-sections in France. While the second chapter also concentrates on hospital incentives and care providers' behavior, it addresses a different reform whereby the Ontario Ministry of Health aimed at introducing activity-based funding in orthopedic surgery via two separate components, namely Quality Based Procedures (QBPs) and the Health Based Allocation Model (HBAM). The last chapter of this dissertation goes beyond the hospital setting to evaluate the effects of a sizable maternity subsidy – the Maternity Capital (MC) –

on a wide array of child health and well-being outcomes, as well as general household consumption and spending patterns in Russia. The contribution of this PhD dissertation consists in providing new insights on factors that may contribute to the success, or lead to the failure, of public policies in the health sector. In particular, we show the need to design reforms with well-aligned incentives for all stakeholders, as well as to ensure a clear communication of policy goals to mitigate possible information asymmetries. The thesis also emphasizes the need to anticipate unintended and collateral consequences of public policies, which result from inherently different motivations of economic agents and which can have considerable distributional consequences.

In theoretical economics, the traditional approach to modelling individual behavior has been to consider the individual as a utility maximizer, faced with constraints such as scarcity of resources and/or time, as well as various degrees of personal tolerance to uncertainty (for example, with regards to the future). Since the 1960s, a new behavioral economics perspective to this issue started to develop. It fully recognized the existence of bounded rationality rooted in stable physiological patterns and cognitive biases that made it impossible for individuals to act purely on utility maximization grounds.

Aggregating individual preferences is an even more challenging, yet necessary, task if one is to come up with policy recommendations for a collective of individuals. Economic theory suggests that this aggregation can occur through a social welfare function, the exact formulation of which depends on such normative aspects of public policy as tolerance for inequality and distributional preferences. Different approaches to this issue can be broadly classified into libertarian (i.e. market mechanisms are posited to be the most desirable), utilitarian (i.e. social welfare is a sum of individual utilities), egalitarian (i.e. each individual's well-being has a diminishing return to the social welfare) and Rawlsian (i.e. inequality is only tolerated as long as it maximizes the total individual utility). In this dissertation, the analyzed institutional settings are likely to feature a combination of these approaches. For instance, similar to France, the Canadian healthcare system, as discussed in Section 2.3.1, favors an egalitarian approach that ensures equality of access to each and every individual, even at the cost of oftentimes decreased efficiency and longer wait times for elective surgery. On the other hand, the Russian government has arguably put a greater emphasis on the utilitarian/Rawlsian approach to social welfare in the Maternity Capital program, which manifests itself in its attempt to increase the overall number of childbirths through massive child subsidies. As discussed in Section 3.6 it also made necessary reform adjustments, such as expanding the list of subsidy-eligible uses, that facilitated achieving this outcome, bringing the program closer to a Rawlsian approach to wealth distribution.

Table 1 – Wealth, Quality of life and Well-being Indicators in Canada, France, Russia and the USA

	Canada	(Rank)	France	(Rank)	Russia	(Rank)	USA	(Rank)
GDP per capita ¹ , USD (2020)	42,080	17/186	39,257	20/186	9,972	61/186	63,051	5/186
GDP per capita ¹ , PPP Int\$	47,589	21/186	45,454	23/186	27,394	50/186	63,051	7/186
Human Development Index ² , 2018	0.922	13/189	0.891	26/189	0.824	49/189	0.92	15/189
Human Development Index (inequality adjusted ²)	0.841	17/189	0.809	25/189	0.743	48/189	0.797	28/189
Life expectancy ³ , 2018	82.3	14/186	82.5	12/186	72.4	106/189	78.9	38/189
Life expectancy, disability adjusted ³	73.2	7/183	73.4	5/183	63.5	104/183	68.5	40/183
Value of statistical life, estimates from <i>Vicusi and Masterman (2017)</i> , in million USD	8.2	–	7	–	2	–	9.6	–
Rule of law Index ⁴ , 2020	0.81	9/128	0.73	20/128	0.47	94/128	0.72	21/128
Where-to-be-born Economist Index ⁵ , 2020	7.81	8/80	7.04	26/80	5.77	59/80	7.38	17/80

Sources: International Monetary Fund¹, United Nations², World Health Organization³, World Justice Project⁴, The Economist⁵

While, as discussed above, social welfare functions can vary considerably between institutional settings, so do the approaches as to what constitutes the inputs, or individual building blocks of these functions. The most general and widespread economics approach has been to approximate utility through individual wealth and income. At the country level, such metrics as GDP per capita (PPP or nominal) are commonly used to draw comparisons between countries. However, in healthcare and public policy contexts, more nuanced and elaborate approaches have been developed to measure the levels of human development, well-being and health. Apart from income and wealth, they incorporate a much broader set of aspects of the quality of life. In particular, the United Nations calculates the Human Development Index which, alongside income per capita, uses metrics of life expectancy and access to education. Alternative indexes, such as the Rule-of-law index of the World Justice Project⁴ and the Where-to-be-born index constructed by The Economist journal, go further and compare the quality of environments and institutions based on an even wider set of country specific characteristics (e.g. Job security, physical security, political rights, etc.). In Table 1, these metrics are presented for the 3 countries analyzed in this Phd dissertation (Canada, France, Russia), and the USA.

Regarding the main dissertation results, all the three chapters tend to feature statistically insignificant estimates, indicating the absence of discernible effects on the analyzed outcomes of interest. This general finding has several explanations, each of which highlights a number of crucial issues that need to be considered when designing new policy interventions.

First, all 3 chapters concentrate on potential side effects of large public policy interventions. In particular, the 2012 DRG reform in France analysed in Chapter 1 sets as the main policy objective to reduce financial risk faced by maternity wards. This was expected to occur through bringing the reimbursement mechanism closer to the real cost of a child delivery ("cost-plus" approach). However, as an unintended spillover, this reform also generated hospital-level financial incentives, which affected the expected financial gains hospitals could receive from administering C-sections and vaginal (normal) child deliveries. In Chapter 1 we investigated whether this change in clinical decisions happened in practice, and we reached a negative conclusion on this question. As discussed in Chapter 2, the fashion in which the QBP/HBAM reform in Ontario was implemented (namely, a selective inclusion of targeted joint replacement procedures, and their gradual implementation over time) allowed us to test a number of potential spillover effects from the targeted QBP procedures (e.g. unilateral hip and knee replacements since 2012) to other medical interventions plausibly competing for the same hospital resources (i.e. bilateral hip and knee replacements between 2012 and 2014, as well as other types of joint surgeries). The last chapter of this PhD dissertation concentrated on welfare effects of the Russian MC subsidy program, whose primary goal, however, was to boost fertility to prevent a looming demographic crisis characterized by long-term projections of low fertility and high mortality. We came to the conclusion that the analyzed spillover effects of these public policies tended to lack in strength for their impact to be statistically detected by micro-econometric estimation methods, such as difference-in-differences and regression discontinuity designs.

It is worth noting that on several occasions in this dissertation, estimation results pointed to the presence of unintended but, but yet anticipated, policy effects. In particular, Chapter 2 presented empirical evidence indicating that the introduction of the QBPs scheme led to a significant reduction in the hospital length of stay for the targeted procedures. However, despite the fact that Ontarian policy-makers did not explicitly regard this effect as their policy objective, it cannot be viewed as completely unexpected. A large body of previous research (for a summary see for example, *Busse et al. 2011*) on activity based funding has consistently pointed to the same effect observed in a number of different healthcare systems that attempted to implement similar reforms. In chapter 3 on the Russian maternity subsidy, the evidence for potentially heterogeneous effects of welfare programs on the psychological well-being of children of different gender was also found in a number of related economics studies, in particular *Kling et al. (2005, 2007)* and *Milligan and Stabile (2011)*.

Second, informational asymmetries and a lack of communication/coordination between

reform stakeholders are likely to reduce policy efficiency and, thus, may lead to insignificant estimation results. Most notably, as evidenced in Chapter 2 on the QBP/HBAM reform, an ambiguous and contradictory policy agenda was likely at the origin of this problem in Ontario. As discussed in *Palmer et al. (2018b)*, policy stakeholders and external experts perceived that the main objectives shifted after QBP implementation started in 2012, moving the funding mechanism from a Pay-for-performance scheme to an activity-based funding. The very term "Quality-based Procedures" was also largely perceived to be increasingly inaccurate, given that the quality component was never funded. The uncertainties with regard to the goals and mechanisms of the 2012 Ontario reform have led to an overall confusion and even anger and distrust among healthcare administrators, one of whom, for example, said: "If you pull back the camera lens, health system funding reform, including quality-based procedures, was entirely designed and primarily implemented by the Ministry in a command-and-control function with virtually no genuine engagement and collaboration with the very sector that is supposed to implement the change. There was a very wide range of meetings and groups that came together under Ministry sponsorship from time to time, but they were inauthentic and simply masquerading as mechanisms of, again, consultation, collection of evidence and guidance, and refinement of tactic and strategy" *Palmer et al. (2018b)*. As a result of this relatively poor communication of policy goals, the intended effect of the QBPs in terms of improvements in care quality and a better adherence to the QBP clinical guidelines in all likelihood was more difficult to achieve for the QBP reform designers. Although on a significantly smaller scale, a similar problem was documented in the 2012 French DRG reform, insofar as many clinicians, and in particular obstetricians, were reported to have limited knowledge of the changes in obstetric tariff schedules. Even though midwives, who accompany the majority of child deliveries in France, were more likely to pay attention to the DRG tariff reform due to the fact that diagnosis encoding is often part of their professional duty, the overall effect of tariff incentives on the choice of the child delivery mode was very likely mitigated by the lack of complete information among healthcare professionals.

Lastly, in many instances, the interactions between reform designs and the institutional context plausibly could mitigate the effect of implemented reforms. Namely, in the first chapter, one of the possible channels that may explain the overall insignificant effect of the French DRG reform on scheduled C-section rates lies in the limited capacity of gynecologists, midwives and obstetricians to accurately predict changes in clinical practice even under complete information, due to a high degree of tariff complexity. As a result, it may have reduced obstetricians' ability to increase hospital expected revenues, even assuming

obstetricians' intend to do so. In addition, in chapter 3, the institutional design of the MC reform (namely, the fact that this subsidy was a one-time transfer with delayed activation and a restricted set of authorized uses), was very likely to affect the effectiveness of the subsidy, as smaller but more frequent payments are known to be more effective than one-time but large transfers. In addition, people's preference for instant gratification over a significantly larger reward in the distant future may have promoted reform designers to limit the number of eligible uses, thus forestalling, at least partially, potential inefficient spending patterns studied in Chapter 3. Finally, it is worthwhile to note that this behavioural economics perspective may also lie at the origin of recent adjustments of the Maternity Capital, which made it possible for Russian families to use the subsidy in the form of an unconditional means-tested regular transfer, thus accounting for cognitive and rationality constraints typically faced by individuals.

The aim of the three studies presented in this PhD dissertation was to better inform policy designs in the analyzed institutional contexts, as well as in institutionally similar setting. In Chapter 1 we concluded that that the unwanted effects of DRG financial incentives on the clinical practice can be mitigated by strong labor protection laws and specific types of work contracts. This result is of particular interest to most other healthcare systems using DRGs as a primary funding mechanism, as in both Europe and North America the degree of risk-adjustment has had a consistent trend towards a greater degree of refinement. Chapter 2 points to the necessity of having consistent and well-articulated policy goals while implementing hospital funding reforms. The lessons learnt from the introduction of the QBP/HBAM in Ontario can be generalized to other healthcare systems relying on activity-based funding, most notably those in Western Europe (such as France, Germany, Belgium). In chapter 3 we came to the conclusion that in order to make the MC fund more accessible to poorer families, the implementation design could be modified such that the subsidy amount would be adjusted with respect to family's financial resources. We also argue that Russian government's initiative to extend MC eligible expenses for poor households was well-aligned with economic incentives, which would be likely to contribute to a better efficiency of the MC spending. These conclusions can be applicable in several other Eastern European countries that have implemented similar reforms (e.g. Hungary, Ukraine), as well as in Singapore, whose childbirth programs bear a lot of resemblance with Russian pro-fertility policies.

The remaining parts of this dissertation are organized as follows: Chapter 1 studied the effect of the 2012 DRG refinement reform on the scheduled C-section rates in France, Chapter 2 investigated the effects of introducing QBPs and HBAM in Ontario on the quality and quantity of orthopedic care procedures, Chapter 3 analysed the impact of the

MC child subsidy on child outcomes and household-level spending in Russia. Each Chapter of this dissertation can be read independently, and can be interchangeably referred to as a “study” or “paper”. At the end of each chapter appendices provide additional information on the data, analytical samples, as well as complementary model estimates. A general conclusion and references are provided at the end of this dissertation.

Chapter 1

Impact of Tariff Refinement on the Choice Between Scheduled Cesarean Section and Normal Delivery: Recent Evidence from France

In this chapter we tested the effects of a considerable DRG tariff refinement that occurred in 2012, designed to reduce financial risks of French maternity wards. To estimate the resulting DRG incentives with regards to the choice between scheduled C-sections and normal deliveries, we predicted, based on pre-admission patient characteristics, the probability of each possible child delivery outcome and calculated expected differences in associated tariffs. After controlling for multiple patient, hospital and regional characteristics and allowing for hospital and year effects, we found that introducing new severity levels and clinical factors into the reimbursement algorithm had no significant effect on the probability of a scheduled C-section being performed. The results were robust to multiple formulations of DRG financial incentives. This study is the first to focus on the consequences of DRG refinement in obstetrics and develops an approach suitable for measuring fee incentives in this setting.

1 Introduction and literature review

The topic of how financial stimuli can affect quality, quantity and composition of health-care services has stirred much debate. Substantial anecdotal and scientific evidence has been pointing to the fact that health care providers may take into consideration factors other than clinical aspects or patients' preferences when prescribing a medical treatment. This Chapter estimates the responsiveness of healthcare providers to a change in hospital level financial stimuli. We investigate the impact of a diagnosis-related group (DRG) refinement, leading to an increase in the number of associated tariffs. Such a change was introduced recently in France, by adding additional criteria and diagnoses to pricing formulae, in order to better account for differences in patient severity. The results of our study provide no evidence that there was a connection between hospital level changes in financial incentives and the rate of scheduled C-sections performed by obstetricians in France between 2010 and 2013.

The risk that financial incentives may lead healthcare providers to change their practices under prospective payment schemes (PPS) has been analyzed extensively in the theoretical literature. *Ellis* (1998) considers healthcare under/over-provision and patient selection as major concerns that may arise under PPS. His model predicts that in a perfect information setting where competitive healthcare providers fix both a maximum severity for admitted patients and a bundle of services available at each severity level, high-severity patients will get a socially sub-optimal amount of treatment ("skimping"), while the opposite should occur to low-severity patients ("cream-skimming"). *Siciliani* (2006) features a more probable information asymmetry setting and shows that the purchaser's lack of knowledge concerning patient's average severity level results in permanently higher levels of a more intensive treatment (i.e. surgery) for low-severity patients. In order to reduce the informational rent, public authorities must pay lump sum transfers to providers who own private information about their average patient severity level. To account for more complex providers' motivations, the assumption of self-interested practitioners was relaxed in *Hafsteindottir and Siciliani* (2009) by introducing a measure of doctors' altruism. It was shown that if prices are differentiated according to severity levels (i.e. refined), hospitals will over-provide the more costly treatment. However, in the case where prices are homogeneous across severity groups, doctors with a low altruism level will under-provide the more costly treatment.

Contrary to these theoretical predictions, the empirical findings show that the effects

of a DRG tariff refinement on health care provision are relatively small or insignificant. Following the approach proposed by *Ellis and McGuire* (1996), *Gilman* (2000) attempts to decompose the effect of DRG tariff refinement for HIV treatment into a moral hazard component (a change in the intensity of treatment induced by tariffs) and a selection component (or "cream-skimming"). The study suggests that financial incentives created by DRG tariff refinement had only a minor effect on the services supply. However, it is likely that this conclusion was caused by reimbursement rates that were already very generous or/and by clinical features specific to HIV treatment. In a more recent study, *Januleviciute et al.* (2015) use fixed effects models to estimate the responsiveness of healthcare providers to changing stimuli within different DRG groups between 2003 and 2007. They found that a 10% rise in DRG reimbursement rates leads to 0.8–1.3 % increase in the number of patients treated for medical DRGs in Norway. However, no such effect was found for surgical DRGs.

Child delivery attracts economists' attention because it fits well into the general framework of incentives theory, where an intrinsic conflict may exist between healthcare practitioners' and mothers' preferences. The healthcare providers' choice between delivery modes may be biased in favor of the option reimbursed more generously, inasmuch as financial profits, personal well-being and altruism play a role in their utility function. In addition, being an elective surgery with little medical guidance regarding appropriateness, C-section arguably allows care providers to increase, depending on the institutional context, individual and/or hospital revenues, with varying risk of facing malpractice claims.¹ Mothers, on the other hand, would rather avoid unnecessary complications and side effects induced by a C-section, especially when the latter could easily have been substituted with a normal delivery. This conflict is a consequence of information asymmetry between healthcare professionals (i.e. agents) and their patients (i.e. principals). From a societal point of view, it can lead to non-optimal levels of medical care and, as a result, redistribution of wealth in favor of a potentially opportunistic agent at the expense of the less informed party.

Part of the explanation for the substantial observed variation in C-section rates, both between countries and within countries' smaller "local markets", has been offered by the demand inducement hypothesis. For instance, *Gruber and Owings* (1996) show that after a fall in fertility in the USA in 1970-1982, physicians increased C-section rates in response to falling revenues. *Gruber et al.* (1999) find a direct positive effect of the price differential

¹See Section 2 for French healthcare context

between C-section and normal childbirth on the rate of C-sections in USA, which itself can be decomposed into the income effect and the substitution effect across services. However, it is worthwhile noting that *Grant's* (2009) replication of Gruber et al. (1999) concludes that much of the difference in C-section and normal delivery rates is in fact attributable to coincidental trends and a sample bias. More recently, *Allin et al.* (2015) studied a sample of Canadian women who gave birth from 2006 to 2011 and come to the same conclusion that increasing the fee differential between C-section and normal delivery provides an incentive for physicians to favor C-sections over vaginal labor. However, the magnitude of the effect is relatively small: doubling the fee differential will on average increase the probability of a C-section only by 5.6%. *Johnson and Rehavi* (2016) provide evidence that physicians perform fewer C-sections on other physicians, stressing the importance of the information asymmetry which exists between physicians and mothers when a decision is made in favor of a child delivery option.

An important behavioral component deemed to mitigate the impact of induced demand channel on medical practice is provider altruism. In particular, *Lefèvre* (2014) showed that the increase in C-section rates before holiday Mondays (e.g. Memorial Day or Labor Day) in the US was attributable to the scheduling effect of C-sections rather than to convenience and leisure reasons. However, it is worth noting that several earlier studies also argued the contrary. For example, *Brown* (1996), C-section rate is not homogeneous across daytime and days of the week. In particular, his results suggest the existence of a “rush-hour“ effect that could explain a surge in unscheduled C-section rates on Friday in the evening. *Phillips et al.* (1982), studying the day of the week frequency of C-section rates found significantly lower C-section rates on weekends for all but one clinical category. Analyzing the available general evidence on provider altruism, the literature review by *Galizzi et al.* (2015) concluded that existing research generally supports the hypothesis that physicians largely behave ‘altruistically’ in their healthcare decisions, while, on the other hand, there likely exists considerable heterogeneity in physicians’ altruistic preferences.

However, *Grytten et al.* (2011) and *Lo* (2008) come to the opposite conclusion. Studying a sample of Norwegian women, they (2011) find a positive relation between patients’ level of education and medical expertise and the odds of a C-section, highlighting the possibility of better communication between better educated patients and medical specialists. The study of child deliveries in Taiwan conducted by *Lo* (2008) shows that the fee equalization policy did not affect C-section rates.

An alternative explanation for higher C-section rates is fear of malpractice claims. As

argued by *Dubay et al.* (1999), low-income families with lower education levels are more likely to sue their physician on malpractice grounds. Studies that investigated this hypothesis include, but are not limited to, *Currie and MacLeod* (2008), *Shurtz* (2013) and *Yang et al.* (2009). The evidence tends to support the hypothesis that higher malpractice fears proxied by court claims or reforms increasing physician liability make physicians resort to defensive medicine, resulting in higher C-section rates. Other explanations also include physicians' preferences for leisure and work (*Brown*, 1996) mothers' idiosyncratic preferences (*Lo*, 2003; *Grytten et al.*, 2013) and different rates of adoption of diagnostic technology (*Grytten et al.*, 2012).

In the French context, *Milcent and Rochut* (2009) aims to explain the variation in C-section probability across public non-profit, private for-profit and private non-profit, controlling for obstetrician, gynecologist and midwife availability in the spirit of *Gruber and Owings* (1996). Milcent and Rochut find that private clinics are significantly more likely to perform more C-sections than both public and private non-profit hospitals. Moreover, the ratio of obstetricians per patient has a positive impact on C-section rates, which may be indicative of physician demand inducement. However, in contrast to our strategy, tariff incentives are not the focus of their analysis. In addition, this study considers the year of 2003 when the DRG system was not yet introduced and global budgets were used instead.²

This Chapter contributes to the existing literature in three ways. First, methodologically, our contribution consists in proposing a more nuanced approach to modeling the impact of hospital-level financial incentives on the obstetrician's choice between delivery modes. In contrast to earlier studies that considered average or aggregated fees for delivery cases (as in *Allin et al.*, 2015) to calculate the corresponding tariff differential, we construct these measures based on expected DRG tariff for each delivery outcomes taking into account the uncertainty with regards to final patient severity. The fact that we rely on solely pre-admission (i.e. antepartum) characteristics for outcome and severity predictions allows us to circumvent potential reverse causality problem – so far not explicitly addressed in related C-section literature – stemming from a likely bidirectional dependence between the chosen child delivery mode and the occurrence of comorbidities. In addition, our approach allows us to mitigate the problem of unobserved actual hospital costs and take into account the fact that, for a given patient severity level and comorbidities (and the associated tariff), C-section and normal delivery tariffs are neither constant nor independent. In other words,

²About work on the demand inducement hypothesis among general practitioners in France, see *Rochaix and Jacobzone* (1997), *Delattre and Dormont* (2003, 2005), *Coudin, Pla and Samson* (2015).

our approach reflects the fact that more severe cases with a C-section would also have had a higher severity level and more comorbidities, had a normal delivery been chosen. Second, this Chapter provides new evidence on the impact of tariff refinement on health care provision in the DRG setting - currently the most widely adopted payment scheme in developed economies. Third, to the best of our knowledge, the consequences of DRG refinement (i.e. extending the number of DRGs to take into account additional criteria in the definition, such as severity) have never been studied in the realm of obstetrics. To shed more light on this topic, we take advantage of the French 2012 refinement of obstetric DRGs, which has led to a significant tariff differentiation.

This Chapter is organized as follows: Section 2 defines the context and presents the motivation of this study, Section 3 presents the data, Section 4 defines the empirical strategy. The main results are presented in Section 5, with robustness checks in Section 6. The results of this study and policy implications are further discussed in Section 7. Section 8 concludes.

2 Obstetric care in France: institutional context

The French healthcare system is characterized by the presence of multiple types of health-care institutions delivering care at highly regulated prices. Based on the US DRG system, in 1986 for statistical purposes France adopted Groupes Homogènes de Malades (GHM), and starting from 2004 this DRG system was gradually introduced to reimburse hospitals. The system splits in a mutually excludable way all diagnoses between major diagnosis categories (CMD). They are further classified by diagnosis roots, which give the main hospitalization cause. Finally, based on a predefined list of complications, a severity level is attributed to each diagnosis root. It reflects the extent to which the patient's general health status, along with delivery-related medical circumstances, can make medical procedures associated with the diagnosis more risky and therefore more costly for the provider.³

Since its introduction, the French classification underwent several revisions. The latest major revision in obstetrics DRGs occurred in 2012 with a significant increase in the num-

³These classifications, which map a set of patient characteristics onto DRGs in a unique way, are established by a state-run agency, Agence Technique de l'Information sur l'Hospitalisation (ATIH), on an annual basis. For example, in DRG version 11d the root is 14Z13 stands for a vaginal single delivery by a multipara mother with no significant complication. In addition, it is followed by a letter which specifies the severity level of the diagnosis, e.g. 14Z13A. Thus, the letter A denotes that there was no significant complication.

ber of diagnosis groups, differentiated to fit more narrowly defined groups of patients. In particular, normal deliveries became differentiated into both single/multiple and primipara/multipara.⁴ For C-sections, however, the only differentiation introduced was that between single/multiple deliveries. Within each DRG group the number of severity levels grew substantially as well. In addition, a considerable number of new diagnoses were introduced into pricing formulas creating an additional variation in tariffs across DRG classifications. Lastly, DRG tariffs started to be adjusted by mother gestational age such that pre-term and post-term pregnancies are reimbursed to hospitals at a higher rate (see Appendix 2 for details).

In France most medical procedures are completely covered by social security or only require minor out-of-pocket payments. Both public hospitals and private clinics are reimbursed nationally by these public funds⁵. Hospitals are compensated for their incurred costs, based on the GHM diagnosis root, its related severity level and the hospital status (public or private).

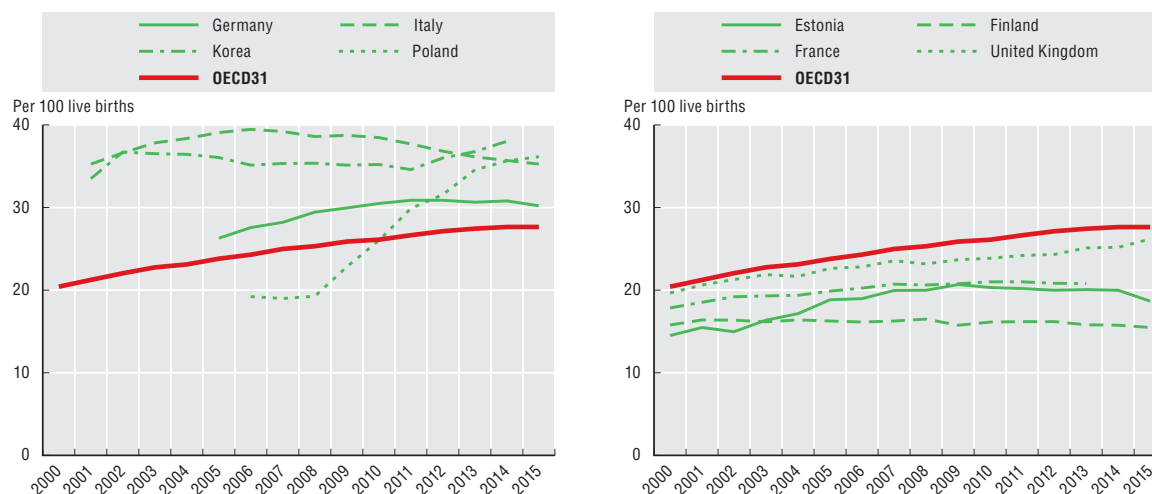
C-sections are considered to be among the most common and well-researched medical procedures. Its popularity has been growing worldwide, with the average C-section rates in OECD countries rising from 20% in 2000 to 28% in 2017 (*OECD, 2019*). Optimal and medically justified C-section rates have been a subject of scientific and political debates for decades. The World Health Organization statement on Cesarean section rates issued in 2015 argued that “CS rates higher than 10 % were not associated with reductions in maternal and newborn mortality rates”, and that medically justified rate may be lying within the range of 10-15% of all child deliveries. In this respect, the relatively stable rate of around 20% observed in France since late 2000s follows the pattern of other industrialized and economically advanced countries, which tend to report figures well above this hypothesized optimal range (see Figure .1.1). However, it is important to stress that this range is dependent on constantly evolving population characteristics. The global trend toward increasing maternal age at childbirth, coupled with increasing rates of chronic diseases related, in particular, to obesity and cardiovascular health, might in part contribute to growing C-section rates worldwide.

Overall, in Europe C-sections tend to be less medicalized procedures, compared to the

⁴In medical terms, a multipara (or pluripara) mother designates “a woman who has given birth at least twice to an infant, liveborn or not, weighing 500 g or more, or having an estimated length of gestation of at least 20 weeks” (Farlex Partner Medical Dictionary, 2012).

⁵The biggest among them, Caisse Nationale de l’Assurance Maladie des Travailleurs Salariés covers around 75% of total healthcare expenditures

Figure .1.1 – C-section rates in 2000-2015 (selected OECD countries and OECD average).
 Source: *OECD Health at a Glance (2017)*



Source: OECD Health Statistics 2017.

StatLink  <http://dx.doi.org/10.1787/888933605331>

USA and Canada, with more emphasis being put on delivery convenience and comfort. In addition, in contrast to the USA and Canada, in France child deliveries are most often performed by midwives. In general, they conduct less complicated child deliveries, while obstetricians handle child deliveries with complex comorbidities. General practitioners rarely perform child deliveries, unless no qualified midwife or obstetrician is available in the area. The only exception to this rule is absence of a qualified midwife/obstetrician in a reasonable proximity. Encoding and entering commodities in PMSI system is performed primarily by admissions secretaries/nurse, and occasionally by midwives. It is not common for doctors to assume this responsibility in French hospitals. While deciding on the mode of child delivery, obstetricians and midwives generally rely on their previous experience and may follow recommended practices summarized, for example, in HAS clinical guidelines (*HAS, 2012*). The French government requires the mother to attend at least 7 pre-natal medical appointments with the mother's physician and/or a midwife, during which risks of each delivery mode are expected to be discussed. A relatively high emphasis is put on mothers' preferences for the mode of child delivery, unless medical examinations reveal at least one of the relatively few strong counter-indications against either C-section or normal delivery.

French care institutions can have either a public or private legal status. While public hospitals are non-profit organizations, private clinics can either be for-profit or non-profit

facilities.⁶ In the non-profit sector, obstetricians and midwives have the status of hospital employees and receive a monthly salary, determined according to a state-mandated salary grid. This grid takes into consideration employees length of employment, his/her qualifications, but is not directly tied to their performance. The nature of the contract may differ for these two professional categories. While obstetricians are normally hired on a permanent contract with a trial period of several months, it is not uncommon for midwives to be employed on a fixed-term contract, especially at the beginning of their careers. This circumstance can potentially make midwives relatively more susceptible to hospital-level incentives relative to physicians and obstetricians. In the for-profit sector, obstetricians' and, to a lesser extent, midwives can sign more flexible contract that includes both a fixed salary amount and an income supplement calculated based on a fee-for-service/per birth basis. Thus, this transmission of hospital-level DRG incentives to individual providers can be somewhat facilitated in private-for-profit facilities, inasmuch as healthcare professionals pursue, at least partially, financial interest in their medical decisions.

The bigger reliance of the French healthcare system on midwives over physicians/obstetricians as compared the US and Canada could have generated additional incentives to hire more midwives to save on labor costs. However, the legal protections against lay-offs for permanent staff, as well as the relative rigidity of the French labor market for medical professionals make such changes implausible over the analytical timespan of this study.

As far as DRG payments are concerned, in private sector they do not cover providers' remuneration and only reflect the average material cost of a procedure. As a result, they tend to be lower than in the private sector, where salaries are determined by individual healthcare facilities and are subject to lower state regulation. Overall, DRG payments are supposed to be disconnected from or have limited impact on providers' income.

Therefore, contrary to the US and Canadian obstetric care markets, changes in DRG tariffs in France directly affect the financial situation of hospitals, and to a much smaller extent that of healthcare providers. In the French context, DRG reforms therefore offer an opportunity to test the extent to which hospital level financial incentives can be transmitted to individual providers, knowing that the financial situation of the latter is not directly impacted. These hospital level incentives may transfer to doctors through explicit internal planning and volume targets/projections set by hospital management,

⁶The latter are usually run by private non-profit complementary insurance funds, commonly referred to as "mutuelle".

as well as more informal communication channels. Although in general empirical studies have not systematically drawn the distinction between the levels at financial incentives are generated, existing evidence on pay-for-performance funding – traditionally the area where this topic has been most well-researched – is suggestive of individual level incentives being more effective in affecting medical practice than hospital level incentives. In addition, the size of these incentives relative to overall provider benefit and the nature of their implementation (e.g. redistribution of revenue vs. providing new sources of funding) may also impact clinical practice (for a systematic reviews see *Cristianson (2008)* and *Van Herck (2010)*). In this respect, the fact that the French 2012 DRG reform was budget-neutral with resulting incentives significantly affecting only a fraction of child delivery cases may have mitigated the impact of the reform on the providers' behavior.

In France, some of the channels that are believed to affect providers' decision for a delivery option, such as fear of malpractice claims, are not likely to be of significant importance. Another important aspect for our study is the fact, according to professional experts in the field, no significant changes have taken place in either child delivery technology or medical guidelines during the period 2010-2013.

One of the main objectives of DRG tariff refinement usually consists in reducing the uncertainty of hospital revenues due to high cost patients. Contrary to fee-for-service schemes, material costs of a procedure are not necessarily fully covered by DRGs due to significant cost differentiation of cases within each DRG. This pattern was noticeable in the French DRG system before March 2012, with only one DRG for C-section and two DRGs for normal delivery, with three severity levels for each of them. Following the 2012 reform, both the number of obstetric DRGs and the DRG severity differentiation were substantially increased. In total, the number of tariff groups increased from 7 to 30 (see Appendix 2).

The 2012 DRG reform followed the general trend of increasing DRG tariff refinement in Europe (*Busse et al., 2011*). One of the major implications of the 2012 DRG reform was bringing the reimbursement scheme closer to a "cost plus" than to a fixed price approach. However, this process is not necessarily always driven by a higher degree of risk adjustment with respect to medical diagnoses. Instead, new DRG schedules often allow tariff adjustment based on medical procedures, which makes DRGs payments schemes lose their PPS character – one of their main theoretical pillars and *raison d'être*. This has had several implications. In particular, risk adjustment based on medical procedures may incentivize overprovision of covered and potentially unnecessary procedures, whose use

DRGs were initially meant to mitigate. In this respect, a legitimate necessity to account for cost differences between C-sections and normal deliveries (i.e. medical procedures) has led French and the majority of other health authorities operating DRG funding schemes to allow for tariff differentiation between the two modes of child delivery, even if it had the potential to produce the aforementioned side effect.

Tariff changes could have made “upcoding” a possible strategy for healthcare providers, generating more profits without affecting the volume or quality of services. Such an effect was observed by *Dafny* (2005), who found that price changes resulting from suppressing age criteria in DRGs (DRG aggregation) in the USA in 1988 and subsequent recalibration of tariffs led to an increase of the share of top-coded patients, while evidence that this price shock affected service volume or intensity was weak. Regarding French context in particular, we are aware of *Milcent* (2016) study that documenting increasingly higher levels of DRG severity claimed by French hospitals following a DRG refinement in 2009. In addition, *Milcent* (2016) results suggest that the strength of the “upcoding” response was the highest among for-profit clinics. A recent study by *Giacomo et al.* (2017) highlights the possibility of the “upcoding” channel in Italian obstetrics markets operating under DRGs. Conditional on the mode of child delivery, they found that hospitals experiencing the highest increases in DRG reimbursement rates (but not necessarily due to refinement) are more likely to subsequently code their patients as more complex in terms of severity. The discussion and potential implications of “upcoding” channel are provided in Section 7.

The 2012 reform had different impacts on institutions providing obstetric care. Since the most complicated (and statistically less common in general population) cases started to be reimbursed at higher rates after the DRG tariff refinement, bigger hospitals and clinics having at their disposal advanced medical equipment (and thus allowed to admit mothers with high severity comorbidities) are likely to have benefited most from the reform. Although revenue uncertainty also decreased for small maternity wards, some may have faced an increased financial pressure due to the fact that the DRG reimbursement rates decreased for the simpler cases (*Vanlerenberghe, 2015*). However, the exact profitability figures are not available for the majority of healthcare institutions since they are confidential.

In addition, political economy considerations also play an important role in healthcare reforms. The existing evidence suggests that state agencies or institutions competing for budgets may have incentives to over-provide services and to increase their size (*Niskanen*

1971, 1975; *Goddard et al.*, 2006). From an administrative point of view, DRG systems are considered to be costly to operate since they require regular monitoring and have a high coding-related workload (*Busse et al.*, 2011). After the French 2012 reform, the costs of running the DRG system very likely increased along with the number of DRGs. From a societal point of view, this could have led to lower public expenditure efficiency and more transactional costs due to an increased regulatory burden, potentially outweighing the benefits of the reform.

3 Data

The primary source of data containing information on patients' hospitalization and diagnoses in France is the *Programme de médicalisation des systèmes d'information* (PMSI) dataset, which served as the main source of data in this Chapter. It is administrative data used to pay hospitals in the PPS system, which covered all French hospitals starting from 2006.

Actual PMSI data were split into 3 different datasets, which separately contained information about:

1. Performed medical procedures
2. DRG main diagnosis, patient and hospital characteristics
3. Complications associated with the main diagnosis⁷

It covers all individuals living in both metropolitan and overseas French departments. Full PMSI data access was given for the years 2010 to 2013.

The resulting data contains a rich set of individual level characteristics (including age, region and community of residence, length of hospital stay, month and year of hospital leave, etc.) and hospital characteristics, which include a hospital unique identifier, ownership information (i.e. public or private), legal status (i.e. profit or non-profit), geographic location, as well as a list of diagnoses and complications (or secondary diagnoses) which occurred during each hospital stay. DRG diagnoses and severity levels are encoded in the current DRG version and, in some cases, both current and previous DRG versions. Each hospital stay also has a unique patient identifier which allows us to track the cases of hospitalization across different years for a given patient. The data provides information on

⁷A full description of the datasets is given in Appendix 1.

Table 1.1 – Summary statistics for obstetrical hospital stays in 2010-2013, by hospital type

Variables	All hospitals		Public		Private (non-profit)		Private (for-profit)	
	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
Age	29.7	5.4	29.4	5.5	30.6	5.0	30.1	5.1
Multiple pregnancy	0.016	0.13	0.019	0.14	0.011	0.11	0.011	0.11
Multiparity (recovered from diagnoses lists and previous years)	0.54	0.5	0.547	0.5	0.507	0.5	0.533	0.5
C-section	0.207	0.41	0.199	0.4	0.203	0.4	0.223	0.42
– scheduled	0.073	0.26	0.064	0.25	0.074	0.26	0.095	0.29
– urgent	0.036	0.19	0.04	0.2	0.028	0.177	0.028	0.17
– unscheduled	0.096	0.29	0.094	0.29	0.1	0.3	0.1	0.3
Normal delivery	0.794	0.4	0.801	0.4	0.797	0.402	0.776	0.42
Number of comorbidities	4.46	2.66	4.96	2.86	4.0	2.21	3.37	1.83
Number of previous C-sections	0.09	0.32	0.09	0.32	0.088	0.31	0.096	0.32
Gynecologist availability index	6.95	3.85	6.53	3.69	8.84	4.34	7.35	3.84
Average income per household per person (by pseudo-postal PMSI code), in thousand euro	19.51	4.41	19.13	4.06	21.45	5.89	19.92	4.5
<i>Tariff incentive measures:</i>								
$\Delta tariff$	-42.5	274.3	-70.48	307.5	-63.81	280.4	34.64	131.7
$r_t - \bar{r}_{2010}$	-0.022	0.25	-0.043	0.27	-0.047	0.26	0.037	0.19
$(r_t - \bar{r}_{2010}) / \bar{r}_{2010}$	-0.009	0.15	-0.019	0.19	-0.021	0.18	0.02	0.13
$\Delta_t - \Delta_{2010}$	-70.34	441.8	-99.38	514.6	-96.53	317.73	11.01	190.8
<i>Gestational age:</i>								
22-35 weeks	0.034	0.18	0.042	0.2	0.023	0.15	0.016	0.13
36-39 weeks	0.513	0.5	0.502	0.5	0.498	0.5	0.544	0.5
≥ 40 weeks	0.452	0.5	0.454	0.5	0.477	0.5	0.44	0.5
<i>Mother comorbidities :</i>								
Infection during pregnancy	0.042	0.2	0.041	0.2	0.034	0.18	0.047	0.21
Diabetes	0.061	0.24	0.067	0.25	0.063	0.24	0.046	0.21
Complicated breech	0.027	0.16	0.027	0.16	0.025	0.16	0.027	0.16
Oligohydramnios	0.015	0.12	0.017	0.13	0.015	0.12	0.008	0.09
Hypertension/Preeclampsia	0.056	0.23	0.057	0.23	0.058	0.23	0.052	0.22
Fetal growth anomaly	0.043	0.2	0.053	0.22	0.037	0.19	0.019	0.14
Hemorrhage/uterine rupture	0.016	0.13	0.017	0.13	0.014	0.12	0.015	0.12
Premature rupture of membrane	0.047	0.21	0.046	0.21	0.052	0.22	0.048	0.21
Observations	3,041,425		2,004,269		243,155		794,001	

DRG reimbursements for every single child delivery case that occurred in French health-care institutions. Thus, we were able to observe diagnoses, comorbidities and in-hospital services registered during childbirth, both before (ante-partum) and after (post-partum) labor begins. These variables are summarized in Table 1.1.

To get information about tariffs on the basis of which hospitals are compensated, we use publicly available data provided by ATIH⁸. Different tariffs scheduled are set for for-profit and non-profit hospitals.

Income data were collected by Institut national de la Statistique et des Etudes Economiques (INSEE) and are available on the level of PMSI pseudo postal codes that largely resemble postal codes. However, in the cases where there are too few people living in the postal code area, for the sake of securing patients' anonymity, PMSI pseudo codes may contain observations from several postal codes. This variable is calculated as a local yearly median income per household per person.

To control for medical obstetric care availability, we use a gynecologist availability index constructed by IRDES and calculated for the year of 2011 at the municipal level. It takes into account gynecologists' volume of activity, service use rates differentiated by population age structure, supply and demand factors in neighboring municipalities and other parameters⁹.

In addition, to obtain information about observed C-section rates and mother multiparity, we use data from the French Enquête Nationale Périnatale 2003 and 2010, collected by Direction de la Recherche, des Études, de l'Évaluation et des Statistiques.¹⁰

⁸See the ATIH data on tariffs on the website <http://www.atih.sante.fr/tarifs-mco-et-had>

⁹For a detailed description how the index was constructed, its interpretation and applicability see: <http://drees.solidarites-sante.gouv.fr/etudes-et-statistiques/open-data/professions-de-sante-et-du-social/1-indicateur-d-accessibilite-potentielle-localise-apl/article/1-indicateur-d-accessibilite-potentielle-localisee-apl#nb2-1>

¹⁰This data is based on a representative survey of mothers who gave birth during these two years and it is available upon request on the "Archives de Données Issues de la Statistique Publique" website <https://www.cmh.ens.fr/greco/enquetes/XML/lil.php?lil=lil-0738>

Table 1.2 – Comparison of normal delivery and C-Section tariffs before and after the 2012 reform

Normal delivery DRG tariffs (No complication, gestational age 37-44 weeks)				
	2011		2012	
	unipara	multipara	unipara	multipara
Single birth	2187.13		2458.69	2070.55
Multiple birth			3301.83	2897.28

C-section DRG tariffs (no complication, gestational age 37-44 weeks)				
	2011		2012	
	unipara	multipara	unipara	multipara
Single birth	2792.71		2850.37	2850.37
Multiple birth			3852.63	3852.63

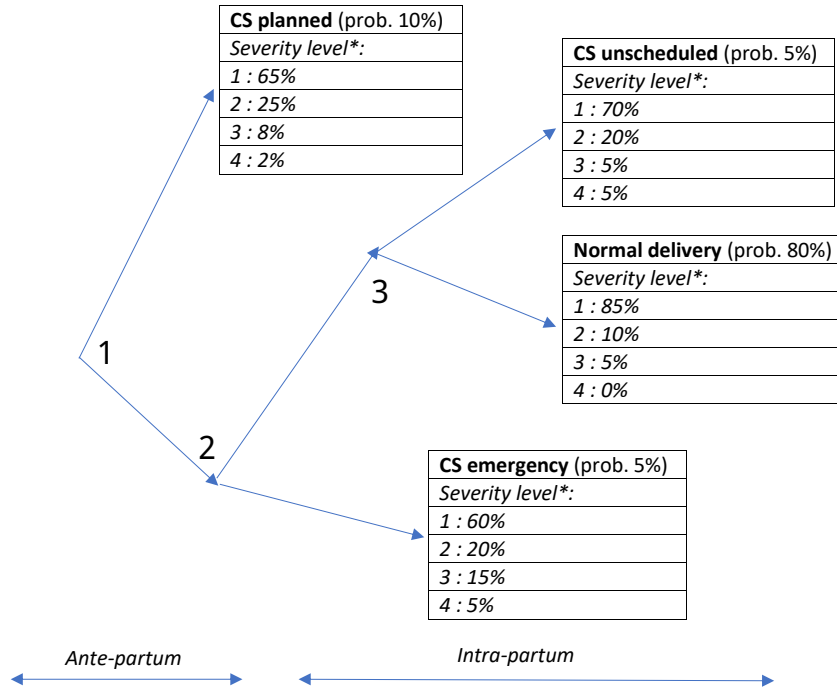
4 Empirical strategy

The individual-level dataset constructed for this study offers a unique opportunity to investigate the healthcare providers’ reaction to changes in hospital-level financial stimuli, following the March 2012 reform. This empirical setting allows us to make use of the fact that after the refinement, clinically identical cases were reimbursed differently, relative to periods preceding the reform (see Table 1.2 for an illustrative example).

More specifically, to estimate the effect of financial incentives, we implement the following empirical strategy:

1. estimate probabilities of each delivery outcome (scheduled, urgent, unscheduled C-sections, and normal delivery) based on patient characteristics observed in antepartum
 - (a) estimation is based on the pre-reform (training) dataset using nested logistic models with the tree structure provided in Figure 1.2.
 - (b) based on obtained estimates, predicted outcome probabilities are assigned to pre- and post-reform observations
2. each stay is classified into both pre- and post-reform DRG classifications (i.e. severity levels and resulting tariffs)
3. estimate probabilities of each severity level for each child delivery outcome and in both DRG classifications

Figure .1.2 – Decision tree for child delivery outcomes, with examples of DRG levels of severity (probabilities given for illustrative purposes)



- (a) based on the pre-reform (training) dataset, estimate ordered logit models with levels of severity as outcome
 - (b) assign predicted severity probabilities to all outcomes in both DRG classifications
4. calculate incentive measure reflecting attractiveness of scheduled C-sections
- (a) for each outcome and both DRG classifications, compute the expected tariff as probability-weighted average of tariffs over s – the set of possible severity levels: $\mathbb{E}(tariff_{outcome}^{DRGpost}) = \sum_s p_s^{DRGpost} \cdot tariff_{outcome,s}^{DRGpost}$ and $\mathbb{E}(tariff_{outcome}^{DRGpre}) = \sum_s p_s^{DRGpre} \cdot tariff_{outcome,s}^{DRGpre}$
 - (b) compute the difference Δ_1 of expected tariffs between unscheduled C-section and of a probability-weighted average tariff of other child delivery modes, separately for the two DRG classifications $\Delta_1^{DRGpre} = \mathbb{E}(tariff_{CSsch}^{DRGpre}) - \mathbb{E}(tariff_{other}^{DRGpre})$, and $\Delta_1^{DRGpost} = \mathbb{E}(tariff_{CSsch}^{DRGpost}) - \mathbb{E}(tariff_{other}^{DRGpost})$ (i.e. difference of expected tariffs between 2 edges at node 1 in Figure .1.2)
 - (c) compute the difference between DRG classifications $\Delta tariff = \Delta_1^{DRGpost} - \Delta_1^{DRGpre}$, which reflects how the attentiveness of scheduled C-sections evolved

between DRG classifications.

5. Run DID models testing the impact of incentives on the provability of scheduled C-section being chosen in ante-partum as child delivery mode
 - (a) based on the value of $\Delta tariff$, separate observations into control (no change in incentives) and two treatment groups, for whom the resulting expected DRG tariff incentive was lower than -100 euros and above 100 euros.
 - (b) testing difference-in-difference models, setting $\Delta tariff$ to zero for pre-reform observations (see Section 6.1 for placebo regression, where pre-reform values of $\Delta tariff$ are kept as originally assigned).

In our analysis, we rely solely on pre-admission (ante-partum) characteristics of a patient, which are used to predict the anticipated delivery outcome, and its severity. This allows us to calculate the expected tariffs that a hospital would receive at the moment when a choice for child delivery modes is made antepartum. The difference between the expected tariff for scheduled C-section delivery and probability-weighted average of other delivery modes in the pre-reform (i.e. Δ_1^{DRGpre}) subsequently serves as a reference point of the degree of financial attractiveness of scheduled C-section before the reform. Since tariffs changed in March 2012, so did the expected tariffs, assuming that factors other than DRG refinement did not affect the underlying patterns of clinical practice. Thus, for every child delivery case, the difference between tariffs for scheduled CS/other modes in post- and pre-reform would reflect how strongly changed the financial incentive to decide in favor of a unscheduled C-section ante-partum.

It is worth noting that the choice of child delivery mode in the post-reform period does not affect the value of the incentive measure, since all predictions are made solely based on pre-existing ante-partum diagnoses over which a gynecologist has no control. For example, risks of high blood pressure during pregnancy are a lot higher when labor starts at node 3 than in ante-partum node 1 of Figure .1.2. In contest to other studies on economic evaluation of obstetric care choices, we also use predicted gestational age instead of the observed one, since the mode of delivery will likely mechanically affect the length of pregnancy. It tends to be shorter in scheduled C-sections compared to, in particular, normal deliveries, since the former are induced by outside interventions. Thus, it allows us to eliminate, to the most possible extent, the possibility of coefficient bias due to reverse causality.

The method we develop enables us to take into account the dependence between types of procedure and in particular, the fact that a more complicated C-section would usually be associated with a more complicated normal delivery, and vice versa. This probabilistic (both in terms of outcomes and severity) approach differs from that used in *Allin et al.* (2015), who evaluate the impact of C-section and vaginal delivery tariff variations on C-section rates across anglophone Canadian provinces between April 1994 and March 2011. In contrast to our study, in *Allin et al.* (2015) counterfactual tariff for child delivery is assumed to be the average fee for a given delivery mode for cases that are relatively uncomplicated in terms of observed comorbidities. In the specific case of DRG refinement, ignoring this relationship may lead to serious bias. For example, if the average tariff remains the same and the tariff variance increases as a result of a refinement, so will the fee differential calculated as the difference between the fee for actual act and the average fee for the alternative mode of delivery. However, DRG refinement would likely result in changes in fees for the alternative delivery mode in the same direction as in the other delivery mode. Thus, if ignored in the price incentive calculation, this dependence will most likely lead to an overestimation of incentives measured as tariff differentials between C-section and vaginal delivery. This can potentially lead to a significant measurement error and biased estimates. This circumstance would warrants an adaptation of the empirical strategy used in *Allin et al.* (2015) to the French context featuring multiple tariff groups for both modes of child delivery though weighing tariffs by predicted probabilities of tariff groups based on the information available at the the ante-partum stage.

Despite the fact that PMSI is the most complete administrative dataset on hospital admissions available to researchers in France, it does not directly record several characteristics important in our context. First, in PMSI dataset the distinction between unipara and multipara mothers was only observable for women who delivered normally. This creates a situation in which one cannot immediately predict tariffs for normal delivery in the post-reform period for women who delivered via a C-section, since those tariffs depended on multiparty status.

As a solution to this missing data problem, we used variables containing DRG roots from previous obstetrics PMSI datasets available from 2005. Since every patient has a unique identification number in these datasets, we are able trace back the history of all hospital stays for a given patient. Similar to *Lo (2008)*, to recover unipara/mutipara distinction on the individual level we look back five consecutive years preceding the actual delivery and see how many mothers already had a previous pregnancy that will subsequently allow

us to classify the mother as multipara.¹¹

Let us denote M_n the fact that a mother was multipara and had a child in the period n , and \tilde{M}_n is multiparity status observed in PMSI data. According to Enquête Périnatale, $P(M_{2010}) = 0.566$. To estimate the number of recovered multiparity cases, we make the assumption that $P(M_t) = \text{const} = P(M_{2010}), t = \{2010, 2011, 2012, 2013\}$. This assumption is plausible because multiparity is directly linked to demographic characteristics such as fertility and the age at delivery. They are stable in countries that have already undergone a demographic transition, such as France. This assumption is also justified by statistical evidence since, according to the same survey, $P(M_{2003}) = 0.568$.

Overall, the share of all multipara cases that were retrieved on the individual level is $\frac{P(\tilde{M}_{2010-2013})}{P(M_{2010-2013})} = 0.947$. The missing 5.3% of cases are comprised of C-section cases only, since the multiparity status is completely observed for women who had a normal delivery. These cases originate from two main sources. First, they are comprised of women for whom the spacing between two consecutive childbirths exceeds 5 years. Second, this method ignores multipara women who had a child outside the French territory and, thus, were not registered in the PMSI dataset between 2005 and 2013. This category of patients can include immigrants and citizens with dual residence.¹²

In order to minimize the potential effect of this data omission on the internal validity of our results, we first restrict our analysis to multipara mothers, on whom the information on this characteristic is exactly known from PMSI or retrieved relying on the procedure described above. Subsequently, in Section 6 devoted to robustness checks, we explore the whole sample of all child deliveries.

Another data issue arises from the fact that in our dataset, except for the year of 2012, a medical procedure can only be observed in either pre- or post reform DRG classification. Without having information on how the same activity would have been classified in different DRG versions in terms of diagnosis root and severity level it would be impossible to predict what would have been the corresponding DRG tariffs for these same cases across different DRG versions. For the sake of this, we reconstructed the DRG decision tree for both pre-and post-reform periods available on ATIH website. The decision tree classifies cases into a specific DRG according to mother relevant medical characteristics (uniparity/multiparity, single/multiple delivery). The severity level is determined by the

¹¹Data limitations regarding PMSI do not allow us to consider a longer time period.

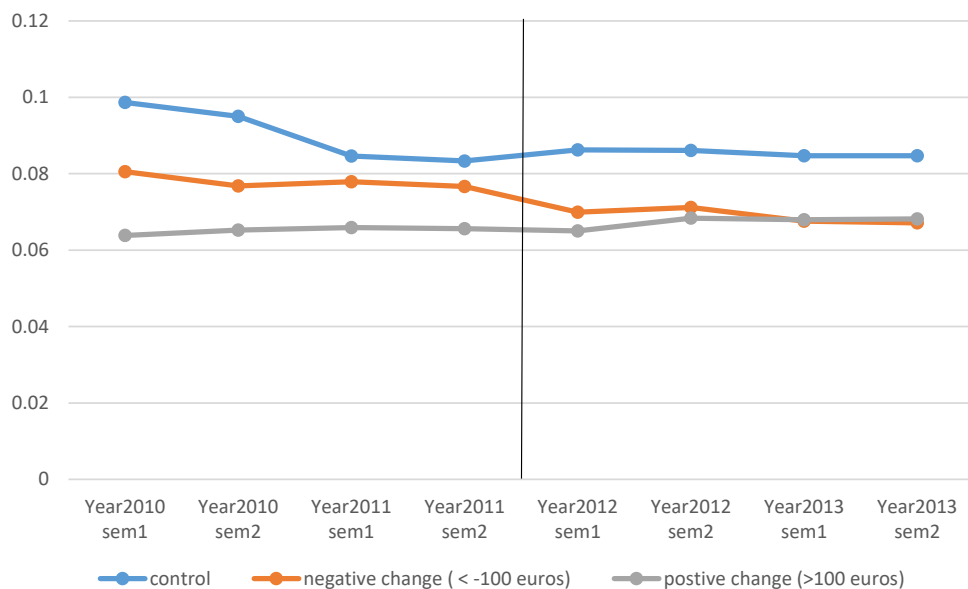
¹²In an attempt to reduce this bias, we tried restricting the sample only to younger women aged 20-25 who, thus, were not likely to have a child earlier. However, this did not significantly change the percentage of retrieved C-section multipara cases

most severe comorbidity observed for a given case.

If a child delivery occurred after (respectively before) the reform, the incentive measure $\Delta tariff$ is positive (respectively, negative) if and only if, compared to pre-reform tariffs and other things kept equal:

1. expectation of scheduled C-section tariff increased (decreased)
2. probability-weighted average of tariffs for the remaining outcomes decreased (increased)
3. a combination of both mentioned tariff changes occurred such that the expectation of C-section tariff increased (decreased) more significantly than that of other delivery modes, compared to the baseline year¹³

Figure .1.3 – Time trend of scheduled C-section rare for patients with positive (> 100 euros), negative (< -100 euros) and near-zero (control group) values of $\Delta tariff$, for all hospitals, 2010-2013.



Time trends for cases associated with positive, negative and small changes of variable $\Delta tariff$ are presented in Figure .1.3. This variable is continuous, which allows us to track later in DiD specifications how different intensities of the change in relative financial profitability of scheduled C-section and normal delivery affected the probability for the obstetrical care provider's choice to be in favor of the remaining delivery options, for

¹³Inflation changes are eliminated in the incentive measure $\Delta tariff$ adjusting tariffs for inflation rate

each child delivery case.¹⁴ DiD specifications require that pre-reform levels of an outcome variable follow the same trend in order for the obtained estimates to reflect the causal impact of the intervention in question. In Figure .1.3 the group of patients associated to a decrease in $\Delta tariff$ variable by more than 100 euros (orange line) saw a minor decrease in scheduled C-section in the 1st semester of 2012, while a group with an analogous positive change (grey line) underwent the opposite, albeit barely noticeable, change one semester later. In the meantime, C-section rates remained stable for observations in the control group (blue line) after the DRG reform was introduced, although it plausibly saw a minor decrease in the pre-reform 1st semester of 2011. Overall, the plot does not reveal significant divergence in outcome trends before the 1st semester of 2012 when the DRG reform introduced, suggesting that the DID estimates can be given causal interpretation. Every hospital incurs a financial cost for every C-section and normal delivery it performs, which is unobserved in the data. These costs are reimbursed through DRG payments to healthcare institutions, but only to a certain degree (which may or may not surpass the real cost). From a financial point of view, C-sections are generally more expensive to perform than normal deliveries. However, in our context, the mere fact that a C-sections, including scheduled ones, are reimbursed more generously by DRG schedules than a normal delivery does not automatically mean that a C-section becomes financially more attractive than a normal delivery since this DRG reimbursement may not cover the actual costs faced by each hospital. In other words, although every given secondary diagnosis is associated to a C-section and a normal delivery severity level (which may or may not coincide) the financial profitability of a specific case will ultimately depend on how well, for a given set of comorbidities and patient characteristics, the reimbursement covers the unobserved cost of the two delivery options. For each child delivery, the evolution of (1) variable between different DRG versions must be reasonably capturing the change in how full was the coverage of two delivery methods by DRG tariffs.

Although a number of previous studies (*Gruber et al.* 1999, *Allin et al.* 2015, etc) used the ratio (or price differentials) of tariffs for C-section and normal delivery as a measure of the financial incentive for providers to opt for either of the two delivery modes, this approach is not informative enough in the case of a DRG refinement leading to multiple tariff groups. First, it does not take into account the heterogeneity of cases in terms of initial DRG reimbursement (and thus the extent to which the DRG payment covers actual hospital costs) and instead takes average fees in a given year and geographical area.

¹⁴See Section 6 for specifications with alternative price incentive measures.

Second, as discussed in Section 2, in France tariffs in public and private sectors are not directly comparable since they include different components. Thus, measuring incentives relative to the baseline DRG classification, rather than their absolute level for a given period, will be a more consistent way to capture the financial incentives generated by the 2012 reform.

In addition, it is worthwhile to note that on average for-profit and non-profit institutions were affected slightly differently by the reform. As shown in Table 1.1, the mean value of $\Delta tariff$ was positive (negative) for for-profit (non-profit) hospitals, indicating that C-section became relatively more (less) profitable to perform in these institutions.

Difference-in-differences (DID) specifications are a common tool to estimate the causal effect of policy changes. Since the outcome variable—scheduled C-section indicator—is binary, we refer to our model as linear probability DID.

To test the effect of the price incentive, we proceed to estimate:

$$CSsch_{iht} = \alpha + X'_{iht}\beta_1 + \beta'_2\Delta tariff POST_t + d'_{iht}\xi + \gamma_t + \theta_h + \nu_t + \epsilon_{iht} \quad (1.1)$$

where $CSsch_{iht}$ is a binary outcome variable, i.e. the probability of a scheduled C-section performed for individual i , in healthcare institution h , in year t . Term α is constant, X_{iht} is a column vector of controls including age, secondary ante-partum (i.e. occurring before labor begins) diagnoses and other observed clinical characteristics, gynecologist availability index and pseudo-postal code average income, $\Delta tariff POST_t$ is the interaction term between the DRG refinement incentive defined earlier in this Section and dummy variables for post-reform periods $POST_t$.¹⁵ d_{iht} is a column vector of dummy variables for treatment groups, and ϵ_{iht} is a random error term. Where indicated, the model also includes year, month and hospital fixed effects γ_t , ν_t and θ_h respectively. In all our models error terms are clustered on the hospital level.¹⁶

Even though, in the general case, the linear probability model is biased and suffers from

¹⁵The inclusion of $POST_t$ variable sets to zero all values of $\Delta tariff$ for observations before March 2012. In general, DRG reimbursement rates underwent a negligible change in 2011 after an inflation and cost-driven re-adjustment of tariffs that were unlikely to influence the medical practice.

¹⁶Recent developments in econometric research into estimating coefficient standard errors under DiD designs has put forward alternative solutions to the widely popular cluster-robust variance estimator (CRVE). As shown, for example, in *Ferman and Pinto* (2020) and *MacKinnon and Webb* (2019), in the presence of few/relatively small treated groups CRVE tends to produce estimates leading to an overrejection of the null hypothesis. To correct for it, *Ferman and Pinto* (2019) suggest a cluster residual bootstrap providing a better to correction for heteroskedasticity. Since the application of this method would lead to an increase in reported standard errors, p-values for β_2 are expected to increase, resulting in an even lower statistical significance of β_2 estimates in our models.

heteroskedasticity, it gives correct average treatment effects, provided the data generating process is correctly specified.

The main coefficients of interest are contained in vector β_2 . In these specifications, it represents the number of percents by which the probability of a scheduled C-section increases if the DRG refinement incentive for scheduled C-section, scaled to thousand of euro, increases by one unit relative to the baseline DRG version. In other terms, if tariffs for other delivery modes are kept fixed, a 1K increase in expected scheduled C-section tariff relative to the pre-reform baseline DRG schedule would, on average, lead to a $100 \cdot \beta_2\%$ increase of the scheduled C-section probability.

In addition, after encoding all hospital stays into both pre- and post-reform DRG classification versions, we find that around 1% of cases are misclassified. We exclude these cases from the sample.

Models are tested for public, private non-profit and for-profit care institutions, separately for multiple and single child deliveries.

5 Main results

The results of various model specifications are presented in Table 1.3. Coefficients presented in models (1)-(9) are in line with general knowledge about obstetric interventions. Namely, having previously delivered via C-sections considerably increases the probability of a C-section in subsequent childbirths. Latter are associated to an increased risk of normal delivery failure, although they are not considered per se as a counter-indication to a normal delivery by the French Health Authority (HAS).¹⁷ In addition, in most cases a lower gestational age anticipated in intra-partum is associated with the risk of a pre-term, potentially more complicated delivery that would require a C-section.

The results also suggest that a higher concentration of gynecologists was associated with a lower probability of a scheduled C-section. Although models presented in *Milcent and Rochut* (2009) and *Gruber and Owings* (1996) reported a positive coefficient for obstetrician and gynecologist concentration, their respective studies concentrated on C-sections regardless of whether or not they were scheduled. However, in the specific case of sched-

¹⁷The French Healthcare Authority (HAS) periodically monitors medical appropriateness for C-sections. It concluded that in order to avoid evening and night-time deliveries, French obstetricians have tended to schedule C-sections too early. Moreover, keeping track of healthcare practitioners' individual record for C-section rates may lead to changes in medical practices. An experiment of this type was carried out in a subset of hospitals and its results show that in these hospitals C-section rates dropped from 22,2% in 2010 to 17% in 2013. (*Vanlerenberghe, 2015*)

Table 1.3 – Difference-in-difference estimation of the impact of DRG refinement incentives ($\Delta tariff$, rescaled to thousands of euro) on scheduled C-section probability in multipara women in 2010-2013

	Public			Private non-profit			Private for profit		
	Single 1	Single 2	Multiple 3	Single 4	Single 5	Multiple 6	Single 7	Single 8	Multiple 9
<i>Dependent variable:</i>									
$\Delta tariff_{POST}$	-0.0071 (0.0026)***	-0.0074 (0.0024)***	-0.006 (0.02)	-0.0066 (0.0069)**	-0.0074 (0.007)	-0.29 (0.22)	-0.012 (0.01)	0.013 (0.01)	-0.034 (0.041)
<i>age</i>	-0.012 (0.004)***	-0.02 (0.005)***	0.08 (0.07)	-0.036 (0.017)**	-0.03 (0.021)	-0.3 (0.26)	0.008 (0.008)	-0.006 (0.1)	0.091 (0.18)
$\ln(age)$	-0.19 (0.06)***	-0.24 (0.06)***	-1.14 (1.1)	0.52 (0.25)**	0.4 (0.26)	0.46 (0.38)	-0.11 (0.12)	0.03 (0.13)	-0.65 (2.85)
age^2	< 0.0001 (< 0.0001)***	-0.015 (0.005)***	-0.008 (0.006)	0.00034 (0.00014)**	0.0002 (0.0002)	0.002 (0.002)	-0.0006 (0.0007)	0.0001 (0.001)	-0.001 (0.001)
$E(gestational\ age)$	0.003 (0.005)	-0.04 (0.018)**	0.07 (0.026)***	-0.014 (0.017)	-0.004 (0.084)	0.1 (0.12)	-0.03 (0.008)***	-0.034 (0.03)	-0.011 (0.049)
<i>number of prev. CS</i>	0.295 (0.005)***	0.296 (0.005)***	0.278 (0.012)***	0.356 (0.011)***	0.353 (0.01)***	0.408 (0.041)***	0.426 (0.005)***	0.422 (0.005)***	0.372 (0.019)***
<i>median income (zip code)</i>	-0.0001 (0.0002)	-0.0003 (0.0002)	0.0012 (0.0006)*	0.0004 (0.0003)	-0.0001 (0.0002)	0.0002 (0.0027)	-0.0006 (0.0003)**	-0.0004 (0.0001)***	0.003 (0.0015)*
<i>Gynecologist acces index</i>	-0.0009 (0.001)***	-0.0005 (0.0001)***	-0.0015 (0.0012)	-0.0006 (0.0003)*	-0.0007 (0.0001)***	0.002 (0.0023)	-0.0002 (0.0002)	-0.0007 (0.0001)***	-0.027 (0.0018)
<i>treated_high</i>	-0.0012 (0.0008)	-0.0022 (0.0008)***	0.045 (0.027)*	-0.01 (0.0067)	-0.012 (0.006)*	-0.089 (0.192)	-0.016 (0.005)***	-0.016 (0.005)***	0.0096 (0.024)
<i>treated_low</i>	0.01 (0.002)	0.009 (0.0023)***	0.006 (0.026)	-0.024 (0.001)**	-0.026 (0.01)**	-0.406 (0.206)	-0.016 (0.005)***	-0.016 (0.006)***	-0.096 (0.069)
<i>(Intercept)</i>	0.253 (0.23)	1.97 (0.75)***	-0.42 (2.3)	-0.4 (0.83)	-0.52 (3.39)	-1.24 (0.94)	1.51 (0.41)***	-1.49 (1.4)	0.931 (5.73)
<i>Month FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Year FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Ante-partum diagnosis FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Hospital FE</i>	NO	YES	NO	NO	YES	NO	NO	YES	NO
N	955,342	955,342	15,356	117,178	117,178	1,167	382,687	382,676	3,692
R2	0.2	0.2	0.34	0.22	0.22	0.37	0.23	0.23	0.38

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at hospital level

uled C-sections, a higher concentration of gynecologists may produce the opposite effect through the channel of reducing the informational asymmetry between patients and care providers due to an easier availability of a second medical opinion. Scheduled C-sections, in contrast to unscheduled ones, leave patients a considerable time window to seek consultation with another specialist. As a result, the fact that an alternative medical opinion is more readily available to a patient may serve as credible deterrence against unjustified C-section decisions made in particular by obstetricians and gynecologists.

Models (1)-(2) estimate the impact of DRG financial incentives on single birth multipara mothers admitted to public facilities, using controls specified in Section 3, without and with hospital fixed effects, respectively. Although in both models the coefficient of interest β_2 representing the change in financial incentives compared to the baseline period, is negative and statistically significant at 1%, its magnitude is small (0.7% decrease in C-section rate with every 1K change in incentives). However, as we will explore in Sec-

tion 6.1, public hospitals likely fail the placebo test, warranting additional caution while interpreting this result. Assuming the magnitude of β_2 of 0.007 and the average size of incentives equaling 42 euros, the change in C-section rates attributable to the 2012 DRG reform is estimated to be lower than 0.1%. These findings are in line with *Allin et al.* (2015) conclusion that the effect of the CS to normal delivery fee ratio on C-section rates is likely to be modest and is statistically insignificant.¹⁸

As far as private non-profit and private for-profit hospitals are concerned, both show no significant reaction to DRG stimuli in terms of their propensity to perform scheduled C-sections. Although in model (4) tested on single birth mothers the coefficient β_2 is significant, it is no longer so after the inclusion of hospital-fixed effects model (5). As for the remaining models (6)-(9), β_2 is consistently insignificant, suggesting that . In general, the introduction of hospital fixed effects in specifications (2), (5) and (7) led to only a negligible increase in the share of explained variation.

In Table 1.4 we present DiD estimation results for the effect of DRG incentives on groups that were affected by them in opposite directions. Thus, those mothers for whom $\Delta tariff > 100$ are considered as subject to hospital incentives to favor scheduled C-section, while $\Delta tariff < -100$ indicates an incentive to favor other modes of childbirth. The set of included covariates is identical to specifications analyzed earlier in this Section presented in Table 1.3.

Overall, across all hospital types results indicate no significant impact of DRG incentives on relatively more strongly affected groups in terms of their probability to give birth via scheduled C-section. In several cases where associated coefficients are marginally significant and are likely to be an artefact of statistical testing. This conclusion holds equally for single and multiple childbirths. As in earlier discussed models, the size of estimated impact in all cases remains very modest and unlikely to impact in a meaningful and noticeable way the pre-existing obstetric practices.

¹⁸In particular, they find that increasing “remuneration for a C-section by 100 percent relative to the baseline fee of a vaginal delivery might increase the probability that a physician opts for Caesarean by 0.6 to 1.1 percentage points”. Assuming French average inflation-adjusted vaginal delivery tariff of 2054 euros, French estimates would provide the corresponding figure of a 0.9-1.9% change in the scheduled C-section rate for an average delivery

Table 1.4 – Difference-in-difference estimation of DRG refinement on treatment groups ($\Delta tariff > 100$ and $\Delta tariff < -100$) on scheduled C-section probability in multipara women in 2010-2013

	Public		Private non-profit		Private for profit	
	Single 1	Multiple 2	Single 3	Multiple 4	Single 5	Multiple 6
Dependent variable: (Intercept)	2.02 (0.78)**	-0.38 (2.3)	-0.5 (3.64)	-0.53 (2.25)	-1.04 (1.41)	0.364 (0.364)
<i>T_high*year 2012 semester 1</i>	0.002 (0.015)	-0.058 (0.088)	-0.01 (0.007)	-0.13 (0.07)*	0.001 (0.009)	-0.025 (0.048)
<i>T_high*year 2012 semester 2</i>	0.002 (0.0015)*	0.123 (0.094)	-0.005 (0.007)	0.145 (0.107)	0.004 (0.009)	0.093 (0.051)*
<i>T_high*year 2013 semester 1</i>	0.0003 (0.002)	-0.026 (0.06)	-0.029 (0.015)*	-0.134 (0.085)	0.008 (0.012)	0.064 (0.057)
<i>T_high*year 2013 semester 2</i>	-0.002 (0.0019)	0.122 (0.072)*	-0.023 (0.015)	0.011 (0.089)	0.01 (0.012)	0.033 (0.044)
<i>T_low*year 2012 semester 1</i>	0.008 (0.005)*	0.019 (0.044)	0.027 (0.018)	0.02 (0.044)	0.017 (0.014)	-0.082 (0.084)
<i>T_low*year 2012 semester 2</i>	0.012 (0.005)**	0.003 (0.043)	-0.027 (0.016)*	0.011 (0.043)	0.002 (0.015)	-0.049 (0.116)
<i>T_low*year 2013 semester 1</i>	0.001 (0.004)	0.037 (0.038)	-0.0227 (0.019)	-0.061 (0.055)	0.021 (0.016)	-0.007 (0.143)
<i>T_low*year 2013 semester 2</i>	-0.0005 (0.005)	0.018 (0.037)	-0.017 (0.022)	-0.078 (0.054)	0.143 (0.017)	-0.138 (0.175)
Other controls	YES	YES	YES	YES	YES	YES
Month FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Ante-partum diagnosis FE	YES	YES	YES	YES	YES	YES
Hospital FE	YES	NO	YES	NO	YES	NO
N	955,342	15,356	117,178	1,125	382,687	3,692
R2	0.2	0.34	0.22	0.36	0.23	0.38

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at hospital level

6 Robustness checks

6.1 Placebo DiD regression

To test whether the estimates obtained in Section 5 are not affected by pre-reform changes uncontrolled for in our models, we perform DiD placebo estimation. These models include as covariates the interaction terms between the DRG incentive measure $\Delta tariff$ and dummies for all semesters in 2010-2013, except for the second semester of 2011 serving as the baseline pre-reform period. Estimation results are presented in Table 1.5.

As mentioned in Section 5, results on single birth mothers who delivered in public hospitals may not stand this test of placebo DiD. In particular, model (1) shows 2 out of 3 statistically significant coefficients for pre-reform periods (for 1st semester of 2010 and 1st semester of 2011, with 5% and 10% level of significance, respectively). As far as other hospital types and mothers groups are concerned, the results indicate no sign that pre-reform periods were affected through channels other uncontrolled for in our functional specifications.

As for the estimated post-reform effects, the obtained results confirm the conclusion that DRG incentives did not significantly affect the probability of scheduled C-section in private non-profit and for-profit clinics. A few sporadically significant coefficients in models (3) and (6) do not exhibit a clear pattern. In public hospitals 2 out of 4 coefficients corresponding to both of 2012 semesters are negative and statistically significant. However, in light of the discussed pre-reform estimates, this estimate should be interpreted with considerable caution.

Finally, it is worth noting that similarly to the main models presented in the previous section and even regardless of the results of placebo DiD on the significance of pre-reform coefficients, the magnitude of the post-reform estimated effects remains low and overall unlikely to impact the clinical practice on a scale posing any significant concern.

6.2 Models with deterministic severity

In this subsection we are testing a different approach to measuring DRG tariff incentives for scheduled C-sections that relies on a set of simplifying assumptions. In line with other studies on tariff incentives in obstetrics (as in, for example, *Allin et al. (2015)*, *Gruber et al. (1999)*), we drop the tree structure of obstetrician/midwife decisions about child delivery modes and consider that tariff can be known with certainty in advance based on ante-partum patient characteristics.

Table 1.5 – Difference-in-difference placebo regressions on scheduled C-section probability in multipara women in 2010-2013 ($\Delta tariff$ rescaled to thousands of euro)

	Public		Private non-profit		Private for profit	
	Single 1	Multiple 2	Single 3	Multiple 4	Single 5	Multiple 6
Dependent variable: (Intercept)	1.67 (0.75)**	-0.42 (2.31)	-0.04 (3.45)	-1.2 (0.91)	-0.88 (1.51)	1.12 (5.83)
$\Delta tariff * year$ 2010 semester 1	0.0051 (0.0018)**	-0.032 (0.034)	0.0055 (0.01)	-0.415 (0.262)	0.016 (0.012)	-0.067 (0.212)
$\Delta tariff * year$ 2010 semester 2	-0.002 (0.0026)	-0.012 (0.027)	0.013 (0.007)*	-0.28 (0.294)	0.013 (0.013)	-0.017 (0.175)
$\Delta tariff * year$ 2011 semester 1	-0.012 (0.0037)***	-0.023 (0.03)	-0.003 (0.013)	-0.262 (0.236)	0.013 (0.014)	0.095 (0.171)
$\Delta tariff * year$ 2012 semester 1	-0.0094 (0.0032)**	-0.019 (0.039)	-0.02 (0.01)*	-0.6 (0.26)**	-0.0054 (0.017)	-0.037 (0.18)
$\Delta tariff * year$ 2012 semester 2	-0.001 (0.0044)***	0.032 (0.045)	0.013 (0.012)	-0.504 (0.25)*	0.0076 (0.015)	0.0045 (0.037)
$\Delta tariff * year$ 2013 semester 1	-0.0043 (0.0038)	-0.05 (0.02)**	0.084 (0.01)	-0.319 (0.31)	-0.019 (0.018)	-0.041 (0.143)
$\Delta tariff * year$ 2013 semester 2	-0.0073 (0.0046)	0.0028 (0.034)	0.0019 (0.0016)	-0.1599 (0.404)	-0.0036 (0.017)	0.228 (0.138)*
Other controls	YES	YES	YES	YES	YES	YES
Month FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Ante-partum diagnosis FE	YES	YES	YES	YES	YES	YES
Hospital FE	YES	NO	YES	NO	YES	NO
N	955,342	15,356	117,178	1,167	382,687	3,692
R2	0.2	0.34	0.22	0.37	0.23	0.38

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at hospital level

These more simplistic, yet more tractable, deterministic models rely on several additional assumptions. First, it is assumed that the main comorbidity (and thus, DRG severity levels) can perfectly predict the final DRG tariff for delivery modes based on the most severe ante-partum comorbidity observed in ante-partum in a given mother. This assumption is likely to hold for scheduled C-sections and normal deliveries, where the most severe patient diagnosis is usually known before the start of labor. However, unscheduled C-sections are less predictable in this respect because they are normally performed when a vaginal delivery is no longer possible due to high risks of severe health consequences. As a result, this circumstance also leads us to focus specifically on the choice between a scheduled C-section versus normal deliveries in this subsection. At the same time, we are aware of *Johnson and Rehavi* (2016) conclusion that information asymmetry may increase the probability of unscheduled C-sections. However, contrary to our study, in *Johnson*

and *Rehavi* the presumed channel which affected the choice of delivery option is parental professional occupation (physician/non-physician) and not changes in fee incentives.

Second, we assume partial altruism of providers manifesting itself in no purposeful prolonging of labor with the goal of aggravating a patient's state and, thus, inducing a higher severity on a patient and increasing the resulting DRG tariff. Although violation of this assumption may lead to reverse causality problem making the incentive measure dependent on the outcome, we strongly believe that it holds for the overwhelming majority of obstetricians and midwives.

A larger computational tractability of these models allows us to test additional formulations of DRG incentives and apply them to a larger sample of mother that include in these models both unipara and multipara¹⁹. In addition, patients sharing a common diagnosis root and a common severity level for the observed and counterfactual child delivery modes in both pre-reform and post-reform DRG classifications are considered to belong to the same treatment group, sharing a common amount of associated financial stimuli to perform scheduled C-sections. Since the number of these groups is in this case finite, it allows to test DiD specifications with control for unobserved heterogeneity. Overall, this gives rise to as many as 850 groups that were subject to a treatment (i.e. financial incentive to perform a specific type of delivery) of different intensity. For simplicity, we excluded small groups with a number of cases below 2500.

The deterministic analogue of Δ_{tariff} is denoted as $\Delta_t - \overline{\Delta}_{2010}$. It is computed following the same steps described in Section 4, except for the fact that no prediction is made neither with regards to levels of severity (which instead are assigned deterministically based on the most severe ante-partum comorbidity) or the (assumed to be either a C-section or a normal delivery, in a way similar to most other existing studies, for example, *Gruber et al. (1999)*). This measure is complemented by its twin, obtained by dividing this expression by the baseline level of tariff differential $\overline{\Delta}_{2010}$.

In a similar vein, we also construct two measures that rely on the ratio between DRG tariffs $r_k = \frac{CS_k}{ND_k}$, instead of their absolute difference $\Delta_k = CS_k - ND_k$. In this notation, CS_t and ND_t stand for DRG tariffs for C-section and normal delivery respectively, observed (or predicted for the counterfactual delivery outcome) using the DRG version of year t , and \overline{CS}_{2010} , \overline{ND}_{2010} are baseline tariffs that would have been observed if the child delivery had taken place in the pre-reform year 2010. For simplicity, we denote the ratio $\frac{CS_k}{ND_k}$ as r_k .

¹⁹More conservative estimates that unipar produce the same results

Thus, four measures for DRG refinement incentives relying on deterministic severity are:

1. $\Delta_t - \overline{\Delta_{2010}}$, where Δ_k is inflation-adjusted price differential $CS_k - ND_k$ in year k , at 2010 prices. It represents the deterministic equivalent of Δ_{tariff}
2. $(\Delta_t - \overline{\Delta_{2010}}) / \overline{\Delta_{2010}}$, which is the same measure as 1, calculated relative to the hypothetical level of the price differential in 2010.
3. $r_t - \overline{r_{2010}}$, or change of tariff ratios between post- and pre-reform DRG schedules, which, compared to measure 1, uses price ratios r_k between C-section and normal delivery tariffs
4. $(r_t - \overline{r_{2010}}) / \overline{r_{2010}}$, which is the same measure as 3, calculated relative to the hypothetical level of the price differential in 2010.

By definition, all these price incentive measures have value 0 in 2010 because of the absence of change, i.e., $\overline{r_{2010}}$ (or $\overline{\Delta_{2010}}$) and r_t (or Δ_t) coincide.

The results for different measures of DRG refinement incentives are presented in Table 1.6. The tested models include a full set of covariates, including multitariff and multiple pregnancy indicators. The specifications also control for year and hospital fixed effects (as in models (5)-(7) in Section 5). As earlier, model errors are clustered at the hospital level. Estimation results are presented in Table 1.6.

Overall, the obtained estimates are in line with those presented in Section 5. The main coefficient of interest β_2 remains statistically significant only at 10% and 5% significance level in regressions run on for institution types and non-profit facilities, respectively. However, when measuring DRG refinement incentives using different approaches, even the minimum conventional 10% significance level of the coefficients β_2 is no longer achieved, suggesting an absence of the effect of DRG incentives on scheduled C-section probability.²⁰

In addition, compared to main models presented in Section 5, the magnitude of coefficients related to $\Delta_t - \overline{\Delta_{2010}}$ decreases to an even lower level. However, we need to note that this drop in coefficients' absolute value may be a consequence of attenuation bias due measurement error, which could have been caused by employing a more simplified severity prediction algorithm.

²⁰additional models with deterministic severity are provided in Appendix 3

Table 1.6 – Difference-in-difference coefficients for deterministically computed incentive measures, by hospital type

Measures of financial incentive	Hospital type		
	All	Non-profit	For-profit
$\Delta_t - \overline{\Delta}_{2010}$	-0.0008* (0.0005) (0.12)	-0.001** (0.0005) (0.03)	0.0046 (0.0038) (0.23)
$(\Delta_t - \overline{\Delta}_{2010}) / \overline{\Delta}_{2010}$	-0.0003 (0.0004) (0.48)	-0.0006 (0.0004) (0.22)	0.0016 (0.0013) (0.23)
$r_t - \overline{r}_{2010}$	-0.0012 (0.001) (0.66)	-0.002 (0.0014) (0.16)	0.004 (0.004) (0.26)
$(r_t - \overline{r}_{2010}) / \overline{r}_{2010}$	-0.0009 (0.002) (0.62)	-0.0024 (0.0025) (0.23)	0.0064 (0.0056) (0.21)

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors and p-values are given in parentheses under the coefficient. DID, linear probability model with year and hospital fixed effects

7 Discussion and policy implications

In this Chapter we have shown the absence of a link between financial incentives and the providers' choice between a scheduled C-section and a normal delivery. The results show that obstetricians working either in non-profit or for-profit hospitals in France did not significantly respond to changes in tariffs caused by the DRG refinement.

The results of presented in this Chapter are most likely to be driven by four channels. First, the results can be explained by the fact that in France healthcare providers' contracts are drawn up in a way that makes them insensitive to hospital-level financial objectives. As a result, it is difficult for healthcare institutions to enforce their financial policies on individual providers who have substantial autonomy and legal guarantees. In addition, healthcare management could potentially face high reputation costs if attempts to exert influence on medical practices became public. Arguably, this situation is even less likely to occur in larger maternity wards where policy coordination between administration and individual healthcare providers is generally considered to be lower. This highlights the importance of giving consideration to incentive compatibility in designing policy across hospitals and providers.

Second, it is important to emphasize that the cost of understanding the financial consequences of DRG reforms can be fairly high for both healthcare management and individual

practitioners. This may have rendered their perception about the arising stimuli imprecise.

Third, since DRG tariff incentives are generated at the hospital level and the transmission of these incentives to the level of individual providers may require their consent – express or implicit – providers’ altruism is an important channel to consider. While the topic of provider altruism is relatively not well-researched, the evidence for at least partially altruistic providers is consistent in economics literature. Thus, this channel could have mitigated (through a decreased uptake of opportunistic behaviors by at least partially altruistic agents, as well as the possibility of whistleblowing by medical staff) potential attempts to use DRG instruments to increase hospital funding.

Lastly, at the level of healthcare institutions, the reform eased the financial pressure on bigger maternity wards since it resulted in an overall revenue increase. In other words, DRG refinement mostly benefited those hospitals that had an increased share of high-cost ‘outlier’ patients presenting with complicated medical comorbidities and who were targeted the most by the policy intervention. Thus, after the reform, bigger maternity wards did not have strong incentives to react to changes in terms of relative DRG reimbursement rates because of their improved financial situation. On the other hand, overall revenues of small obstetric care facilities generally remained stable or even decreased, making this channel less relevant for them.

The evidence presented in this Chapter suggests that the main objective of the DRG reform - decreasing financial risk for obstetric healthcare providers - was plausibly met without significant impact on associated obstetric practices. However, the overall impact of the DRG refinement reform in terms of public welfare is not clear. Although unobservable, transnational and administrative costs related to encoding medical procedures are likely to have increased as a consequence of the DRG reform; the exact magnitude of this effect is cannot be observed.

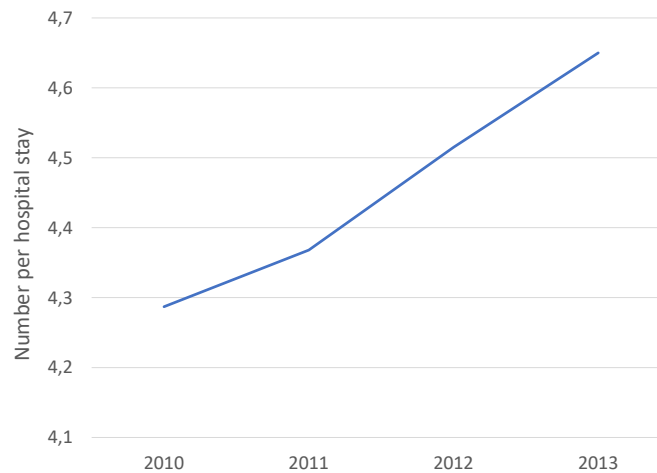
This Chapter contributes to the previous research in three ways. First, our methodology introduced a considerable improvement to measuring tariff differentials. Unlike previous studies, we make use of probability-weighted DRG tariffs, which are predicted for each mode of child delivery solely based on pre-admission (ante-partum) patient characteristics. This allows us, in particular, to take into account the uncertainty an obstetrician/midwife are facing with regards to post-admission level of severity, and to exploiting more fully the resulting variation in tariff incentives due to elimination of likely reverse causality and attenuation biases. Second, our study is the first to analyze the impact of tariff refinement

in the context of Diagnosis-Related Groups - currently the most widely adopted payment scheme in Europe. Third, to the best of our knowledge, effects of DRG tariff refinement on care provision have never been previously studied in the realm of obstetrics.

There are several channels that can be hypothesized to affect the internal validity of obtained estimates. First, DID strategy allows eliminating only time-invariant component in group unobserved heterogeneity. Such an event could be the publication of practice guidelines by HAS in 2012, which was coincidental with the DRG reform. However, despite the importance of this publication, it contained no strong indications or counter-indications for scheduled C-sections, instead listing in a non-exhaustive fashion about 7 common medical conditions for which a scheduled C-section can be recommended. Hence, the extent to which it could affect the actual obstetrician's practice appears to be restricted to cases where obstetric patients presented with only relatively common comorbidities, the final decision still remaining at the obstetrician's end, and in some cases, at the mother's discretion.

Second, the "upcoding channel" introduced in Section 2, if present, can lead to non-random time-dependent measurement error in comorbidity variables, which could have the potential to cause estimates' bias and inconsistency. However, the percentage of top coded patients did not change considerably over the analyzed time period. These patients with severe and major ante-partum comorbidities represented only 0.21% of C-section cases in 2010 vs 0.26% in 2013, while corresponding figures for mothers who delivered normally stood at 0.31% and 0.37% in 2010 and 2013, respectively. There are two main reasons to believe that "upcoding" channel did not significantly impact obtained estimates. First, the upward trend in the number of encoded comorbidities existed before the 2012 reform, making it unlikely for this channel to significantly bias the results of the estimates. In obstetrics, this trend was likely due to increasing women's average age at child delivery (and, thus, higher frequency of age-related complications) and changes in comorbidity classification in French DRG manuals. The latter occurred as a result of a consistently increasing number and level of thoroughness in comorbidities coding with each subsequent DRG manual in France. Second, our identification strategy relies solely on ante-partum comorbidities, which are less likely to be manipulated to due to patients' ability to better evaluate their health status in ante-partum and, in case of doubt, to seek a second medical opinion.

Figure .1.4 – Average number of encoded comorbidities per child delivery (2010-2013)



The efficiency of the estimates could have increased if additional explanatory variables had been included. For example, our database lacks information about the mother's occupation which can affect provider's choice between delivery options if the mother is perceived by a physician as lacking relevant obstetrical knowledge (*Johnson and Rehavi, 2016*). Similar to other related studies, we cannot observe variables based on which patients likely self-select into private/public healthcare providers, such as a mother's risk-aversion and idiosyncratic preferences. In France, under certain conditions, women have freedom of choice between a C-section and a normal delivery. This freedom is generally larger in private hospitals, where women can have more influence on the treatment plan they receive.

8 Conclusion

This Chapter attempted to estimate the impact of DRG refinement and subsequent tariff changes on the choice between scheduled C-section and normal delivery. This policy shock occurred in 2012 in an attempt to reduce financial risk for obstetric care providers by introducing additional parameters into the price formula, such as multiparity, multiple birth, gestational age and new comorbidities, bringing the payment system closer to a cost-plus approach. Thus, obstetric patients who were identical in terms of observed characteristics would have had different DRG reimbursement rates before and after the reform, changing the relative profitability between child delivery options. To exploit this exogenous variation in DRG tariffs, we adopted both probabilistic and deterministic

approaches to measuring DRG price incentives and tested them in difference-in-difference models. The results suggest that the reform did not have a significant impact on healthcare providers' obstetrical practice. Models featuring alternative formulations for measures of financial incentives led to the same conclusion. Moreover, there is no evidence that effects varied across years and for different types of care institutions (for-profit or non-profit). The results can be explained by three main channels. Firstly, the transmission of hospital-level financial stimuli to individual care providers appears to be limited. Secondly, the transaction costs associated with understanding changes in DRG tariff stimuli can be relatively high for hospital administrators. Lastly, larger obstetric care institutions benefited from increased tariffs, due to recalibration of tariffs for the most severely coded obstetric patients targeted by the reform. As a result of an improved financial position, they faced less pressure to raise profits by encouraging changes in medical practice.

This study fills a relative lack in the literature on the impact of changes in financial incentives due to DRG refinements. Methodologically, our contribution consists in proposing an approach to modeling changes in DRG financial incentives at the patient level. In contrast to earlier studies that considered *average* fees for a restricted sample of uncomplicated deliveries (as in *Allin et al. (2015)*) to calculate the corresponding fee incentives, we construct these measures based on probability-weighted DRG tariff groups for child delivery modes according to DRG classifications before and after the refinement reform. The resulting estimates exploit tariff variations in a nuance fashion, thus alleviating a potential source of attenuation bias due to measurement error. To the best of our knowledge, the consequences of DRG refinement and subsequent tariff changes have never been studied specifically in the realm of obstetrics.

9 Appendixes

Appendix 1

The 'Acts' dataset provides information on all medical procedures that were performed on a patient during their hospital stay in a given year. In addition, it includes characteristics that provide details about specific circumstances in which a medical act was performed (e.g. late in the night, on Sunday or on a holiday). These specific characteristics lead to a modification of the price associated with a procedure. For example, if a procedure was performed by a gynecologist between 0am and 6am, this would lead to an increase in prices by 40 euros. Reflecting the two main delivery options: vaginal labor and C-section, medical procedures used to denote the fact of giving birth consist of two groups and are coded with prefixes 'JQGD' and 'JQGA' respectively. However, this dataset does not allow to identify a patient, and only characterizes a hospital stay.

In CMD each diagnosis is associated with a medical procedure. The former are provided in the 'RSA' dataset. In 'RSA' each hospital stay has a unique patient identifier which allows us to track the cases of hospitalization across different years for a given patient. In addition, it allows me to find a corresponding medical activities contained in the previous dataset. It is worthwhile to note that since in France each version of DRG comes into effect on the 1st of March of each year, 'RSA' dataset classifies diagnoses based on two versions of DRG, that were operational before and after the 1st of March. Apart from diagnoses, 'RSA' dataset contains a rich set of individual level characteristics (including age, region and community of residence, length of hospital stay, month of exit, etc.) and hospital characteristics, which include hospital unique identifiers, ownership information (i.e. public or private), legal status (i.e. profit or non-profit), geographic location, etc.

The final 'DIAG' dataset lists complications (or secondary diagnoses) that were observed during each hospital stay. These complications describe medical conditions that arise both before and after hospital admission. The data are presented in a way similar to 'Actes' dataset. In DRG, these complications have different severity levels, the number of which may vary from 1 to 3 depending on year and diagnosis roots. It is worth noting that with the introduction of DRG version 11d in 2012, not only diagnosis roots underwent a major revision, but also severity levels.

Appendix 2

Table 1.7 – Correlation matrix of main variables

	Sch. C-section	$r_k - \bar{r}_{2010}$	$\Delta_t - \bar{\Delta}_{2010}$	Age	Num. of comorb.	Gest. age	Multiple pregn.	Num. of previous C-sect.	Sunday or holiday	Multipara	Average pseudo postal code income	Gyn. availability
Scheduled C-section	1											
$r_k - \bar{r}_{2010}$	0.001	1										
$\Delta_t - \bar{\Delta}_{2010}$	-0.001	0.958	1									
Age	0.112	0.071	0.041	1								
Num. of comorb.	0.006	-0.164	-0.187	0.013	1							
Gest. age	-0.09	0.006	0.072	-0.016	-0.126	1						
Multiple pregn.	0.06	-0.067	-0.096	0.036	0.102	-0.27	1					
Num. of previous C-sect.	0.389	0.068	0.038	0.104	0.055	-0.057	-0.003	1				
Week-end	-0.07	0.009	0.01	-0.013	-0.01	0.014	-0.001	-0.025	1			
Multipara	-0.02	0.306	0.207	0.268	-0.098	0.024	-0.023	0.24	0.006	1		
Average pseudo postal code income	0.012	-0.01	-0.006	0.14	>-0.001	0.013	0.008	0.001	0.005	-0.033	1	
Gyn. availability	-0.0016	-0.018	-0.016	0.062	0.038	-0.009	0.003	-0.009	0.001	-0.05	0.171	1

Average pseudo postal code income	Gyn. availability
1	
0.171	1

Table 1.8 – Obstetrical DRGs before and after 2012 reform

DRG before reform (versions 11b, 11c)			DRG after reform (versions 11d, 11e)		
DRG code	Num. of severity levels	Name of procedure	DRG code	Num. of severity levels	Name of procedure
14C02	3	Cesarean section	14C06	4	Cesarean section, with child dead
			14C07	4	Cesarean section, multiple pregnancy
			14C08	4	Cesarean section, single pregnancy
14C03	1	Normal delivery, with other interventions	14C03	4	Normal delivery, with other interventions
			14Z10	2	Normal delivery, with child dead
			14Z11	2	Normal delivery, unipara mother with multiple pregnancy
14Z02	3	Normal delivery	14Z12	2	Normal delivery, multipara mother with multiple pregnancy
			14Z13	4	Normal delivery, unipara mother with single pregnancy
			14Z14	4	Normal delivery, multipara mother with single pregnancy

Table 1.9 – Effect of gestational age on severity level of a normal delivery with single pregnancy, after 2012 DRG reform

Severity level	Gestational age				
	22-31	32	33-35	36	37-44
D	D	D	D	D	D
C	C	C	D	C	C
B	B	C	D	C	B
A	A	B	C	B	A

Table 1.10 – Effect of gestational age on severity level of a Cesarean section, after 2012 DRG reform

Severity level	Gestational age				
	22-31	32	33-35	36	37-44
D	D	D	D	D	D
C	D	D	D	D	C
B	C	C	C	C	B
A	B	B	C	B	A

Appendix 3

Appendix 3.1. Additional DiD models with deterministic severity

In this appendix we are testing additional DiD models featuring deterministically computed DRG refinement incentives computed as $r_t - \bar{r}_{2010}$. As in Sections 5 and 6 of this Chapter, the main coefficient of interest is β_2 . In these specifications, it represents the number of percents by which the probability of a scheduled C-section increases if the DRG relative tariff ratio for scheduled C-section increases by one unit relative to the baseline period level. In other terms, if normal delivery tariffs are kept fixed, a 100% increase in C-section tariff relative to the DRG schedule of the baseline year would, on average, lead to a $\beta_2 \cdot \bar{r}_{2010}\%$ increase of the scheduled C-section probability.

Table 1.11 – Effect of gestational age on severity level of a normal delivery with multiple pregnancy, after 2012 DRG reform

Severity level	Gestational age				
	22-31	32	33-35	36	37-44
B	B	B	B	B	B
A	A	A	B	A	A

Table 1.12 – Difference-in-difference estimation of impact of DRG incentives $r_t - \overline{r}_{2010}$ computed deterministically on scheduled C-section probability in multipara women in 2010-2013

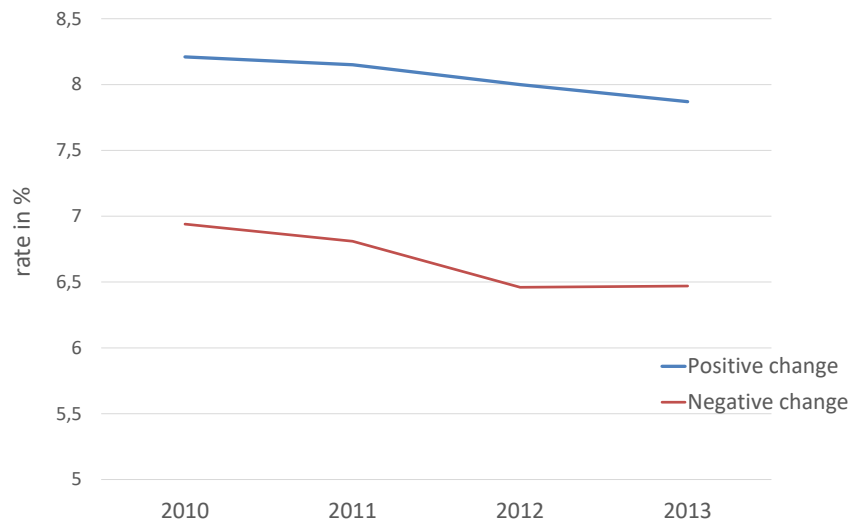
Variable	All hospitals			Non-profit		For-profit	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Diff. of priceratios (rel.to version 11b)</i>	—	-0.0011 (0.0011)	-0.0012 (0.0011)	-0.0017 (0.0011)	-0.002* (0.001)	0.0048 (0.0044)	0.004 (0.004)
<i>Age</i>	0.04*** (0.0006)	0.0023*** (< 0.0001)	0.0023*** (< 0.0001)	0.0024*** (< 0.0001)	0.002*** (< 0.0001)	0.0023*** (0.0001)	0.002*** (0.0001)
<i>Gest. age 22-35 weeks</i>	-0.72*** (0.02)	-0.018*** (0.002)	-0.017*** (0.002)	-0.0131*** (0.0024)	-0.01*** (0.002)	-0.0466*** (0.0061)	-0.04*** (0.006)
<i>Gest. age 40-44 weeks</i>	-1.35*** (0.008)	-0.0431*** (0.0003)	-0.042*** (0.0002)	-0.0442*** (0.0003)	-0.04*** (< 0.0001)	-0.0368*** (0.0005)	-0.03*** (0.0005)
<i>Number of previous C-sections</i>	0.64*** (0.003)	0.0992*** (0.0005)	0.097*** (0.0005)	0.0958*** (0.0006)	0.093*** (0.0005)	0.0975*** (0.001)	0.094*** (0.001)
<i>Gynecologist concentration index</i>	-0.01*** (0.001)	-0.0008*** (< 0.0001)	-0.0004*** (< 0.0001)	-0.0009*** (< 0.0001)	-0.0004*** (< 0.0001)	-0.0006*** (< 0.0001)	-0.0004*** (-0.0001)
<i>Medain zip code income</i>	0.0001 (< 0.0001)	-0.0001*** (< 0.0001)	-0.0001*** (< 0.0001)	-0.0001*** (< 0.0001)	-0.0001*** (< 0.0001)	0.0001** (< 0.0001)	0.0001*** (< 0.0001)
<i>Antepartum diagnosis controls</i>	Y	Y	Y	Y	Y	Y	Y
<i>Treatment group controls</i>	Y	Y	Y	Y	Y	Y	Y
<i>Year fixed effects</i>	Y	Y	Y	Y	Y	Y	Y
<i>Hospital fixed effects</i>	Y	N	Y	N	Y	N	Y
<i>Num. of observations</i>	2,612,472	2,369,472	2,369,472	1,784,737	1,784,737	584,735	584,735
<i>R-squared</i>	—	0.438	0.444	0.41	0.416	0.536	0.541
<i>Pseudo R-squared</i>	0.15	—	—	—	—	—	—

*Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. In DiD linear probability models error terms are clustered at hospital level*

The main models of interest are (2)-(7). Models (2)-(4) estimate the impact of financial incentives using all the controls specified in Section 3 without hospital fixed effects. Compared to model (1), the signs of most coefficients remain unchanged. The main coefficient of interest β_2 representing the change in financial incentives compared to the baseline period is insignificant, suggesting that the financial incentives associated with the 2012 DRG reform did not affect obstetric providers' practice. We will explore models with alternative formulations of price incentive in Section 6.

The introduction of hospital fixed effects in specifications (5)-(7) led to a very moderate increase in the share of explained variation, but did not alter the main results. Coefficients for $r_t - \overline{r}_{2010}$ remain insignificant in the model tested on all hospitals (5) and for-profit clinics (7), while β_2 becomes minimally significant at 10% for non-profit hospitals. However, given that \overline{r}_{2010} is comprised within the interval (0.83; 2.6), the magnitude of the

Figure .1.5 – Scheduled C-section trend 2010-2013, for DRG groups with positive and negative changes in financial incentives $r_t - \overline{r_{2010}}$



effect of a 100% C-section tariff increase, $\beta_2 \cdot \overline{r_{2010}}$, is negligible.

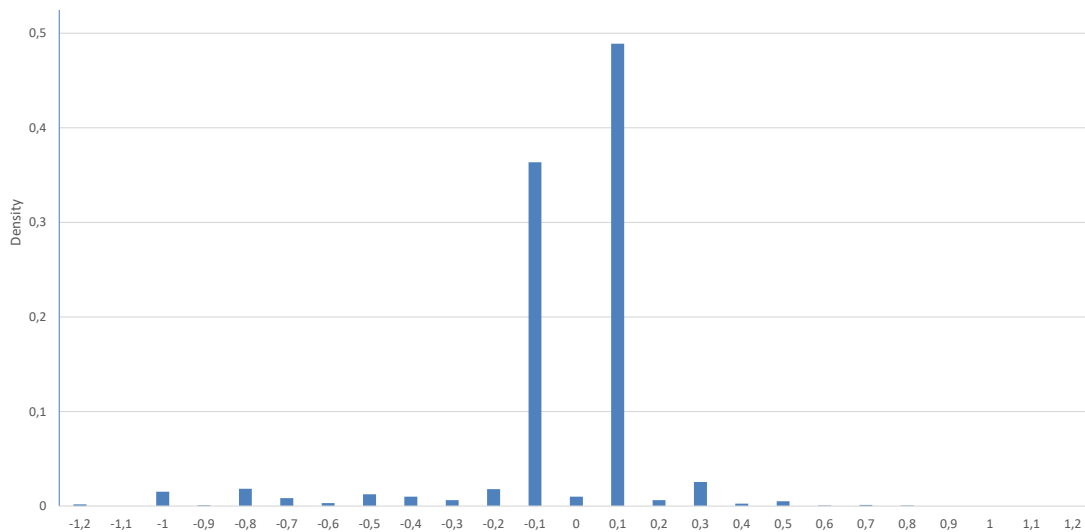
Although insignificant, in the different specifications the coefficient for the price incentive tends to be positive for models tested on private hospitals, and negative for models run on all hospitals or on non-profit hospitals.

Appendix 3.2. Bigger changes in DRG incentives (DiD models with deterministic severity)

The changes in tariffs resulting from DRG refinement concerned, with varying degree, all patient groups. However, as noted earlier, some of them were more affected by the reform than others. As shown in Figure .1.6, after the DRG refinement, only a subset of cases underwent substantial changes in relative tariffs, while most cases were not unaffected or very weakly affected by the reform. Thus, it may be of interest to investigate the impact of the DRG refinement on cases where financial incentives to switch to a different child delivery mode were the strongest.

It is worthwhile to note that the 2012 DRG refinement most strongly impacted financial incentives related to assisting childbirth in the most complicated outlier cases (notably, through an increased risk adjustment with respect to rare and more severe commodities, as well as gestational age anomalies). Thus, the subsample analyzed in this Appendix subsection is primarily comprised of patients with a high level of severity.

Figure .1.6 – Distribution of the change in financial incentive $r_{2012,2013} - \bar{r}_{2010}$, for all hospitals



Under the hypothesis that only sizeable DRG tariff incentive changes would lead to changes the medical practice, or even be noticed by care providers, we impose additional criteria on the sample. In this sample we include only those treatment groups for which the change in the relative price relative to its level in 2010 exceeded 15% , that is $(r_t - \bar{r}_{2010}) / \bar{r}_{2010} > 0.15$, $t = \{2012, 2013\}$. By construction, these treatment groups also contain pre-reform observations for which $(\bar{r}_{2012} - r_t) / r_t > 0.15$, $t = \{2010, 2011\}$, where

Table 1.13 – Difference-in-difference estimation of impact of DRG incentives $r_t - \overline{r}_{2010}$ computed deterministically on scheduled C-section probability in multipara women in 2010-2013 (subsample of patients with a greater effect of incentives $|r_t - \overline{r}_{2010}| > 0.15$)

Variable	All hospitals			Non-profit		For-profit	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Diff. of price ratios (rel.11b)</i>	-0.0084*** (0.0023)	-0.0028 (0.0021)	-0.003 (0.002)	-0.0042* (0.0023)	-0.004* (0.0022)	-0.0139 (0.0104)	0.004 (0.004)
<i>Age</i>	0.0051*** (0.0001)	0.003*** (0.0001)	0.0032*** (<0.0001)	-0.0038*** (0.0001)	0.0036*** (0.0001)	0.0012*** (0.0004)	0.002*** (0.00005)
<i>Gest. age 22-35 weeks</i>	-0.0919*** (0.0094)	-0.0543*** (0.0085)	-0.054*** (<0.0001)	-0.0601*** (0.0122)	-0.058*** (0.0121)	-0.057*** (0.002)	-0.04*** (0.006)
<i>Gest. age 40-44 weeks</i>	-0.1179*** (0.0014)	-0.0846*** (0.0014)	-0.083*** (<0.0001)	-0.085*** (0.0014)	-0.0842*** (0.0013)	-0.068*** (0.0007)	-0.03*** (0.0005)
<i>Number of previous C-sections</i>	0.2729*** (0.0018)	0.0957*** (0.0021)	0.0921*** (0.002)	0.0941*** (0.0023)	0.091*** (0.0023)	0.1064*** (0.0062)	0.094*** (0.001)
<i>Gyn concentration index</i>	-0.0015*** (0.0001)	-0.0009*** (0.0002)	-0.0005*** (0.0002)	-0.001*** (0.0002)	-0.0006*** (0.0002)	-0.001* (0.0006)	-0.0004*** (< 0.0001)
<i>Medain zip code income</i>	(< 0.0001) (< 0.0001)	(< 0.0001) (0.0001)	(< 0.0001) (< 0.0001)	(< 0.0001) (0.0001)	(< 0.0001) (< 0.0001)	-0.0015*** (0.0005)	-0.0009** (0.0006)
<i>Antepartum diagnosis controls</i>	N	Y	Y	Y	Y	Y	Y
<i>Treatment group controls</i>	Y	Y	Y	Y	Y	Y	Y
<i>Year fixed effects</i>	Y	Y	Y	Y	Y	Y	Y
<i>Hospital fixed effects</i>	Y	N	Y	N	Y	N	Y
<i>Num. of observations</i>	188,438	188,438	188,438	177,084	177,084	11,354	11,354
<i>R-squared</i>	0.199	0.357	0.368	0.359	0.368	0.355	0.374

*Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at hospital level*

\overline{r}_{2012} is a hypothetical ratio of DRG tariffs in the post reform year of 2012.²¹

These inclusion criteria lead to a reduction in the size of the sample 188,438 observations. The estimation results for this subsample are presented in Table 1.13. Models (2), (3) and (4) are run on obstetric patients treated in all, non-profit and for-profit hospitals respectively. Results do not provide evidence that financial incentives influence the clinical decision in favor of a scheduled C-section. As in Section 5, the inclusion of hospital fixed effects in models (5)-(7) does not change the results, the main coefficient of interest β_2 remaining insignificant for models (5) and (7), minimally significant in the model (6) tested on obstetric patients treated in non-profit hospitals.

It is important to note that, despite the fact that main conclusions presented in Section 5 remained identical for the analyzed subsample comprised of more severe and outlier patients, there may be additional mechanisms at play driving the absence of observed impact of DRG incentives. Concretely, results provided in *Allin et al. (2015)* suggest that relatively more complicated child delivery cases (for example, deliveries featuring breech

²¹Using either 2012 or 2013 DRG tariffs gives the same result, since in 2013 tariffs were only adjusted by an inflation rate that cancels out after calculating the expression

fetus presentation, for which C-section was clearly a more preferable option according to available clinical evidence) were not as strongly affected by financial incentives as “marginal”, borderline cases for which the decision in favor of a C-section was not as clear. This argument posits that had a physician been motivated by financial gain, s/he would have preferred to choose simpler cases where indications are not unequivocal in order to minimize the probability of being accused of malpractice.

The models presented in this Subsection of Appendix 3 are partially concordant with *Allin et al.* (2015) explanation for non-significant β_2 estimates. However, it is worthwhile to note that, apart from being non-binding, 2012 HAS clinical guidelines for obstetric procedures differ significantly from their Canadian analogs. In particular, in contrast to Canada, French health authorities do not consider breech presentation in itself as a counterindication against vaginal labor (*HAS, 2012*), and provide a list of conditions which, if met, can justify a normal child delivery. HAS issued similar recommendations for a number of diagnoses, including twin birth, HIV infection, herpes infection, scarred uterus, macrosomia, etc. This circumstance, along with a perceived lack of a strong medical litigation culture in France, arguably makes this channel less significant compared to the institutional setting featured in *Allin et al.* (2015) study.

Appendix 4. HAS recommendations for C-sections

Cette recommandation de bonne pratique a été élaborée selon la méthode résumée dans l'argumentaire scientifique et décrite dans le guide méthodologique de la HAS disponible sur son site :

Élaboration de recommandations de bonne pratique – Méthode Recommandations pour la pratique clinique

Les objectifs de cette recommandation, la population et les professionnels concernés par sa mise en œuvre sont résumés en dernière page (fiche descriptive) et décrits dans l'argumentaire scientifique. Ce dernier ainsi que la synthèse de la recommandation sont téléchargeables sur www.has-sante.fr.

Grade des recommandations

A	<p style="text-align: center;">Preuve scientifique établie</p> <p>Fondée sur des études de fort niveau de preuve (niveau de preuve 1) : essais comparatifs randomisés de forte puissance et sans biais majeur ou méta-analyse d'essais comparatifs randomisés, analyse de décision basée sur des études bien menées.</p>
B	<p style="text-align: center;">Présomption scientifique</p> <p>Fondée sur une présomption scientifique fournie par des études de niveau intermédiaire de preuve (niveau de preuve 2), comme des essais comparatifs randomisés de faible puissance, des études comparatives non randomisées bien menées, des études de cohorte.</p>
C	<p style="text-align: center;">Faible niveau de preuve</p> <p>Fondée sur des études de moindre niveau de preuve, comme des études cas-témoins (niveau de preuve 3), des études rétrospectives, des séries de cas, des études comparatives comportant des biais importants (niveau de preuve 4).</p>
AE	<p style="text-align: center;">Accord d'experts</p> <p>En l'absence d'études, les recommandations sont fondées sur un accord entre experts du groupe de travail, après consultation du groupe de lecture. L'absence de gradation ne signifie pas que les recommandations ne sont pas pertinentes et utiles. Elle doit, en revanche, inciter à engager des études complémentaires.</p>

Haute Autorité de Santé
 Service documentation – information des publics
 2 avenue du Stade de France - F 93218 Saint-Denis La Plaine CEDEX
 Tél. :+33 (0)1 55 93 70 00 - Fax :+33 (0)1 55 93 74 00

Ce document a été validé par le Collège de la Haute Autorité de Santé en janvier 2012
 © Haute Autorité de Santé – 2012

Recommandations

1 Quelles sont les indications de la césarienne programmée ?

1.1 Utérus cicatriciel

Un utérus est dit cicatriciel lorsqu'il comporte, en un endroit quelconque de l'isthme ou du corps, une ou plusieurs cicatrices myométriales. Cette définition exclut les cicatrices cervicales, les cicatrices purement muqueuses et les cicatrices purement séreuses.

L'augmentation ces dernières années du taux de césariennes dans les pays industrialisés s'accompagne d'une augmentation du nombre de femmes enceintes porteuses d'un utérus cicatriciel. Dans ces pays, la césarienne est la principale étiologie de la cicatrice utérine. Compte tenu de la littérature disponible, les recommandations concernent uniquement l'antécédent de césarienne.

Utérus cicatriciel	
Antécédents de césarienne	
C	L'utérus cicatriciel n'est pas en lui-même une indication de césarienne programmée.
AE	Les comptes rendus des interventions antérieures sur l'utérus et de l'éventuel travail (partogramme) ayant abouti à une césarienne sont utiles pour choisir le mode d'accouchement. Cependant, l'absence de ces éléments n'est pas en elle-même une indication de césarienne programmée.
AE	L'utérus cicatriciel n'impose pas la réalisation d'une radiopelvimétrie.
AE	Une échographie de la cicatrice utérine n'est pas nécessaire.
C	En cas d'antécédent d'une césarienne, au vu des risques maternels et périnataux, il est raisonnable de proposer une tentative de voie basse, sauf en cas de cicatrice corporeale.
C	En cas d'antécédent de 3 césariennes ou plus, il est recommandé de proposer une césarienne programmée.
C	Dans les autres situations, dont l'utérus bicicatriciel, le choix entre une tentative de voie basse ou une césarienne programmée se fait au cas par cas.

1.2 Grossesse gémellaire

Par rapport à une grossesse monofoetale, la grossesse gémellaire est associée à des risques plus importants de prématurité avant 37 SA (risque multiplié par 7) et de faible poids (risque de poids de naissance inférieur à 2 500 g multiplié par 9 environ). Le taux de césariennes avant le travail est trois fois plus important pour la grossesse gémellaire que pour la grossesse monofoetale.

Les pratiques obstétricales françaises se caractérisent par une prise en charge active de la naissance du 2^e jumeau (accouchement dirigé de J2) afin de réduire l'intervalle de temps entre les deux naissances.

Grossesse gémellaire

- | | |
|----------|--|
| C | Une grossesse gémellaire avec le premier jumeau (J1) en céphalique n'est pas en elle-même une indication de césarienne programmée. |
| C | Dans le cas d'une grossesse gémellaire avec le premier jumeau (J1) en siège, les données actuelles ne permettent pas de recommander une voie d'accouchement plutôt qu'une autre. |

1.3 Présentation par le siège

La présentation par le siège est définie par la présence en regard du détroit supérieur du pôle fœtal représenté par le sacrum dans le mode décompleté et par les pieds dans le mode complet. Le taux de présentations en siège lors d'un accouchement à terme d'un enfant unique se situe aux alentours de 3 à 5 % des accouchements.

Présentation par le siège**Critères d'acceptabilité de la tentative de voie basse**

- | | |
|----------|---|
| B | La présentation par le siège n'est pas en elle-même une indication de césarienne programmée. |
| C | Les critères d'acceptabilité de la voie basse, sont : <ul style="list-style-type: none"> • confrontation favorable entre la pelvimétrie et l'estimation des mensurations fœtales ; • absence de déflexion de la tête fœtale ; • coopération de la patiente. L'évaluation des critères d'acceptabilité de la voie basse définis ci-dessus, doit être réalisée, avant toute décision de césarienne programmée. |

Césarienne programmée

- | | |
|-----------|--|
| C | En cas d'indication de césarienne programmée pour le siège, il est recommandé de proposer à la patiente une version par manœuvres externes (VME) en accord avec les recommandations nationales et internationales traitant de la prise en charge d'une présentation par le siège et dans le respect des modalités de pratique édictées dans celles-ci. |
| C | En conséquence, en cas de présentation par le siège il est recommandé de réaliser une césarienne programmée, dans les situations suivantes : <ul style="list-style-type: none"> • confrontation défavorable entre la pelvimétrie et l'estimation des mensurations fœtales ; • déflexion persistante de la tête fœtale ; • non-coopération de la patiente. |
| AE | Il est nécessaire de contrôler par échographie la persistance de la présentation juste avant l'accès au bloc opératoire pour pratiquer la césarienne. |
| C | Pour toutes les autres situations, le choix entre la césarienne et la voie basse est à discuter au cas par cas. |

1.4 Macrosomie

L'évaluation du poids fœtal reste difficile à déterminer de façon précise malgré l'évolution des examens prénatals : échographie, méthode clinique (manœuvre de Léopold), méthode maternelle (basée sur l'expérience maternelle en cas de grossesse multipare).

La macrosomie fœtale est généralement définie par un poids de naissance supérieur à 4 000 g ou par un poids de naissance supérieur au 90^e percentile d'une courbe de référence de la

population donnée. La fréquence rapportée est variable selon les seuils choisis et les populations étudiées et se situe entre 5 et 10 % des naissances.

Dans le cas de ces recommandations une macrosomie correspond à une estimation du poids fœtal à partir de 4 000 g, en tenant compte des difficultés d'évaluation du poids fœtal.

La macrosomie fœtale constitue l'une des complications du diabète maternel (essentiellement diabète gestationnel).

Macrosomie	
Macrosomie hors diabète	
C	En l'absence de diabète, la macrosomie n'est pas en elle-même une indication systématique de césarienne programmée.
C	La césarienne programmée est recommandée en cas de poids fœtal estimé supérieur ou égal à 5 000 g.
C	En raison de l'incertitude de l'estimation du poids fœtal, pour une suspicion de macrosomie comprise entre 4 500 g et 5 000 g la césarienne programmée est à discuter au cas par cas.
Macrosomie liée au diabète	
C	En présence d'un diabète, la césarienne programmée est recommandée en cas d'estimation du poids fœtal supérieur ou égal à 4 500 g.
C	En raison de l'incertitude de l'estimation du poids fœtal, pour une suspicion de macrosomie comprise entre 4 250 g à 4 500 g la césarienne programmée est à discuter au cas par cas en tenant compte des autres critères liés à la pathologie et au contexte obstétrical.
Antécédents de césarienne	
C	La suspicion de macrosomie n'est pas en elle-même une indication systématique de césarienne programmée en cas d'utérus cicatriciel.
Antécédents de dystocie	
C	Les antécédents de dystocie des épaules sont à rechercher et à renseigner en détail.
C	En cas de suspicion de macrosomie et d'antécédents de dystocie des épaules compliquée d'élongation du plexus brachial, la césarienne programmée est recommandée

1.5 Transmissions mère-enfant d'infections maternelles

Dans ce contexte, la césarienne programmée vise à réduire le risque de transmission verticale des infections virales de la mère à l'enfant (transmission mère-enfant [TME]). Les autres interventions utilisées dans ce contexte, tels les traitements médicamenteux, ne font pas partie du thème de ces recommandations.

La TME du virus de l'immunodéficience humaine (VIH) a lieu le plus souvent au moment de l'accouchement. Elle fait alors suite aux contractions utérines et à la rupture des membranes, par exposition directe du fœtus au sang maternel et aux sécrétions cervicovaginales. Une charge virale plasmatique maternelle élevée est le facteur de risque le plus important de la TME du VIH. Un contrôle précoce de la charge virale maternelle au moment de l'accouchement permet d'éviter la transmission du VIH. La TME se produit plus rarement pour des charges virales indétectables (< 50 copies/ml).

Le virus de l'herpès, *Herpes simplex* virus de type 1 (HSV-1) ou de type 2 (HSV-2), peut se transmettre de la mère à l'enfant dans la majorité des cas au cours de l'accouchement à

partir des lésions génitales. L'herpès néonatal cause une morbi-mortalité importante, notamment liée à des anomalies du neurodéveloppement.

Transmission mère-enfant d'infection maternelle	
Virus de l'immunodéficience humaine (VIH1, VIH2)	
C	Il est recommandé de réaliser une césarienne programmée en cas de charge virale supérieure à 400 copies VIH/ml.
AE	Dans le cas d'une charge virale comprise entre 50 et 400 copies VIH/ml une discussion entre le gynécologue-obstétricien et l'infectiologue est nécessaire. <ul style="list-style-type: none"> • Si la cinétique de la charge virale est décroissante, une voie basse est proposée. • Si la cinétique de la charge virale est croissante, une césarienne programmée est proposée.
Virus de l'hépatite C (VHC)	
C	Il n'est pas recommandé de réaliser une césarienne programmée dans le cas d'une mono-infection par le VHC.
AE	Dans le cas d'une co-infection VIH/VHC, la césarienne programmée est recommandée en raison de l'augmentation du taux de transmission maternofoetale du VHC.
Virus de l'hépatite B (VHB)	
AE	Il n'est pas recommandé de réaliser une césarienne programmée dans le cas d'une mono-infection VHB, ou d'une co-infection VHB/VIH.
Virus de l'herpès (HSV-1, HSV-2)	
C	Si une primo-infection herpétique se déclare après 35 SA, une césarienne programmée est recommandée à 39 SA en raison du risque d'herpès néonatal.
C	Des récurrences herpétiques pendant la grossesse ne sont pas une indication de césarienne programmée.

1.6 Autres indications

Les objectifs de la RBP ont été de cibler les indications majeures de la césarienne programmée.

Il existe par ailleurs des situations variées et plus rares où la réalisation d'une césarienne programmée est à discuter au cas par cas. Il s'agit en particulier des situations suivantes :

- défaut de placentation (*placenta accreta*, *placenta prævia*, etc.) ;
- malformations fœtales et fœtopathies (anasarque, hernie diaphragmatique, thrombopénie fœtale, omphalocèle, laparoschisis, etc.) ;
- antécédents et pathologies maternelles intercurrentes (ex. : accident vasculaire cérébral, etc.) ;
- problèmes périnéaux (antécédents de périnée complet compliqué, cure de prolapsus et/ou incontinence urinaire, etc.).

Indications de fréquence plus rare	
AE	Il existe des situations à risque pour lesquelles la césarienne programmée est nécessaire (par exemple : <i>placenta prævia</i> recouvrant).
AE	Il existe des situations de fréquence plus rare pour lesquelles la césarienne programmée peut être discutée au cas par cas, en concertation en équipe avec les spécialistes concernés par une pathologie fœtale et/ou maternelle.

1.7 Césarienne sur demande

Selon la définition la plus souvent retenue dans la littérature, la césarienne sur demande concerne la césarienne sur demande maternelle en l'absence d'indications médicales ou obstétricales.

La césarienne sur demande maternelle sans indication médicale et/ou obstétricale n'est pas une situation clinique parfaitement identifiable à partir d'activité de codage ou dans les études observationnelles. Son incidence est difficilement estimable.

La césarienne sur demande maternelle pourrait être la résultante de facteurs multiples comprenant des facteurs internes (expérience personnelle, perception de la naissance, etc.), des facteurs externes (informations obtenues auprès de la famille, des amis, des médias, des professionnels de santé) mais aussi plus spécifiques (impression d'une sécurité pour l'enfant et pour la mère, notion de contrôle et de préservation).

Selon la littérature, les principales causes associées à une demande maternelle de césarienne seraient la peur de la naissance, en particulier chez les primipares, ainsi que les mauvaises expériences lors d'une précédente grossesse.

Les débats et les controverses actuels autour de la césarienne sur demande sont aussi liés à de nouvelles problématiques : la demande accrue des femmes de participation active au choix du mode d'accouchement (autonomie de décision de la patiente) et l'éventuelle augmentation du risque de procédure juridique envers le médecin quant au choix du mode d'accouchement en cas de complications.

Césarienne sur demande

Information de la patiente sur la césarienne sur demande

	La demande maternelle n'est pas en soi une indication à la césarienne. Il est recommandé de rechercher les raisons spécifiques à cette demande, de les discuter et de les rapporter dans le dossier médical.
	Lorsqu'une femme demande une césarienne par peur de l'accouchement par voie basse, il est recommandé de lui proposer un accompagnement personnalisé. Une information sur la prise en charge de la douleur peut constituer une réponse efficace à cette peur.
AE	L'information et la discussion doivent intervenir le plus tôt possible, en particulier lors de l'entretien prénatal précoce, en accord avec les recommandations HAS sur la préparation à la naissance et à la parentalité.
	Les bénéfices attendus par la patiente doivent être mis en regard des risques associés à la césarienne et expliqués à la patiente (en particulier, le risque accru de <i>placenta prævia</i> et <i>accreta</i> associé à la césarienne pour les grossesses futures).
	Après recueil des raisons de la demande initiale, suivi d'une information sur les différents modes d'accouchement, l'acceptation d'une césarienne sur demande nécessite une analyse concertée avec la patiente de ses motivations.
Attitude du professionnel	
AE	Un médecin peut décliner la réalisation d'une césarienne sur demande. Il doit alors orienter la patiente vers un de ses confrères.

2 Quelles sont les informations à transmettre à la femme enceinte lorsqu'une césarienne programmée est envisagée ?

La décision du mode d'accouchement (césarienne ou accouchement par les voies naturelles, encore appelé accouchement par voie basse) est réévaluée tout au long de la grossesse en fonction de différents éléments médicaux et obstétricaux.

La décision de pratiquer une césarienne peut être prise avant l'accouchement si, à l'examen, des difficultés dans le déroulement de l'accouchement sont prévisibles et susceptibles d'entraîner des conséquences pour l'enfant ou sa mère.

En fonction du déroulement de la grossesse, le mode d'accouchement effectif, c'est-à-dire celui réellement intervenu, sera conforme ou non à celui initialement programmé.

Information de la patiente	
Modes d'accouchement	
AE	Pour toute césarienne programmée, une information de la femme enceinte est nécessaire (indication, rapport bénéfices/risques, conditions de réalisation). Les éléments d'information communiqués doivent être mentionnés dans son dossier médical.
AE	Il est recommandé d'informer les femmes enceintes de l'existence d'événements pouvant intervenir avant ou pendant le travail et modifier le mode d'accouchement initialement programmé.
Conséquences de la césarienne programmée	
C	En raison de la morbidité néonatale (notamment détresse respiratoire), il est recommandé de ne pas faire une césarienne programmée avant 39 SA pour les grossesses monofoetales.
C	Il est recommandé d'informer la patiente de la balance bénéfices/risques de la césarienne programmée par rapport à la tentative de voie basse dans sa situation particulière. En particulier : <ul style="list-style-type: none"> la césarienne programmée augmente les risques de complications pour les grossesses futures ;
AE	<ul style="list-style-type: none"> la césarienne programmée augmente les accidents thromboemboliques veineux.

Un document d'information destiné aux femmes enceintes est proposé en annexe de ces recommandations.

Perspectives

Selon l'enquête nationale périnatale de 2010, le taux de césariennes s'est stabilisé depuis le début des années 2000. En France, une femme sur cinq donne naissance par césarienne. Dans moins de la moitié des cas, la césarienne est programmée.

Une hétérogénéité des pratiques a été cependant relevée par différentes enquêtes, en particulier entre les différents établissements et les différents départements.

Plusieurs types de facteurs, liés entre eux, peuvent influencer sur le nombre d'actes de césarienne : les caractéristiques des femmes, celles des maternités, et les pratiques médicales. Sans qu'aucun d'entre eux ne puisse être la seule variable explicative des taux de césariennes programmées et des variations rapportées dans les différentes enquêtes.

Le taux global de césariennes programmées ne semble pas le critère le plus pertinent à la mise en place d'un programme d'amélioration des pratiques. En effet, il ne permet pas de

prendre en compte la complexité et la diversité des situations médicales, l'existence de populations maternelles différentes, la variabilité de l'offre de soins selon les territoires, et autres éléments organisationnels. En outre, il ne rend pas compte de l'existence possible de césariennes programmées pour des raisons autres que des indications médicales, que cela soit les césariennes sur demande maternelle mais aussi les césariennes programmées pour des raisons organisationnelles, de pratiques professionnelles, d'offre de soins hétérogène sur le territoire, etc.

Il serait intéressant de disposer de données de pratiques et/ou de réaliser une enquête prospective pour se doter d'outils d'analyse sur le chemin clinique des patientes enceintes selon chacune des situations cliniques et de s'interroger sur les critères d'évaluation les plus pertinents.

Chapter 2

Impact of Quality-based Procedures on orthopedic care quantity and quality in Ontario Hospitals

In this chapter we evaluated the effects of the 2012 healthcare reform in Ontario, whereby for a selected set of medical interventions hospitals started to be reimbursed based on the price by volume formula, with the expectation that payments would be subsequently adjusted with respect to hospital performance on quality indicators. In complement to QBPs, in 2012 the Health-Based Allocation Model (HBAM) was phased in allowing part of hospital funding to be tied to municipality-level patient and hospital characteristics. Using patient-level data from the Canadian Discharge Abstract Database (DAD), we evaluated through a difference-in-difference approach the impact of QBPs/HBAM on the volume and quality of targeted procedures and other types of joint replacements plausibly competing for hospital resources. After controlling for patient, hospital and regional characteristics, we found a significant decrease in acute length of stay associated with QBPs, as well as a marked shift towards patients being discharged home with/without post-operative supporting services. However, evidence with regards to spillover effects and quality improvement across all joint replacement types was weak. Results are robust to various model specifications, and different estimation techniques, including matching methods and synthetic control groups.

1 Introduction and literature review

In an attempt to improve the quality of care at reasonable cost and minimal unintended consequences, many countries have adopted pay-for-performance (P4P) and prospective payment (PPS) mechanisms in their healthcare systems. Designs under which they operate and their ultimate success differ widely, with regulators targeting individual providers and/or hospitals, setting different quality indicators and ensuring different degrees of monitoring and control. Outcomes and implications, both expected and unintended, of reforms that comprise both of these elements - have so far been poorly understood in the economics literature. In addition, even less is known about the impact of reforms undergoing a policy drift, where one of the initially designed components ends up being gradually retracted.

The objective of this Chapter is to comprehensively evaluate the impact of a funding reform introduced in Ontario aiming to replace global budget funding with a new and desirably more efficient system. Two of its main components - the Health-Based Allocation Model (HBAM) and Quality-based procedures (QBPs) - were phased in from 2012 with the goal to incorporate PPS and P4P incentives into Ontario inpatient care. The reform also envisaged preserving around 30% of hospital funding through global budgets based on hospital budgets from previous years.

The first component of the reform, HBAM, at its core is a mechanism designed to distribute a fixed provincial envelope between hospitals based on expected spending of each hospital. This component determines a hospital's share of the envelope by forecasting future hospital budgets, relying on a volume by unit cost approach. The volume part of the formula is adjusted, in particular, based on severity of admitted patients and other relevant characteristics of recorded hospital stays, such as number of interventions, patient transfer, and mode of discharge. Unit costs are modulated with respect to hospital-specific characteristics justifying a different cost structure of a hospital (for example, size, teaching status, geographical isolation etc.) (*Ontario Hospital Association, 2019*). In addition, the calculation method takes into account historical service volumes, expected population growth and regional patterns of health care access, etc. (*Born & Dhalla, 2012*).

At the inception of the program, the second component of the reform - QBPs - were supposed to encourage adoption of better clinical practices by affecting financial stimuli at the hospital level. Hospital costs were planned to be reimbursed on the basis of prices negotiated by expert panels and fixed for all care facilities, with final payments being

adjusted with respect to a list of quality indicators. However, due to a presumed lack of coordination and communication between the designers of the reform and its various participants and because of inconsistency of policy objectives over time, the substance of QBPs changed from financially rewarding providers for quality to a risk-adjusted volume by price funding for every eligible procedure performed, supplemented with an array of clinical guidelines to which hospital practitioners were expected to adhere. Thus, in practice, no monetary bonus or penalty was put in place after the introduction of the reform. In addition, as argued by *Palmer et al. (2018a, 2018b)*, reform conditions and mechanisms were understood neither in a timely manner nor accurately by key stake holders affected by the reform.

Despite having a broadly common volume by unit price approach to funding, HBAM and QBP have a number of important differences. First, HBAM payments are calculated as a hospital share of a pre-defined spending envelope, whereas in QBPs unit prices are fixed, paid for each episode of care and do not depend on other hospitals' activity. Second, contrary to HBAM budgets that are decided based on future estimated health care needs within individual hospitals and the community it serves, QBPs are paid fully prospectively taking into consideration only procedures that have already been carried out. Lastly, in relation with the second point, characteristics on which procedure volumes and unit prices are adjusted differ for these two funding mechanisms, HBAM having a stronger focus on hospital- and community-level factors than QBP.

In 2012, during the announcement of the reform, among the anticipated effects were cited "shorter wait times and better access to care in their communities", "more services, where they are needed" and "better quality care with less variation between hospitals" (*Government of Ontario, 2012*). Despite the proclaimed goals, our study shows that although clinical practice patterns on average changed in accordance with the guidelines for most affected procedures, there is little evidence to support the claim that quality of care significantly improved as a result of the reform.

The remainder of this Chapter is organized as follows. Section 2 provides a literature review of the existing research on PPS and P4P, Section 3 presents in more detail the clinical context of knee and hip replacements, as well as the institutional environment, conditions and evolution of HBAM and QBP in Ontario. Research questions and a summary of results are provided in Section 4. Section 5 provides descriptive statistics for data used in this Chapter. The main results of this Chapter are presented and interpreted in Section 6. Section 7 presents robustness checks, Section 8 discusses the significance and

limitations of the results and concludes.

2 Pay-for-performance (P4P) and Prospective Payment Systems (PPS): review of existing research

2.1 Prospective Payment Systems (PPS)

A considerable body of research has been produced on the effects of Prospective Payment Systems (PPS) on a multitude of healthcare indicators. In the economics literature on PPS, it has been argued that such systems can generate a number of positive outcomes. In particular, PPS systems can encourage care providers to favor cost-effective treatments by limiting, for example, hospital length of stay (thus, mitigating the problem of long wait times and addressing the lack of care accessibility, which has long been a major concern for many countries, including Canada) and prescribing medications and interventions with a proven clinical benefit to patients. This same consideration may also push providers acting under PPS to specialize in procedures in which they are the most cost-effective and referring other patients to a more suitable care provider. As a result, patients may receive a well-coordinated high-quality treatment. In addition, since PPS payments are known to care providers in advance, as long as a given treatment is reimbursed at a level even marginally higher than expected patient costs, hospitals would have a financial interest to admit the maximum number of such patients. The fact that most PPS payments are adjusted based on severity of reported diagnoses motivates personnel to implement more rigorous and systematic practices with regards to coding diagnoses, which may contribute to a higher degree of transparency and information accuracy (*Busse et al., 2011*).

However, all these organizational benefits may be compromised by a number of perverse incentives inherent to PPS. In this respect, *Ellis (1998)* studied a particular market environment, wherein patients and care providers act under complete information with respect to patient severity and hospital practice patterns. In addition, providers form a duopolistic market for the procedure demanded by fully-insured patients, while a third-party payer (insurer) is myopic to patient individual severity and sets in advance a payment schedule as a combination of a global budget and reimbursement on a per-patient basis. *Ellis (1998)* showed that in this setting the maximization of hospital profits can be achieved through decreasing costs per hospital stay by “skimping” on care quality, that is leaving certain patients with a sub-optimal level of care. A relatively broad interpretation of the

term “quality” makes it possible to consider any preventable event susceptible to compromise treatment quality as “skimping” (for example, an unjustifiably early patient discharge from hospital). Finally, arguably the most radical strategy may involve denying hospital care altogether to patients with a perceived higher-than-average cost burden, a situation referred to as “dumping”. *Ellis (1998)* also suggests that “skimping” and “dumping” can be expected to be practised together.

Empirical studies overall find moderate to no effects of PPS on care quality. For example, a descriptive study by *Schwartz and Tatter (1998)* on patients who underwent colorectal cancer surgery in Mount Sinai Hospital in New York from 1983 to 1987 pointed to, among other things, a significantly lower rate of blood loss (down to 387cc from 550cc), insignificant changes in margins of surgical resection (down to 13cm from 18cm) and a 5 percentage point decrease in the rate of post-operative complications. A more recent study by *Shin (2018)* analyzing DRG reimbursement changes within Medicare comes to the conclusion that increasing payments under PPS did not translate into improvements in healthcare quality measured by 30-day in-hospital mortality and 30-day readmission rates.

Under PPS hospital profits can be boosted not only by reducing cost but also by increasing revenue. Thus, hospitals professionals may “up-code” their patients, that is purposefully inflate the severity of diagnoses and/or adding more diagnoses than otherwise would be medically justified. Empirical evidence for recourse to such a strategy by hospitals is strong. For example, *Dafny (2005)* and the earlier mentioned *Shin (2018)* show that hospitals operating under Medicare and experiencing hikes in DRG reimbursement rates (caused by elimination of age criterion in the first study, and a reformulation of geographic areas in the latter) may act under the influence of presumed perverse PPS incentives and eventually alter DRG coding practices and/or, as also suggested by *Shin (2018)*, shift patients into a higher paying tariff group. In Europe, evidence for DRG upcoding was recently provided by *Januleviciute et al. (2016)*, who find, in particular, that surgical DRGs tend to be more liable to upcoding than medical ones.

In some cases, hospitals may also be incentivised to over-provide well-reimbursed services, even despite them having little to no expected clinical benefit for the patient, leading to what is referred to in the economics literature as supplier-induced demand. Evidence from Japan on the use of neonatal intensive care units (NICU) provided by *Shigeoka & Fushimi (2014)* suggests that introduction of PPS reimbursements for a vast majority of procedures except for those related to NICU (which remained to be reimbursed on a doctor fee-for-

service basis), and the ensuing discrepancy between reimbursement mechanisms within an episode of care resulted in a protracted NICU utilization likely achieved by manipulating infants' reported birth weights. Due to the fact that both Ontario and Japan currently combine hospital PPS and doctor fee-for-service payment schedules, policy implications of this Chapter can be of particular significance in the Canadian context. Anecdotal and statistical evidence also suggests that, for example, extensive use of X-ray scanners and other types of medical imaging (*Baker, 2010*), as well as a more frequent recourse to C-sections as compared to normal deliveries (*Johnson & Rehavi, 2016*) may at times be accounted for by profit-making considerations, at least in the US context.

Perhaps due to the aforementioned potential negative ramifications associated with PPS systems, it is very uncommon to see them implemented in their pure form. In most countries, they are amalgamated with a form of global budgets (e.g. Germany from 2003, France from 2004), P4P (e.g. QOF in UK from 2004, HQID in USA from 2004) or fee-for-service arrangements (e.g. Ontario from 2012, Japan from 2003).

In practice, PPS systems are usually implemented through a set of tariffs unique to each Diagnosis-related group (DRG) oftentimes adjusted by patient severity. Although this approach may arguably render hospital funding more structured and manageable, *Geissler et al. (2012)* have expressed concerns that in Europe DRGs may not predict patient costs as precisely as a basic set of patient characteristics, such as indicators for age groups, Charlson index, patient transfer during hospital stay and occurrence of an adverse event. However, in the context of the Ontario reform, it is difficult to compare effectiveness of HBAM funding formulas since they were not made publicly available (for more information on HBAM see Subsection 3.2)

2.2 Pay-for-performance (P4P)

In theory, healthcare providers' behaviour has been traditionally considered through the lens of contract theory with the goal of describing environments where principals, agents and final service consumers can have areas of both overlapping and conflicting interest. *Holmstrom and Milgrom (1991)* initiated the debate on the multitasking problem and its implications. Their model suggests that in the presence of noise in a set of performance indicators determining agents' pay-offs, in order to prevent agents from skimping on poorly measured activities and to align agents' behaviours with social interests, a risk-neutral regulator can exert its influence through a broad range of leverages, such as influencing asset ownership structure, imposing restrictions on performing tasks, influencing incentives

to perform outside non-contractual activities, as well as imposing specialization among providers. A partial adaptation of this general framework to the healthcare context was proposed by *Eggleston (2005)*, who models a benevolent purchaser reimbursing a single provider with a mixed payment scheme calculated as a fraction of incurred cost and a performance premium. The model predicts that under a pay-for-performance scheme reallocation of effort towards enhancing performance with respect to a relatively more incentivized quality measure can be attenuated by introducing mixed payment systems, such as partial capitation, insofar as it reduces providers' financial burden (i.e. supply-side cost-sharing) of ensuring unrewarded dimensions of care.

However, this rather intuitive conclusion may no longer hold true if other features specific to healthcare markets are allowed for in a model. Notably, as suggested, for example, by *Mullen (2010)*, if dimensions of care are related one to another, such that efforts in incentivized dimensions of care can compensate for a decrease in unrewarded ones, then an overall negative impact of multitasking on unrewarded quality dimensions can be considerably dampened or even reversed.

Including providers' altruism motivation can have considerable theoretical implications on providers' optimal behavior. Under perfect observability of all parameters and a set-up featuring providers with a heterogeneous level of altruism, *Siciliani (2009)* shows that an increase in prices for services may discourage production outputs by providers with a medium level of altruism, while increasing service volumes by doctors with low and high degrees of altruism. However, since these changes occur in opposite directions, the impact of price incentives on the overall production volume is ambiguous. Nevertheless, under a different set of assumptions, predictions can drastically change. For example, according to *Markis and Siciliani (2013)*, if a limited financial liability is ensured to partially altruistic providers who can select patients and whose cost functions are not observed by a purchaser, providers with low and high degrees of altruism will systematically skew their production levels compared to the first-best solution, the direction of this output distortion depending on agents' unobservable efficiency type.

Building on *Eggleston's (2005)* model, *Kaarboe and Siciliani (2011)* provide an insight into the optimal size of performance incentives. According to their model, in environments featuring an unobservable quality dimension and a sole partially altruistic provider the strength of incentives depends on whether quality dimensions are substitutes or complements in both provider's disutility and the patient's benefit functions. In the case of substitutes, incentives should either be low-powered or be set to the marginal benefit of

the contractible quality dimensions, while complementary quality measures would always necessitate high-powered incentives. However, in this model the influence of the provider's degree of altruism on the optimal incentive size is ambiguous and depends on the form of the patient's benefit function.

Finally, not only can providers react to stimuli by adjusting their output volume and its quality, but they can also react strategically in dimensions that do not directly relate to care. In line with a growing empirical evidence on presence of "upcoding" in pay-for-performance schemes, *Kuhn and Siciliani (2011)* propose a model wherein multiple providers having private knowledge about their ability must attain a specific level of a quality indicator set by the purchaser, and to this end can engage in costly efforts to bias this performance measure. Depending on the degree of concavity of purchaser's preferences with respect to patient benefits from treatment, it may be optimal to set higher quality benchmarks for high performers (i.e. separating equilibrium) or to pool the highest performing providers together and remunerate them based on a universal quality target and the associated payment, while keeping a P4P scheme for low-performers. In the context of our study, an important implication of this model consists in the necessity for the purchaser to set higher benchmark performance levels if a quality measure is easily manipulable.

Empirical studies on pay-for-performance tend to report inconclusive results that largely depend on institutional contexts of the reforms in question. Systematic literature reviews conducted by *Cristianson (2008)*, *Emmert (2011)*, *Van Herck (2010)* conclude that, on average, the effects of P4P schemes on most relevant care indicators are likely zero to modest in magnitude. In general, the literature tends to support the conclusion that hospital-level reforms are less likely to produce sizable changes than those implemented at the care provider level; that higher-powered incentives are more likely to bring about stronger changes in indicators; that procedural indicators are more likely to move following P4P reforms than outcome indicators; and, finally, that most P4P reforms, even despite their seemingly modest results, turn out to be cost-effective in terms of QALY gains per unit of cost. These reviews also report a high level of heterogeneity in methodological rigor in the analyzed studies, whereof the approach varies from purely descriptive ones to randomized controlled experiments.

Our study builds on the existing literature on hospital-level P4P incentive reforms. One of the earliest studies of this type was conducted by *Norton (1992)*. This randomized experiment looked into the effects of a simultaneous introduction of payments (lump-sum

and per diem) made to US nursing homes to prevent severity-based patient selection, encourage a timely discharge of healthy patients and reward facilities that demonstrate improvements in patients' health status. Although the author reported improvements in all of the three targeted care dimensions, the validity of the Markov chain model that he tested critically relied on a set of strong assumptions, such as patient homogeneity, independence from past transitions between states and constant time spells between observations taken for a given individual.

Later research on hospital-level P4P produced mixed results. *Ryan (2008)* investigates the impact of Premier Inc. and Centers for Medicare and Medicaid Services Hospital Quality Incentive Demonstration (PHQID), a voluntary hospital-level P4P and public reporting reform implemented in 2003 to incentivize quality performance in a set of selected surgeries and conditions with a bonus of up to 2 percent of a hospital budget. Using fixed effects panel model to predict in particular patients' risk-adjusted mortality and 60-day risk-adjusted day cost, *Ryan (2008)* concludes that there was no significant impact on either of these two parameters, although an earlier study on PHQID by *Lindenauer et al. (2007)* reported a positive effect of the PHQID reform vis-à-vis process quality measures.

One of the most closely related studies to our research is *Mullen et al. (2010)*, who investigated the impact of a successive introduction of Quality Incentive Payment (QIP) and Integrated Healthcare Association program (IHA) in California in 2002 and 2003 respectively. In the early stages, both reforms rewarded hospitals based on virtually perfectly overlapping sets of quality indicators, although the actual incentive power assigned to each of them varied significantly. In addition, the mechanism of these two reforms was quite different. QIP calculated payment amounts conditional on achieving performance thresholds set relative to providers' performance distribution in the previous year. On the other hand, in most participating care plans IHA payments were determined as a function of relative performance and were more attainable for lower-performing groups. Using a difference-in-difference approach with hospitals in the Northwestern US states serving as a control group, the authors conclude that even despite a significant size of monetary incentives (15% and 60% of total capitated revenues of studied medical groups), the reform had very modest effects overall in terms of both rewarded and unrewarded dimensions of care. Moreover, quality improvements in cervical cancer screening – the only measure that clearly responded to P4P – did not discernibly spill over to closely related measures, whether they were incentivized or not. Finally, the authors also did not find a stable pattern in the relationship between potential incentive size and hospital performance on

quality measures, which altogether questions the presence of a multitasking channel (i.e. discretionary effort to improve specific dimensions of care) in this setting. However, potential drawbacks of this study come from, first, the fact that it was impossible to disentangle the effects of QIP and the anticipation effect of IHA and, second, a speculative nature of hypothesized interactions between the effects of providers' multitasking effort (if such was present) and an unobservable degree of commonality between quality dimensions, which, according to their model, should produce spill-over effects.

Somewhat more optimistic conclusions were made by *Nahra et al.(2006)*, who studied the impact of hospital-level P4P in Michigan from 2000 to 2003 which were aimed at increasing the rate of prescription of aspirin, beta-blockers and ACE inhibitors used to treat patients with heart disease. Their findings suggest that the rates of prescription steadily increased during the time span of the study for all of three of the drugs. In addition, this procedural improvement generated cost-effective gains in QALY for targeted patients. However, a serious limitation of this study comes from the absence of a control group, its main conclusions being based solely on post-reform time trends. Finally, in a study on tobacco users in the US state of Minnesota, *An(2009)* finds that payments incentivizing smoking quitline referrals were effective in achieving this goal since clinics subject to the P4P incentive demonstrated higher levels of referrals compared to similar clinics that did not participate in the program (11.4% vs. 4.2%).

In the context of joint replacements, to the best of our knowledge, so far *Papanicolas and McGuire (2015)* has been the only study that looked into a closely-related topic of how hospital-level tariff incentives can affect the rates of procedure uptake. To identify the impact of financial incentives on the choice between cemented and uncemented hip prostheses, the authors used the introduction of the Payment-for-Results system that replaced global budget funding in England. However, the subsequent differentiation in prices between these two alternative treatments was not adopted in Scotland, which served as a control group. Their results indicate that the rate of the more expensive uncemented hip replacement increased significantly due to payment incentives in England, despite the fact that it went against existing clinical guidance.

3 Canadian healthcare context

3.1 Hip and knee replacements: clinical evidence and quality measures

Hip and knee replacements are currently considered to be one of the most well-researched and commonplace clinical procedures. Their primary goals are to reduce joint pain and to improve mobility status. The vast majority of hip and knee replacements are caused by degenerative disease or physical trauma, and are performed on elderly populations (see Table 2.2).

During the operation, which normally lasts around 3 hours, a surgeon is supposed to completely remove the affected joint(s) and fix different parts of the prosthesis onto the remaining bone tissues. Major post-operative complications are most commonly caused by ensuing joint fractures and dislocations, and infections of the operated site. (see Table 2.1) In addition, implanted prostheses may wear out earlier than expected. In all of these cases, a revision surgery may be carried out attempting to ensure proper functioning of the prosthesis.

Table 2.1 – Frequencies and rates of diagnoses in joint replacements in Ontario, Alberta and British Columbia, 2008/08-2017/18

Post admit diagnoses (complications)	Rate	Count
Anemia, unspecified	0.018	13145
Acute posthemorrhagic anemia	0.016	11685
Retention of urine	0.0082	5988
Hypotension, unspecified	0.0067	4893
Other delirium	0.0063	4601
Urinary tract infection, site not specified	0.0051	3725
Vascular complications following a procedure	0.0051	3725
Haemorrhage and haematoma complicating a procedure	0.005	3652
Acute pain	0.0041	2994
Other complications of procedures	0.0038	2775
Sample size		730301

Rehabilitation from surgery can occur in a hospital setting, in a specialized facility and/or at home. Recent evidence has been pointing to the fact that the out-of-hospital setting can be no less effective and more cost-effective than in-hospital recovery, which is also reflected in recommendations issued for orthopedic surgeons in Ontario (see Appendix B; section Post-acute care for more details). Most patients undergoing hip and knee replacements

are able to stand/walk several days after the operation and in 6-12 weeks they can return to work and most daily activities. (*NHS, 2019a, 2019b*)

There exist non-surgical alternatives to joint replacements which include prescription of painkilling and anti-inflammatory medications, and physiotherapy. Joint surgery is considered to be an intervention of the last resort and is normally carried out when non-invasive types of treatment fail. If a decision is eventually made to operate on a joint, a surgeon can choose between a complete joint replacement and a joint repair. The latter may involve replacement of only part of the affected cartilage/bone tissue (e.g. hip resurfacing). However, although oftentimes less costly and requiring shorter recovery times, joint repairs can cause additional complications over time (e.g. femoral neck fracture in hip) and are generally not recommended for specific groups of patients (e.g. age over 60 years old, patients with weak bones).

Before a hip/knee replacement, surgeons normally decide on the type of prostheses and materials used to affix the prosthesis to the bone. Prostheses can be cemented or be cementless. In the former case, a prosthesis is fixed using a bone cement that dries during the operation. In the latter case a prosthesis has a porous coating with a rough surface, which allows the patient's bone to grow naturally onto the device. In general, cemented prostheses are more likely chosen for patients with weaker bones, lower bone regenerative potential and an increased risk of infection. At the surgeon's discretion, an antibiotic can be added into cement material in addition to a prophylactic administration of an antibiotic both pre- and post-operatively. However, cemented joint replacements can have more long-term complications due to a risk of pieces of cement breaking off the surgery site - called cement debris, which may lead to a loosening of the prosthesis, irritation of surrounding tissues and the debris entering the bloodstream. The Ontario Health Technology Advisory Committee (OHTAC) recommends that the prosthesis fixation technique be decided unilaterally by the surgeon, although in practice available options may be discussed with the patient (see Appendix, section Surgery for more details).

Prosthetic parts themselves are usually made either from a high-density plastic material or from a metal alloy, normally titanium. As of 2017, it is most common to use a metal alloy for shaft and head of the prosthesis and a plastic material for the socket, which in some cases may prolong prosthesis life time and reduce the risk of release of metal particles into the body.

There exists a number of widely accepted indicators used to evaluate quality of joint replacements. They can be roughly divided into procedural and outcome subgroups. The

former focus on the adequacy and appropriateness of provided treatment and procedures, while the latter evaluate the ultimate success of the attempted intervention in terms of patient performance status, rate of complications, satisfaction etc.

In 2014 the Ontario Ministry of Health and Long Term Care proposed seven quality indicators, summarized in the Orthopedic Quality Scorecard, aiming to evaluate the success of introducing orthopedic QBPs. These metrics were meant to evaluate hospital performance on three dimensions – efficiency, effectiveness/safety and accessibility – by measuring length of stay, proportion of patients discharged home, rates of 30-day readmission and 1-year revision, as well as wait times before surgery (see Appendix C). At different points in time, governmental authorities and medical associations in countries outside Canada put forward their metrics evaluating the hip/knee replacement quality, as part of ongoing funding reforms (for example, France from 2019) or in an effort to inform better clinical practices (for example, a measure set by the American Academy of Orthopaedic Surgeons). To obtain a more balanced and well-rounded picture of the evolution of orthopedic care quality in Ontario, we rely, where possible, on both domestic and international quality measure sets provided in Appendix C.

3.2 Institutional context of HBAM and QBPs

The Canadian healthcare system is characterized by a quasi-complete domination of publicly funded hospitals and providers, whose activities are regulated in a decentralized fashion by provincial authorities, heavily incentivized to follow federally established basic principles. Healthcare public expenditures are jointly financed by federal, provincial and territorial authorities. The federally adopted *Canada Health Act (1984)* requires, in particular, that to be eligible for federal contributions – a crucial source of public funding – healthcare services must be universal, reasonably accessible and be regulated by a public non-profit authority(ies) designated by each province. Provinces and territories complement these federal transfers with funds raised primarily via provincial taxes.

On the level of individual providers, the entirety of federal and provincial legislation either explicitly prohibits (e.g. Ontario) or makes it financially unsustainable for the vast majority of them to opt out of public insurance plans in favor of solely private practice. Thus, physicians and specialists earn their income almost exclusively by serving as independent contractors and directly billing a provincial healthcare insurance plan in which they are registered. Historically, these payments were made on a fee-for-service basis according to fee schedules set on the provincial level. However, recent developments in

several provinces allow for more complex income formulas featuring per capita payments, fixed salary arrangements or mixed plans (*Sarma et al., 2018*).

In general, in comparison to the USA for example, the Canadian healthcare system puts a relatively strong emphasis on the issues of equitable access and affordability, and less so on provider competition, care timeliness and clinical innovation. This focus on equality and the presence of a universal payer in many respects brings it closer to healthcare systems typically observed in Europe. A similar pattern is present in terms of total healthcare spending per capita and its structure. According to the OECD, Canadian levels (\$4974 US in 2018) are closer to those in France and Germany (\$4965 US and \$5986 US in 2018, respectively), and are considerably lower than in the USA (\$10586 US in 2018). In Canada, only \$749 of this spending was funded out-of-pocket, compared to \$463 and \$738 in France and Germany, respectively, and \$1122 in the US (*OECD, 2019*)

Within a given province, designated provincial authorities decide on the allocation formula of pooled funding between care facilities. The most common scheme of redistributing these funds was and still remains global budgets, although several provinces have undertaken attempts to reform it. Notably, in 2010 British Columbia was the first Canadian province to introduce activity-based funding for most of its hospitals.

Hip and knee replacements fall into the category of medically necessary procedures covered by all Canadian provincial healthcare plans. From a patient's perspective, during a hospital stay, at no point is there liability for the patient to pay out-of-pocket, unless the admitted person voluntarily chooses to purchase non-medical services, such as stay in a private/semi-private room or TV access.

However, while free at the point of service, hospital services have been reported, both anecdotally and in published research, to suffer from long wait times, high rates of complications and providers' choice of inappropriate and/or unjustifiably costly treatments. In particular, according to Canadian Institute for Health Information (CIHI), in 2013 only one out of ten Canadian provinces managed to attain the benchmark of 182 days of wait time from the booking date, to the date the patient received a planned total hip/knee replacement.

In 2012 a hospital funding reform started to be gradually implemented in Ontario, whereby, instead of completely relying on global budgets, a new payment mechanism was introduced stipulating that an increasingly large share of hospital funds be tied to patient characteristics, their clinical outcomes and other quality measures. This patient-based funding (PBF) was comprised of two elements: Quality-Based Procedures (QBP)

and Health Based Allocation Model (HBAM) payments.

In the beginning, QBP payments were designed to be a P4P mechanism aimed at encouraging adoption of better clinical practices by introducing financial stimuli at the hospital level. QBPs are supposed to be reimbursed on the basis of prices negotiated by expert panels and fixed for all care facilities, with hospital payment adjusted by their performance on a set of relevant clinical quality indicators. These quality measures were expected to include, but not limited to, variables reflecting patient health outcomes, timeliness and accessibility of care, compliance with care pathways, and others. Nevertheless, to the best of our knowledge, the list of orthopedic quality indicators was never publicly released. Moreover, *Palmer et al. (2018b)* argue that many hospital providers and even some reform architects were not in possession of that knowledge.

The QBP reform appears to have created general expectations which eventually did not match reality, from a majority of stakeholders' perspectives. As the reform was deployed, perceptions regarding QBP gradually started to shift along with reform objectives. After the introduction, the QBP design gradually drifted towards funding an episode of care, where selected medical conditions were reimbursed based on a pre-set price per episode of care calculated at the 40th percentile of average costs observed in Ontario hospitals (Palmer, 2018a), multiplied by service volume. Subsequently, these payments were adjusted with respect to a DRG-based Case Mix Index (CMI), which modified the volume component of the formula based on observed patient clinical characteristics, with the goal of reflecting the intensity of hospital resource use during the performed procedure ¹ (*Ontario Hospital Association, 2019*). For each of the selected procedures, a QBP manual was issued summarizing the state of existing medical research and putting forward recommendations concerning good clinical practice standards to which doctors were encouraged to adhere. Thus, despite the original plan, payments under QBPs were actually never tied to patient outcomes or adherence to best practice (i.e. no financial penalties for noncompliance with the best practices were ever put in place).

QBPs had nearly universal coverage for eligible procedures in Ontario. Despite this, a group of small-sized hospitals accounting for less than 1% of hip/knee replacement surgeries were not included in the reform since their cost structures were considered too

¹For example, a hospital performing 100 unilateral hip replacements reimbursed at 5214\$ would earn $100 \cdot 5214 = 521400$ \$ annually. After correcting for the fact that patients of this hospital had a higher than average severity, the volume would be adjusted through CMI, for example, by a factor of 1.5. Thus, the final annual payment provided for the QBP component of hospital funding would amount to $150 \cdot 5214 = 782100$ \$.

different from those of larger providers.

At the beginning of the implementation stage only four selected types of interventions were reimbursed through QBPs: in 2012 they were introduced for primary unilateral hip replacement, primary unilateral knee replacement, unilateral cataract and chronic kidney disease. Since the introduction of QBPs, their number progressively increased - from 4 (accounting for 6% of funding) in 2012 to 22 procedures in 2018 (see Appendix D). Since the introduction of QBPs, their number and share has progressively increased - from 4 (accounting for 6% of funding) in 2012 to 22 procedures in 2018. However, the share of funding ensured by QBPs, as shown by Palmer et al. (2018a), did not keep up with initial expectations. Instead the benchmark of 30% of hospitals funded through QBPs by 2014/15, their share stagnated at 12-15% after 2013/14.

The second component – HBAM payments – are annual monetary transfers made to Ontario healthcare institutions to account for differences in case-mix of patients across hospitals. To this end, each hospital contributes an agreed percentage of its total budget², the resulting common envelope being shared between hospitals based on their share of total provincial expected costs. As mentioned in the introduction, this funding mechanism primarily aims to reflect differences in costs incurred by hospitals due to their specific status and disease burden with the population they serve. Thus, when predicting hospital budgets, volume by price tends to be modulated with respect to observed and projected hospital and municipality level factors. Starting from 2016/17 HBAM contributions to the common funding pot started to be determined based on data lagged by two years (Ontario Hospital Association, 2019). Since 2012 the HBAM share of hospital funding stabilized at a level of around 34%, slightly below the initial target of 40%. (*see Palmer et al., 2018a and Appendix D*)

It is important to note that physician payments are included in neither QBPs/HBAM nor, more generally, in any other hospital funding system in Canada. In Canada physicians bill separately for each service provided based on a specific payment schedule developed at the provincial level, which was the case in both the pre- and post-reform as of November 2019. (*Wettstein et al., 2019*).

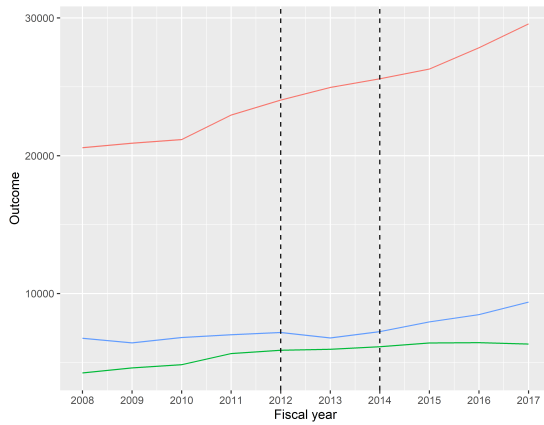
Hip and knee replacements are one of the fastest-growing procedures not only in Canada, but most developed countries. Such a rise can be attributed, among other causes, to an aging population and an increased awareness that joint replacements can dramatically

²However, several years into the reform, only revenues stemming from Ontario Ministry of Health were counted as the base of calculating contributions to the common envelope

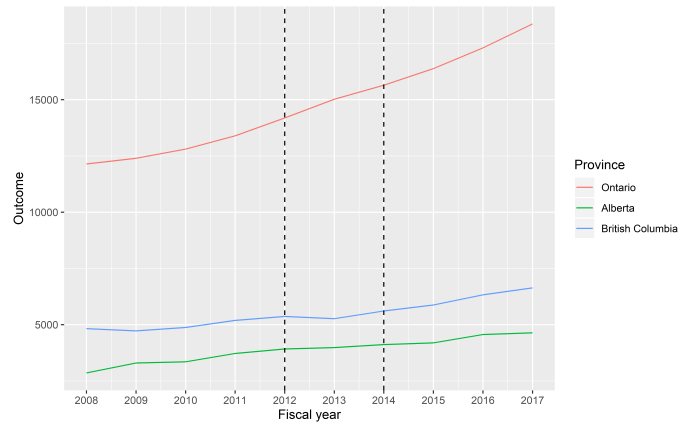
increase quality of life for patients relative to non-invasive medical treatments. According to the Canadian Joint Replacement Registry (CJRR) Annual Report by Canadian Institute for Health information (CIHI), the demand for hip and knee replacements in Canada has been steadily increasing over the last two decades. As of 2017/18, 58,492 hip replacements and 70,502 knee replacements were performed, which is around 17% higher than in 2012/13. The total cost of surgeries has kept up with their rate and reached 1.2 billion Canadian dollars (equivalent to 910 million US dollars in November 2019) in 2017/18 alone (*CJRR Annual Report, 2019*). A costing analysis conducted by *Sutherland et al. (2012)* for the province of Ontario reveals that in 2007-2009 the total expected costs for an episode of care was estimated at \$15,863 CAD for hip replacements and \$14,192 CAD for knee replacements, of which the largest part was generated in-hospital (\$12,535 CAD and \$11,609 CAD respectively for hip and knee replacements).

In the last two decades a number of initiatives have been undertaken at both the federal and provincial levels to impact different aspects of joint replacement provision. In 2004, during an annual meeting of provincial and territorial premiers with the Canadian Prime Minister, joint replacements were identified as a priority area for reducing wait times. In 2005, the 10-Year Plan to Strengthen Health Care identified the federal wait time benchmark of 26 weeks, within which patients should receive hip/knee replacement surgery (*Health Canada, 2005*). Although after this federal initiative wait times improved in Canada, as of 2018 the proportion of hip replacements carried out meeting the 26-week wait time reached 75%, with provinces demonstrating vastly different results (e.g. 84% of timely hip replacements in Ontario vs. 49% in Manitoba, Nova Scotia and Prince Edward Island) (*CIHI, 2019*).

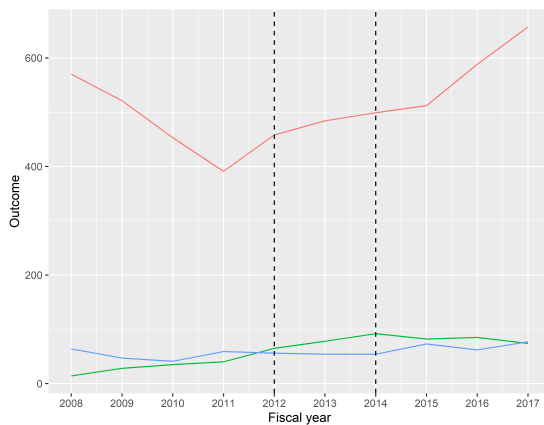
In Ontario this was in part achieved through an effort to increase volumes of hip/knee replacements, whose growth has consistently outpaced that of the rest of Canada (see Figure .2.1). However, there is evidence that in the last decade in Ontario there has been a gradual change in policy priorities, which increasingly started to focus on joint replacement appropriateness rather than their brute volume. (*QBP, 2013*). In addition, in light of new clinical evidence, since the late 2010s there has been a transition in Ontario from in-hospital towards less resource-intensive home and community-based rehabilitation. In particular, to this end the Ontario Health Technology Advisory Committee formulated a recommendation in the QBP manual for hip/knee replacements (see Appendix C).



(a) Unilateral knee replacements



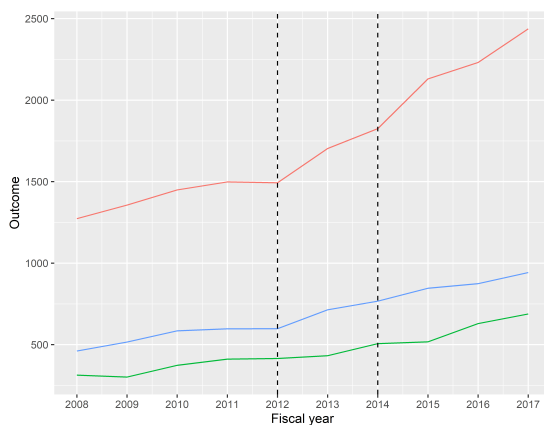
(b) Unilateral hip replacements



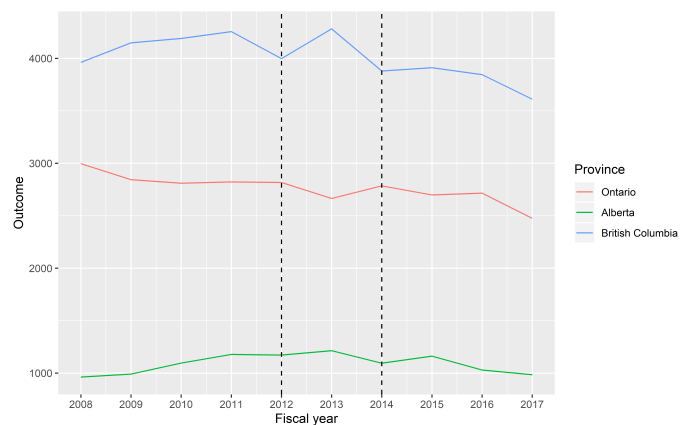
(c) Bilateral knee replacements



(d) Bilateral hip replacements



(e) Other joint replacements



(f) Surgery unrelated to arthroplasty

Figure .2.1 – Total number of joint surgeries by type in Ontario, Alberta and British Columbia between 2008-2017

EFFECTS OF QBPs AND HBAM ON ORTHOPEDIC CARE IN ONTARIO

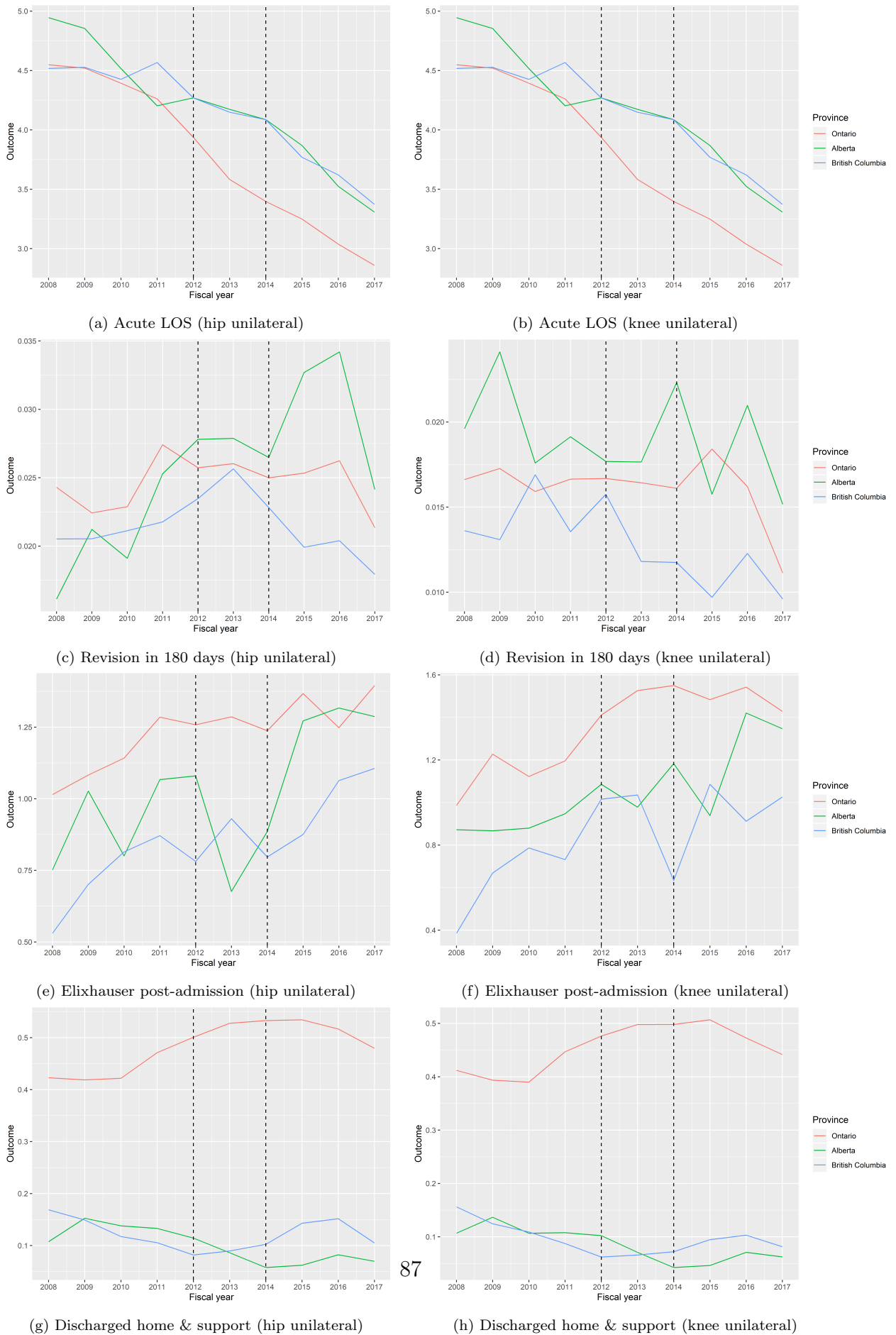


Figure .2.2 – Outcome trends unilateral knee/hip replacements in Ontario, Alberta and British Columbia between 2008-2017

4 Research questions

The purpose of this Chapter is to evaluate how incentives resulting from QBP and HBAM implemented in Ontario in 2012 affected the main orthopedic care outcomes. First, we attempt to answer the question as to whether incentives created by introducing QBP and HBAM impacted the quality and quantity of knee and hip replacement surgeries, making, where possible, separate conclusions for both components of the reform.

In 2012 QBP and HBAM payments were simultaneously introduced for unilateral hip and knee replacements, making it difficult to separate the effects of the two without comparing pre- and post-reform outcomes to similar populations of patients who did not experience the same policy changes. In difference-in-difference models, under the assumption that both the affected (or ‘treated’) and control populations follow the same pre-reform trend in the outcome variable of interest and are not separately affected by time-variant factors other than the reform, the obtained results can be interpreted as a causal impact of the reform on a variable of interest (e.g. quality of care).

Since QBP for bilateral hip and knee replacements was introduced in 2014, (i.e. two years after the HBAM reform was launched) these procedures were directly affected only by HBAM (and not by QBP) in 2012-2013. The fact that QBPs were phased in gradually for different procedures thus allows us to separately estimate the impact of HBAM and QBP for bilateral hip and knee replacements as long as no spillover effects are assumed from closely related unilateral orthopedic procedures.

In our models, as control populations we use comparable patients residing in the Canadian provinces of Alberta (AB) and British Columbia (BC). This choice is dictated by two main reasons. First, their funding mechanics are similar to Ontario and remained unchanged during the study period. In 2007-2018 Alberta had neither QBP nor HBAM payment mechanisms in orthopedic care, while British Columbia had only HBAM analogous to Ontario. Second, BC and AB are the third and fourth biggest provinces in Canada, respectively after French-speaking Quebec, which runs a largely different healthcare system and collects health-related data according to standards different from the rest of Canada. Secondly, we study the question whether stimuli that arose in unilateral and bilateral hip and knee replacement surgeries affected quality, process outcomes and appropriateness of other types of closely-related joint replacement surgeries (e.g. ankle and shoulder replacements). Theoretical justification for this question lies in the fact that resources used to meet the goals set for QBP procedures, such as increased care quality, are prone to

being diverted from other un-incentivized types of joint replacements and procedures. As a result, the appropriateness and quality of surgeries performed on other joints could have suffered. In order to test if this effect took place, it is necessary to look into subsequent changes in rates, and quality indicators for other orthopedic surgeries, comparing these changes to similar control populations outside Ontario in a fashion similar to those used for replacements included in QBPs.

5 Data

The Discharge Abstract Database (DAD) constitutes the main source of data in our study. It is a data set which is comprised of patient-level characteristics of every hospital stay in all Canadian provinces, except for Quebec. Data are submitted to the Canadian Institute of Health Information (CIHI) by acute care facilities or by their regional authority.

To obtain access to selected abstracts of this data set, a data request was submitted to Canadian Institute of Health Information (CIHI) in 2018, with a prior authorization by University of Toronto ethics commission. The received data set contained all records of adult Canadians residing in Alberta, British Columbia or Ontario, who either underwent an orthopedic procedure or had a diagnosis potentially indicating a need for it. More specifically, inclusion/exclusion criteria were specified as:

- patient must have had an admission date between 1st April 2008 and 31st March 2018.
- patient must be a resident of Alberta, British Columbia or Ontario at the time of admission
- patient's age must be at least 18 years old
- patient's discharge record must contain at least any of the following:
 - a diagnosis code indicative of a potential need for a joint replacement
 - a record of a medical procedure linked to joint replacement
 - a Case Mix Group (CMG) code identifying a joint replacement
- trauma-related, cancer-related and obstetric care cases are excluded³

³Diagnosis codes, procedure codes, CMG codes for included cases and diagnosis codes for excluded cases are provided in Appendix A

All records also include a unique but meaningless patient identifier, which allows us to track the trajectory of care and demand for services of every person admitted to a care facility. For every stay, records contain observed diagnoses and comorbidities for which it was specified when they occurred - before or after hospital admission. The latter feature allows us to compare health outcomes of patients before and after hospital admission. Every performed procedure is coded with information provided on its start and end time, time length, targeted location on patient's body, as well as de-identified number of the medical professional responsible for the procedure and the de-identified medical facility in which it took place.

In total, the merged data set with all DAD records contained a total of 700 variables and 656,626 observations. The most important variables in the context of our analysis are summarized by province in Table 2.2.

Overall, the descriptive statistics reveal very similar distributions of most variables related to patient case-mix, including patient age, sex and their diagnosis profiles. However, one can observe noticeable differences in several hospital-level care parameters. In particular, preferred modes of patient discharge and propensities to favor bilateral replacement over unilateral ones vary significantly depending on the analyzed province. Compared to Alberta and British Columbia, Ontario care providers were significantly more likely to offer post-operative supporting care to their patients, be it in a home setting or in a specialized facility, and on average were more inclined to opt for bilateral joint replacements. Due to this fact, estimation results, in particular for different modes of patient discharge, should be considered with additional caution and regarded as tentative evidence of the QBP/HBAM effect that may not have a causal interpretation.

Table 2.2 – Descriptive statistics for whole sample and by province

Variables	By province:									
	All sample				Ontario		Alberta		British Columbia	
	Mean	SD	Min	Max	Mean	SD	Mean	SD	Mean	SD
Patient age	66.5	11.6	18	107	67.3	11.0	65.6	11.6	65.0	13.0
Male	0.42	0.49	0	1	0.42	0.49	0.43	0.49	0.44	0.49
QBP Flag	0.58	0.48	0	1	0.94	0.23	0	0	0	0
Post 2012	0.64	0.48	0	1	0.64	0.48	0.66	0.48	0.62	0.48
Post 2014	0.45	0.5	0	1	0.44	0.5	0.45	0.5	0.43	0.5
<i>Length of stay (days):</i>										
total	4.8	9.4	0	31	4.4	8.14	5.2	12.07	5.4	10.2
acute	4.4	6.8	0	28	4.1	5.28	4.9	9.05	5.0	8.4
Num. of diagnoses (post-admission)	0.2	0.7	0	19	0.2	0.7	0.2	0.7	0.2	0.6
Num. of interventions (post-admission)	2.3	0.8	0	20	2.3	0.8	2.3	0.8	2.2	0.6
<i>Responsibility for payment:</i>										
province	0.989	0.03	0	1	0.991	0.09	0.992	0.09	0.987	0.13
work insurance	0.01	0.1	0	1	0.01	0.08	0.01	0.01	0.01	0.12
<i>Stay leads to:</i>										
Readmission all-cause (in 180 days)	0.011	0.1	0	1	0.009	0.095	0.015	0.124	0.01	0.1
Replacement revision: (in 180 days)	0.017	0.13	0	1	0.018	0.13	0.019	0.14	0.013	0.11
<i>By replacement :</i>										
unilateral hip (QBP 2012-2017)	0.32	0.466	0	1	0.334	0.471	0.346	0.476	0.309	0.462
unilateral knee (QBP 2012-2017)	0.512	0.5	0	1	0.551	0.497	0.505	0.5	0.419	0.493
bilateral hip (QBP 2014-2017)	0.002	0.041	0	1	0.001	0.036	0.004	0.062	0.001	0.035
bilateral knee (QBP 2014-2017)	0.008	0.092	0	1	0.012	0.107	0.005	0.073	0.003	0.058
other replacements	0.039	0.195	0	1	0.039	0.194	0.041	0.195	0.039	0.194
<i>Discharged to:</i>										
home	0.88	0.32	0	1	0.85	0.35	0.88	0.32	0.94	0.21
home, with support care services	0.303	0.46	0	1	0.44	0.497	0.09	0.288	0.08	0.279
support care facility	0.08	0.28	0	1	0.13	0.33	0.017	0.13	0.024	0.15
another acute care facility	0.028	0.167	0	1	0.015	0.12	0.095	0.293	0.023	0.14
<i>Diagnoses (observed pre- and post-admission):</i>										
Gonarthrosis, unspecified(M179)	0.26	0.44	0	1	0.27	0.44	0.25	0.43	0.22	0.42
Benign hypertension (I100)	0.22	0.41	0	1	0.24	0.42	0.31	0.46	0.13	0.34
Coxarthrosis, unspecified(M169)	0.26	0.44	0	1	0.16	0.37	0.18	0.38	0.14	0.35
Type 2 diabetes (E119)	0.091	0.29	0	1	0.09	0.3	0.087	0.28	0.076	0.26
Primary coxarthrosis, bilateral (M160)	0.066	0.25	0	1	0.062	0.24	0.078	0.27	0.067	0.25
Presence of artificial knee(Z9661)	0.051	0.22	0	1	0.062	0.24	0.064	0.25	0.017	0.13
Anaemia, unspecified (D649)	0.03	0.18	0	1	0.037	0.19	0.038	0.19	0.024	0.15
<i>Comorbidity indexes (pre- and post-admission)</i>										
Charlson (pre-admit)	0.05	0.31	0	11	0.06	0.32	0.06	0.31	0.05	0.3
Elixhauser (pre-admit)	0.09	1.35	-17	40	0.01	1.38	0.01	1.37	0.04	1.27
Charlson (post-admit)	0.06	0.06	0	11	0.07	0.35	0.06	0.33	0.06	0.32
Elixhauser (post-admit)	0.22	1.89	-17	49	0.26	1.99	0.22	1.86	0.12	1.62
Number of observations:	730,301				442,263		111,592		176,446	

6 Models and main results

6.1 Patient-level models

The difference-in-difference (DiD) estimation approach is a standard and widely applied technique used to evaluate the impact of public policies. In addition to classic OLS requirements, it demands that treatment and control groups follow the same trend in the pre-reform period with respect to analyzed outcome variables, and that no unobserved time-variant differences exist between observations in treatment and control groups. If these conditions are respected, one might be able to argue that the model estimate for the effect of the reform is unbiased and retains causal properties.

To evaluate the impact of QBP and HBAM, we ran DiD models of the functional form:

$$y_{ipht} = \alpha + \beta'_1 postQBP_{ipht} + \beta'_2 X_{ipht} + \beta'_3 Z_{ipht} + \psi_h + \gamma_t + \omega_i + \epsilon_{ipht}, \quad (2.1)$$

where i denotes a treated patient, p - province, h - a health care facility, t - a year between 2007/08 and 2017/18; y_{ipht} is the outcome variable of interest; QBP_{iph} is a flag for hospitals participating in QBP/HBAM; β'_1 is a vector of coefficients for interactions $postQBP_{ipht}$ between post-reform periods and QBP/HBAM participation dummy, and contains main coefficients of interest which represent average post-reform treatment effects. X_{ipht} is a column vector of patient-specific characteristics, including age, sex, pre-admit co-morbidity Charlson/Elixhauser indexes, Z_{ipht} are zip-code-specific characteristics, such as median household income and type of settlement (urban/rural) of the patient; ψ_h are hospital-specific fixed effects; γ_t are year-specific fixed effects; ω_i are pre-admit diagnoses fixed effects; ϵ_{ipht} is a random error term. Standard errors on all models are clustered at the hospital level. In all tested models, error terms are clustered at the hospital level.

The results of difference-in-differences models for clinical outcomes of unilateral hip and knee replacement surgeries are summarized in Table 2.3. As control variables, models include pre-admit, year and hospital fixed effects, as well as a full set of patient and hospital-level controls. Obtained estimates indicate that the reform led to a significant decrease in acute length of hospital stay (LOS) in the first two years after its introduction (by 0.27 and 0.33 days on average for unilateral and knee replacements, respectively). However, after 2014 no statistically significant difference was found compared to the period 2012-2013, suggesting that acute LOS stabilized after 2014. The coefficients post-admission Elixhauser comorbidity indexes show a marginally significant decrease of 0.05 points and a strongly significant decrease of 0.15 index points for hip and knee unilateral replacements,

Table 2.3 – Summary of difference-in-difference estimates for unilateral replacements (quality measures)

Outcome	Hip unilateral						Knee unilateral					
	Acute LOS	Revision(180d)	Readm.(180d)	N. diag.(post admit)	Elkhauser index	N.of interv.	Acute LOS	Revision(180d)	Readm.(180d)	N. diag.(post admit)	Elkhauser index	N.of interv.
Dependent var.	1	2	3	4	5	6	7	8	9	10	11	12
<i>(Intercept)</i>	2.545 (0.578)***	-0.01 (0.016)	0.001 (0.013)	0.255 (0.069)***	0.136 (0.24)	1.732 (0.067)***	7.318 (0.373)***	0.048 (0.012)***	0.042 (0.01)***	1.927 (0.168)***	1.405 (0.26)***	2.691 (0.071)***
<i>post2012 · QBP</i>	-0.273 (0.101)***	-0.001 (0.002)	0 (0.001)	-0.013 (0.01)	-0.059 (0.034)*	0.001 (0.009)	-0.326 (0.043)***	0.002 (0.001)	0.002 (0.001)*	0.272 (0.018)***	-0.152 (0.025)***	0.025 (0.007)***
<i>post2014 · QBP</i>	-0.132 (0.114)	-0.003 (0.002)	-0.002 (0.002)	0.029 (0.011)**	0.015 (0.037)	-0.021 (0.01)**	-0.042 (0.043)	0 (0.001)	0 (0.001)	-0.077 (0.018)***	0.022 (0.025)	0.111 (0.007)***
<i>age</i>	-0.045 (0.013)***	0.002 (0)***	0 (0)***	-0.016 (0.001)***	-0.036 (0.005)***	0.009 (0.001)***	-0.138 (0.011)***	0 (0)	-0.001 (0)***	-0.014 (0.004)***	-0.072 (0.006)***	-0.015 (0.002)***
<i>age²</i>	0.001 (0)***	0 (0)***	0 (0)***	0 (0)***	0 (0)***	0 (0)***	0.001 (0)***	0 (0)	0 (0)***	0 (0)***	0.001 (0)***	0 (0)***
<i>Male</i>	-0.231 (0.035)***	-0.001 (0.001)	0 (0)	-0.01 (0.004)***	0.041 (0.013)***	0 (0.003)	-0.225 (0.014)***	0.004 (0)***	0.004 (0)***	0.007 (0.007)	0.039 (0.009)***	0.016 (0.002)***
Resp. for payment:												
<i>Worker insurance</i>	-0.386 (0.288)	-0.001 (0.009)	0.005 (0.006)	-0.031 (0.025)	-0.082 (0.098)	-0.073 (0.029)**	0.104 (0.05)**	0.002 (0.003)	0.001 (0.002)	-0.037 (0.033)	0.044 (0.043)	0.021 (0.014)
<i>Other</i>	0.586 (0.791)	-0.018 (0.003)***	-0.009 (0.001)***	0.178 (0.148)	-0.122 (0.271)	-0.059 (0.08)	-0.087 (0.172)	-0.012 (0.006)*	-0.007 (0.003)**	-0.184 (0.077)**	0.145 (0.094)	-0.019 (0.038)
Income quintile												
<i>Lowest</i>	0.515 (0.229)**	-0.01 (0.014)	-0.012 (0.012)	0.013 (0.048)	0.194 (0.099)**	0.041 (0.034)	0.311 (0.103)***	0.011 (0.003)***	0.001 (0.005)	0.328 (0.101)***	-0.269 (0.157)*	0.045 (0.035)
<i>Medium-low</i>	0.305 (0.226)	-0.011 (0.014)	-0.013 (0.012)	-0.002 (0.048)	0.235 (0.098)*	0.03 (0.034)	0.166 (0.103)	0.011 (0.003)***	0.001 (0.005)	0.266 (0.101)***	-0.24 (0.157)	0.04 (0.035)
<i>Medium</i>	0.311 (0.228)	-0.011 (0.014)	-0.013 (0.012)	-0.002 (0.048)	0.235 (0.098)**	0.029 (0.034)	0.136 (0.103)	0.011 (0.003)***	0.001 (0.005)	0.224 (0.101)**	-0.218 (0.157)	0.045 (0.035)
<i>Medium-high</i>	0.249 (0.225)	-0.012 (0.014)	-0.013 (0.012)	-0.007 (0.048)	0.288 (0.098)***	0.023 (0.034)	0.107 (0.102)	0.011 (0.003)***	0.001 (0.005)	0.195 (0.101)*	-0.2 (0.157)	0.053 (0.035)
<i>Other controls</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Year FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Diagnosis FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Hospital FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>N</i>	167322	167322	167322	167322	167322	167322	265620	265620	265620	265620	265620	265620
<i>R2</i>	0.36	0.1	0.011	0.21	0.22	0.41	0.25	0.15	0.01	0.31	0.13	0.26
<i>AdjR2</i>	0.36	0.1	0.009	0.21	0.22	0.41	0.25	0.15	0.01	0.31	0.13	0.26

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at hospital level

respectively. However, as for 180-day revision and 180-day readmission rates, there was no meaningful or statistically strong change after 2012 nor after 2014. This may suggest only a modest, if any, short-term improvement in care quality of targeted procedures following the introduction of QBPs/HBAM.

Within this class of models, these results are robust with respect to significant variations in included control variables. Additional examples of these models are provided in Appendix E.

As expected, major risk factors, such as older age, male sex or precarious financial situation, are associated with a higher probability of adverse clinical outcomes. The fact that a given procedure was funded by worker insurance does not significantly affect the characteristics of provided care. However, other funding sources, which predominately include foreigners required to pay out-of-pocket, are associated with on average worse

Table 2.4 – Summary of difference-in-difference estimates for unilateral replacements, with slope terms after 2012 and 2014

Hip unilateral:							
Outcomes:	QBPpostlinear12	SD	QBPpostlinear14	SD	N	R2	AdjR2
Acute LOS	-0.246	(0.013)***	0.272	(0.023)***	222047	0.298	0.297
Discharged home	0.05	(0.002)***	-0.057	(0.003)***	222047	0.218	0.218
Discharged support service	-0.053	(0.001)***	0.057	(0.002)***	222047	0.203	0.203
N. all diagnoses	0.057	(0.01)***	-0.088	(0.017)***	222047	0.381	0.381
N.of post-admit diag	-0.005	(0.004)	0.007	(0.006)	222047	0.096	0.095
N.of interventions	-0.008	(0.003)***	0.011	(0.005)**	222047	0.325	0.324
Readmission (180d)	-0.001	(0.001)**	0.002	(0.001)**	222047	0.007	0.006
Revision (180d)	-0.001	(0.001)	0.001	(0.001)	222047	0.067	0.066
Total LOS	-0.26	(0.015)***	0.28	(0.025)***	222047	0.291	0.291
Tranferred	0.004	(0.001)***	-0.002	(0.002)	222047	0.119	0.118
Elixhauser post-admission	-0.012	(0.007)*	0.001	(0.012)	222047	0.02	0.019
Elixhauser pre-admission	0.009	(0.005)*	-0.008	(0.009)	222047	0.162	0.161
Knee unilateral:							
Outcomes:	QBPpostlinear12		QBPpostlinear14		N	R2	AdjR2
Acute LOS	-0.239	(0.009)***	0.27	(0.016)***	361361	0.255	0.255
Discharged home	0.048	(0.001)***	-0.055	(0.002)***	361361	0.177	0.176
Discharged support service	-0.052	(0.001)***	0.056	(0.002)***	361361	0.183	0.182
N. all diagnoses	0.063	(0.008)***	-0.101	(0.013)***	361361	0.353	0.353
N.of post-admit diag	-0.007	(0.003)***	0.009	(0.004)**	361361	0.079	0.079
N.of interventions	0.023	(0.002)***	0.022	(0.004)***	361361	0.302	0.301
Readmission (180d)	0.001	(0)	-0.001	(0.001)	361361	0.008	0.008
Revision (180d)	0.001	(0.001)	-0.001	(0.001)	361361	0.114	0.114
Total LOS	-0.254	(0.01)***	0.282	(0.017)***	361361	0.252	0.252
Tranferred	0.005	(0.001)***	-0.003	(0.001)**	361361	0.089	0.089
Elixhauser post-admission	0.007	(0.006)	-0.022	(0.01)**	361361	0.019	0.018
Elixhauser pre-admission	-0.003	(0.004)	0.004	(0.006)	361361	0.231	0.23

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at hospital level

care outcomes, likely due to unobserved patient characteristics such as official language ability/familiarity with local healthcare systems and ability to effectively navigate them. Additional model specifications featuring slope terms for QBP/HBAM participating facilities after 2012 and 2014 of the form:

$$y_{ipht} = \alpha + \beta_1 \cdot QBPpost12linear_{ipht} + \beta_2 \cdot QBPpost14linear_{ipht} + \beta_3' X_{ipht} + \beta_4' Z_{ipht} + \psi_h + \gamma_t + \omega_i + \epsilon_{ipht}, \quad (2.2)$$

are presented in Table 2.4, where $QBPpost12linear_{ipht}$ and $QBPpost14linear_{ipht}$ are interaction terms between the QBP/HBAM participation flag, years dummies $post12$ and $post14$, and a linear trend.

The results confirm that both acute and total length of hospital stay decreased in the first two years (by around 0.25 days in both 2012 and 2013) after the introduction of QBPs/HBAM for hip and knee unilateral replacements. However, this trend was mitigated since 2014 (i.e. $\beta_2 \cong \beta_3$), which suggests that the impact of the QBP/HBAM on acute/total LOS took two years to fully set in, and/or that it was reversed by including bilateral replacements as QBPs in 2014. As for other measures of quality, estimates for hospital readmissions, joint replacement revisions do not show a consistent change after 2012. In the meantime, the fact that the total number of diagnoses significantly increased by on average 0.05 units may be suggestive of the presence of up-coding incentives that are typical of most activity-based funding payment schemes.

Table 2.5 – Difference-in-difference estimates by post-reform year for bilateral knee and hip replacements (quality measures)

Outcome	Bilateral Hip				Bilateral knee			
	Acute LOS	N.diag. (post)	Elixhauser index	N.of interv.	Acute LOS	N.diag. (post)	Elixhauser index	N.of interv.
Dep.var./Model	1	2	3	4	5	6	7	8
<i>Intercept</i>	2.604 (4.128)	-0.149 (0.48)	0.663 (1.286)	1.144 (0.454)***	6.589 (1.803)***	3.673 (0.934)***	1.499 (1.151)	1.34 (0.394)***
<i>post2012 · QBP</i>	1 (1.103)	0.393 (0.209)**	0.631 (0.546)	0.086 (0.119)	0.274 (0.351)	-0.196 (0.166)	0.012 (0.219)	-0.023 (0.057)
<i>post2014 · QBP</i>	-2.049 (1.067)**	-0.376 (0.202)**	-0.667 (0.528)	-0.122 (0.105)	-0.702 (0.322)***	0.106 (0.151)	-0.174 (0.218)	0.098 (0.056)**
<i>age</i>	-0.004 (0.142)	-0.002 (0.018)	-0.078 (0.047)**	0.015 (0.013)	-0.076 (0.057)	-0.087 (0.03)***	-0.067 (0.037)*	0.027 (0.012)***
<i>age²</i>	0.001 (0.001)	0 (< 0.001)	0.001 (0)**	< 0.001 (< 0.001)	0.001 (< 0.001)***	0.001 (< 0.001)***	0.001 (< 0.001)***	< 0.001 (< 0.001)***
<i>male</i>	-0.164 (0.338)	-0.074 (0.053)	-0.226 (0.131)**	0.068 (0.045)	-0.198 (0.088)***	-0.076 (0.048)	0.149 (0.066)***	0.062 (0.019)***
<i>QBP flag</i>	-0.409 (0.579)	0.004 (0.09)	0.116 (0.243)	0.15 (0.083)**	-1.016 (0.23)***	0.219 (0.108)***	-0.148 (0.12)	0.038 (0.037)
<i>Year FE</i>	YES	YES	YES	YES	YES	YES	YES	YES
<i>Diagnosis FE</i>	YES	YES	YES	YES	YES	YES	YES	YES
<i>Hospital FE</i>	NO	NO	NO	NO	NO	NO	NO	NO
<i>N</i>	1225	1227	1227	1227	6310	6313	6313	6313
<i>R²</i>	0.17	0.07	0.06	0.19	0.17	0.24	0.02	0.17
<i>AdjR²</i>	0.15	0.06	0.05	0.17	0.17	0.23	0.01	0.17

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at hospital level

Results for bilateral hip and knee replacement are presented in Table 2.5. Compared to models testing unilateral replacement, the former do not include hospital fixed effects due to a much smaller sample size of bilateral procedures and ensuing insufficient variation in

most diagnosis controls. In addition, because of a very low frequency count for 180-day readmissions and revisions in bilateral replacements, the associated models could not be estimated.

The results indicate that acute LOS increased, albeit insignificantly, in 2012-2013 (i.e. immediately after both QBP and HBAM were introduced for unilateral replacement, and only HBAM for bilateral replacements). However, from 2014 onward this indicator decreased (2.04 and 0.7 days for bilateral hip and knee replacements, respectively) when bilateral replacements were included as QBPs. This fact suggests that the observed fall in length of hospital stay in both unilateral and bilateral replacement is due to the QBP component of the reform, and not HBAM.

The impact of QBP/HBAM on other care parameters, such as post-admission diagnoses and post-admission Elixhauser index, albeit significant on certain occasions in bilateral hip replacement, overall had a sporadic and irregular impact across other procedures.

6.2 Hospital-level models

To check additional care outcomes and to verify whether the results hold on a more global level of analysis, a set of hospital-level models reflecting parameters of appropriateness, care process and quality were tested.

Similarly, tested models have the functional form:

$$y_{ipht} = \alpha + \beta_1' postQBP_{ipht} + \beta_2 \cdot QBP_{iph} + \beta_3' X_{ipht} + \beta_4' Z_{ipht} + \psi_h + \gamma_t + \epsilon_{ipht}, \quad (2.3)$$

where variables and effects are denoted identically to equation (2.2), except for the fact that, in contrast to patient-level models, all included variables represent hospital-level means. In hospital-level models each observation stands for a hospital in a given year, and is weighed based on the observed volume of a given procedure output in a given hospital-year. This allows us to make sure that bigger hospitals are assigned a proportionally greater weight while deriving model estimates.

Overall, the results in Tables 2.6 and 2.7 confirm that QBP led to a moderate reduction in acute length of stay in unilateral replacement (around 0.3 drop for both unilateral hip and unilateral knee replacement after 2012, accompanied by an additional reduction by 0.2-0.25 days from 2014 onward). However, in bilateral replacements in Table 2.7, unlike in patient-level models, results are not significant even at minimally acceptable significance level of 10%, even though the signs of coefficients are preserved.

In the meantime, in line with individual-level models, the remaining parameters of care in

unilateral hip/knee replacements did not exhibit any significant shift. In particular, coefficients reflecting appropriateness through the use of antibiotic agent and admission based on pain symptoms did not undergo any considerable change, as suggested by coefficients β_1 in Tables 2.6 and Appendix F. A similar conclusion seems most likely with regards to quality outcomes, such as Charlson/Elixhauser post-admission comorbidity indexes and 180-day rates of replacement revision and hospital readmission.

As far as unincentivized types of joint replacement are concerned (i.e. bilateral hip and knee until end 1st March 2014, and all replacements on other joints), results for the respective procedures provided in Table 2.7 and Table 2.20 in Appendix F suggest that no spill-over effect from QBP/HBAM- targeted unilateral hip and knee replacements occurred with respect to virtually all care related parameters. None of the coefficients for QBP impact were significant in replacements other than hip and knee, while bilateral replacement demonstrated sporadic and marginally significant changes.

As in previous models, no significant change was found with regards to preferences for bilateral versus unilateral replacements in models on each QBP replacement type (i.e. share of unilateral/bilateral hip/knee replacements, as evidenced by the coefficient for the share of procedures of the same type in Table 2.7), since none of related coefficients fall below the minimal significance level threshold of 10%. On the extensive margin, the results on the impact of QBP/HBAM on shares of each joint replacement type in the total joint replacement output are also presented in Appendix F. No evidence was found in favor of spill-over effects from incentivized joint replacement types on the frequency of other types of replacement, which include all replacements other than hip and knee (e.g. ankle, shoulder).

Despite a likely violation of the common trend assumption and significant differences in pre-reform levels (see Figure 2.2, we tentatively tried to estimate the impact of QBP and HBAM on the likelihood of different modes of patient discharge. Tables 2.19 -2.21 provided in Appendix F offer tentative evidence that QBP and HBAM had an impact on the mode of discharge from hospital. Results on unilateral QBP procedures suggest that discharge home accompanied by post-rehabilitation services could have been rising in importance relative to other options, including discharge home with no support and being discharged to a post-operative rehabilitation facility. A similar finding is observed in hospital-level models run on all joint replacements, presented in Table 2.22 Appendix F. A more detailed analysis of these variables is provided in subsections 7.1 and 7.2 devoted to covariate and propensity matching techniques.

Table 2.6 – Hospital-level difference-in-difference estimates for unilateral hip replacements

Outcome	Appropriateness		Process		Quality			
	Antibiotic	Pain	Acute LOS	Transfer	Readm. (180d)	Revision (180d)	Charlson post	Elixhauser post
Dependent variable: (Intercept)	-0.023 (0.051)	0.179 (0.079)**	1.868 (6.063)	-1.047 (0.312)***	-0.006 (0.054)	-0.148 (0.069)**	1.481 (0.496)***	-1.504 (5.183)
post2012 QBP	-0.002 (0.001)*	0.002 (0.002)	-0.286 (0.165)**	< 0.001 (0.013)	-0.002 (0.002)	-0.001 (0.002)	-0.003 (0.012)	-0.06 (0.13)
post2014 QBP	0.001 (0.001)	-0.001 (0.002)	-0.265 (0.154)**	0.013 (0.009)	0.001 (0.002)	0.001 (0.002)	-0.002 (0.011)	-0.001 (0.126)
age	0.001 (0.002)	0.006 (0.003)**	0.195 (0.204)	0.031 (0.01)***	< 0.001 (0.002)	0.006 (0.002)***	-0.04 (0.016)**	0.049 (0.172)
age2	< 0.001 (< 0.001)	< 0.001 (< 0.001)**	< 0.001 (0.002)	< 0.001 (< 0.001)***	< 0.001 (< 0.001)	< 0.001 (< 0.001)**	< 0.001 (< 0.001)**	< 0.001 (0.001)
female	0.003 (0.006)	-0.018 (0.01)*	-0.193 (0.769)	-0.073 (0.032)**	-0.018 (0.007)***	-0.007 (0.009)	0.126 (0.056)**	1.027 (0.589)*
urban	0.001 (0.001)	0.001 (0.001)	0.666 (0.11)***	-0.01 (0.005)*	0.001 (0.001)	-0.001 (0.001)	-0.007 (0.008)	0.221 (0.087)**
QBP flag dummy	-0.002 (0.001)*	-0.002 (0.001)	-0.442 (0.121)***	-0.043 (0.009)***	-0.002 (0.001)	0.002 (0.001)	0.009 (0.009)	0.302 (0.096)***
Other controls	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Diagnosis FE	YES	NO	YES	YES	YES	YES	YES	YES
Hospital FE	NO	NO	NO	NO	NO	NO	NO	NO
N	1138	1138	1128	1138	1138	1138	1106	1106
R2	0.626	0.025	0.794	0.53	0.352	0.684	0.209	0.205
AdjR2	0.588	0.02	0.773	0.483	0.286	0.652	0.127	0.122

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at hospital level

Table 2.7 – Hospital-level difference-in-difference estimates for unilateral knee, bilateral knee and bilateral hip replacements

	Variable (SE)				Model statistics:		
	post2012 QBP	Sd	post2014 QBP	Sd	N	R2	AdjR2
Knee bilateral							
<i>Acute LOS</i>	0.226	(0.529)	-0.558	(0.503)	554	0.194	0.167
<i>Imaging procedures</i>	0.052	(0.046)	-0.001	(0.054)	265	0.13	0.066
<i>N.of post-admit diag</i>	-0.127	(0.089)	-0.039	(0.085)	557	0.09	0.059
<i>N.of interventions</i>	0.019	(0.139)	0.069	(0.101)	557	0.126	0.097
<i>Pain symptoms</i>	-0.003	(0.013)	0.005	(0.012)	557	0.028	-0.005
<i>Share bilateral repl.</i>	-0.004	(0.028)	-0.011	(0.027)	557	0.475	0.458
<i>Total LOS</i>	0.314	(0.538)	-0.666	(0.543)	556	0.172	0.144
<i>Transferred</i>	-0.069	(0.055)	0.029	(0.056)	557	0.205	0.179
<i>Charlson post</i>	-0.061	(0.214)	-0.121	(0.216)	557	0.168	0.14
<i>Elixhauser post</i>	0.203	(0.448)	-0.441	(0.396)	329	0.077	0.023
Hip bilateral							
<i>Acute LOS</i>	1.03	(1.348)	-1.938	(1.308)	263	0.254	0.199
<i>Imaging procedures</i>	0.052	(0.046)	-0.001	(0.054)	265	0.13	0.066
<i>N.of post-admit diag</i>	0.404	(0.23)*	-0.401	(0.219)*	265	0.149	0.087
<i>N.of interventions</i>	0.092	(0.15)	-0.089	(0.128)	265	0.084	0.017
<i>Pain symptoms</i>	-0.072	(0.049)	0.049	(0.042)	265	0.021	0.019
<i>Share bilateral repl.</i>	-0.005	(0.005)	0.004	(0.007)	265	0.06	-0.008
<i>Total LOS</i>	2.14	(1.587)	-3.033	(1.594)*	265	0.268	0.214
<i>Elixhauser post</i>	0.159	(0.372)	-0.337	(0.363)	265	0.51	0.475
Knee unilateral							
<i>Acute LOS</i>	-0.304	(0.117)***	-0.18	(0.107)*	1119	0.678	0.646
<i>Antibiotic use</i>	-0.001	(0.002)	0.001	(0.001)	1152	0.593	0.555
<i>Inhospital death</i>	0	(0)	0	(0)	1152	0.435	0.382
<i>Diagnostic procedures</i>	-0.002	(0.002)	0.003	(0.003)	1152	0.287	0.219
<i>Imaging procedures</i>	-0.004	(0.004)	-0.002	(0.005)	1152	0.374	0.315
<i>N.of post-admit diag</i>	-0.007	(0.015)	0.001	(0.015)	1152	0.457	0.406
<i>N.of interventions</i>	-0.116	(0.052)**	0.17	(0.043)***	1152	0.511	0.465
<i>Pain symptoms</i>	0	(0.001)	0.001	(0.001)	1152	0.054	0.047
<i>Readmission (180d)</i>	0.002	(0.001)	0	(0.001)	1152	0.349	0.288
<i>Revision (180d)</i>	0.002	(0.001)	0	(0.001)	1152	0.749	0.726
<i>Share unilateral repl.</i>	0.002	(0.005)	0	(0.005)	1152	0.221	0.148
<i>Total LOS</i>	-0.355	(0.122)***	-0.197	(0.111)*	1140	0.673	0.642
<i>Transferred</i>	0.005	(0.009)	0.01	(0.006)	1152	0.45	0.398
<i>Charlson post</i>	0.005	(0.011)	-0.018	(0.011)*	1094	0.153	0.069
<i>Elixhauser post</i>	0.033	(0.132)	0.142	(0.142)	1094	0.273	0.2

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses to the right of the coefficient column. Errors are clustered at hospital level

7 Robustness checks

Some of the classic techniques used to validate the robustness of difference-in-difference estimates are based on different matching algorithms. Their goal is to take into account observed differences between observations in control and treatment groups, which is achieved by reweighing observations in the sample such that more similar observations are assigned a greater weight. In practice, matching methods vary widely with respect to, in particular, variables/covariates used for finding matches, metrics reflecting the degree of similarity between observations, matching rules and available techniques to infer standard errors of obtained estimates. This same logic of sample reweighing can be applied to constructing a control group within the framework of the classic difference-in-difference design. In this section, we explore three additional estimation strategies that approach the task of causal inference from the mentioned standpoints.

7.1 Covariate matching (nearest-neighbor Mahalanobis distance matching)

To date, nearest neighbor matching (NNM) with replacement is by far the most well-researched and well-understood matching algorithm in applied econometric research. In essence, the goal of this matching technique consists in pairing each observation in the treatment group with a pre-specified number N (in our analysis, equalling 1) of observations in the control group, such that the dissimilarity distance between the two is the smallest.

In the context of our matching models, each observation corresponds to a hospital observed during one fiscal year. To ensure that bigger hospitals have a greater weight in estimating treatment effects, all observations were assigned weights corresponding to the number of patients admitted to hospital in a given year for the type of joint replacement in question. To measure the degree of dissimilarity between observations in our models, we rely on Mahalanobis distance applied to mean hospital-level patient (age, age squared, sex and Charlson and Elixhauser pre-admission comorbidity indexes, year of admission), and hospital characteristics (urban and rural populations and procedure volume). This measure is calculated as $M_{ij} = \sqrt{(x_i - x_j)'S^{-1}(x_i - x_j)}$, where for each observation i , j is taken from treatment/control group such that i and j are from opposite groups, x_i denotes vector of covariate values, and S is co-variance matrix from distributions of the treatment/control group opposite to i . Providing a comparison metric for each observation in terms of the number of standard deviations relative to an observation from the different group, it has the advantage of solving the issue of choosing appropriate measurement

units for included variables, which would normally have posed a problem in, for example, Euclidean or Manhattan metrics.

After M_{ij} is obtained for all i , each observation i ⁴ is matched to one observation k from the opposite group, such that k has minimal $M_i - M_k$. When an observation is chosen, it returns to the observation set of potential matches (i.e. matching with replacement). Subsequently, observations appear in the analytical data set as many times as they were drawn by the NNM algorithm.

This procedure is repeated for one pre-reform (2008-2011) and two post-reform periods (namely, 2012-2013 and 2014-2017)⁵. For each of them, both average treatment effects (ATE) $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$ and average treatment effects on the treated (ATT) $\tau^{tr} = \mathbb{E}[Y_i(1) - Y_i(0)|W_i = 1]$ are estimated as:

$$\hat{\tau} = \frac{1}{\sum_i w_i N_i} \sum_{i=1}^N w_i [\hat{Y}_i(1) - \hat{Y}_i(0)]$$

$$\hat{\tau}^{tr} = \frac{1}{\sum_{W_i=1} w_i N_i} \sum_{W_i=1} w_i [\hat{Y}_i(1) - \hat{Y}_i(0)],$$

where W_i is indicator of QBP and HBAM participation, $Y_i(S_i)$ is outcome dependent on states $S_i \in \{0, 1\}$; w_i are hospital-specific weights reflecting the volume of performed procedures, N and N_{tr} are total number of hospitals and the number of hospitals participating in QBP/HBAM, respectively. The estimated outcome values $\hat{Y}(S_i)$ are assigned as observed values for the observed state. For the unobserved counterfactual state, NNM-matched value corrected by *Abadie and Imbens (2011)* linear regression bias-correcting term is assigned, ensuring that the obtained estimates retain the property of $N^{1/2}$ consistency important for finite small and medium-size samples, although it comes at the cost of losing some efficiency.

In their earlier studies, *Abadie & Imbens (2006, 2008)* derived analytical asymptotic formulas for effect standard errors for the specific case of NNM matching with replacement. Finally, they also showed that, for this estimator, popular bootstrap estimates for standard errors of treatment effects are, in general, biased and inconsistent, although widely used in other types of matching techniques based on covariates or propensity score.

The corresponding effect of the QBP and HBAM is represented by the difference between estimated effects in post-reform and pre-reform periods, as showed in four last columns of Table 2.8. It reflects the difference in outcome attributable to QBP and HBAM, as long as

⁴Or only the ones from the treatment group, in the case where average effect on the treated is estimated

⁵Yearly analysis has also been performed, producing very similar general results with regards to QBP/HBAM impact. Estimation output is more voluminous and available on request

- as in the case of DiD inference- relevant time-varying characteristics were appropriately controlled for.

To ensure that covariate balance between treatment and control groups was achieved after the matching algorithm was applied, we visually examined distributions and checked Kolmogorov-Smirnov bootstrapped tests on equality of distributions.

The results of the NNM covariate matching models confirm the main conclusions drawn from the difference-in-difference models, namely, with regards to the reduction of observed hospital length of stay and plausible changes in the mode of patient discharge.

As far and QBP procedures are concerned, Tables 2.8 and 2.9 suggest that acute length of stay decreased on average by 0.4-0.6 days for unilateral hip and by around 0.2 for unilateral knee replacements. Results for bilateral replacement provided in Appendix G, however, are no longer significant, although the coefficients retained similar magnitudes and signs for the estimated effects (an temporary increase in 2012-13, followed by a decrease to the level 0.5-0.7 days below the pre-reform). However, different trends are observed for bilateral hip and bilateral knee replacements with regard to the probability to be discharged to a support care setting (the former saw a consistent and sharp fall by 20-30 percentage points, while the latter experienced a significant but a more moderate growth of 15-25 percentage points).

These results are in line with QBP recommendations published in 2012, which stipulated that being discharged home, preferably with additional rehabilitative support, should be prioritized by practitioners in most circumstances, since there was no associated reduction in rehabilitation quality based on available evidence.

In addition, NNM covariate matching models suggest that Ontario hospitals may have seen a very marginal but statistically significant decrease (by around 0.01 units per patient admission) in the number of imaging procedures after 2012 for unilateral hip replacements, and an increase by 2-3 percentage points of the probability of being transferred to another acute care facility. However, as far as bilateral replacements are concerned, the latter decreased by 6-10%, as shown in Table 2.23 of Appendix G. This result may be indicative of a higher retention of patients recommended for bilateral joint replacements, which goes along with the recommendations set out by the Ontario Health Technology Advisory Committee participating in the development of QBP procedure manuals. The evidence relative to changes in other parameters of care of QBP procedures remains sporadic and weak.

Finally, as evidenced by the results reported in Table 2.9, other joint replacements were unaffected by spillover effects from QBPs with respect to the vast majority of parameters of care, including length of hospital stay. The only aspect of care that consistently showed

Table 2.8 – Estimates of nearest-neighbor covariate matching for unilateral hip replacements

Variable	Effect	Estimate	SE	Estimate	SE	Estimate	SE	Effect	SE	Effect	SE
		2008-11 [1]		2012-13 [2]		2014-17 [3]		2012-13 [2] – [1]		2014-17 [3] – [1]	
Acute LOS	ATE	-0.803	0.187***	-1.214	0.192***	-1.387	0.153***	-0.41	0.268	-0.584	0.241**
	ATT	-0.831	0.252***	-1.379	0.221***	-1.448	0.197***	-0.548	0.335	-0.617	0.32*
Antibiotic use	ATE	-0.001	0.001	-0.002	0.002	-0.002	0.001	-0.001	0.002	-0.001	0.002
	ATT	0	0.001	-0.003	0.002	-0.002	0.002	-0.002	0.003	-0.001	0.002
Discharged home	ATE	-0.121	0.014***	-0.034	0.01***	-0.041	0.007***	0.088	0.017***	0.08	0.016***
	ATT	-0.117	0.019***	-0.041	0.013***	-0.045	0.009***	0.076	0.023***	0.072	0.021***
Discharged homewith support	ATE	0.288	0.023***	0.426	0.03***	0.417	0.026***	0.139	0.038***	0.129	0.035***
	ATT	0.256	0.032***	0.431	0.045***	0.387	0.037***	0.176	0.055***	0.131	0.049***
Discharged support service	ATE	0.178	0.013***	0.085	0.008***	0.076	0.005***	-0.092	0.015***	-0.102	0.014***
	ATT	0.183	0.018***	0.088	0.012***	0.078	0.007***	-0.095	0.021***	-0.104	0.019***
Inhospital death	ATE	-0.001	0.001	0.001	0.001	0	0.001	0.002	0.001	0.001	0.001
	ATT	-0.002	0.001	0	0.001	0	0.001	0.002	0.002	0.001	0.001
Diagnostic procedures	ATE	0.003	0.002	0.001	0.003	0.009	0.002***	-0.002	0.004	0.006	0.003*
	ATT	0.003	0.003	-0.001	0.004	0.009	0.003**	-0.004	0.005	0.006	0.004
Imaging procedures	ATE	0.057	0.006***	0.043	0.004***	0.044	0.003***	-0.014	0.007**	-0.013	0.006**
	ATT	0.052	0.006***	0.043	0.005***	0.045	0.004***	-0.01	0.008	-0.008	0.007
N. of post-admit diag	ATE	0.041	0.017**	0.033	0.019*	0.044	0.013***	-0.008	0.025	0.003	0.021
	ATT	0.02	0.021	0.028	0.023	0.04	0.017**	0.008	0.031	0.021	0.027
N. of interventions	ATE	0.076	0.021***	0.031	0.033	0.105	0.021***	-0.045	0.039	0.029	0.03
	ATT	0.07	0.025***	0.024	0.043	0.102	0.027***	-0.046	0.05	0.033	0.037
Pain symptoms	ATE	0.019	0.007***	0.029	0.009***	0.018	0.007**	0.01	0.011	-0.002	0.01
	ATT	0.017	0.008**	0.026	0.01**	0.021	0.009**	0.008	0.013	0.004	0.012
Readmission (180d)	ATE	-0.002	0.001*	-0.004	0.001***	-0.003	0.001***	-0.002	0.002	-0.002	0.001
	ATT	-0.002	0.001**	-0.003	0.002	-0.002	0.001**	-0.001	0.002	0	0.002
Revision (180d)	ATE	0.002	0.002	0	0.003	0.001	0.002	-0.002	0.003	-0.001	0.002
	ATT	0.001	0.002	0.002	0.004	0.001	0.002	0.001	0.004	0	0.003
Share unilateral repl.	ATE	0.002	0***	0.002	0.001***	0.002	0.001**	0	0.001	0	0.001
	ATT	0.001	0.001**	0	0.001	0	0.001	-0.001	0.001	-0.002	0.001
Total LOS	ATE	-0.705	0.222***	-1.13	0.216***	-1.481	0.208***	-0.425	0.31	-0.776	0.304**
	ATT	-0.754	0.298**	-1.314	0.264***	-1.596	0.233***	-0.56	0.398	-0.842	0.378**
Transferred	ATE	-0.059	0.007***	-0.05	0.007***	-0.033	0.004***	0.009	0.01	0.026	0.008***
	ATT	-0.069	0.01***	-0.045	0.01***	-0.032	0.006***	0.025	0.014*	0.037	0.011***
Charlson post	ATE	0.006	0.008	0.009	0.009	-0.001	0.007	0.003	0.012	-0.008	0.011
	ATT	0.001	0.011	0	0.013	-0.006	0.011	-0.001	0.017	-0.007	0.016
Elixhauser post	ATE	0.228	0.069***	0.413	0.083***	0.204	0.077***	0.185	0.108*	-0.024	0.104
	ATT	0.217	0.097**	0.378	0.124***	0.213	0.106**	0.161	0.157	-0.005	0.144

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given to the right of estimated effects.

significant results was related to the mode of patient discharge after 2014. As in the case of the QBP procedures, more patients were recommended for discharge to home with accompanying rehabilitation support (7 percent point increase post-2014 relative to pre-reform), and moderately fewer (almost 2 percentage point decrease post-2014) patients were discharged to a specialized care facility.

7.2 Propensity score matching

One of disadvantages of NNM consists in the fact that no adjustment is made with regard to degree of similarity between observations after matches are found. In addition, unless a matching caliper is imposed while searching matches (which can have unpredictable consequences on the validity of Abadie-Imbens asymptotic variance formulas), the NNM algorithm can pick vastly different pairs of observations, making the estimator asymptotically less efficient. Moreover, in the context of finite samples, a potentially more pernicious

spin-off of this problem can be an increased risk of regression to the mean, whereby observations are more likely to be chosen as the other's conjugate the closer they are situated to the distribution mean. To address these concerns and to provide additional robustness to our analysis, we test propensity score matching (PSM) with control observations weighed by a kernel function.

Despite having a lot in common, covariate NNM and kernel PSM matching have several important implementation differences and ramifications with regard to interpretation of obtained results. From the implementation standpoint, the propensity score is obtained by running logistic regression models on a set of covariates identical to NNM models in subsection 7.1, whose estimates are used to predict probabilities of being classified as a treatment group observation in the tested PSM models. Treatment effects $\hat{\tau}$ and $\hat{\tau}^{tr}$ are computed in a fashion identical to NNM covariate matching described above, except for the fact that in kernel-weighted PSM the predicted outcome values in unobserved counterfactual state are a weighed average of all hospitals from the opposite (treatment or control) group⁶

Thus, the estimation of $\hat{Y}_i(S_i)$ is modified to incorporate kernel weights, such that in treated and control states :

$$\hat{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0 \\ \frac{1}{\sum_{W_m \neq W_i} w_m k_m} \sum_{W_m \neq W_i} w_m k_m Y_m & \text{if } W_i = 1 \end{cases}$$

$$\hat{Y}_i(1) = \begin{cases} \frac{1}{\sum_{W_m \neq W_i} w_m k_m} \sum_{W_m \neq W_i} w_m k_m Y_m & \text{if } W_i = 0 \\ Y_i & \text{if } W_i = 1, \end{cases}$$

where Y_i and w_i follow the notation provided in Subsection 7.1. Kernel weights k_i are obtained by applying the Epanechnikov kernel function K to estimated propensity scores, such that $k_i = \frac{1}{n_c h} K\left(\frac{x_c - x_t}{h}\right)$, where t and c are indexes for observations in the treatment group population and the other is in the control group of size n_c ⁷. In our PSM models, the Epanechnikov kernel is used with the plug-in bandwidth parameter $h = 1.06\hat{\sigma}n^{-1/5}$.

To the best of our knowledge, unlike in covariate NNM, analytical approximations for standard errors are not available for this type of estimator. Hence, bootstrap remains the

⁶Thus, NNM is a particular, akin to degenerate, case of sample weight whose functional form is $\mathbb{I}(X = \text{argmin} \{M_j\})$

⁷In case of calculating ATE for the part of observations belonging to the control group, indexes c and t are reversed

sole available option for standard error estimation, although it is generally advised that it be used with caution (*Abadie and Imbens, 2008*)⁸. To take into account this shortfall of kernel PSM, coefficient estimates should be assumed to be more reliable than their bootstrapped standard errors while interpreting the results.

As in the case of covariate nearest-neighbor matching, both ATE and ATT effects are reported for all QBP replacements, as well as for all replacements combined.

The estimation results for unilateral hip, unilateral knee and other (i.e. non-QBP) replacements are presented in Table 2.10, while results for the remaining procedure types and for all replacements types combined can be found in Table 2.26 of Appendix G. Obtained coefficients are in line with the general conclusion that the introduction of QBP led to a decrease in acute LOS for the eligible procedures, although the size of the effects was more mitigated (0.15-0.5 days for hip and knee replacements). In a similar vein, results on modes of patient discharge show the earlier observed pattern wherein increasingly more patients spend their post-operational rehabilitation period at home with or without supporting services. Despite the fact that statistical significance tended to be less strong within this class of models, this conclusion should be considered with caution, due to the aforementioned issue of obtaining consistent standard errors.

7.3 Synthetic kernel-weighted control group

This strategy incorporates characteristics of matching techniques while preserving the framework of DiD analysis. In particular, hospitals in the control group are assigned a kernel weight based on the degree of their closeness to hospitals in the treatment group represented by logit propensity score, in the way described in subsection 7.2. The important difference of this method relative to covariate/propensity matching consists in relying only on pre-reform hospital characteristics. More specifically, the weight is computed as an average of Epanechnikov kernel weights used in the aforementioned subsection for each pre-reform year spanning across 2008-2012. This implies that hospital-specific weights remain constant over the whole available analytical time frame of the data.

After obtaining hospital weights, the same specification used with the DiD regression models are used as in the subsection 6.2. As earlier, errors are clustered at the hospital level.

The results for unilateral hip and knee replacements are presented in Table 2.11, while tables with treatment effect estimates for other procedures can be found in Appendix

⁸To obtain estimates, we bootstrap values of ATT and ATE (with replacement), with the number of bootstrap replications $B = 499$

H. The results are consistent, albeit statically less significant, and corroborate the previous finding that QBP led to a decrease of 0.3-0.5 days in acute length of stay in the two first years after the reform, and an additional 0.3-0.4 days in the following years. In addition, these models provide additional confirmatory evidence that patient discharge mode shifted towards rehabilitation at home with/without supporting services, while the resource-intensive support care facilities showed a consistent decline. Finally, in line with previous evidence, other types of care parameters generally used to monitor appropriateness and quality of care remained largely unaffected.

Tables with estimates on bilateral procedures and other joint replacement are also available in Appendix H. Overall, they preserve the same pattern in terms of the direction and the magnitude of observed change. Although for bilateral hip replacement a few outcome variables, namely the number of post-admission diagnoses and admission with pain symptom, showed a significant increase of 0.72 units and a decrease of 0.15 percentage points in 2012-13, respectively, these effects subsided very quickly in the following years. In the meantime, no such effects were observed in the closely related bilateral knee replacements. On the extensive margin, no shift was noticeable with regards to preference for bilateral/unilateral replacement, nor regarding the volume of individual joint replacement types. Finally, estimation results did not indicate meaningful changes in replacements other than hip and knee with respect to all mentioned parameters.

Table 2.9 – Estimates of nearest-neighbor covariate matching for unilateral knee and other replacements

Variable	Effect	Knee unilateral				Other replacements			
		Effect 2012- 13	SE	Effect 2014- 17	SE	Effect 2012- 13	SE	Effect 2014-17	SE
<i>Acute LOS</i>	ATE	-0.211	0.121*	-0.175	0.097*	-0.229	0.261	0.07	0.211
	ATT	-0.277	0.152*	-0.212	0.119*	-0.587	0.417	0.123	0.29
<i>Antibiotic use</i>	ATE	-0.002	0.003	0.003	0.002*	-0.001	0.001	-0.002	0.001
	ATT	-0.002	0.005	0.004	0.002*	0.001	0.001	0.001	0.001
<i>Discharged home</i>	ATE	0.085	0.017***	0.078	0.016***	-0.002	0.01	0.013	0.008
	ATT	0.077	0.024***	0.07	0.022***	0.004	0.013	0.013	0.011
<i>Discharged home with support</i>	ATE	0.113	0.043***	0.113	0.039***	0.04	0.036	0.079	0.027***
	ATT	0.16	0.059***	0.136	0.054**	0.027	0.051	0.091	0.036**
<i>Discharged support service</i>	ATE	-0.101	0.015***	-0.109	0.014***	0.001	0.007	-0.014	0.006**
	ATT	-0.106	0.021***	-0.114	0.02***	-0.002	0.01	-0.017	0.008**
<i>Inhospital death</i>	ATE	-0.001	0.001	0	0	0	0.003	-0.001	0.001
	ATT	-0.001	0.002	0	0	-0.004	0.004	-0.001	0.001
<i>Diagnostic procedures</i>	ATE	-0.001	0.002	0.009	0.003***	-0.006	0.007	0.007	0.007
	ATT	-0.002	0.003	0.006	0.004	-0.011	0.011	-0.001	0.009
<i>Imaging procedures</i>	ATE	-0.006	0.004	-0.001	0.005	-0.014	0.012	-0.023	0.011**
	ATT	-0.006	0.007	-0.003	0.007	-0.011	0.012	-0.031	0.014**
<i>N.of post-admit diag</i>	ATE	0.001	0.017	0.002	0.014	-0.006	0.024	0.007	0.02
	ATT	0.011	0.023	0.004	0.019	-0.016	0.036	0.004	0.027
<i>N.of interventions</i>	ATE	-0.116	0.055**	0.138	0.038***	0.052	0.076	0.075	0.067
	ATT	-0.15	0.076**	0.093	0.048*	0.031	0.099	0.021	0.083
<i>Pain symptoms</i>	ATE	-0.004	0.004	-0.009	0.003***	0.012	0.031	-0.037	0.023
	ATT	-0.005	0.005	-0.012	0.004***	0.026	0.043	-0.061	0.03*
<i>Readmission (180d)</i>	ATE	0.003	0.001***	0.005	0.002***	0	0.004	-0.001	0.003
	ATT	0.005	0.002***	0.006	0.001***	-0.002	0.005	-0.004	0.004
<i>Revision (180d)</i>	ATE	0.002	0.002	0.008	0.002***	0.005	0.005	0.004	0.005
	ATT	0.005	0.003	0.009	0.002***	0.005	0.007	0	0.007
<i>Share unilateral repl.</i>	ATE	0.004	0.005	0.005	0.004				
	ATT	0.003	0.007	0.006	0.006				
<i>Total LOS</i>	ATE	-0.288	0.144**	-0.198	0.099**	-0.281	0.33	-0.468	0.32
	ATT	-0.438	0.208**	-0.229	0.122*	-0.495	0.557	-0.614	0.461
<i>Transferred</i>	ATE	0.022	0.008***	0.036	0.007***	0.006	0.007	0.004	0.006
	ATT	0.034	0.012***	0.047	0.01***	0.006	0.009	0.01	0.007
<i>Charlson post</i>	ATE	0.007	0.011	-0.014	0.01	-0.012	0.029	0.014	0.029
	ATT	0.01	0.015	-0.008	0.013	-0.016	0.045	-0.017	0.039
<i>Elixhauser post</i>	ATE	-0.088	0.134	0.15	0.109	0.425	0.369	-0.037	0.309
	ATT	-0.138	0.188	0.232	0.154	0.456	0.459	-0.192	0.428

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given to the right of estimated effects.

Table 2.10 – Propensity score matching estimates for unilateral hip, unilateral knee and other replacements

Variable	Hip unilateral			Knee unilateral			Other replacement		
	Effect	SE	Effect	SE	Effect	SE	Effect	SE	
Acute LOS	ATE	-0.151	0.426	-0.404	0.426	-0.147	0.188	-0.243	0.143*
	ATT	-0.185	0.58	-0.398	0.58	-0.188	0.225	-0.236	0.164
Antibiotic use	ATE	-0.002	0.004	-0.002	0.004	0.001	0.003	0.003	0.002
	ATT	-0.001	0.002	-0.001	0.002	0.001	0.004	0.003	0.002
Discharged home	ATE	0.097	0.027***	0.093	0.027***	0.1	0.025***	0.1	0.023***
	ATT	0.087	0.03***	0.095	0.03***	0.098	0.024***	0.1	0.022***
Discharged home with support	ATE	0.123	0.051**	0.137	0.051**	0.102	0.064	0.11	0.052**
	ATT	0.145	0.055***	0.155	0.055***	0.122	0.062*	0.104	0.049**
Discharged support service	ATE	-0.101	0.018***	-0.098	0.018***	-0.103	0.017***	-0.112	0.017***
	ATT	-0.099	0.021***	-0.105	0.021***	-0.106	0.017***	-0.113	0.016***
Inhospital death	ATE	0.001	0.006	0.001	0.006	0	0.001	0	0
	ATT	0.001	0.001	0.001	0.001	0	0.002	0	0
Diagnostic procedures	ATE	-0.003	0.004	0.007	0.004	-0.001	0.004	0.004	0.004
	ATT	-0.002	0.004	0.008	0.004	-0.001	0.006	0.003	0.005
Imaging procedures	ATE	-0.005	0.008	-0.011	0.008	-0.006	0.005	-0.006	0.006
	ATT	-0.009	0.008	-0.011	0.008	-0.006	0.005	-0.006	0.004
N. of post-admit diag	ATE	-0.018	0.036	0.004	0.036	-0.014	0.027	-0.009	0.019
	ATT	-0.015	0.034	0.008	0.034	-0.001	0.031	-0.005	0.021
N. of interventions	ATE	-0.005	0.051	0.029	0.051	-0.048	0.075	0.122	0.062*
	ATT	-0.019	0.054	0.039	0.054	-0.064	0.081	0.115	0.067*
Pain symptoms	ATE	0.011	0.019	-0.001	0.019	-0.002	0.007	-0.003	0.006
	ATT	0.01	0.02	-0.004	0.02	-0.002	0.008	-0.005	0.008
Readmission (180d)	ATE	-0.002	0.002	-0.001	0.002	0.003	0.002	0.003	0.001**
	ATT	0	0.002	0	0.002	0.003	0.003	0.003	0.001**
Revision (180d)	ATE	-0.002	0.005	-0.002	0.005	0.003	0.003	0.004	-0.004
	ATT	0.001	0.005	-0.001	0.005	0.003	0.004	0.005	-0.001
Share unilateral repl.	ATE	0.001	0.001	0.002	0.001	0.004	0.005	0.004	0.004
	ATT	0	0.002	0.002	0.002	0.003	0.006	0.002	0.005
Total LOS	ATE	-0.354	0.87	-0.441	0.87	-0.238	0.224	-0.41	0.161**
	ATT	-0.442	1.309	-0.365	1.309	-0.263	0.292	-0.442	0.186**
Transferred	ATE	0.011	0.018	0.011	0.018	0.008	0.013	0.016	0.01
	ATT	0.019	0.021	0.015	0.021	0.011	0.016	0.016	0.013
Charlson post	ATE	-0.004	0.016	-0.011	0.016	0.003	0.012	-0.017	0.011
	ATT	0.001	0.017	-0.006	0.017	0.002	0.012	-0.02	0.013
Eichlauser post	ATE	-0.021	0.136	-0.105	0.136	-0.065	0.156	0.118	0.136
	ATT	0.004	0.149	-0.112	0.149	-0.086	0.178	0.198	0.152
									0.695
									0.582
									0.23
									0.418
									0.318
									0.455
									0.008
									0.01
									0.04
									0.034
									0.385
									0.418

Legend: ***- 1 % sign., **-5% sign., *-10% sign. Coefficient std. errors are given to the right of estimated effects.

Table 2.11 – Difference-in-difference estimates with synthetic kernel-weighted control group, for unilateral hip and unilateral knee replacements

	Variable				Model statistics:		
	post2012 QBP	Sd	post2014 QBP	Sd	N	R2	AdjR2
Hip unilateral:							
<i>Acute LOS</i>	-0.51	(0.29)*	-0.39	(0.311)	1100	0.921	0.913
<i>Antibiotic use</i>	0.007	(0.007)	-0.004	(0.007)	1104	0.793	0.771
<i>Discharged home</i>	0.094	(0.02)***	-0.02	(0.017)	1104	0.849	0.833
<i>Discharged home with support</i>	0.131	(0.053)**	0.025	(0.056)	1104	0.493	0.44
<i>Discharged support service</i>	-0.087	(0.016)***	0.006	(0.011)	1104	0.749	0.723
<i>Inhospital death</i>	0.001	(0.002)	-0.001	(0.002)	1104	0.782	0.759
<i>Diagnostic procedures</i>	-0.008	(0.005)	0.012	(0.005)**	1104	0.745	0.719
<i>Imaging procedures</i>	-0.008	(0.012)	-0.006	(0.011)	1104	0.651	0.615
<i>N.of post-admit diag</i>	0.018	(0.035)	0.002	(0.035)	1104	0.83	0.812
<i>N.of interventions</i>	-0.048	(0.046)	0.024	(0.048)	1104	0.85	0.834
<i>Pain symptoms</i>	0.003	(0.005)	-0.006	(0.005)	1104	0.025	0.021
<i>Readmission (180d)</i>	0.002	(0.003)	0.002	(0.003)	1104	0.634	0.596
<i>Revision (180d)</i>	0.012	(0.008)	-0.007	(0.008)	1104	0.789	0.767
<i>Share unilateral repl.</i>	0.002	(0.002)	-0.002	(0.002)	1104	0.583	0.539
<i>Total LOS</i>	-0.42	(0.412)	-0.616	(0.398)	1101	0.909	0.899
<i>Transferred</i>	-0.004	(0.016)	0.013	(0.015)	1104	0.768	0.744
<i>Charlson post</i>	0.006	(0.017)	-0.005	(0.018)	1082	0.42	0.358
<i>Elixhauser post</i>	-0.09	(0.188)	-0.011	(0.2)	1082	0.462	0.404
Knee unilateral:							
<i>Acute LOS</i>	-0.306	(0.192)	-0.307	(0.185)*	1102	0.834	0.818
<i>Antibiotic use</i>	-0.002	(0.003)	0.005	(0.005)	1127	0.897	0.887
<i>Discharged home</i>	0.099	(0.018)***	0.012	(0.012)	1127	0.83	0.813
<i>Discharged home with support</i>	0.17	(0.059)***	-0.066	(0.056)	1127	0.404	0.346
<i>Discharged support service</i>	-0.095	(0.016)***	-0.019	(0.01)**	1127	0.819	0.802
<i>Inhospital death</i>	0.001	(0.001)	0.002	(0.004)	1127	0.957	0.953
<i>Diagnostic procedures</i>	0.002	(0.003)	0.002	(0.003)	1127	0.469	0.418
<i>Imaging procedures</i>	-0.006	(0.004)	0.007	(0.007)	1127	0.621	0.584
<i>N.of post-admit diag</i>	0.016	(0.025)	-0.005	(0.026)	1127	0.748	0.724
<i>N.of interventions</i>	-0.071	(0.052)	0.17	(0.05)***	1127	0.716	0.688
<i>Pain symptoms</i>	0.005	(0.004)	-0.004	(0.004)	1127	0.038	0.031
<i>Readmission (180d)</i>	0.003	(0.002)	0	(0.002)	1127	0.773	0.751
<i>Revision (180d)</i>	0.001	(0.002)	0	(0.002)	1127	0.94	0.935
<i>Share unilateral repl.</i>	0.002	(0.005)	0.003	(0.005)	1127	0.46	0.408
<i>Total LOS</i>	-0.367	(0.194)*	-0.301	(0.185)	1115	0.828	0.812
<i>Transferred</i>	-0.002	(0.011)	0.005	(0.011)	1127	0.644	0.61
<i>Charlson post</i>	0.019	(0.017)	-0.016	(0.015)	1078	0.234	0.156
<i>Elixhauser post</i>	0.265	(0.157)*	0.224	(0.182)	1078	0.398	0.337

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses to the right of coefficient columns. Errors are clustered at hospital level

8 Discussion and conclusion

Using patient-level data from Canadian Discharge Abstract Database (DAD), we evaluate through both a difference-in-difference approach and a range of matching covariate/propensity score matching techniques the impact of QBPs/HBAM on the volume and quality of targeted procedures and other types of joint replacements plausibly competing for hospital resources. After controlling for patient, hospital and regional characteristics, we found a significant decrease in acute length of stay associated with QBPs, as well as a marked shift towards patients being discharged home with/without post-operative supporting services. However, evidence for quality improvement across all joint replacement types was weak, inconsistent and at best short-lived.

To test robustness, we first ran the more conservative and, from a few perspectives, less flexible NNM algorithm with replacement and analytical Abadie-Imbens bias-corrected standard errors. To provide additional robustness, this analysis was further complemented with propensity score matching with control observations weighed by Epanechnikov kernel function, and bootstrapped (with replacement) standard errors, and kernel-bases control groups in difference-in-difference models.

There are several results presented in this Chapter that may have policy implications. First, this Chapter provides an insight with regards to the impact of activity-based funding, potentially supplemented with weak or lapsed pay-for-performance incentives. Our results indicate that such reforms can lead to a reduction in hospital stay, likely in an attempt to minimize financial loss and/or maximize hospital operational revenue. In addition, we find evidence that non-monetary and soft mechanisms aimed at improving care, in and of themselves, are unlikely to translate into meaningful, let alone long-lasting, clinical changes with regards to virtually any quality dimension of care.

However, there are some considerations that can arguably affect the external and internal validity of the results. First, the tested models assume unidirectional causal impact, while it might not entirely be the case at least for certain outcome variables. For example, premature patient discharge may decrease the number registered comorbidities and affect the observed level of severity. On the other hand, a prolonged hospital stay might be associated with an additional risk of hospital-acquired (nosocomial) infection, which—directly or through pre-admission comorbidity indexes—are controlled for in the tested models. Thus, despite being widely used in applied economics, estimates obtained through DiD and matching methods might inherently suffer from reverse causality problem, whose impact is difficult to estimate.

Another general concern in difference-in-difference estimation is the degree to which the

internal validity of the results may be compromised by omitted time-variant variables. However, the datasets that were at our disposal contained an extremely rich set of patient- and hospital-level variables, that take into account nearly all relevant and proven clinical risk factors. We expect this circumstance to minimize this risk of this channel negatively affecting the validity of our estimates.

Since data used in this study are obtained from hospital administrative registers and went through mandatory data quality controls by Canadian Institute for Health Information (CIHI), the risk of measurement errors affecting model estimates is likely minimal. In addition, due to the QBP/HBAM being tied exclusively to providers in Ontario, the risk of patient self-selection into treatment and control populations can be expected to be very low.

As far as external validity is concerned, the results presented in this Chapter can— with due diligence— be generalized to most Canadian jurisdictions and to countries having a similar institutional setting. To name a few points, these systems should feature universal health coverage for elective joint replacements and generate little to no out-of-pocket expenses passed on to the patient. In addition, such jurisdictions would be expected to have a comparable level of per capita healthcare expenditures. Among potential examples of these counties one could include many Western European states, such as France, Germany and the Netherlands. In addition, due to the presence of two distinct reform components —prospective funding and allocation of a global budget— the empirical set-up also bears a considerable degree of similarity with the US. In particular, the QBP component of the reform resembles US Medicare and Medicaid plans, in which, thanks to their large enrollee pool, the government has enough power to set prices to providers. This brings hospital payments received for treating patients insured by Medicare and Medicaid closer to fully prospective DRG funding. On the other hand, the majority of private employer-sponsored insurance plans negotiate prices with hospitals on their own; different plans setting their own premiums, providing various degree of coverage and featuring different patient case-mix. This type of insurance plan has similarities with HBAM in that they allow for a risk adjustment with respect to the characteristics of the enrollee pool and its size.

Further lines of research in which our research agenda can be expanded and elaborated are suggested in the General Conclusion of this dissertation.

9 Appendixes

Appendix A. Data specifications

Project Title: The impact of introduction of Quality Based Procedures (QBP) for hip and knee replacement on orthopedic care quality, intensity and care substitution in Ontario.

Database(s):

- Discharge Abstract Database (DAD) – (excluding Quebec)
- National Ambulatory Care Reporting System (NACRS)

Level(s) of Care (facility type):

- Discharge Abstract Database (DAD)
 - All levels of care
- National Ambulatory Care Reporting System (NACRS)
 - All levels of care

Fiscal Year(s):

- 2008/09 to 2017/18

Classification:

- ICD-10-CA/CCI

Scope:

- All records of Canadians with a recorded orthopedic procedure of interest that were submitted by Alberta, British Columbia, and Ontario to DAD/NACRS during the study period. **NOTE:** The patient must be a resident of the submitting province/territory.

Details of Request:

DAD – All Levels of Care

Inclusions:

- Fiscal Year = 2008/09 to 2017/18
 - Patient must have an admission date within the study period.
- Submitting Province/Territory = Alberta, British Columbia, and Ontario
 - Please note that the patient must be a resident of the submitting province. Province/Territory Issuing Health Card Number will be used to determine residency (include code = CA - Canada (Penitentiary Inmates, Indian Affairs, Veteran Affairs)).
- Analytical institution Type = All
- Patient's age GE 18 years old.
- Include all deaths
- Canadian Residents only = based on Postal Code (include Transient/homeless).
- Record must have at least one of the following to be selected:
 - Diagnosis Code of interest – please refer to Appendix A – Diagnosis Codes
 - Diagnosis Type = MRDx
 - Procedure Code of interest – please refer to Appendix B – Procedure Codes

- Procedure Code Position = any
- Include Out of Hospital (OOH) procedures
- Include abandon procedures
- Case Mix Group (CMG) of interest – please refer to Appendix C – CMG/CACS Codes

Exclusions:

- Postal Code = US States, Other Country, Unknown.
- Exclude newborns, stillbirths, and cadaveric donations.
- Trauma related records:
 - ICD-10-CA Codes: S00.^ to T32.^
 - Diagnosis Type = MRDx
- Cancer related records:
 - ICD-10-CA Codes: C.^ to D.^
 - Diagnosis Type = MRDx
- All therapeutic abortions (TA) records, defined as:
 - *ICD-10-CA/CCI Codes*
 - ICD-10-CA code = [O04.^, O07.^, P96.4], in any position within the diagnosis fields **OR**
 - CCI Intervention codes = [5.CA.20.^, 5.CA.24.^, 5.CA.88.^, 5.CA.89.^, 5.CA.90.^], in any position within the intervention fields.

NACRS – All Levels of Care

Inclusions:

- Fiscal Year = 2008/09 to 2017/18
- Submitting facility provinces = ON, BC*, NS*, YK*, SK*, PE*, MB*, AB
- Ambulatory care group = All
- Include all deaths
- Canadian Residents only = based on Postal Code (include Transient/homeless).
- Record must have at least one of the following to be selected:
 - Diagnosis Code of interest – please refer to Appendix A – Diagnosis Codes
 - Diagnosis Type = Main Problem
 - Procedure Code of interest – please refer to Appendix B – Procedure Codes
 - Procedure Code Position = any
 - Include Out of Hospital (OOH) procedures
 - Include abandon procedures
 - Comprehensive Ambulatory Classification System (CACS) of interest – please refer to Appendix C – CMG/CACS Codes

* Partial submission only, in some of the fiscal years

Exclusions:

- Postal Code = US States, Other Country, Unknown.
- Exclude instances of interventions performed out of hospital, and abandoned/cancelled.
- All therapeutic abortions (TA) records, defined as:
 - *ICD-10-CA/CCI Codes*
 - ICD-10-CA code = [O04.^, O07.^, P96.4], in any position within the diagnosis fields **OR**
 - CCI Intervention codes = [5.CA.20.^, 5.CA.24.^, 5.CA.88.^, 5.CA.89.^, 5.CA.90.^], in any position within the intervention fields.

A meaningless but unique number (MBUN) will be assigned to identify all hospitalization records that belong to the same individual. These IDs will be released upon approval by the CIHI Privacy, Confidentiality and Security (PC&S) committee.

PRIVACY ROUTINE: To comply with CIHI's Privacy and Confidentiality guidelines, the following modified privacy routine level 3 will be applied:

DAD Records

Field Names	Privacy Routine Level 3 (modified)
Institution/facility number	De-identified
Chart/health record number	Not Provided
2 nd Chart/health register number	Not Provided
Register/registration number	Not Provided
Maternal/newborn chart/register number	Not Provided
Health care/card number	Not Provided
Provincial Ancillary Data	Not Provided
Postal code	Truncated – FSA only (1 st three digits)
Birthdate	Not Provided
Institution/facility transferred from	De-identified
Institution/facility transferred to	De-identified
OOH (out of hospital) institution/facility number	De-identified
Provider number	De-identified
Intervention provider number	De-identified
Anaesthetist number	De-identified
Previous therapeutic abortions	Not Provided
Mental health source of referral	Not Provided
Mental health method of admission	Not Provided
Mental health change in legal status	Not Provided
Mental health AWOL	Not Provided
Mental health suicide	Not Provided
Mental health previous psych admission	Not Provided
Mental health referred to	Not Provided
Mental health ECT indicator	Not Provided
Mental health number of ECT	Not Provided
Mental health education	Not Provided
Mental health employment status	Not Provided
Mental health financial support	Not Provided

NACRS Records

Field Name	Privacy Routine Level 3 (Modified)
Institution/facility number	De-identified
Chart/health record number	Not Provided

Field Name	Privacy Routine Level 3 (Modified)
Ambulatory registration number	Not Provided
Health care/card number	Not Provided
Postal code	Truncated – FSA only (1 st three digits)
Birthdate	Not Provided
Institution/facility transferred from	De-identified
Institution/facility transferred to	De-identified
OOH (out of hospital) institution/facility number	De-identified
Ambulance call number	Not Provided
Provider identification	De-identified
Living arrangement	Not Provided
Highest level of education	Not Provided
Previous therapeutic abortions	Not Provided

Yukon, Northwest Territories and Nunavut Records

Due to the low number of facilities in the Yukon, Northwest Territories and Nunavut all provincial/territorial identifiers must be either changed to the letter “A” or suppressed to prevent any possibility of identifying the institutions within these provinces/territories. The following data fields need to be changed/suppressed:

- Changed to the letter “A”
 - Province
 - Institution number (Province/Territory code)
 - Institution From Number (Province/Territory code)
 - Institution To Number (Province/Territory code)
 - Intervention 1 – 20 OOH Institution Number (Province/Territory code)
- Suppressed
 - Province Issuing health Care/Card Number
 - Postal Code Forward Sortation Area (FSA)
 - Residence Code

Diagnosis Codes

ICD-10-CA DIAGNOSIS DESCRIPTION	DIAGNOSIS CODES OF INTEREST				NOTES
	ICD-10-CA DIAGNOSIS CODES				
	v2006	v2009	v2012	v2015	
Arthropathic psoriasis	L405	-	-	-	No change
Pyogenic arthritis NOS lower leg	M0096	-	-	-	No change
Other specified rheumatoid arthritis	M068	-	-	-	No change
Rheumatoid arthritis unspecified	M069	-	-	-	No change
Juvenile rheumatoid arthritis	M080	-	-	-	No change
Other specified arthritis lower leg	M1386	-	-	-	No change
Arthritis unspecified pelvis & thigh	M1395	-	-	-	No change
Arthritis unspecified lower leg	M1396	-	-	-	No change
Polyarthrosis	M15	-	-	-	No change
Primary generalized (osteo)arthrosis	M150	-	-	-	No change
Secondary multiple arthrosis	M153	-	-	-	No change
Erosive (osteo)arthrosis	M154	-	-	-	No change
Other polyarthrosis	M158	-	-	-	No change

Polyarthrosis, unspecified	M159	-	-	-	No change
Primary coxarthrosis bilateral	M160	-	-	-	No change
Other primary coxarthrosis	M161	-	-	-	No change
Bil coxarthrosis result from dysplasia	M162	-	-	-	No change
Other dysplastic coxarthrosis	M163	-	-	-	No change
Other post-traumatic coxarthrosis	M165	-	-	-	No change
Other secondary coxarthrosis bilateral	M166	-	-	-	No change
Other secondary coxarthrosis	M167	-	-	-	No change
Coxarthrosis unspecified	M169	-	-	-	No change
Primary gonarthrosis bilateral	M170	-	-	-	No change
Other primary gonarthrosis	M171	-	-	-	No change
Post-traumatic gonarthrosis bilateral	M172	-	-	-	No change
Other post-traumatic gonarthrosis	M173	-	-	-	No change
Other secondary gonarthrosis bilateral	M174	-	-	-	No change
Other secondary gonarthrosis	M175	-	-	-	No change
Gonarthrosis unspecified	M179	-	-	-	No change
Arthrosis of first carpometacarpal joint	M18	-	-	-	No change
Primary arthrosis of first carpometacarpal joints, bilateral	M180	-	-	-	No change
Other primary arthrosis of first carpometacarpal joint	M181	-	-	-	No change
Post-traumatic arthrosis of first carpometacarpal joints, bilateral	M182	-	-	-	No change
Other post-traumatic arthrosis of first carpometacarpal joint	M183	-	-	-	No change
Other secondary arthrosis of first carpometacarpal joints, bilateral	M184	-	-	-	No change
Other secondary arthrosis of first carpometacarpal joint	M185	-	-	-	No change
Arthrosis of first carpometacarpal joint, unspecified	M189	-	-	-	No change
Other arthrosis	M19	-	-	-	No change
Primary arthrosis of other joints	M190	-	-	-	No change
Post-traumatic arthrosis of other joints	M191	-	-	-	No change
Other secondary arthrosis	M192	-	-	-	No change
Other specified arthrosis	M198	-	-	-	No change
Arthrosis, unspecified	M199	-	-	-	No change
Osteoporosis NOS w path fx pelvis thigh	M8095	-	-	-	No change
Nonunion fx [pseudarthrosis] pelvis thigh	M8415	-	-	-	No change
Pathological fracture NEC pelvis thigh	M8445	-	-	-	No change
Idiopath aseptic necrosis bone pelv thigh	M8705	-	-	-	No change
Osteonecrosis due to drugs pelvis thigh	M8715	-	-	-	No change
Osteonecrosis dt prev trauma pelv thigh	M8725	-	-	-	No change
Other osteonecrosis lower leg	M8786	-	-	-	No change

Unspecified osteonecrosis pelvis thigh	M8795	-	-	-	No change
Unspecified osteonecrosis lower leg	M8796	-	-	-	No change
Mech comp of int fix device of femur	T8413	-	-	-	No change
Infect & infl reaction dt knee prosth	T8454	-	-	-	No change
Oth comp int ortho prosth dev impl gft	T848	-	-	-	No change

Procedure Codes

PROCEDURE CODES OF INTEREST					
CCI PROCEDURE DESCRIPTION	CCI CODES				NOTES
	v2006	v2009	v2012	v2015	
Implantation of internal device, temporomandibular joint [TMJ]	1EL53	-	-	-	1EL53PNQD retired in v2012 1EL53PNQDA retired in v2012 1EL53PNQDK retired in v2012 1EL53PNQDN retired in v2012 1EL53PNQDQ retired in v2012 1EL53PNQE retired in v2012 1EL53PNQEA retired in v2012 1EL53PNQEK retired in v2012 1EL53PNQEN retired in v2012 1EL53PNQEQ retired in v2012
Implant internal device, pelvis	1SQ53	-	-	-	No change
Implantation of internal device, shoulder joint	1TA53	-	-	-	1TA53LAPQ added in v2009 1TA53LAPQA added in v2009 1TA53LAPQK added in v2009 1TA53LAPQN added in v2009 1TA53LAPQQ added in v2009
Implantation of internal device, elbow joint	1TM53	-	-	-	No change
Implantation of internal device, wrist joint	1UB53	-	-	-	No change
Implantation of internal device, distal radioulnar joint and carpal joints and bones	1UC53	-	-	-	1UC53LAPN added in v2015 1UC53LAQH added in v2015
Implantation of internal device, other metacarpophalangeal joint(s)	1UG53	-	-	-	1UG53LAPM added in v2015 1UG53LAPMN added in v2015
Implantation of internal device, first metacarpophalangeal joint		1UH53	-	-	1UH53LAPM added in v2015 1UH53LAPMN added in v2015
Implantation of internal device, other interphalangeal joints of hand	1UK53	-	-	-	1UK53LAPN added in v2015 1UK53LAPNN added in v2015 1UK53LAQH added in v2015
Implantation of internal device, first interphalangeal joint of hand		1UM53	-	-	1UM53LAPN added in v2015 1UM53LAPNN added in v2015 1UM53LAQH added in v2015
Implant internal device, hip joint	1VA53	-	-	-	1VA53LLPM added in v2015 1VA53LLPMA added in v2015 1VA53LLPMK added in v2015 1VA53LLPMN added in v2015 1VA53LLPMQ added in v2015 1VA53LLPN added in v2015 1VA53LLPNA added in v2015 1VA53LLPNK added in v2015 1VA53LLPNN added in v2015 1VA53LLPNQ added in v2015 1VA53LLSLN added in v2015

					1VA53PNPM retired in v2012 1VA53PNPMA retired in v2012 1VA53PNPMK retired in v2012 1VA53PNPMN retired in v2012 1VA53PNPMQ retired in v2012 1VA53PNPN retired in v2012 1VA53PNPNA retired in v2012 1VA53PNPNK retired in v2012 1VA53PNPNN retired in v2012 1VA53PNPNQ retired in v2012
Implant sing comp prosth hip OA	1VA53L APM	-	-	-	No change
Implant sing comp prosth hip OA &autogr	1VA53L APMA	-	-	-	No change
Implant sing comp prosth hip OA &synth mat	1VA53L APMN	-	-	-	No change
Implant sing comp prosth hip OA &combo tis	1VA53L APMQ	-	-	-	No change
Implant dual comp prosth hip OA	1VA53L APN	-	-	-	No change
Implant dual comp prosth hip OA &autogr	1VA53L APNA	-	-	-	No change
Implant dual comp prosth hip OA &homogr	1VA53L APNK	-	-	-	No change
Implant dual comp prosth hip OA &synth mat	1VA53L APNN	-	-	-	No change
Implant dual comp prosth hip OA &combo tis	1VA53L APNQ	-	-	-	No change
Implant dev hip OA &spacer synth mater	1VA53L ASLN	-	-	-	No change
Implant internal device, knee joint	1VG53	-	-	-	No change
Implant sing comp prosth knee OA	1VG53L APM	-	-	-	No change
Implant sing comp prosth knee OA &syn mat	1VG53L APMN	-	-	-	No change
Implant sing comp prosth knee OA &comb tis	1VG53L APMQ	-	-	-	No change
Implant dual comp prosth knee OA	1VG53L APN	-	-	-	No change
Implant dual comp prosth knee OA &autogr	1VG53L APNA	-	-	-	No change
Implant dual comp prosth knee OA &synth mat	1VG53L APNN	-	-	-	No change
Implant dual comp prosth knee OA &comb tis	1VG53L APNQ	-	-	-	No change
Implant tri comp prosth knee OA	1VG53L APP	-	-	-	No change
Implant tri comp prosth knee OA &autogr	1VG53L APPA	-	-	-	No change
Implant tri comp prosth knee OA &homogr	1VG53L APPK	-	-	-	No change

Implant tri comp prosth knee OA & synth mat	1VG53L APPN	-	-	-	No change
Implant tri comp prosth knee OA & comb tis	1VG53L APPQ	-	-	-	No change
Implant cement spacer knee OA	1VG53L ASLN	-	-	-	No change
Implant internal device, patella	1VP53	-	-	-	1VP53LAPN added in v2012 1VP53LAPNN added in v2012
Implant dev patella OA & prosthesis synth mater	1VP53L APMN	-	-	-	No change
Implantation of internal device, ankle joint	1WA53	-	-	-	1WA53LAPM added in v2015 1WA53LAPMA added in v2015 1WA53LAPMK added in v2015 1WA53LAPMN added in v2015 1WA53LAPMQ added in v2015 1WA53LAPP added in v2015 1WA53LAPPA added in v2015 1WA53LAPPK added in v2015 1WA53LAPPN added in v2015 1WA53LAPPQ added in v2015
Implantation of internal device, tarsal bones and intertarsal joints [hindfoot, midfoot]	1WE53	-	-	-	No change
Implantation of internal device, first metatarsal bone and first metatarsophalangeal joint		1WI53	-	-	1WI53LAPN added in v2015 1WI53LAPNN added in v2015
Implantation of internal device, tarsometatarsal joints, other metatarsal bones and other metatarsophalangeal joints [forefoot]	1WJ53	-	-	-	No change
Implantation of internal device, other interphalangeal joints of toe	1WM53	-	-	-	No change
Implantation of internal device, first interphalangeal joint of toe		1WN53	-	-	No change

CMG/CACS Codes

CASE MIX GROUPING (CMG) / Comprehensive Ambulatory Classification System (CACS)					
CMG/CACS DESCRIPTION	CMG/CACS CODES				NOTES
	v2006	v2009	v2012	v2015	
<i>CMG CODES</i>					
Bilateral Hip/Knee Replacement		315	-	-	CMG+ 315 was introduced in F2007/08. Therefore use this code for data extraction.
Revised Hip Replacement with Infection		316	-	-	CMG+ 316 was introduced in F2007/08. Therefore use this code for data extraction.
Revised Hip Replacement without Infection		317	-	-	CMG+ 317 was introduced in F2007/08. Therefore use this code for data extraction.
Revised Knee Replacement with Infection		318	-	-	CMG+ 318 was introduced in F2007/08. Therefore use

					this code for data extraction.
Revised Knee Replacement without Infection		319	-	-	CMG+ 319 was introduced in F2007/08. Therefore use this code for data extraction.
Unilateral Hip Replacement		320	-	-	CMG+ 320 was introduced in F2007/08. Therefore use this code for data extraction.
Unilateral Knee Replacement		321	-	-	CMG+ 321 was introduced in F2007/08. Therefore use this code for data extraction.
Shoulder Replacement		326	-	-	CMG+ 326 was introduced in F2007/08. Therefore use this code for data extraction.
Other Joint Replacement		327	-	-	CMG+ 327 was introduced in F2007/08. Therefore use this code for data extraction.
Replacement/Fixation/Repair of Tibia/Fibula/Knee		729	-	-	CMG+ 729 was introduced in F2007/08. Therefore use this code for data extraction.
<i>CACS CODES</i>					
Joint Replacement				C325	CACS C325 was introduced in F2013/14. Therefore use this code for data extraction.

Appendix B

B1. Quality measures for hip knee replacements measures by Ontario Ministry of Health and Long-term Care



Provincial Orthopaedic Quality Scorecard - HIP Replacement Surgery Reporting Period - Q2 FY 13/14				
PROVINCE				
Quality Dimensions	Indicator	Current Performance	Target	% Completed Within Target
Efficiency	Average length of stay (days) - All patients	3.9	-	-
	Average length of stay (days) - Patients Discharged Home ¹	3.4	4.4	88.0
	90th percentile for 7 day length of stay - Patients Discharged Home ¹	97.8	90%	97.8
	Proportion of Patients Discharged Home	89.2	90% ±9%	89.2
Effectiveness/Safety	Rate of Readmission within 30 days after primary joint replacement	3.4	-	-
	Rate of Revisions within 365 days after primary joint replacement	1.6	-	-
Accessibility	Replacement Wait Time (90th Percentile Days)	190	182	-

Provincial Orthopaedic Quality Scorecard - KNEE Replacement Surgery Reporting Period - Q2 FY 13/14				
PROVINCE				
Quality Dimensions	Indicator	Current Performance	Target	% Completed Within Target
Efficiency	Average length of stay (days) - All patients	3.5	-	-
	Average length of stay (days) - Patients Discharged Home ¹	3.3	4.4	89.5
	90th percentile for 7 day length of stay - Patients Discharged Home ¹	98.1	90%	98.1
	Proportion of Patients Discharged Home	92.2	90% ±9%	92.2
Effectiveness/Safety	Rate of Readmission within 30 days after primary joint replacement	3.2	-	-
	Rate of Revisions within 365 days after primary joint replacement	1.1	-	-
Accessibility	Replacement Wait Time (90th Percentile Days)	225	182	-

Note:

¹ Discharge destination Home includes Home Care, Senior's Care, Attendant Care (Discharge type = 04, 05)
Please refer to indicator definitions worksheet for full description

B2. French quality for hip knee replacements adopted in Article 51 reform in 2019



Synthèse des indicateurs
Expérimentation d'un paiement à l'épisode de soins
Prothèse totale de hanche et prothèse totale du genou

Intitulé	Nature	Objectifs	Utilisation dans l'expérimentation
% des patients EDS pour lesquels une évaluation pré opératoire du statut nutritionnel a été réalisée	Processus	Qualité de la prise en charge pré opératoire et action en vue de faciliter la récupération	Suivi / Evaluation
% des patients EDS dénutris pour lesquels une action correctrice pré opératoire a été mise en place	Processus		Financement 2 nd temps
% des patients EDS pour lesquels une évaluation pré opératoire de l'anémie et du statut martial a été réalisée	Processus		
% des patients EDS anémiés pour lesquels une actions correctrice pré opératoire a été mise en place	Processus		
Profil de risque des patients en entrée	Processus	Evolution du case mix durant l'expérimentation	Suivi / Evaluation
Nb moyen d'actes paramédicaux pré séjour index 3 indicateurs à décliner pour Infirmier d'état, Masseur kiné. Et Médecin physique et réadaptation	Processus	Evolution de la prise en charge pré opératoire du patient	Suivi / Evaluation
Nb moyen d'actes paramédicaux post séjour index (et écart entre prescrit et observé) 3 indicateurs à décliner pour Infirmier d'état, Masseur kiné. Et Médecin physique et réadaptation	Processus	Evolution de la prise en charge pré opératoire du patient et impact de la coordination sur la prise en charge	Suivi / Evaluation
Mode de sortie des patients	Résultat	Suivi du modèle de prédiction des parcours et identification pour étude des écarts	Suivi / Evaluation
Adéquation entre le mode de sortie initial prévu en pré-opératoire et le mode de sortie réalisé	Processus	Suivi de l'impact du séjour initial sur la prise en charge du patient dans l'EDS et test du modèle	Suivi / Evaluation
Part des patients pour lesquels le mode de sortie prescrit lors du séjour Index est conforme au mode de sortie observé dans les 7 jours suivant la sortie du séjour Index	Processus	Tester l'évolution des pratiques, de la coordination et de la pertinence	Suivi / Evaluation
ETE ORTHO (cliquer ici) ISO ORTHO (cliquer ici)	Résultat	Suivi de l'impact des pratiques et organisations sur la qualité et la sécurité des soins	Financement Evaluation
% de séjours index avec au moins une complication	Résultat		Evaluation

Intitulé	Nature	Objectifs	Utilisation dans l'expérimentation
DMS de l'ensemble des séjours MCO	Résultat	Impact d'EDS sur la prise en charge des patients	Suivi / Evaluation
% de séjours avec au moins une complication médicale en lien avec réhospitalisation	Résultats	Impact sur la qualité et la sécurité des soins	Suivi / Evaluation
% de séjours avec au moins une complication chirurgicale en lien avec réhospitalisation	Résultats		
Questionnaire d'expérience patient EDS (PREMs) : -Année 2 : Taux de collecte des mails patients de l'EDS -Année 3 et suivant : Résultats du PREMs	Résultats	Prise en compte du retour patient et impact d'EDS sur sa perception	Financement Evaluation
PROMs : EQ-5D - % d'administration du questionnaire (PTH) HOOS PS - % d'administration du questionnaire (PTG) KOOS PS - % d'administration du questionnaire	Résultats	Prise en compte du ressenti du patient	Suivi Et financement
Année 2 et 3 : Taux d'exhaustivité de collecte des données identifiées pour calculer les indicateurs de l'expérimentation	Activité	Suivi, l'évaluation de l'expérimentation sur la base de données fiables et complètes	Financement
% de patients inclus dans l'expérimentation / % des patients de l'établissement relevant du périmètre EDS	Activité	Suivi de l'activité des porteurs de projet	Suivi
% de patients relevant du mécanisme de sauvegarde	Activité	Suivi des profils atypiques	Suivi
ETP consacré à la coordination <i>Nb d'ETP dédié à la coordination des épisodes de soins / Nb d'épisodes de soins</i>	Ressources	Qualité de la prise en charge globale du patient pré, per et post intervention	Suivi
ETP dédié à la collecte des données pour l'expérimentation EDS	Ressources	Suivi de la charge liée à l'expérimentation	Suivi

B3. 2017 Orthopaedic Preferred Specialty Measure Set by American Academy of Orthopaedic Surgeons (AAOS)

2017 Orthopaedic Preferred Specialty Measure Set

Quality #	NQF #	Data Submission Method	Measure Type	High Priority?	National Quality Strategy Domain	Measure Title	Measure Description
General Orthopaedic Measures							
024	0045	Claims, Registry	Process	Yes	Communication and Care Coordination	Communication with the Physician or Other Clinician Managing On-going Care Post-Fracture for Men and Women Aged 50 Years and Older	Percentage of patients aged 50 years and older treated for a fracture with documentation of communication, between the physician treating the fracture and the physician or other clinician managing the patient's on-going care, that a fracture occurred and that the patient was or should be considered for osteoporosis treatment or testing. This measure is reported by the physician who treats the fracture and who therefore is held accountable for the communication
046	0097	Claims, CMS Web Interface, Registry	Process	Yes	Communication and Care Coordination	Medication Reconciliation Post-Discharge	The percentage of discharges from any inpatient facility (e.g. hospital, skilled nursing facility, or rehabilitation facility) for patients 18 years and older of age seen within 30 days following discharge in the office by the physician, prescribing practitioner, registered nurse, or clinical pharmacist providing on-going care for whom the discharge medication list was reconciled with the current medication list in the outpatient medical record. This measure is reported as three rates stratified by age group: • Reporting Criteria 1: 18-64 years of age • Reporting Criteria 2: 65 years and older • Total Rate: All patients 18 years of age and older
047	0326	Claims, Registry	Process	Yes	Communication and Care Coordination	Care Plan	Percentage of patients aged 65 years and older who have an advance care plan or surrogate decision maker documented in the medical record or documentation in the medical record that an advance care plan was discussed but the patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan
109	N/A	Claims, Registry	Process	Yes	Person and Caregiver-Centered Experience and Outcomes	Osteoarthritis (OA): Function and Pain Assessment	Percentage of patient visits for patients aged 21 years and older with a diagnosis of osteoarthritis (OA) with assessment for function and pain
128	0421	Claims, EHR, CMS Web Interface, Registry	Process	No	Community / Population Health	Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up Plan	Percentage of patients aged 18 years and older with a BMI documented during the current encounter or during the previous six months AND with a BMI outside of normal parameters, a follow-up plan is documented during the encounter or during the previous six months of the current encounter Normal Parameters: Age 18 years and older BMI => 18.5 and < 25 kg/m2
130	0419	Claims, EHR, Registry	Process	Yes	Patient Safety	Documentation of Current Medications in the Medical Record	Percentage of visits for patients aged 18 years and older for which the eligible professional attests to documenting a list of current medications using all immediate resources available on the date of the encounter. This list must include ALL known prescriptions, over-the-counters, herbals, and vitamin/mineral/dietary (nutritional) supplements AND must contain the medications' name, dosage, frequency and route of administration.
131	0420	Claims, Registry	Process	Yes	Communication and Care Coordination	Pain Assessment and Follow-Up	Percentage of visits for patients aged 18 years and older with documentation of a pain assessment using a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present

2017 Orthopaedic Preferred Specialty Measure Set

Quality #	NQF #	Data Submission Method	Measure Type	High Priority?	National Quality Strategy Domain	Measure Title	Measure Description
134	0418	Claims, EHR CMS Web Interface, Registry	Process	No	Community / Population Health	Preventive Care and Screening: Screening for Clinical Depression and Follow-Up Plan	Percentage of patients aged 12 years and older screened for depression on the date of the encounter using an age appropriate standardized depression screening tool AND if positive, a follow-up plan is documented on the date of the positive screen
154	0101	Claims, Registry	Process	Yes	Patient Safety	Falls: Risk Assessment	Percentage of patients aged 65 years and older with a history of falls that had a risk assessment for falls completed within 12 months
155	0101	Claims, Registry	Process	Yes	Communication and Care Coordination	Falls: Plan of Care	Percentage of patients aged 65 years and older with a history of falls that had a plan of care for falls documented within 12 months
178	N/A	Registry	Process	No	Effective Clinical Care	Rheumatoid Arthritis (RA): Functional Status Assessment	Percentage of patients aged 18 years and older with a diagnosis of rheumatoid arthritis (RA) for whom a functional status assessment was performed at least once within 12 months
179	N/A	Registry	Process	No	Effective Clinical Care	Rheumatoid Arthritis (RA): Assessment and Classification of Disease Prognosis	Percentage of patients aged 18 years and older with a diagnosis of rheumatoid arthritis (RA) who have an assessment and classification of disease prognosis at least once within 12 months
180	N/A	Registry	Process	No	Effective Clinical Care	Rheumatoid Arthritis (RA): Glucocorticoid Management	Percentage of patients aged 18 years and older with a diagnosis of rheumatoid arthritis (RA) who have been assessed for glucocorticoid use and, for those on prolonged doses of prednisone \geq 10 mg daily (or equivalent) with improvement or no change in disease activity, documentation of glucocorticoid management plan within 12 months
226	0028	Claims, EHR CMS Web Interface, Registry	Process	No	Community / Population Health	Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention	Percentage of patients aged 18 years and older who were screened for tobacco use one or more times within 24 months AND who received cessation counseling intervention if identified as a tobacco user
318	0101	EHR, CMS Web Interface	Process	Yes	Patient Safety	Falls: Screening for Future Fall Risk	Percentage of patients 65 years of age and older who were screened for future fall risk during the measurement period.
358	N/A	Registry	Process	Yes	Person and Caregiver-Centered Experience and Outcomes	Patient-Centered Surgical Risk Assessment and Communication	Percentage of patients who underwent a non-emergency surgery who had their personalized risks of postoperative complications assessed by their surgical team prior to surgery using a clinical data-based, patient-specific risk calculator and who received personal discussion of those risks with the surgeon
374	N/A	EHR	Process	Yes	Communication and Care Coordination	Closing the Referral Loop: Receipt of Specialist Report	Percentage of patients with referrals, regardless of age, for which the referring provider receives a report from the provider to whom the patient was referred
408	N/A	Registry	Process	No	Effective Clinical Care	Opioid Therapy Follow-up Evaluation	All patients 18 and older prescribed opiates for longer than six weeks duration who had a follow-up evaluation conducted at least every three months during Opioid Therapy documented in the medical record
412	N/A	Registry	Process	No	Effective Clinical Care	Documentation of Signed Opioid Treatment Agreement	All patients 18 and older prescribed opiates for longer than six weeks duration who signed an opioid treatment agreement at least once during Opioid Therapy documented in the medical record.

2017 Orthopaedic Preferred Specialty Measure Set

Quality #	NQF #	Data Submission Method	Measure Type	High Priority?	National Quality Strategy Domain	Measure Title	Measure Description
414	N/A	Registry	Process	No	Effective Clinical Care	Evaluation or Interview for Risk of Opioid Misuse	All patients 18 and older prescribed opiates for longer than six weeks duration evaluated for risk of opioid misuse using a brief validated instrument (e.g. Opioid Risk Tool, SOAPP-R) or patient interview documented at least once during Opioid Therapy in the medical record
418	0053	Claims, Registry	Process	No	Effective Clinical Care	Osteoporosis Management in Women Who Had a Fracture	The percentage of women age 50-85 who suffered a fracture and who either had a bone mineral density test or received a prescription for a drug to treat osteoporosis in the six months after the fracture
458	1789	Administrative Claims	Outcome	No	Communication and Care Coordination	All-cause Hospital Readmission	The 30-day All-Cause Hospital Readmission measure is a risk-standardized readmission rate for beneficiaries age 65 or older who were hospitalized at a short-stay acute care hospital and experienced an unplanned readmission for any cause to an acute care hospital within 30 days of discharge.
Specialty Specific Measures - Hip							
021	0268	Claims, Registry	Process	Yes	Patient Safety	Perioperative Care: Selection of Prophylactic Antibiotic – First OR Second Generation Cephalosporin	Percentage of surgical patients aged 18 years and older undergoing procedures with the indications for a first OR second generation cephalosporin prophylactic antibiotic who had an order for a first OR second generation cephalosporin for antimicrobial prophylaxis
023	0239	Claims, Registry	Process	Yes	Patient Safety	Perioperative Care*: Venous Thromboembolism (VTE) Prophylaxis (When Indicated in ALL Patients)	Percentage of surgical patients aged 18 years and older undergoing procedures for which venous thromboembolism (VTE) prophylaxis is indicated in all patients, who had an order for Low Molecular Weight Heparin (LMWH), Low- Dose Unfractionated Heparin (LDUH), adjusted-dose warfarin, fondaparinux or mechanical prophylaxis to be given within 24 hours prior to incision time or within 24 hours after surgery end time
376	N/A	EHR	Process	Yes	Person and Caregiver-Centered Experience and Outcomes	Functional Status Assessment for Total Hip Replacement	Percentage of patients 18 years of age and older with primary total hip arthroplasty (THA) who completed baseline and follow-up patient-reported functional status assessments

2017 Orthopaedic Preferred Specialty Measure Set

Quality #	NQF #	Data Submission Method	Measure Type	High Priority?	National Quality Strategy Domain	Measure Title	Measure Description
Specialty Specific Measures - Knee							
021	0268	Claims, Registry	Process	Yes	Patient Safety	Perioperative Care: Selection of Prophylactic Antibiotic – First OR Second Generation Cephalosporin	Percentage of surgical patients aged 18 years and older undergoing procedures with the indications for a first OR second generation cephalosporin prophylactic antibiotic who had an order for a first OR second generation cephalosporin for antimicrobial prophylaxis
023	0239	Claims, Registry	Process	Yes	Patient Safety	Perioperative Care*: Venous Thromboembolism (VTE) Prophylaxis (When Indicated in ALL Patients)	Percentage of surgical patients aged 18 years and older undergoing procedures for which venous thromboembolism (VTE) prophylaxis is indicated in all patients, who had an order for Low Molecular Weight Heparin (LMWH), Low- Dose Unfractionated Heparin (LDUH), adjusted-dose warfarin, fondaparinux or mechanical prophylaxis to be given within 24 hours prior to incision time or within 24 hours after surgery end time
350	N/A	Registry	Process	Yes	Communication and Care Coordination	Total Knee Replacement: Shared Decision-Making: Trial of Conservative (Non-surgical) Therapy	Percentage of patients regardless of age undergoing a total knee replacement with documented shared decision-making with discussion of conservative (non-surgical) therapy (e.g., non-steroidal anti-inflammatory drug (NSAIDs), analgesics, weight loss, exercise, injections) prior to the procedure
351	N/A	Registry	Process	Yes	Patient Safety	Total Knee Replacement: Venous Thromboembolic and Cardiovascular Risk Evaluation	Percentage of patients regardless of age undergoing a total knee replacement who are evaluated for the presence or absence of venous thromboembolic and cardiovascular risk factors within 30 days prior to the procedure (e.g. history of Deep Vein Thrombosis (DVT), Pulmonary Embolism (PE), Myocardial Infarction (MI), Arrhythmia and Stroke)
352	N/A	Registry	Process	Yes	Patient Safety	Total Knee Replacement: Preoperative Antibiotic Infusion with Proximal Tourniquet	Percentage of patients regardless of age undergoing a total knee replacement who had the prophylactic antibiotic completely infused prior to the inflation of the proximal tourniquet
353	N/A	Registry	Process	Yes	Patient Safety	Total Knee Replacement: Identification of Implanted Prosthesis in Operative Report	Percentage of patients regardless of age undergoing a total knee replacement whose operative report identifies the prosthetic implant specifications including the prosthetic implant manufacturer, the brand name of the prosthetic implant and the size of each prosthetic implant
375	N/A	EHR	Process	Yes	Person and Caregiver-Centered Experience and Outcomes	Functional Status Assessment for Total Knee Replacement	Percentage of patients 18 years of age and older with primary total knee arthroplasty (TKA) who completed baseline and follow-up patient-reported functional status assessments

*The PMC acknowledges the past controversy around the use of aspirin for DVT prophylaxis. Quality #023 does not specify the use of aspirin, however, the use of “mechanical prophylaxis” is specified in the numerator of the measure specification. Because aspirin is usually given in combination with mechanical prophylaxis the PMC deemed the measure appropriate to include in the OPS Set.

Appendix C. Recommendations of Ontario Health Technology Advisory Committee (OHTAC) regarding hip and knee replacements in 2013

Full text with references to supporting clinical evidence can be found in QBP unilateral hip and knee replacement manual (2013):

http://www.health.gov.on.ca/en/pro/programs/ecfa/funding/hs_funding_qbp.aspx

OHTAC Episode of Care Recommended Practices:

- Referral from Primary Care
 - The referring practitioner should provide standard radiograph investigations of the affected joints.
 - Pre-consultation MRIs are rarely indicated and should not be routinely ordered.
 - The primary care provider (PCP) should make the referral for surgery consultation and be the coordinator of patient care.
 - Self-referral should be considered for patients who do not have a PCP.
 - Referrals should be made through a standardized template that includes the reason for referral, radiographs of the affected joint(s), and relevant patient comorbidities.
- Coordinated Intake and Assessment
 - Hip and knee referrals should be managed through a coordinated intake and assessment process.
 - Patient assessments should be completed by an appropriate health care practitioner qualified and trained to assess patients and to make decisions regarding the appropriateness of surgeon consultation or surgery.
 - Every patient scheduled to undergo joint replacement should receive a functional assessment.
- Decision to Treat Clinical Assessment Node
 - Surgical patients need to be assessed by a surgeon to make the final decision regarding appropriateness for surgery.

- The risks and benefits of surgery should be explained to the patient, and the patient should be charged with the decision whether or not to proceed with surgery.
 - If it is determined that surgery is not appropriate for a patient, the coordinated intake should provide “outbound” care back to the appropriate health care practitioner.
 - The coordinated intake process should ensure that non-surgical options are explained to the patient.
 - Results of the assessment and plan for treatment should be communicated back to the patient’s PCP.
- Preparation for Surgery
 - Preparation for surgery should occur with adequate time before surgery to address modifiable patient risk factors.
 - Patients should receive education addressing the entire continuum of care.
 - Discharge planning should begin at the time of the decision to treat.
 - Lifestyle or behaviour modification may be necessary before surgery to optimize the benefit and reduce the risks of surgery
 - Smoking cessation counselling prior to surgery should be recommended for people who smoke.
 - Weight loss counselling prior to surgery should be recommended for obese and morbidly obese people.
 - Exercise should be recommended, as tolerated, in preparation for hospital admission if indicated by lifestyle risk factors.
 - The following OHTAC recommendation should be considered on preoperative physiotherapy exercise
 - Pre-Admission Screening
 - Pre-admission screenings should be conducted in an appropriate time frame before surgery to avoid empty operating room time due to late cancellations
 - A multi-disciplinary team is necessary to optimize patient preparation for surgery.
 - Patients should be medically optimized before elective surgery.

- Specific investigations for medical preparation need to follow evidence-based best practices.
- A multidisciplinary blood management program adaptable to individualized patient needs should be implemented.
- The Hip and Knee Expert Panel suggest the use of tranexamic acid for prevention of blood loss. Because the use of tranexamic acid is off-label, the decision should rest with the Pharmacy and Therapeutics committee of the hospital.
- Admission and Preoperative Management
 - Hospitals should use a structured clinical care pathway
- Surgery
 - The World Health Organization (WHO) surgical safety checklist, in addition to other surgical safety tools and supports, should be referenced prior to surgery.
 - The choice of anesthesia should involve the anesthesiologist and surgeon, as well as patient preference.
 - Neuraxial anesthesia is recommended when appropriate.
 - Individual hospitals should develop and implement an implant matching program, where appropriate prostheses are determined based on best available, current evidence applied to individual patient characteristics.
 - Evidence of clinical effectiveness should be held to national and international standards
 - a) The benchmark set by the National Institute for Clinical Excellence (NICE) for primary total hip arthroplasty prosthesis selection is a revision rate of 10% or less at 10 years.
 - b) Prosthesis selection should also take into consideration patient characteristics, surgeon recommendations, cost effectiveness, and the ability to maximize early rehabilitation potential.
 - If metal-on-metal (MOM) hip resurfacing arthroplasty (HRA) is to be used, the following OHTAC recommendation should be adhered to:
 - a) Metal-on-metal HRA is a reasonable treatment option for osteoarthritis patients who meet appropriate criteria. Expert opinion informed that the appropriate criteria for patient selection are: male patients under 60 years

of age with osteoarthritis, good bone quality, no significant acetabular deformity, and a large diameter femoral head to accommodate a femoral component of 50 mm or larger. Selection of female patients for this procedure requires very careful consideration.

- b) Metal-on-metal HRA should only be performed by surgeons who have appropriate training and who have acquired a high level of experience by performing a high annual volume of THAs and MOM HRAs. Expert opinion, informed that the appropriate volume is considered to be performing at least 100 THAs and at least 20 HRAs per year.
 - c) There is evidence of increased cobalt and chromium levels in the blood and urine of patients who receive MOM HRA; however, there is no conclusive evidence that exposure to high metal ion levels has harmful biological consequences. As such, OHTAC recommends that patients receiving these implants be informed of the potential for exposure to metal ions, and that the adverse effects and long-term implications of elevated metal ion exposure in patients who receive these implants are not known at this time.
 - d) Since cobalt and chromium can pass the placental barrier, OHTAC recommends that non-MOM-bearing surfaces be used in women of childbearing ages who require hip arthroplasty.
- When bilateral joint replacements are required, they can be performed sequentially under the same anesthetic or staged over two separate hospitalizations.
 - a) The treatment decision should be at the surgeon's discretion.
 - b) The potential increased risk of mortality and pulmonary embolism associated with simultaneous bilateral replacements needs to be recognized, and appropriate patient selection and rationale should be applied.
 - The decision to use cemented or cementless fixation should be at the surgeon's discretion.
 - There is insufficient evidence to make a recommendation for or against the use of Antibiotic-laden Bone Cement (ALBC) for primary joint replacement.
 - Routine antibiotic administration is recommended as a prophylaxis against infection. It is recommended that patients receive 1 dose of antibiotic preoperatively and 3 subsequent doses postoperatively over the course of 24 hours.
 - The use of chlorhexidine for surgical site infection prevention should follow the Institute for Healthcare Improvement enhanced surgical practice recommenda-

tions.

- Venous thromboembolism (VTE) prevention is recommended. Care providers should consider following the American College of CHEST Physicians guidelines on the prevention of VTE in orthopedic surgery patients.
- Postoperative Care
 - A multimodal approach to postoperative pain management should be employed. This may include systemic analgesics (both non-opioid and opioid), nerve blocks (peripheral or neuraxial), and/or local infiltration analgesia (LIA).
 - Early postoperative mobilization is recommended. There should be input from a multidisciplinary rehabilitation team and a structured mobilization plan for postoperative rehabilitation.
 - The optimal intensity of rehabilitation during the acute hospitalization period is unknown.
 - Continuous passive motion is not recommended.
- Post-Acute Care: Inpatient Rehabilitation, Home Care Rehabilitation, and Outpatient Rehabilitation
 - Rehabilitation is required for successful recovery of patients after hip or knee replacement surgery.
 - OHTAC recommends the health system support the move towards community-based physiotherapy after primary total knee or hip replacement and discharge from acute care. In regards to location of physiotherapy within the community, the health system should allow for flexibility, depending on the local care context and patients' needs. Current initiatives that are underway in the province to improve allocation of physiotherapy services for primary hip and knee replacement patients should be supported by the health care system.
 - All patients discharged home should be provided an independent home exercise program.
 - For patients who could attend an outpatient physiotherapy clinic, consideration may be given to a self-managed home exercise program with a physiotherapist monitoring through phone calls.
 - Patients should have access to the Community Care Access Centres (CCACs) for assessment of eligibility for supportive services. CCAC eligibility algorithms should be standardized across the province

- Inpatient rehabilitation should be restricted to patients who meet specific eligibility criteria. Eligibility criteria for inpatient rehabilitation should be standardized.
- There is insufficient evidence to make a recommendation regarding the restricting of high-impact activities.
- Patients should have follow-up appointments with their surgical team after primary hip or knee replacement.

Appendix C. Timeline and funding share of QBP and HBAM in Ontario 2011/12 – 2016/17 (from Palmer et al., 2018a)

Funding Components	2011/12	2012/13	2013/14
<i>Added QBP procedures</i>	Pre-reform, no procedures	primary unilateral hip replacement, primary unilateral knee replacement, unilateral cataract, chronic kidney disease	chronic obstructive pulmonary disease, congestive heart failure stroke, non-cardiac vascular surgery, systemic chemotherapy, gastrointestinal endoscopy systemic chemotherapy, gastrointestinal endoscopy
<i>QBPs intended funding in %</i>	0	6	15
<i>QBPs actual funding in %</i>	0	6	12
<i>HBAM intended funding in %</i>	1,5	40	40
<i>HBAM actual funding in %</i>	0	34	34
<i>Global budget intended funding in %</i>	98,5	54	45
<i>Global budget actual funding in %</i>	100	60	54

Funding Components	2014/15	2015/16	2016/17
<i>Added QBP procedures</i>	hip fracture pneumonia tonsillectomy neonatal jaundice bilateral hip and knee replacement	knee arthroscopy cancer surgery (prostate, colorectal)	cancer surgery (breast, thyroid) non-routine and bilateral cataract
<i>QBPs intended funding in %</i>	30	30	30
<i>QBPs actual funding in %</i>	13	14	15
<i>HBAM intended funding in %</i>	40	40	40
<i>HBAM actual funding in %</i>	33	32	32
<i>Global budget intended funding in %</i>	30	30	30
<i>Global budget actual funding in %</i>	54	54	54

Appendix D. Difference-in-difference regression tables (patient-level)

Table 2.12 – Difference-in-difference estimates for unilateral hip replacements (quality measures)

Outcome	Acute LOS	Acute LOS	Acute LOS	Revision(180d)	Revision(180d)	Revision(180d)	Readm.(180d)	Readm.(180d)	Readm.(180d)
Dependent var./Model	1	2	3	4	5	6	7	8	9
<i>Intercept</i>	6.24 (0.37)***	4.621 (0.479)***	2.545 (0.578)***	-0.004 (0.007)	-0.017 (0.007)*	-0.01 (0.016)	0.001 (0.004)	-0.01 (0.005)*	0.001 (0.013)
<i>post2012 · QBP</i>	-0.3 (0.084)***	-0.266 (0.086)**	-0.273 (0.101)**	-0.002 (0.002)	-0.003 (0.002)	-0.001 (0.002)	-0.003 (0.001)*	-0.002 (0.001)*	< 0.001 (0.001)
<i>post2014 · QBP</i>	-0.359 (0.079)***	-0.35 (0.08)***	-0.132 (0.114)	0.001 (0.002)	< 0.001 (0.002)	-0.003 (0.002)	0.001 (0.001)	0.001 (0.001)	-0.002 (0.002)
<i>age</i>	-0.08 (0.011)***	-0.076 (0.011)***	-0.045 (0.013)***	0.001 (0.001)***	0.002 (0.001)***	0.002 (0.001)***	< 0.001 (0.001)**	< 0.001 (0.001)***	< 0.001 (0.001)***
<i>age²</i>	0.001 (0.001)***	0.001 (0.001)***	0.001 (0.001)***	< 0.001 (0.001)***	< 0.001 (0.001)***	< 0.001 (0.001)***	< 0.001 (0.001)**	< 0.001 (0.001)***	< 0.001 (0.001)***
<i>male</i>	-0.273 (0.03)***	-0.271 (0.03)***	-0.231 (0.035)***	< 0.001 (0.001)	< 0.001 (0.001)	-0.001 (0.001)	< 0.001 (0.001)	< 0.001 (0.001)	< 0.001 (0.001)
<i>QBP flag</i>	-0.418 (0.052)***			0.002 (0.001)*			-0.001 (0.001)*		
Resp. for payment:									
<i>Worker insurance</i>			-0.386 (0.288)			-0.001 (0.009)			0.005 (0.006)
<i>Other</i>			0.586 (0.791)			-0.018 (0.003)***			-0.009 (0.001)***
Income quintile:									
<i>Lowest</i>			0.515 (0.229)*			-0.01 (0.014)			-0.012 (0.012)
<i>Medium-low</i>			0.305 (0.226)			-0.011 (0.014)			-0.013 (0.012)
<i>Middle</i>			0.311 (0.228)			-0.011 (0.014)			-0.013 (0.012)
<i>Medium-high</i>			0.249 (0.225)			-0.012 (0.014)			-0.013 (0.012)
<i>Highest</i>			0.121 (0.225)			-0.012 (0.014)			-0.013 (0.012)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	NO	NO	YES
<i>Year FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Diagnosis FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Hospital FE</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>N</i>	240954	240954	167322	240971	240971	167322	240971	240971	167322
<i>R2</i>	0.35	0.36	0.36	0.08	0.09	0.1	0.007	0.011	0.011
<i>AdjR2</i>	0.35	0.35	0.36	0.08	0.09	0.1	0.007	0.009	0.009

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Errors are clustered at hospital level.

Table 2.13 – Difference-in-difference estimates for unilateral hip replacements (quality measures, continued)

Dep. var./Model	Outcome N. diag.(post admit)	N. diag.(post admit)	N. diag.(post admit)	Elixhauser index	Elixhauser index	Elixhauser index	N.of interv.	N.of interv.	N.of interv.
	1	2	3	4	5	6	7	8	9
<i>Intercept</i>	0.411 (0.034)***	0.317 (0.044)***	0.255 (0.069)***	0.707 (0.125)***	0.526 (0.205)*	0.136 (0.24)	1.961 (0.036)***	1.915 (0.053)***	1.732 (0.067)***
<i>post2012 · QBP</i>	-0.015 (0.009)	-0.013 (0.009)	-0.013 (0.01)	-0.048 (0.031)	-0.033 (0.031)	-0.059 (0.034)	-0.027 (0.008)***	-0.025 (0.008)**	0.001 (0.009)
<i>post2014 · QBP</i>	0.013 (0.008)	0.007 (0.008)	0.029 (0.011)**	-0.109 (0.029)***	-0.093 (0.029)**	0.015 (0.037)	0.054 (0.007)***	0.012 (0.007)	-0.021 (0.01)*
<i>age</i>	-0.017 (0.001)***	-0.017 (0.001)***	-0.016 (0.001)***	-0.042 (0.004)***	-0.04 (0.004)***	-0.036 (0.005)***	0.003 (0.001)**	0.006 (0.001)***	0.009 (0.001)***
<i>age²</i>	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***	< 0.001 (< 0.001)*	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***
<i>male</i>	-0.014 (0.003)***	-0.012 (0.003)***	-0.01 (0.004)**	0.036 (0.011)***	0.037 (0.011)***	0.041 (0.013)**	-0.005 (0.003)	-0.003 (0.003)	< 0.001 (0.003)
<i>QBP flag</i>	0.037 (0.005)***			0.136 (0.019)***			0.068 (0.005)***		
Resp. for payment:									
<i>Worker insurance</i>			-0.031 (0.025)			-0.082 (0.098)			-0.073 (0.029)*
<i>other</i>			0.178 (0.148)			-0.122 (0.271)			-0.059 (0.08)
Income quintile:									
<i>Lowest</i>			0.013 (0.048)			0.194 (0.099)*			0.041 (0.034)
<i>Medium-low</i>			-0.002 (0.048)			0.235 (0.098)*			0.03 (0.034)
<i>Middle</i>			-0.002 (0.048)			0.235 (0.098)*			0.029 (0.034)
<i>Medium-high</i>			-0.007 (0.048)			0.288 (0.098)**			0.023 (0.034)
<i>Highest</i>			-0.023 (0.048)			0.313 (0.098)**			0.013 (0.034)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	NO	NO	YES
<i>Year FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Diagnosis FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Hospital FE</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>N</i>	240971	240971	167322	240971	240971	167322	240971	240971	167322
<i>R²</i>	0.2	0.21	0.21	0.22	0.22	0.22	0.37	0.4	0.41
<i>AdjR²</i>	0.2	0.21	0.21	0.22	0.22	0.22	0.37	0.4	0.41

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Errors are clustered at hospital level.

Table 2.14 – Difference-in-difference estimates for unilateral knee replacements (quality measures)

Outcome	Acute LOS	Acute LOS	Acute LOS	Revision(180d)	Revision(180d)	Revision(180d)	Readm.(180d)	Readm.(180d)	Readm.(180d)
Dependent var.	1	2	3	4	5	6	7	8	9
<i>Intercept</i>	8.203 (0.336)***	7.562 (0.344)***	7.318 (0.373)***	0.062 (0.009)***	0.066 (0.01)***	0.048 (0.012)***	0.062 (0.007)***	0.065 (0.008)***	0.042 (0.01)***
<i>post2012 · QBP</i>	-0.307 (0.039)***	-0.271 (0.04)***	-0.326 (0.043)***	0.002 (0.001).	0.002 (0.001).	0.002 (0.001)	0.002 (0.001)*	0.003 (0.001)**	0.002 (0.001).
<i>post2014 · QBP</i>	-0.206 (0.035)***	-0.183 (0.035)***	-0.042 (0.043)	< 0.001 (0.001)	< 0.001 (0.001)	< 0.001 (0.001)	< 0.001 (0.001)	< 0.001 (0.001)	< 0.001 (0.001)
<i>age</i>	-0.158 (0.01)***	-0.149 (0.01)***	-0.138 (0.011)***	< 0.001 (< 0.001)	< 0.001 (< 0.001)	< 0.001 (< 0.001)	-0.001 (< 0.001)***	-0.001 (< 0.001)***	-0.001 (< 0.001)**
<i>age²</i>	0.001 (< 0.001)***	0.001 (< 0.001)***	0.001 (< 0.001)***	< 0.001 (< 0.001)	< 0.001 (< 0.001)	< 0.001 (< 0.001)	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***	< 0.001 (< 0.001)**
<i>male</i>	-0.237 (0.013)***	-0.228 (0.013)***	-0.225 (0.014)***	0.004 (< 0.001)***	0.004 (< 0.001)***	0.004 (< 0.001)***	0.004 (< 0.001)***	0.004 (< 0.001)***	0.004 (< 0.001)***
<i>QBP flag</i>	-0.089 (0.025)***			-0.001 (0.001).			-0.005 (0.001)***		
Respons. for payment:									
<i>Worker insurance</i>			0.104 (0.05)*			0.002 (0.003)			0.001 (0.002)
<i>other</i>			-0.087 (0.172)			-0.012 (0.006).			-0.007 (0.003)*
Income quintile:									
<i>Lowest</i>			0.311 (0.103)**			0.011 (0.003)***			0.001 (0.005)
<i>Medium-low</i>			0.166 (0.103)			0.011 (0.003)***			0.001 (0.005)
<i>Middle</i>			0.136 (0.103)			0.011 (0.003)***			0.001 (0.005)
<i>Medium-high</i>			0.107 (0.102)			0.011 (0.003)***			0.001 (0.005)
<i>High</i>			0.009 (0.103)			0.011 (0.003)***			< 0.001 (0.005)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	NO	NO	YES
<i>Year FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Diagnosis FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Hospital FE</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>N</i>	374165	374165	265620	374317	374317	265620	374317	374317	265620
<i>R2</i>	0.24	0.26	0.25	0.14	0.14	0.15	0.01	0.02	0.01
<i>AdjR2</i>	0.24	0.26	0.25	0.14	0.14	0.15	0.01	0.02	0.01

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Errors are clustered at hospital level.

Table 2.15 – Difference-in-difference estimates for unilateral knee replacements (quality measures, continued)

Dependent var.	Outcome N. diag. (post admit)	N. diag. (post admit)	N. diag. (post admit)	Elixhauser index	Elixhauser index	Elixhauser index	N. of interv.	N. of interv.	N. of interv.
	1	2	3	4	5	6	7	8	9
<i>Intercept</i>	0.527 (0.043)***	0.473 (0.044)***	1.927 (0.168)***	1.025 (0.173)***	1.114 (0.183)***	1.405 (0.26)***	2.622 (0.053)***	2.275 (0.052)***	2.691 (0.071)***
<i>post2012 · QBP</i>	-0.009 (0.006)	-0.009 (0.006)	0.272 (0.018)***	-0.107 (0.023)***	-0.101 (0.023)***	-0.152 (0.025)***	-0.067 (0.006)***	-0.071 (0.005)***	0.025 (0.007)***
<i>post2014 · QBP</i>	0.001 (0.005)	-0.002 (0.005)	-0.077 (0.018)***	-0.072 (0.022)***	-0.04 (0.022)	0.022 (0.025)	0.172 (0.005)***	0.159 (0.005)***	0.111 (0.007)***
<i>age</i>	-0.022 (0.001)***	-0.023 (0.001)***	-0.014 (0.004)***	-0.069 (0.005)***	-0.072 (0.005)***	-0.072 (0.006)***	-0.013 (0.001)***	-0.009 (0.001)***	-0.015 (0.002)***
<i>age²</i>	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***	0.001 (< 0.001)***	0.001 (< 0.001)***	0.001 (< 0.001)***	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***	< 0.001 (< 0.001)***
<i>male</i>	-0.006 (0.002)***	-0.004 (0.002)*	0.007 (0.007)	0.059 (0.008)***	0.055 (0.008)***	0.039 (0.009)***	0.019 (0.002)***	0.022 (0.002)***	0.016 (0.002)***
<i>QBP flag</i>	0.023 (0.003)***			0.113 (0.013)***			0.071 (0.003)***		
Resp. for payment:									
<i>Worker insurance</i>			-0.037 (0.033)			0.044 (0.043)			0.021 (0.014)
<i>other</i>			-0.184 (0.077)*			0.145 (0.094)			-0.019 (0.038)
Income quintile:									
<i>Lowest</i>			0.328 (0.101)**			-0.269 (0.157)			0.045 (0.035)
<i>Medium-low</i>			0.266 (0.101)**			-0.24 (0.157)			0.04 (0.035)
<i>Middle</i>			0.224 (0.101)*			-0.218 (0.157)			0.045 (0.035)
<i>Medium-high</i>			0.195 (0.101)			-0.2 (0.157)			0.053 (0.035)
<i>Highest</i>			0.157 (0.101)			-0.158 (0.157)			0.07 (0.035)*
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	NO	NO	YES
<i>Year FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Diagnosis FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Hospital FE</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>N</i>	374317	374317	265620	374316	374316	265620	374317	374317	265620
<i>R²</i>	0.09	0.1	0.31	0.11	0.13	0.13	0.26	0.36	0.26
<i>AdjR²</i>	0.09	0.1	0.31	0.11	0.13	0.13	0.26	0.36	0.26

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Errors are clustered at hospital level.

Table 2.16 – Difference-in-difference estimates by post-reform year for unilateral replacements (quality measures)

Outcome	Unilateral Hip						Unilateral knee					
	Acute LOS	Readmission(180d)	Revision(180d)	N diag.(post)	Elixhauser index	N.of interv.	Acute LOS	Readmission(180d)	Revision(180d)	N diag.(post)	Elixhauser index	N.of interv.
Dependent var.	1	2	3	4	5	6	7	8	9	10	11	12
<i>qbp · year2012</i>	-0.107 (0.128)	0 (0.002)	-0.002 (0.003)	0 (0.013)	0.037 (0.043)	0.022 (0.012)*	-0.218 (0.054)***	0.002 (0.001)	0 (0.002)	-0.009 (0.009)	-0.125 (0.031)***	0.004 (0.008)
<i>qbp · year2013</i>	-0.438 (0.126)***	0 (0.002)	-0.001 (0.003)	-0.027 (0.013)*	-0.156 (0.044)***	-0.02 (0.012)	-0.401 (0.053)***	0.001 (0.001)	0.003 (0.002)	-0.009 (0.008)	-0.182 (0.032)***	0.04 (0.008)***
<i>qbp · year2014</i>	-0.339 (0.138)**	0 (0.002)	0 (0.003)	-0.002 (0.014)	-0.127 (0.049)***	-0.032 (0.013)**	-0.47 (0.087)***	0.002 (0.001)	0.002 (0.002)	-0.001 (0.008)	-0.152 (0.033)***	0.051 (0.009)***
<i>qbp · year2015</i>	-0.18 (0.191)	-0.006 (0.002)**	-0.007 (0.003)**	0.025 (0.013)*	0.012 (0.047)	-0.023 (0.013)*	-0.331 (0.062)***	0.002 (0.002)	0.003 (0.002)	0.004 (0.008)	-0.141 (0.032)***	0.053 (0.009)***
<i>qbp · year2016</i>	-0.664 (0.171)***	0 (0.002)	-0.005 (0.003)*	0.02 (0.014)	-0.047 (0.046)	0.002 (0.012)	-0.421 (0.058)***	0 (0.002)	0 (0.002)	0.006 (0.008)	-0.123 (0.032)***	0.096 (0.009)***
<i>Individual controls</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Year FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Diagnosis FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Hospital FE</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>N</i>	167322	167322	167322	167322	167322	167322	265620	265620	265620	265620	265620	265620
<i>R2</i>	0.36	0.01	0.1	0.21	0.22	0.41	0.27	0.01	0.15	0.11	0.13	0.38
<i>AdjR2</i>	0.36	0.01	0.1	0.21	0.22	0.41	0.27	0.01	0.15	0.11	0.13	0.38

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Errors are clustered at hospital level.

Table 2.17 – Difference-in-difference estimates by post-reform year for bilateral knee and hip replacements (quality measures)

Outcome	Bilateral Hip				Bilateral knee			
	Acute LOS	N.diag.(post)	Elixhauser index	N.of.interv.	Acute LOS	N.diag.(post)	Elixhauser index	N.of.interv.
Dep.var./Model	1	2	3	4	5	6	7	8
<i>Intercept</i>	2.463 (4.083)	-0.2 (0.479)	0.585 (1.301)	1.136 (0.459)**	6.552 (1.795)***	3.647 (0.933)***	3.647 (0.933)***	1.344 (0.395)***
<i>QBP · year2012</i>	2.934 (1.808)	0.775 (0.354)**	1.072 (0.922)	-0.031 (0.145)	0.606 (0.522)	-0.183 (0.21)	-0.183 (0.21)	-0.065 (0.069)
<i>QBP · year2013</i>	-0.815 (1.017)	0.034 (0.184)	0.215 (0.523)	0.195 (0.15)	-0.029 (0.348)	-0.208 (0.208)	-0.208 (0.208)	0.015 (0.073)
<i>QBP · year2014</i>	-2.797 (1.376)**	-0.072 (0.181)	0.098 (0.556)	-0.005 (0.127)	-0.338 (0.525)	0.036 (0.2)	0.036 (0.2)	0.003 (0.111)
<i>QBP · year2015</i>	-0.931 (0.739)	-0.081 (0.134)	-0.22 (0.425)	-0.01 (0.153)	-0.383 (0.368)	0.036 (0.224)	0.036 (0.224)	0.061 (0.095)
<i>QBP · year2016</i>	-0.254 (0.856)	0.21 (0.192)	0.197 (0.372)	0.019 (0.184)	-0.17 (0.351)	-0.188 (0.187)	-0.188 (0.187)	0.137 (0.057)**
<i>QBP · year2017</i>	-0.721 (1.034)	-0.007 (0.135)	-0.155 (0.391)	-0.1 (0.12)	-0.787 (0.469)*	-0.23 (0.175)	-0.23 (0.175)	0.097 (0.06)
<i>age</i>	0.002 (0.14)	0.001 (0.018)	-0.074 (0.048)	0.015 (0.014)	-0.075 (0.056)	-0.086 (0.03)***	-0.086 (0.03)***	0.027 (0.012)**
<i>age²</i>	0.001 (0.001)	< 0.001 (< 0.001)	0.001 (< 0.001)*	< 0.001 (< 0.001)	0.001 (< 0.001)**	0.001 (< 0.001)***	0.001 (< 0.001)***	< 0.001 (< 0.001)**
<i>male</i>	-0.187 (0.339)	-0.076 (0.053)	-0.226 (0.131)*	0.07 (0.044)	-0.198 (0.088)**	-0.076 (0.048)	-0.076 (0.048)	0.062 (0.019)***
<i>QBP flag</i>	-0.414 (0.58)	0.004 (0.09)	0.117 (0.244)	0.15 (0.083)*	-1.015 (0.23)***	0.22 (0.108)**	0.22 (0.108)**	0.038 (0.037)
<i>Year FE</i>	YES	YES	YES	YES	YES	YES	YES	YES
<i>Diagnosis FE</i>	YES	YES	YES	YES	YES	YES	YES	YES
<i>Hospital FE</i>	NO	NO	NO	NO	NO	NO	NO	NO
<i>N</i>	1225	1227	1227	1227	6310	6313	6313	6313
<i>R2</i>	0.17	0.08	0.06	0.19	0.18	0.24	0.24	0.17
<i>AdjR2</i>	0.16	0.06	0.04	0.17	0.17	0.23	0.23	0.17

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Errors are clustered at hospital level.

Appendix E. Difference-in-difference regression tables (hospital-level)

Table 2.18 – Hospital-level difference-in-difference estimates for unilateral hip and unilateral knee replacements, by year pre- and post-reform years

	QBP year 2008	QBP year 2009	QBP year 2010	QBP year 2011	QBP year 2012	QBP year 2013	QBP year 2014	QBP year 2015	QBP year 2016	QBP year 2017	N	R ²	Adj.R ²				
Hip unilateral:																	
Acute LOS	-0.024	0	(0.285)	(0.286)	-0.133	(0.274)	-0.513	(0.287)*	-0.654	(0.275)**	-0.718	(0.274)***	-0.581	(0.272)**	1128	0.795	0.773
Antibiotic use	-0.001	-0.001	(0.002)	(0.002)	-0.002	(0.002)	-0.004	(0.0287)*	-0.003	(0.002)	-0.003	(0.002)	-0.001	(0.003)	138	0.027	0.586
Discharged home	-0.08	-0.079	(0.039)**	(0.039)**	-0.075	(0.031)	0.047	(0.028)**	0.05	(0.026)**	0.053	(0.026)**	0.036	(0.025)	138	0.596	0.508
Discharged home with support	-0.064	-0.126	(0.069)*	(0.069)*	-0.046	(0.067)	0.07	(0.066)	0.131	(0.063)**	0.106	(0.069)	-0.063	(0.072)	138	0.518	0.466
Discharged support service	0.073	(0.036)**	(0.034)**	(0.029)*	-0.044	(0.022)**	-0.061	(0.021)***	-0.064	(0.021)***	-0.063	(0.021)***	-0.063	(0.021)***	138	0.564	0.517
Hospital death	0	0	(0.001)	(0.001)	0.002	(0.001)	0	(0.001)	0	(0.001)	0.002	(0.001)	0.002	(0.001)	138	0.675	0.64
Diagnostic procedures	-0.003	-0.003	(0.005)	(0.005)	-0.006	(0.005)	-0.008	(0.005)	-0.003	(0.005)	0.002	(0.005)	0.007	(0.007)	138	0.409	0.345
Imaging procedures	0.004	(0.01)	(0.01)	(0.01)	0.003	(0.008)	-0.012	(0.009)	-0.016	(0.009)	-0.007	(0.011)	-0.006	(0.008)	138	0.539	0.489
N of post-admit diag	-0.02	-0.02	(0.037)	(0.039)	-0.028	(0.038)	-0.035	(0.034)	-0.029	(0.034)	-0.015	(0.036)	0.001	(0.033)	138	0.482	0.426
N of interventions	0.043	0.05	(0.054)	(0.059)	-0.005	(0.054)	-0.04	(0.054)	-0.042	(0.06)	0.06	(0.052)	0.071	(0.051)	138	0.587	0.542
Pain symptoms	-0.003	-0.001	(0.003)	(0.004)	0.001	(0.003)	0.003	(0.004)	0	(0.003)	0.001	(0.003)	0	(0.003)	138	0.042	0.037
Readmission (180d)	-0.005	-0.004	(0.002)*	(0.002)*	-0.001	(0.003)	-0.004	(0.003)	-0.003	(0.003)	-0.007	(0.003)**	-0.002	(0.002)	138	0.361	0.292
Readmission (180d)	0	-0.002	(0.003)	(0.003)	-0.003	(0.003)	-0.001	(0.003)	-0.003	(0.003)	-0.003	(0.003)	0	(0.003)	138	0.686	0.652
Share unilateral repl.	-0.001	-0.002	(0.002)	(0.002)	-0.003	(0.002)	-0.001	(0.002)	0.001	(0.002)	0.002	(0.003)	-0.001	(0.002)	138	0.39	0.324
Total LOS	0.26	0.139	(0.337)	(0.348)	0.025	(0.321)	-0.469	(0.35)	-0.509	(0.323)*	-0.917	(0.34)***	-0.639	(0.335)*	134	0.654	0.616
Transferred	-0.005	0.004	(0.024)	(0.023)	-0.003	(0.022)	0.014	(0.02)	0.014	(0.018)	0.016	(0.017)	-0.019	(0.017)	138	0.533	0.483
Charlson post	-0.022	0.033	(0.019)*	(0.019)	-0.001	(0.018)	-0.002	(0.017)	0	(0.018)	0.003	(0.018)	-0.007	(0.017)	106	0.216	0.128
Ethnicity post	-0.08	-0.07	(0.233)	(0.224)	-0.152	(0.223)	-0.088	(0.212)	-0.069	(0.21)	0.1	(0.231)	-0.257	(0.215)	106	0.208	0.12
Knee unilateral:																	
Acute LOS	0.089	0.137	(0.185)	(0.167)	-0.057	(0.157)	-0.349	(0.189)	-0.514	(0.172)***	-0.393	(0.151)***	-0.326	(0.149)**	1119	0.68	0.647
Antibiotic use	0.001	-0.003	(0.004)	(0.002)	-0.002	(0.002)	-0.001	(0.002)	-0.001	(0.002)	0	(0.002)	-0.001	(0.002)	152	0.594	0.553
Discharged home	-0.007	-0.076	(0.036)**	(0.029)	-0.042	(0.029)	0.055	(0.024)**	0.002	(0.023)***	0.069	(0.023)***	0.063	(0.023)***	152	0.51	0.46
Discharged home with support	-0.07	-0.081	(0.079)	(0.079)	-0.048	(0.078)	0.112	(0.077)	0.144	(0.081)*	0.107	(0.081)	0.061	(0.081)	152	0.421	0.363
Discharged support service	0.085	0.068	(0.033)**	(0.025)	-0.064	(0.019)***	-0.074	(0.02)***	-0.075	(0.019)***	-0.087	(0.019)***	-0.083	(0.019)***	152	0.519	0.471
Hospital death	0	0	(0.001)	(0.001)	0	(0)	0	(0)	0	(0.001)	0	(0)	0	(0.001)	152	0.437	0.38
Diagnostic procedures	-0.005	-0.006	(0.004)	(0.003)	-0.004	(0.003)	-0.006	(0.004)*	-0.004	(0.003)	-0.004	(0.005)	0.001	(0.005)	152	0.288	0.216
Imaging procedures	0.009	0.007	(0.006)	(0.007)	0.006	(0.005)	-0.008	(0.005)	-0.008	(0.005)	-0.003	(0.006)	0	(0.008)	152	0.377	0.313
N of post-admit diag	-0.019	-0.013	(0.028)	(0.027)	-0.012	(0.027)	-0.019	(0.025)	-0.02	(0.026)	-0.016	(0.027)	-0.015	(0.025)	152	0.458	0.403
N of interventions	-0.002	0.024	(0.092)	(0.09)	-0.016	(0.086)**	-0.06	(0.077)	-0.033	(0.079)	0.032	(0.071)	0.106	(0.07)	152	0.517	0.468
Pain symptoms	0	0	(0.002)	(0.002)	0	(0.002)	0	(0.002)	0	(0.002)	0.001	(0.002)	0	(0.002)	152	0.039	0.037
Readmission (180d)	0.001	0	(0.002)	(0.002)	-0.001	(0.002)	0.001	(0.002)	0.001	(0.002)	0.002	(0.002)	0.001	(0.002)	152	0.351	0.285
Share unilateral repl.	-0.008	-0.006	(0.008)	(0.008)	-0.004	(0.008)	-0.004	(0.008)	0.001	(0.009)	-0.001	(0.009)	-0.003	(0.008)	152	0.222	0.143
Total LOS	0.086	0.108	(0.194)	(0.175)	-0.097	(0.165)	-0.414	(0.197)**	-0.557	(0.179)***	-0.454	(0.158)***	-0.415	(0.157)***	140	0.675	0.642
Transferred	0.003	0.004	(0.019)	(0.017)	0.009	(0.016)	0.013	(0.016)	0.013	(0.015)	0.016	(0.015)	-0.001	(0.014)	152	0.451	0.395
Charlson post	0.003	0.019	(0.019)	(0.017)	0.011	(0.015)	0.011	(0.017)	0.001	(0.021)	-0.006	(0.015)	0.022	(0.016)	1004	0.155	0.064
Ethnicity post	-0.124	0.14	(0.196)	(0.206)	0.021	(0.211)	0.127	(0.233)	0.352	(0.235)	0.324	(0.236)	0.213	(0.211)	1004	0.277	0.2

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients' column. Errors are clustered at hospital level.

Table 2.19 – Hospital-level difference-in-difference estimates for unilateral knee, bilateral knee, bilateral hip replacements, (mode of discharge)

	Variable (SE)				Model statistics:		
	post2012 QBP	Sd	post2014 QBP	Sd	N	R2	AdjR2
Knee bilateral:							
<i>Discharged home</i>	0.029	(0.109)	-0.008	(0.116)	557	0.304	0.281
<i>Discharged home with support</i>	0.121	(0.065)*	-0.015	(0.064)	557	0.089	0.059
<i>Discharged support service</i>	0.042	(0.111)	-0.022	(0.117)	557	0.365	0.343
<i>Transferred</i>	-0.069	(0.055)	0.029	(0.056)	557	0.205	0.179
Knee unilateral:							
<i>Discharged home</i>	0.112	(0.016)***	0.003	(0.011)	1152	0.502	0.455
<i>Discharged home with support</i>	0.152	(0.048)***	-0.005	(0.047)	1152	0.42	0.365
<i>Discharged support service</i>	-0.113	(0.015)***	-0.013	(0.009)	1152	0.513	0.467
<i>Transferred</i>	0.005	(0.009)	0.01	(0.006)	1152	0.45	0.398
Hip bilateral:							
<i>Discharged home</i>	0.464	(0.151)***	-0.226	(0.146)	265	0.269	0.216
<i>Discharged home with support</i>	0.209	(0.136)	0.057	(0.143)	265	0.304	0.253
<i>Discharged support service</i>	-0.37	(0.142)***	0.252	(0.137)*	265	0.319	0.269
<i>Transferred</i>	-0.073	(0.041)*	-0.042	(0.044)	265	0.267	0.213

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients column, in parentheses. Errors are clustered at hospital level.

Table 2.20 – Hospital-level difference-in-difference estimates for other replacements

	Variable (SE)				Model statistics:		
	post2012 QBP	Sd	post2014 QBP	Sd	N	R2	AdjR2
Other replacement:							
<i>Acute LOS</i>	0.082	(0.208)	-0.056	(0.201)	1011	0.362	0.337
<i>Total LOS</i>	-0.009	(0.309)	-0.394	(0.324)	1143	0.201	0.173
<i>Discharged home</i>	-0.003	(0.009)	0.01	(0.008)	1201	0.196	0.169
<i>Discharged home with support</i>	0.029	(0.038)	0.046	(0.038)	1201	0.274	0.249
<i>Discharged support service</i>	0.003	(0.008)	-0.012	(0.006)*	1201	0.243	0.218
<i>Transferred</i>	0.005	(0.006)	0.002	(0.005)	1201	0.145	0.116
<i>N.of post-admit diag.</i>	0.01	(0.027)	0	(0.025)	1201	0.224	0.198
<i>N.of interventions</i>	0.046	(0.073)	-0.034	(0.072)	1201	0.427	0.408
<i>Pain symptoms</i>	0	(0.009)	0.005	(0.007)	1201	0.047	0.039
<i>Readmission (180d)</i>	0.001	(0.003)	-0.002	(0.003)	1201	0.119	0.089
<i>Revision (180d)</i>	0.007	(0.005)	-0.005	(0.005)	1201	0.235	0.21
<i>Elixhauser post</i>	0.272	(0.461)	-0.326	(0.448)	690	0.082	0.027
<i>Diagnostic procedures</i>	-0.004	(0.008)	0.004	(0.008)	1201	0.122	0.092
<i>Imaging procedures</i>	-0.023	(0.013)*	-0.008	(0.012)	1201	0.148	0.119

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients column, in parentheses. Errors are clustered at hospital level.

Table 2.21 – Hospital-level difference-in-difference estimates for unilateral hip replacements (mode of discharge)

Outcome	Discharged home	Discharged home with support	Discharged support service
Dependent variable			
<i>(Intercept)</i>	0.906 (0.733)	-2.519 (1.621)	0.499 (0.674)
<i>post2012 QBP</i>	0.104 (0.02)***	0.11 (0.043)**	-0.099 (0.015)***
<i>post2014 QBP</i>	-0.004 (0.014)	0.047 (0.042)	-0.011 (0.011)
<i>age</i>	0.003 (0.024)	0.046 (0.054)	-0.016 (0.021)
<i>age2</i>	0 (0)	0 (0)	0 (0)
<i>female</i>	-0.043 (0.071)	-0.219 (0.198)	0.122 (0.065)*
<i>urban</i>	-0.048 (0.009)***	-0.018 (0.029)	0.062 (0.008)***
<i>QBP flag dummy</i>	-0.144 (0.017)***	0.208 (0.032)***	0.181 (0.016)***
Other controls	YES	YES	YES
Year FE	YES	YES	YES
Diagnosis FE	YES	YES	YES
Hospital FE	NO	NO	NO
N	1138	1138	1138
R2	0.55	0.516	0.559
AdjR2	0.504	0.467	0.514

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are below coefficients in parentheses. Errors are clustered at hospital level.

Table 2.22 – Hospital-level difference-in-difference estimates for all replacements

	Variable (SE)				Model statistics:		
	post2012 QBP	Sd	post2014 QBP	Sd	N	R2	AdjR2
All replacement:							
<i>Share bilat. hip</i>	0	(0.001)	0	(0.001)	1402	0.242	0.232
<i>Share unilat. hip</i>	0.004	(0.012)	0.004	(0.013)	1402	0.286	0.276
<i>Share bilat. knee</i>	-0.003	(0.004)	0	(0.004)	1402	0.107	0.095
<i>Share unilat. Knee</i>	0.006	(0.013)	0	(0.013)	1402	0.363	0.354
<i>Share other repl.</i>	-0.001	(0.002)	0	(0.002)	1402	0.126	0.114
<i>Acute LOS</i>	-0.38	(0.148)**	-0.178	(0.136)	1199	0.556	0.549
<i>Total LOS</i>	-0.444	(0.166)***	-0.253	(0.151)*	1341	0.421	0.413
<i>Discharged home</i>	0.112	(0.02)***	-0.004	(0.013)	1402	0.381	0.373
<i>Discharged home with support</i>	0.125	(0.046)***	0.005	(0.046)	1402	0.32	0.311
<i>Discharged support service</i>	-0.106	(0.016)***	-0.009	(0.01)	1402	0.421	0.414
<i>Transferred</i>	-0.001	(0.014)	0.013	(0.011)	1402	0.199	0.189
<i>Antibiotic use</i>	-0.001	(0.002)	0.001	(0.002)	1402	0.207	0.197
<i>Inhospital death</i>	0	(0)	0	(0)	1402	0.165	0.154
<i>Diagnostic procedures</i>	-0.003	(0.003)	0.007	(0.003)**	1402	0.134	0.122
<i>Imaging procedures</i>	-0.01	(0.005)**	-0.003	(0.004)	1402	0.26	0.25
<i>N. of post-admit diag</i>	-0.024	(0.019)	0.011	(0.018)	1402	0.231	0.221
<i>N. of interventions</i>	-0.07	(0.047)	0.118	(0.045)***	1402	0.225	0.215
<i>Pain symptoms</i>	0.008	(0.008)	-0.002	(0.009)	1402	0.045	0.041
<i>Readmission (180d)</i>	0	(0.001)	0	(0.001)	1402	0.097	0.086
<i>Revision (180d)</i>	0	(0.002)	0.001	(0.002)	1402	0.249	0.239
<i>Charlson post</i>	-0.001	(0.009)	-0.013	(0.009)	1172	0.07	0.055
<i>Elixhauser post</i>	0.022	(0.111)	0.01	(0.114)	1172	0.177	0.164

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients' column in parentheses. Errors are clustered at hospital level.

Appendix F. Matching estimators (additional tables)

Table 2.23 – Estimates of nearest-neighbor covariate matching for bilateral hip and bilateral knee replacements

variable	Effect	Hip bilateral				Knee bilateral			
		Effect 2012-13	SE	Effect 2014-17	SE	Effect 2012-13	SE	Effect 2014-17	SE
<i>Acute LOS</i>	ATE	0.881	1.429	-0.659	0.958	0.183	0.692	-0.481	0.583
	ATT	1.303	1.782	-0.369	1.082	0.287	0.811	-0.359	0.641
<i>Antibiotic use</i>	ATE	-0.01	0.028	-0.009	0.029	0.008	0.011	0.005	0.012
	ATT	0.014	0.035	0.012	0.036	0.009	0.014	0.003	0.014
<i>Discharged home</i>	ATE	0.44	0.122***	0.289	0.098***	-0.14	0.123	-0.059	0.108
	ATT	0.334	0.134**	0.292	0.114**	-0.151	0.152	-0.072	0.128
<i>Discharged home with support</i>	ATE	0.251	0.116**	0.3	0.075***	0.025	0.08	0.063	0.077
	ATT	0.175	0.142	0.258	0.089***	0.048	0.093	0.076	0.087
<i>Discharged supportservice</i>	ATE	-0.365	0.116***	-0.197	0.089**	0.249	0.112**	0.125	0.104
	ATT	-0.234	0.124*	-0.188	0.105*	0.273	0.136**	0.149	0.125
<i>Inhospital death</i>	ATE	-0.006	0.004	0	0	-0.001	0.003	-0.002	0.004
	ATT	0	0	0	0	-0.001	0.004	-0.002	0.005
<i>Diagnostic procedures</i>	ATE	-0.003	0.004	0.039	0.017**	0.001	0.017	0.003	0.018
	ATT	-0.007	0.006	0.035	0.024	0.003	0.022	0.005	0.022
<i>Imaging procedures</i>	ATE	0.059	0.048	0.054	0.045	-0.03	0.03	-0.052	0.031*
	ATT	0.042	0.053	0.051	0.051	-0.03	0.033	-0.046	0.036
<i>N.of post-admit diag</i>	ATE	0.263	0.256	-0.021	0.16	-0.169	0.112	-0.328	0.109***
	ATT	0.455	0.322	0.109	0.14	-0.19	0.13	-0.34	0.13***
<i>N.of interventions</i>	ATE	0.101	0.2	0.004	0.171	-0.005	0.161	0.21	0.156
	ATT	0.196	0.229	0.111	0.199	-0.046	0.201	0.186	0.192
<i>Pain symptoms</i>	ATE	-0.131	0.073*	-0.026	0.044	-0.008	0.008	-0.004	0.009
	ATT	-0.097	0.071	-0.041	0.053	-0.007	0.009	-0.004	0.009
<i>Readmission (180d)</i>	ATE	0.003	0.028	0.017	0.028	0.007	0.011	0.003	0.014
	ATT	0.021	0.035	0.022	0.038	0.007	0.013	0.003	0.016
<i>Revision (180d)</i>	ATE	-0.022	0.017	-0.017	0.017	0.003	0.005	-0.002	0.008
	ATT	-0.014	0.015	-0.008	0.016	0.003	0.006	-0.002	0.01
<i>Share bilateralreplacements</i>	ATE	-0.008	0.002***	-0.008	0.012	0.011	0.023	-0.021	0.017
	ATT	-0.009	0.003***	-0.008	0.017	0.021	0.029	-0.021	0.021
<i>Total LOS</i>	ATE	1.362	1.513	-0.77	1.019	0.314	0.741	-0.676	0.691
	ATT	1.871	1.91	-0.706	1.078	0.478	0.847	-0.451	0.769
<i>Transferred</i>	ATE	-0.063	0.035*	-0.092	0.036**	-0.106	0.046**	-0.064	0.035*
	ATT	-0.08	0.042*	-0.092	0.044**	-0.116	0.054**	-0.073	0.04*
<i>Charlson post</i>	ATE	-0.042	0.091	-0.069	0.056	-0.016	0.036	-0.037	0.032
	ATT	-0.048	0.108	-0.071	0.069	-0.007	0.044	-0.029	0.036
<i>Elixhauser post</i>	ATE	0.286	1.793	-0.074	0.819	0.486	0.523	-0.112	0.597
	ATT	-0.063	2.071	0.631	0.898	0.589	0.597	-0.054	0.72

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients' column.

Table 2.24 – Estimates of nearest-neighbor covariate matching for all joint replacements

Variable	Effect	Effect	SE	p-val	Effect	SE	p-val
		2012-13			2014-17		
<i>Acute LOS</i>	ATE	-0.074	0.147	0.617	-0.174	0.153	0.254
	ATT	-0.039	0.203	0.846	-0.272	0.217	0.21
<i>Antibiotic use</i>	ATE	-0.001	0.002	0.47	0.001	0.001	0.648
	ATT	0	0.003	0.887	0.002	0.002	0.447
<i>Discharged home</i>	ATE	0.055	0.018***	0.003	0.047	0.016***	0.004
	ATT	0.055	0.026**	0.033	0.05	0.022**	0.026
<i>Discharged home with support</i>	ATE	0.114	0.048**	0.018	0.143	0.037***	0
	ATT	0.223	0.065***	0.001	0.232	0.052***	0
<i>Discharged support service</i>	ATE	-0.06	0.016***	0	-0.071	0.015***	0
	ATT	-0.087	0.022***	0	-0.095	0.02***	0
<i>Elbow replacement</i>	ATE	-0.001	0.001	0.367	0	0.001	0.928
	ATT	-0.002	0.001	0.12	-0.001	0.001	0.334
<i>Share bilateral hip</i>	ATE	-0.001	0	0.183	-0.001	0*	0.096
	ATT	0	0.001	0.609	-0.001	0.001*	0.083
<i>Share unilateral hip</i>	ATE	0	0.013	0.999	-0.005	0.01	0.589
	ATT	-0.001	0.014	0.928	-0.018	0.012	0.138
<i>Inhospital death</i>	ATE	-0.002	0.001	0.102	-0.002	0.001*	0.056
	ATT	0	0.001	0.99	0	0.001	0.638
<i>Diagnostic procedures</i>	ATE	-0.002	0.003	0.456	0.01	0.003***	0
	ATT	-0.002	0.004	0.686	0.005	0.003	0.128
<i>Imaging procedures</i>	ATE	-0.011	0.006**	0.048	-0.01	0.005*	0.063
	ATT	-0.004	0.007	0.606	-0.001	0.007	0.86
<i>Share bilateral knee</i>	ATE	-0.002	0.005	0.722	-0.002	0.003	0.548
	ATT	0	0.005	0.928	-0.003	0.004	0.43
<i>Share unilateral knee</i>	ATE	0.012	0.014	0.386	0.004	0.011	0.719
	ATT	0.002	0.016	0.898	0.022	0.014	0.112
<i>N.of post-admit diag</i>	ATE	-0.014	0.02	0.477	-0.001	0.014	0.935
	ATT	0.008	0.026	0.762	0.02	0.02	0.302
<i>N.of interventions</i>	ATE	-0.011	0.046	0.814	0.148	0.033***	0
	ATT	0.001	0.059	0.98	0.132	0.043***	0.002
<i>Share other replacements</i>	ATE	0.005	0.006	0.42	0.005	0.005	0.312
	ATT	0.01	0.006*	0.097	0.01	0.006*	0.096
<i>Pain symptoms</i>	ATE	0.009	0.007	0.213	0	0.006	0.949
	ATT	0.009	0.009	0.281	0.004	0.007	0.533
<i>Readmission (180d)</i>	ATE	0.003	0.002*	0.083	0.004	0.001***	0
	ATT	0.004	0.002**	0.048	0.005	0.001***	0
<i>Revision (180d)</i>	ATE	-0.002	0.002	0.489	0.003	0.002*	0.072
	ATT	0	0.003	0.886	0.005	0.002**	0.014
<i>Share shoulder replacement</i>	ATE	-0.009	0.008	0.25	0.005	0.008	0.55
	ATT	-0.005	0.007	0.501	-0.004	0.007	0.625
<i>Total LOS</i>	ATE	-0.136	0.173	0.433	-0.236	0.167	0.158
	ATT	-0.064	0.242	0.791	-0.242	0.226	0.283
<i>Transferred</i>	ATE	0.019	0.01**	0.045	0.037	0.008***	0
	ATT	0.04	0.014***	0.004	0.05	0.012***	0
<i>Charlson post</i>	ATE	0.003	0.009	0.771	-0.016	0.008**	0.043
	ATT	0.019	0.012	0.102	-0.006	0.011	0.575
<i>Elixhauser post</i>	ATE	-0.039	0.118	0.738	-0.103	0.1	0.301
	ATT	-0.035	0.146	0.809	0.076	0.124	0.538

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients' column.

Table 2.25 – Propensity score matching estimates with kernel weights (bilateral hip and bilateral knee replacements)

Variable	Effect	Hip bilateral				Knee bilateral			
		Effect 2012-13	SE	Effect 2014-17	SE	Effect 2012-13	SE	Effect 2014-17	SE
<i>Acute LOS</i>	ATE	0.421	2.502	-0.507	1.591	1.424	1.826	1.101	1.867
	ATT	1.048	2.546	-0.872	1.698	1.744	2.22	1.54	2.27
<i>Antibiotic use</i>	ATE	-0.01	0.016	-0.007	0.018	0.003	0.005	0	0.004
	ATT	-0.007	0.017	-0.01	0.018	0	0.004	-0.001	0.004
<i>Discharged home</i>	ATE	0.242	0.239	0.293	0.111***	-0.318	0.191*	-0.24	0.168
	ATT	0.361	0.184*	0.302	0.124**	-0.454	0.227**	-0.33	0.206
<i>Discharged home with support</i>	ATE	0.206	0.215	0.308	0.099***	-0.108	0.17	0.055	0.119
	ATT	0.194	0.172	0.264	0.095***	-0.191	0.207	0.026	0.144
<i>Discharged support service</i>	ATE	-0.164	0.227	-0.177	0.108	0.099	0.194	0	0.171
	ATT	-0.317	0.169*	-0.174	0.116	0.161	0.228	0.021	0.206
<i>Diagnostic procedures</i>	ATE	-0.009	0.008	0.043	0.027	-0.001	0.027	0.003	0.028
	ATT	-0.007	0.007	0.035	0.026	-0.001	0.033	0.005	0.033
<i>Imaging procedures</i>	ATE	0.024	0.047	0.056	0.055	-0.037	0.016**	-0.047	0.017***
	ATT	0.042	0.051	0.045	0.042	-0.031	0.015**	-0.041	0.016**
<i>N.of post-admit diag</i>	ATE	0.091	0.678	0.042	0.186	0.053	0.219	0.007	0.211
	ATT	0.291	0.311	0.003	0.149	0.1	0.271	0.059	0.262
<i>N.of interventions</i>	ATE	0.351	0.228	0.015	0.171	-0.024	0.169	-0.127	0.176
	ATT	0.14	0.234	-0.008	0.219	-0.05	0.195	-0.221	0.212
<i>Pain symptoms</i>	ATE	-0.389	0.167**	-0.019	0.036	-0.004	0.009	-0.001	0.009
	ATT	-0.074	0.066	-0.007	0.05	-0.002	0.011	0.001	0.011
<i>Readmission (180d)</i>	ATE	-0.014	0.041	0.003	0.033	0.004	0.006	-0.01	0.028
	ATT	-0.016	0.07	0.009	0.056	0.004	0.006	-0.013	0.036
<i>Revision (180d)</i>	ATE	-0.013	0.019	-0.007	0.02	0.001	0.001	0	0.002
	ATT	-0.014	0.011	-0.008	0.012	0	0.001	-0.001	0.002
<i>Share bilateral repl.</i>	ATE	-0.009	0.004**	-0.005	0.013	0.019	0.046	-0.015	0.041
	ATT	-0.004	0.004	-0.003	0.025	0.028	0.056	-0.017	0.05
<i>Total LOS</i>	ATE	1.321	2.565	-0.428	1.599	1.549	1.54	1.055	1.569
	ATT	2.192	2.771	-0.9	1.692	1.869	1.86	1.482	1.947
<i>Transferred</i>	ATE	-0.053	0.053	-0.104	0.042**	0.225	0.162	0.245	0.157
	ATT	-0.024	0.064	-0.118	0.05**	0.301	0.197	0.315	0.192

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients' column.

Table 2.26 – Propensity score matching estimates with kernel weights (all joint replacements)

Variable	Effect	Effect 2012-13	SE	p-val	Effect 2014-17	SE	p-val
<i>Acute LOS</i>	ATE	-0.39	0.284	0.17	-0.39	0.226*	0.085
	ATT	-0.506	0.391	0.195	-0.433	0.326	0.184
<i>Antibiotic use</i>	ATE	-0.001	0.002	0.655	0.001	0.001	0.48
	ATT	-0.001	0.003	0.724	0	0.002	1
<i>Day surgery</i>	ATE	0	0.008	1	0	0.004	1
	ATT	-0.001	0.013	0.939	-0.001	0.008	0.896
<i>Discharged home</i>	ATE	0.105	0.024***	0	0.101	0.02***	0
	ATT	0.102	0.027***	0	0.097	0.021***	0
<i>Discharged home with support</i>	ATE	0.115	0.054**	0.033	0.125	0.044***	0.004
	ATT	0.115	0.058**	0.047	0.124	0.042***	0.003
<i>Discharged support service</i>	ATE	-0.101	0.019***	0	-0.109	0.017***	0
	ATT	-0.102	0.019***	0	-0.105	0.017***	0
<i>Elbow replacement</i>	ATE	-0.001	0.001	0.48	-0.001	0.001	0.48
	ATT	-0.001	0.001	0.48	0	0.001	1
<i>Share bilateral hip</i>	ATE	0	0.001	1	-0.001	0.001	0.317
	ATT	0	0.001	1	-0.001	0.001	0.317
<i>Share unilateral hip</i>	ATE	0.006	0.023	0.792	0.02	0.018	0.278
	ATT	0.009	0.028	0.752	0.022	0.023	0.348
<i>Diagnostic procedures</i>	ATE	-0.003	0.004	0.405	0.003	0.004	0.405
	ATT	-0.003	0.005	0.549	0.004	0.004	0.346
<i>Imaging procedures</i>	ATE	-0.009	0.013	0.489	-0.011	0.006*	0.059
	ATT	-0.009	0.006	0.16	-0.01	0.005**	0.046
<i>Share bilateral knee</i>	ATE	-0.002	0.004	0.579	-0.004	0.003	0.157
	ATT	-0.001	0.004	0.782	-0.003	0.003	0.289
<i>Share unilateral knee</i>	ATE	0.008	0.026	0.755	-0.004	0.02	0.841
	ATT	0.006	0.026	0.819	-0.007	0.021	0.744
<i>N.of post-admit diag</i>	ATE	-0.009	0.025	0.724	-0.006	0.021	0.776
	ATT	-0.01	0.029	0.732	-0.009	0.026	0.733
<i>N.of interventions</i>	ATE	-0.056	0.058	0.334	0.083	0.045*	0.067
	ATT	-0.069	0.068	0.31	0.066	0.053	0.214
<i>Share other replacements</i>	ATE	-0.001	0.006	0.864	0	0.004	1
	ATT	-0.002	0.009	0.832	-0.001	0.006	0.864
<i>Pain symptoms</i>	ATE	0.009	0.019	0.633	-0.001	0.012	0.935
	ATT	0.013	0.015	0.394	-0.001	0.011	0.93
<i>Readmission (180d)</i>	ATE	0	0.002	1	0.001	0.001	0.48
	ATT	0	0.002	1	0.001	0.001	0.48
<i>Revision (180d)</i>	ATE	0	0.003	1	0.002	0.002	0.371
	ATT	-0.001	0.004	0.782	0.002	0.003	0.48
<i>Share shoulder replacement</i>	ATE	-0.007	0.013	0.576	-0.006	0.008	0.442
	ATT	-0.008	0.011	0.474	-0.005	0.007	0.48
<i>Total LOS</i>	ATE	-0.393	0.315	0.213	-0.631	0.264**	0.017
	ATT	-0.42	0.427	0.325	-0.595	0.381	0.118
<i>Transferred</i>	ATE	0.003	0.016	0.847	0.014	0.012	0.247
	ATT	0.005	0.017	0.768	0.012	0.013	0.371
<i>Charlson post</i>	ATE	0.001	0.011	0.93	-0.013	0.01	0.189
	ATT	-0.001	0.012	0.935	-0.011	0.011	0.301
<i>Elixhauser post</i>	ATE	0.073	0.132	0.58	0.024	0.114	0.833
	ATT	0.086	0.145	0.553	0.003	0.133	0.982

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients' column.

Appendix G. Synthetic control group estimators (additional tables)

Table 2.27 – Difference-in-difference estimates with synthetic kernel-weighted control group, for bilateral hip and bilateral knee replacements

	Variable (SE)				Model statistics:		
	post2012 QBP	Sd	post2014 QBP	Sd	N	R2	AdjR2
Hip bilateral:							
<i>Acute LOS</i>	2.498	(1.619)	-2.415	(1.323)*	242	0.398	0.35
<i>Discharged home</i>	0.16	(0.159)	0.133	(0.152)	242	0.127	0.056
<i>Discharged support service</i>	-0.092	(0.148)	-0.035	(0.136)	242	0.214	0.15
<i>Imaging procedures</i>	0.02	(0.044)	0.033	(0.052)	242	0.116	0.045
<i>N.of post-admit diag</i>	0.723	(0.333)**	-0.728	(0.297)**	242	0.335	0.282
<i>N.of interventions</i>	0.362	(0.272)	-0.266	(0.193)	242	0.138	0.068
<i>Pain symptoms</i>	-0.151	(0.084)*	0.156	(0.069)**	242	0.024	0.019
<i>Share bilateral repl.</i>	-0.003	(0.003)	0.003	(0.003)	242	0.159	0.091
<i>Total LOS</i>	2.703	(1.67)	-3.748	(1.645)**	242	0.366	0.315
<i>Transferred</i>	-0.056	(0.055)	-0.114	(0.065)*	242	0.229	0.167
<i>Elixhauser post</i>	0.019	(0.111)	0.025	(0.111)	242	0.328	0.277
Bilateral knee :							
<i>Acute LOS</i>	0.341	(0.953)	-0.92	(0.996)	538	0.114	0.083
<i>Discharged home</i>	0.177	(0.12)	-0.1	(0.123)	538	0.098	0.066
<i>Discharged support service</i>	-0.17	(0.128)	0.122	(0.131)	538	0.082	0.05
<i>Diagnostic procedures</i>	-0.024	-0.033	0.023	(0.025)	538	0.062	0.029
<i>Imaging procedures</i>	-0.039	(0.034)	0.013	(0.047)	538	0.113	0.082
<i>N.of post-admit diag</i>	-0.14	(0.115)	-0.021	(0.118)	538	0.134	0.104
<i>N.of interventions</i>	0.031	(0.223)	0.464	(0.209)**	538	0.131	0.101
<i>Pain symptoms</i>	-0.013	(0.019)	0.023	(0.033)	538	0.038	0.025
<i>Share bilateral repl.</i>	0.004	(0.008)	-0.006	(0.008)	538	0.082	0.051
<i>Total LOS</i>	0.105	(0.965)	-0.776	(1.084)	538	0.129	0.099
<i>Transferred</i>	-0.005	(0.036)	-0.017	(0.036)	538	0.086	0.054
<i>Elixhauser post</i>	0.094	(0.502)	0.119	(0.572)	538	0.1	0.046

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients' column, in parentheses. Errors are clustered at hospital level.

Table 2.28 – Difference-in-difference estimates with synthetic kernel-weighted control group, all joint replacements

	Variable (SE)				Model statistics:		
	post2012 QBP	Sd	post2014 QBP	Sd	N	R2	AdjR2
Other replacement:							
<i>Acute LOS</i>	0.08	(0.262)	-0.139	(0.262)	988	0.467	0.445
<i>Discharged home</i>	-0.003	(0.014)	0.004	(0.013)	1158	0.265	0.239
<i>Discharged home with support</i>	0.017	(0.042)	0.086	(0.041)**	1158	0.421	0.401
<i>Discharged support service</i>	0.002	(0.012)	-0.011	(0.01)	1158	0.231	0.204
<i>Diagnostic procedures</i>	-0.004	(0.009)	0.011	(0.009)	1158	0.243	0.217
<i>Imaging procedures</i>	-0.01	(0.012)	0.003	(0.013)	1158	0.165	0.136
<i>N.of post-admit diag</i>	0.007	(0.036)	0.012	(0.031)	1158	0.264	0.239
<i>N.of interventions</i>	0.134	(0.075)*	0.022	(0.074)	1158	0.59	0.576
<i>Pain symptoms</i>	0.006	(0.015)	-0.001	(0.013)	1158	0.025	0.022
<i>Readmission (180d)</i>	0.003	(0.006)	-0.005	(0.006)	1158	0.396	0.374
<i>Revision (180d)</i>	0.008	(0.007)	0.003	(0.008)	1158	0.306	0.281
<i>Total LOS</i>	0.025	(0.347)	-0.476	(0.352)	1100	0.277	0.25
<i>Transferred</i>	0.002	(0.008)	0.009	(0.008)	1158	0.211	0.183
<i>Elixhauser post</i>	0.011	(0.496)	0.033	(0.492)	677	0.119	0.065
All replacements:							
<i>Acute LOS</i>	-0.38	(0.148)**	-0.178	(0.136)	1199	0.556	0.549
<i>Antibiotic use</i>	-0.001	(0.002)	0.001	(0.002)	1402	0.207	0.197
<i>Discharged home</i>	0.112	(0.02)***	-0.004	(0.013)	1402	0.381	0.373
<i>Discharged home with support</i>	0.125	(0.046)***	0.005	(0.046)	1402	0.32	0.311
<i>Discharged support service</i>	-0.106	(0.016)***	-0.009	(0.01)	1402	0.421	0.414
<i>Elbow replacement</i>	0	(0.001)	0	(0.001)	1402	0.065	0.053
<i>Hip bilateral replacement</i>	0	(0.001)	0	(0.001)	1402	0.242	0.232
<i>Hip unilateral replacement</i>	0.004	(0.012)	0.004	(0.013)	1402	0.286	0.276
<i>Inhospital death</i>	0	(0)	0	(0)	1402	0.165	0.154
<i>Diagnostic procedures</i>	-0.003	(0.003)	0.007	(0.003)**	1402	0.134	0.122
<i>Imaging procedures</i>	-0.01	(0.005)**	-0.003	(0.004)	1402	0.26	0.25
<i>Knee bilateral replacement</i>	-0.003	(0.004)	0	(0.004)	1402	0.107	0.095
<i>Knee unilateral replacement</i>	0.006	(0.013)	0	(0.013)	1402	0.363	0.354
<i>N.of post-admit diag</i>	-0.024	(0.019)	0.011	(0.018)	1402	0.231	0.221
<i>N.of interventions</i>	-0.07	(0.047)	0.118	(0.045)***	1402	0.225	0.215
<i>Other replacement</i>	-0.001	(0.002)	0	(0.002)	1402	0.126	0.114
<i>Pain symptoms</i>	0.008	(0.008)	-0.002	(0.008)	1402	0.042	0.039
<i>Readmission (180d)</i>	0	(0.001)	0	(0.001)	1402	0.097	0.086
<i>Revision (180d)</i>	0	(0.002)	0.001	(0.002)	1402	0.249	0.239
<i>Shoulder replacement</i>	-0.004	(0.007)	-0.002	(0.008)	1402	0.038	0.026
<i>Total LOS</i>	-0.444	(0.166)***	-0.253	(0.151)*	1341	0.421	0.413
<i>Transferred</i>	-0.001	(0.014)	0.013	(0.011)	1402	0.199	0.189
<i>Elixhauser post</i>	0.022	(0.111)	0.01	(0.114)	1172	0.177	0.164

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are to the right of coefficients column in parentheses. Errors are clustered at hospital level.

Chapter 3

Impact of Child Subsidies on Child Health, Well-being and Investment in Child Human Capital: Evidence from Russian Longitudinal Monitoring Survey 2011-2017

In this chapter we studied a maternity subsidy – called Maternity Capital (MC) – introduced in Russia in January 2007 by the Russian government in order to overcome a long-standing Russian demographic problem of low fertility and high mortality. This subsidy had an unprecedented size of 250,000 Rub (7,150 euros or 10,000 USD, in 2007) and was granted to all Russian families for giving birth to /adopting a 2nd or subsequent child since January 2007. The reform made it possible for eligible Russian families to use this fund to improve family housing conditions, to sponsor children’s education, or to invest it in the mother’s retirement fund. The objective of this chapter was to evaluate the impact of the MC claim eligibility on various child health and developmental outcomes, as well as household consumption patterns. Using data from the representative Russian Longitudinal Monitoring Survey 2011-2017, we tested regression discontinuity models and concluded that the MC did not significantly affect the vast majority of analyzed child and household outcomes. In the meantime, there was tentative evidence that the MC effect may have affected boys and girls differently, in that it may have resulted in improved socialization of female children. In addition, wealthier households may have benefited more fully from the MC by accumulating more savings that could plausibly be used for improving housing conditions.

1 Introduction and literature review

Faced with an aging population and the global trend of decreasing fertility, many governments increasingly regard pro-natalist policies and their impact on households as an issue of primary concern. While most developed and developing countries have put in place welfare systems that support fertility and childbirth, these programs vary greatly in terms of incentive size, implementation mechanisms, and their perceived success. To date, despite the growing popularity of pro-natalist policies, little is known about their medium- and long-term consequences on child health, development and well-being, as well as their general impact on household economic decisions.

At the beginning of 2007, the Russian government announced a reform whereby the second and subsequent childbirths/child adoptions would be incentivized by a government-sponsored subsidy of around \$10,000 US. This program, called Maternity Capital (MC), made it possible for eligible Russian families to use this fund on condition that it be spent on three eligible purposes (namely, improving housing conditions, providing education to household children or investing in the mother's state pension fund account). The objective of this Chapter is to evaluate how this pro-natalist measure affected multiple measures of child health and well-being, parents' willingness to invest in child human capital, as well as major household consumption patterns.

For the majority of tested child health, development and well-being outcomes, the results presented in this Chapter do not support the hypothesis that the MC claim eligibility had a significant impact at the aggregate level. However, estimation results are suggestive of the fact that MC may have differentially impacted several demographic subgroups. More concretely, this Chapter provides tentative evidence that MC subsidy enhanced school metrics of socialization in girls, while the opposite pattern was observed for boys. At the household level, wealthier households may have benefited more fully from the MC subsidy insofar as they could more readily afford to accumulate savings/take out a loan to purchase new real estate.

The contribution of our study to the existing economics literature is threefold. First, it studies a broad set of child well-being outcomes, some of which, to the best of our knowledge, have never been investigated in the economics literature. In particular, no earlier study looked at the impact of child subsidies on leisure time spending patterns by children. Some of these metrics, such as time spent on extracurricular study and arts, serve as very likely proxies of parental willingness to invest in child human capital. Second, this paper is the first to study the link between child subsidies and household diets, which are believed to have long-lasting effects on children's health and development. Lastly, our study fills a relative lack of research on the impact of pro-natalist reforms in middle-income transitioning countries, whose distinct features (e.g. weaker social security support, higher degree of uncertainty with respect to future income) may create a set of incentives different from that commonly observed in developed countries.

As for related research on Maternity Capital, an earlier study by *Slonimczyk and Yurko (2014)* concentrated on the impact of MC on fertility and labor market participation using Russian Longitudinal Monitoring Survey (RLMS) data from 1994 to 2011. Their before-and-after analysis, complemented with difference-in-difference, regression discontinuity and nearest-neighbor matching models, suggests that financial incentives provided by the MC program could lead to 1.4-2.4% increase in the probability of childbirth (i.e. an estimated 0.47 - 0.81 increase of the number of children born per woman). *Slonimczyk and Yurko (2014)* also tested a dynamic stochastic model of fertility, wherein women simultaneously make decisions as to whether to have a child and with regards to their participation in the labor market. The authors concluded that the MC subsidy affected married and less educated women more strongly. However, the authors did not find sufficient evidence for a differential impact of MC on fertility based on women's current employment status or urban/rural residence.

A closely related study by *Gonzalez (2013)* evaluated the impact of a largely similar reform, whereby the Spanish government introduced a 2500 euro (based on the exchange rate of 2007) subsidy upon the birth of a child after 1st July 2007, universally on all of its territory. The presence of this exact cut-off date made the author focus on models using regression discontinuity design applied to time series for fertility, employment and household spending outcomes aggregated on national level. The estimates obtained from parametric models with high-order polynomial approximations suggest that, while reform led to a reduction in the number of abortions and a considerable jump in the number of conceptions around the cut-off date, which subsequently translated into a 6% increase in the observed number of births, it did not significantly affect either total or child-related household expenditures. In addition, the intervention likely reduced women's labor supply by an estimated 0.2 months of work during the first year after childbirth. *Gonzalez (2013)* suggests that this time was likely spent on providing informal care for the newborn, which was reflected by the fact that both the fraction of families using private day care services and the monetary amount of these expenditures decreased by 4-12 % and 200-400 euros respectively in families with children born immediately after the cut-off date and a decrease on private childcare expenditures.

However, it is worthwhile to note that several studies have found evidence of heterogeneity in responsiveness to pro-natalist incentives, in particular with respect to age, ethnicity and religious affiliations. (*Cohen et al., 2013; Baughman and Dickert-Conlin, 2003*). As far as income is concerned, the evidence has been somewhat inconclusive, with no consensus even as to the sign of its impact. Several authors point out that income levels are likely endogenous to fertility choices and, thus, conclusions about the heterogeneity of reform impact with respect to the mother's and household income should be interpreted with caution (*Cohen et al., 2013; Milligan, 2005*).

The economics theory has investigated women's choices with respect to childbirth and investment in human capital in a dynamic contexts, taking into account the long-term

and interconnected character of the effects they engender for the mother, her children and the household in general. A series of papers, including *Kalemlı-Ozcan(2003)* and *Tamura(2006)*, predict that higher child mortality and uncertainty as to the survival of offspring can discourage parental investment into child human capital. A large body of research across various disciplines points to the fact that the first years of children's lives have a crucial, and perhaps irreversible, impact on their subsequent physical and cognitive development. In the context of public policy, creating conditions favorable to child development and well-being has been considered, or at least declared, as one of the main priorities of most current governments, including that of Russia.

So far, various metrics have been proposed by both governmental bodies and scholars to measure child well-being, which, in line with *Conti and Heckman(2012)*, can be classified into indicators of (with several examples cited in parentheses):

1. material well-being (experience of absolute and relative poverty or reported deprivation, availability of basic durable goods and services, such as mobile phone and Internet)
2. health and safety (mortality, chronic somatic and psychological illness, presence of health risk-factors, such as body weight abnormalities and low physical activity, psychological and emotional stress, immunizations)
3. educational well-being (school attendance, scores in tests on reading skills and numeracy, school grades, extracurricular activities)
4. family and peer relationships (raised in a single parent household, time spent on activities shared by family, violence and child abuse in family, number and intensity of peer relationships, exposure to peer violence and abuse)
5. housing and environment (livable space available to a child, presence amenities, degree of contamination of ingested water and inhaled air)
6. risk behaviors (smoking, alcohol and drug consumption, early unprotected sex, teenage pregnancies, juvenile felony)
7. subjective well-being (e.g.s self-reported life satisfaction measures, as well as those reported by parents)

There is a prevailing consensus that none of these metrics can by themselves serve as ultimate and solely correct units of measurement of child development. Instead, they should be considered in a comprehensive and holistic manner while making conclusions on the quality of environments in which children are raised and to which they are exposed on a day-to-day basis. In this Chapter, we intend to follow this approach as much as possible while analyzing the impact MC policy on child well-being.

Despite their likely good intent and purpose, reforms such as the Maternity Capital oftentimes have unexpected distributional consequences. In particular, studying the effect of introducing a low-fee childcare network in Quebec in 1997, *Haeck et al. (2014)* show that this policy resulted in a significant reduction in the share of alcohol and tobacco consumption in families' budget. Similar findings are reported in studies focusing on a broader concept of mothers' share of income in the global family budget (see *Schultz, 1990; Hoddinot and Haddad, 1995*). However, the evidence on the effectiveness of pro-fertility programs on a broader set of family health and well-being outcomes is less well-documented and consistent.

As for child health indicators, empirical research tends to support the idea that a family's financial wealth and material investments made in early childhood could generate benefits in terms of health outcomes (*Baughman and Duchovny, 2016; Case, Lubotsky, and Paxson, 2002; Currie, 2009*). Similar effects are observed with respect to child cognitive development and educational attainment, as evidenced, for example, by *Dahl and Lochner (2012)*, who find that the introduction of tax benefits in USA between 1993 and 1997 helped raise scores in reading and math tests taken by children living in affected households. Using parental income variation instrumented by trade tariffs that exogenously changed after trade liberalization across multiple sectors in India in 1991, *Ajefu (2018)* finds that higher levels of household income translated into better school attendance by children and lower rates of child labor, which could overall contribute to a long-term child human capital formation. *Attanasio (2015)* evaluated the impact of a policy intervention whereby, as a treatment, a group of young children residing in multiple Colombian villages received micronutrient supplements, cognitive development training or combination of both. It was found that, in comparison to children who did not receive any treatment, only children who underwent cognitive training sessions experienced significant improvements in cognition abilities. However, such an effect was not observed in those children who received only micro-nutrient supplementation.

There are few studies concentrating specifically on the effects of child subsidies on child well-being. In 2014 the Australian government introduced \$3000 lump-sum all-purpose cash transfers to Australian families who had a childbirth or adopted a child. Similar to the MC, this subsidy was not means tested. Starting from 2007 the subsidy implementation design changed such that families had to meet AUS\$75,000 eligibility family income threshold, with sums being paid fortnightly during the first 26 weeks after the childbirth/child adoption. This reform was studied by *Gaitz and Schurer (2017)* and *Deutscher and Breunig (2018)*, who, using similar difference-in-difference designs and data sources, come to the conclusion the reform had little effect on various educational, physiological and physical health outcomes of pre-school children. In the meantime, *Deutscher and Breunig (2018)* suggest that this impact may be positive for children living in families where parents have a weaker educational background in terms of the highest completed level of education.

Lebihan & Takongmo (2018) analyzed the universal child benefit program introduced in 2006 in all Canadian provinces outside Quebec, and also found no significant impact of the reform on child or mother well-being. Mirroring the argument by *Gaitz & Schurer (2017)* and *Deutscher and Breunig (2018)*, the authors suggest that the universal and, thus, untargeted character of the reform, and not the size of the program – \$1200 CAD annually or \$100 CAD monthly as of 2006 – was likely to be the primary explanation for the absence of observable statistical effects.

In a closely related Canadian study, using the share of plausibly exogenous variation in both provincial tax credit and supplemental child benefit programs in Canada, *Milligan and Stabile (2011)* conclude that more generous child benefits can positively impact educational achievement, psychological traits and well-being measures of the child. In addition, the results suggest that in boys improvements are more likely to be observed with regards to education and learning, whereas for girls the effect was stronger regarding their mental and emotional well-being.

As far as the impact of family income on child development and well-being is concerned, the economics literature has produced mixed results, which are largely dependent on model specifications and hypothesized channels through which income can translate into well-being improvements. On one side of the spectrum, a multidisciplinary literature review by *McEwen and Stewart (2014)* draws a conclusion that higher income improves a range of child outcomes (cognitive, behavioural, social and emotional) but that improvement is small in magnitude. Much of the correlation between lower income and worse child outcomes is explained by other factors that are often associated with having a low income, while income itself has a relatively small influence on child outcomes when researchers account for other factors. From this point of view, the causal effect of income on child outcomes is, in and of itself, quite small. From a policy perspective, this finding suggests that one cannot expect income transfers to low-income families to vastly improve child outcomes, and that focusing exclusively on income to close the gap between children from low- and higher-income families, via transfers or otherwise, is unlikely to be effective. In terms of public policy, these studies also suggest that income transfers could be more effective if they were targeted. Finally, they provide evidence that income effect on child outcomes displays diminishing marginal returns, each dollar spent having a bigger impact in the early years of life. On the other hand, several previous studies, including *Duncan et al. (2011)* and earlier mentioned *Dahl and Lochner (2012)*, have chosen to consider income as the primary contributing factor that makes it possible to support, for example, child education, health and recreation, which, in turn, improve child development and well-being outcomes. In any case, as argued by *Cooper and Stewart (2018)*, estimation results can be sensitive to the type of data and statistical methods used to make inference, the biggest effect sizes being typically reported in experimental studies.

The unique implementation design of the MC also places this study at the intersection of different strands of behavioral economics literature. Namely, MC key features of:

1. being a lump-sum payment (as compared to a payment flow)
2. having a delayed activation of payment rights (as opposed to immediate availability of funds)
3. several eligible purposes (i.e. conditional transfers vs. all-purpose payments)
4. providing the option to allocate the benefit to individuals other than the MC certificate recipient

have the potential to elicit a number of behavioral responses in participating individuals. In particular, people's preference for instant gratification over a significantly bigger reward in the distant future has been shown to lead to sub-optimal behavioral choices. This hypothesis has been modelled theoretically (for example, through hyperbolic discounting by *Laibson (1997)*) and confirmed experimentally (see, for example, *Epper et al. (2011)*; *Richards and Green (2014)*). Restricting the number of available MC uses and delaying MC application plausibly allowed reform designers to overcome, at least partially, potential inefficient spending patterns. In addition, the literature review by *Carmo de Souza Cruz et al. (2017)* concluded that conditional payments are likely to be more efficient than all-purpose cash transfers in improving vaccination coverage and improving general child morbidity. However, such evidence is weaker as far as objective biometric health outcomes are concerned. In the meantime, existing empirical evidence is also suggestive of the fact that smaller but more frequent payments can be more effective in affecting individuals' behavior than one-time lump-sum transfers (see, for example, *Diamond (2009)*), which may diminish the impact strength of the MC program. However, since MC certificates were generally used as one-time transfers as opposed to a stream of revenue, it can have a different effects on consumption to meet current need (such as food items) as compared to durable goods. This distinction between goods is drawn in Subsection 4.4 looking into the patterns of household diets and spending.

The remaining parts of this Chapter are organized as follows. Section 2 describes the conditions for Maternity Capital receipt and provides the Russian institutional context. Section 3 describes RLMS data used in this Chapter, Section 4 explains the empirical design of the study and provides the main estimation results. Robustness check are conducted in Section 5. Section 6 discusses the contribution, limitations and external validity of this study and concludes.

2 Maternity Capital institutional context

From the collapse of the Soviet Union in the early 1990s and until the turn of 20th century, Russia underwent a series of profound economic transformations aimed at increased privatization and liberalization of economic activity. Despite their partial success, one of

major and long-standing consequences of these changes was a progressive loss of public social welfare systems.

In particular, child care facilities have experienced a decline in funding and accessibility. In 2016, only 57.4 % of Russian children had access to public pre-school daycare and education (*Abankina and Filatova, 2018*), compared to an anecdotally almost universal coverage in Soviet era. In addition, child benefits in the form of monthly payments, tax refunds and child care subsidies are virtually non-existent in modern Russia. After childbirth, maternity leave lasts for up to 86 days for the majority of childbirths, and unlike in a number of European states, cannot be shared between parents¹. Similar problems are pervasive in other areas of social security, in particular, healthcare system which has difficulty ensuring universal access.

In a broader context, the size of MC – 250 000 Rubles in 2007 (equivalent to around \$9800 US in 2007) – is a very significant sum of money for the majority of Russian families. To put it into perspective, in 2011 the average salary in Russia was estimated to be 23369 Rubles (725 USD in 2011 prices), while the maximum amount of monthly unemployment benefit was set at 4900 Rubles (163 USD in 2011 prices, less than 1/4 of the average salary amount). Thus, MC certificates are approximately worth one year’s average salary amount. According to ROSSTAT data, in 2013 an average Russian household spent around 26% of its income on food, 10% on clothes and shoes, 12% on transport, and only 3.7 and 1.3% on healthcare and education, respectively, savings representing around 10% of household income (*ROSSTAT, 2020*). However, it is important to note that in Russia both income and cost structure vary considerably in different regions. For instance, Moscow residents spend only 20% of their income on food items, this share reaches 28.2% in Samara region. Overall, due to differences in living costs, the MC subsidy may have a higher purchasing power in rural areas and small regional centers than in Russian cities. For instance, in 2014 the average price of a 1-bedroom apartment in Moscow stood at 155,000 USD, while prices can be 10 times lower in rural areas (*CIAN, Rosrealt, 2020*). *Sorvachev and Yakolev (2020)* show that fertility grew faster in regions where the subsidy to housing price ratio was higher, which suggests that the MC program could have had a stronger impact in low cost areas.

However, despite an overall trend towards privatization and a less broad access to free leisure and extracurricular services on the 2000s and 2010s, free or affordable options continued to exist in many Russian regions. Available choices vary from sports clubs to Sunday schools and are extremely heterogeneous depending on the region; residents of larger and richer cities enjoying a lot more accessibility and diversity.

In the 2000s Russia’s economic situation revived, in part thanks to booming oil and gas prices and to structural market reforms implemented in the preceding decade, which was accompanied by a steady growth in disposable income and an improved living standard. In an attempt to boost falling fertility rates, and, plausibly, to insure an increased public

¹The 86 day leave can be extended by an additional 26 days for complicated child deliveries

support, in 2007 the Russian government announced the introduction of a subsidy targeted for mothers with multiple children, called Maternity Capital.

Eligibility conditions for Maternity Capital certificates are stipulated in the federal law 26.12.2006 N 256 “On complementary measures of state support for families with children”. It defines as potential beneficiaries of Maternity Capital any individual holding Russian citizenship and falling into one of the three categories :

1. woman who gave birth to or adopted a second child after the 1st January 2007
2. woman who gave birth to or adopted a third or consecutive child after 1st January 2007, on the condition that the claimant earlier did not exercise her right to Maternity Capital
3. man who from 1 January 2007 onward is recognized as the sole adopter of a second, third or subsequent children and who did not earlier exercise his right to Maternity Capital ²

These eligibility criteria imply that, first, a family must give birth to/adopt at least 2 children to be able to benefit from the policy. Second, MC can only be claimed once. Third, child adoptions are considered to be an equally valid circumstance on par with births entitling to MC.³

During the time period analyzed in this study, MC funds were allowed to be used towards:

1. improving housing conditions (through co-funding purchase of real estate, paying part of mortgage and investing in repairs of existing lodgings)
2. paying for education expenses of (any) child in the family
3. investing in mother’s pension fund ⁴

The MC subsidy certificate has to be granted to an adult family member (for the sake of simplicity, we will further refer to potential MC beneficiaries as mother, leaving the male population outside the scope of the analysis). There is a 3-year delay before MC subsidy can be issued. However, it does not apply to cases where funds are used to make a downpayment, or fund part of mortgage, in which case the MC subsidy can be used almost

²The law also describes statistically unlikely but yet possible situations in which the right to MC can be lost by, or transferred to, one of the two parents or to the child himself/herself. These cases include death of one or all of the custodians, loss of child custody rights, etc. This study does not take into account these changes due to lack of available information and/or practical difficulty to establish the occurrence of these events

³Although individuals falling into the category of child adopters would be an extraordinarily interesting case to investigate, their very limited sample size in the data does not allow us to make valid inference

⁴Starting from 1st January 2016, MC it was permitted to use MC to purchase goods and services for a child suffering from disabilities. From 1st January 2018, it also became possible for families to spend MC funds in the form of monthly monetary benefits

immediately after the eligibility criteria are met. MC funds may be used partially, with multiple motives, and allocated between several users. It is important to stress that MC certificates are not tied to the child whose birth/adoption was at the origin of MC rights eligibility. In other words, any child in a given family can benefit from MC certificate funds (for example, MC can be used to fund education of a previously born child), or, if used for improving housing conditions, the whole family will then benefit from it. We use this feature of MC to perform robustness checks in Section 5.

Due to instability of the exchange rate of Russian currency, the size of the MC subsidy converted into main foreign currencies - euro and US dollar - fluctuated considerably since its implementation in January 2007. Although terms and conditions of MC did not allow for using it to purchase foreign real estate or to subsidize child's education outside the Russian educational system, those mothers who eventually opted for investing the funds into their pension plans could nevertheless have been impacted by such fluctuations. It appears plausible that relatively financially well-off families in Russia who can afford savings may base their expectation of the worth of MC in part on the value of the Russian rouble. Historically, after the collapse of the Soviet Union, even ordinary Russians regarded investments into foreign currencies as one of the few - if not the only - accessible ways to safeguard their savings in the face of economic instability; relying on the exchange rate as a general indicator of Russia's economic performance.

In general, Russian political institutions are considered to be lacking in accountability and rule of law. Partially as a result of this, a relatively small fraction of the Russian population claims to have trust in public institutions, be they federal or local, executive, legislative or judicial, with one major exception from this rule being the Russian president (*Levada Center, 2014*).

In particular, both economic theory and existing empirical evidence suggest that in the presence of distrust in public institutions individuals will likely have high discount rates for their future. As a result, it will affect households' decisions in a way that will make them favor behavioral strategies providing a certain and safe reward in the present, rather than immediately less profitable but more attractive strategies in the longer term. In the specific case of MC, this implies that the use of the MC certificate to buy real estate may be regarded as a more attractive option relative to the other two, since it can be benefited from almost immediately, with no need to wait for retirement age or the moment where the newborn reaches childhood or adolescence. In the meantime, since Soviet times a considerable number of families have been able to benefit from free or highly subsidized real estate purchases. Although this housing is usually provided in low-cost houses and entails a wait period of several years or even decades, it anecdotally helped to address housing shortages in many regions. However, considered altogether, the institutional instability described above could render the decision to raise a child a financially risky one, which could be particularly the case for low to average income families and single-parent families, given the relative lack of robust and institutionalized social security nets

nationwide.

Among the most prominent characteristics of modern Russia has also been the massive heterogeneity in terms of peoples' cultural norms, accompanied by a considerable inequality in income, living standards, and levels of economic development. Overall, members of religious confessions tend to have different stances towards such issues as fertility, abortion and child rearing. The onset of a recent trend towards state promotion of religiosity, dating from the beginning of 2010, could plausibly have had an impact on fertility and upbringing preferences of Russian families, although its exact magnitude is unclear and is arguably modest due to Russia's relatively high percentage of atheists and non-religious populations and an overall lack of trust in state institutions⁵ (*Levada Center, 2017*).

The Russian MC reform bears a great deal of resemblance with several current pro-fertility policies, most notably that introduced in Singapore. In particular, in addition to tax rebates and various housing subsidy programs available for large-sized Singaporean families, in 2012 their government introduced a cash gift for childbirths, the size of which varied with respect to the child order of the newborn. The maximum payment size reached 10,000 Singaporean dollars (around US\$7300) for giving birth to a third and subsequent children. A year later, Singapore introduced a US\$2,200 Medisave Grant for newborns, which made it possible to fund newborn's health-related expenses through the child's enrollment in approved health insurance plans. Finally, since 2015 the Singaporean government has offered the opportunity to match on a dollar-to-dollar basis parental investments into so-called Child Development Accounts, provided that the funds are spent towards education and healthcare-related eligible uses. The maximum subsidy amount is also largely comparable to MC and reaches US\$13200.

3 Data and descriptive statistics

The main source of data for this study is the Russian Longitudinal Monitoring Survey (RLMS). It is a panel survey conducted on a yearly basis since 1994 using a representative sample of Russian households. RLMS features three core modules: adult, children and for households. The resulting datasets contain an extremely rich set of variables that cover virtually all areas of respondents' lives. In particular, adults provide current and past information on their employment, income, health, education, attitudes towards social issues and family relationships. Overall, more than 800 and 1500 variables are available in the most recent RLMS waves at the individual and household level, respectively.

Each module has a separate questionnaire to which only adult members are eligible to answer. Thus, the information on child questionnaire is provided by an adult member present at home during the interview. The sample size in one RLMS wave (i.e. year of study) contains around 6500 households and 18000 individuals. Households that did

⁵Only 35% of population considered themselves at least somewhat religious in 2014, although this figure rose to 53 % in 2017 (*Levada Center, 2017*)

not participate in a given RLMS wave can return to the sample in the following year. RLMS interviews take place every year at respondent's home, normally from October to December, as long as participating families continue to live at the same address and are willing to participate in the study.

In total, 14 child outcomes related to health and development, educational achievement, material conditions and the degree of parental involvement in child upbringing were analyzed to evaluate the impact of MC on child well-being. In addition, this analysis is complemented with a household-level study of the MC impact on dietary habits and spending patterns, which are likely to affect all family members directly or through spillover effects.⁶ This impact is liable to be particularly strong for young children who are likely to heavily depend on parental choices for their diets (see Section 4 and Appendix 1 for variable description).

The main analytical sample is restricted to 2nd children born between 2004 and 2010 (i.e. 3 years before and after the introduction of Maternity capital in 2007) and aged from 6 to 8 years old, whose parents were surveyed in RLMS waves from 2011-2017 and living in households with 2 children. The year of birth of the second child determines whether the family is eligible to receive MC subsidy (that is, if the second child is born on or after the 1st January of 2007) or is ineligible for this type of support. In the analytical dataset information on each child is merged with corresponding mother and household characteristics, which allows us to take into account changes occurring at the household level that could have an impact on the analysed child outcomes.

The observations were collected from RLMS waves 2011 to 2017, with the median year of 2013. However, due to the fact that in the analytical sample we choose a 3-year window around the cut-off date of January 1st 2007, the measurements for MC eligible and ineligible families are separated by on average 3 years (median survey years are 2012 for MC ineligible families, and 2015 for MC eligible ones). In order to account for these differences, we included a wide range of child and household characteristics in our models, and use smaller window widths around the cut-off date, as discussed in Section 4. As mentioned in Section 1, information on the MC spending is not consistently collected and is not available in the RLMS for the vast majority of observations. However, available information indicates that on average time lapse between the moment MC eligibility rights were acquired and the measurements in RLMS analytical sample is around 7.1 years. Additional tables on time distribution of observations, as well as other distributive statistics are provided in Tables 3.11–3.15 of Appendix 2. Nationwide statistics collected by the Russian Pension Fund also indicate that as of 2017 around 91.1% of families who

⁶The impact of MC program on the probability of child birth has been studied by *Slonimczyk and Yurko (2014)*, who using similar regression discontinuity specifications found that financial incentives provided by MC program could have led to 1.4-2.4% increase in the probability of childbirth. In the context of this study it would be particularly interesting to study the impact of MC on the decision to have a 3rd child. However, in the RLMS cohorts this event is very rare, which does not allow for meaningful estimation

Table 3.1 – Descriptive statistics for the 2nd children (answers from RLMS 2011-2017 respondents with 2nd children born between 2004 and 2010 and aged 6-8 years, in households with two children), by period of birth

Variable	All sample				MC eligible		MC ineligible	
	Mean	SD	Min	Max	Mean	SD	Mean	SD
<i>Child characteristics:</i>								
Age	6.984	0.802	6	8	6.938**	0.804	7.037**	0.798
Sex (male)	0.521		0	1	0.578***		0.455***	
In good/excellent health	0.764		0	1	0.786*		0.739*	
Health score (1-best, 5-worst)	2.201	0.503	1	4	2.177*	0.499	2.23*	0.506
Health problem in last 30d	0.355		0	1	0.37		0.337	
Num. of chronic conditions	0.095	0.35	0	3	0.057***	0.254	0.14***	0.433
Good/Excellent GPA	0.683		0	1	0.664		0.698	
Went to excursion/gallery	0.812		0	1	0.812		0.812	
Sees friends > 2 times per week	0.693		0	1	0.685		0.701	
Has cell phone	0.664		0	1	0.615***		0.715***	
Vacation with parent in 1yr	0.768		0	1	0.762		0.774	
School homework/assignments	200.703	223.864	0	1200	216.569*	230.14	184.703*	216.5
Extracurricular study	19.933	75.617	0	1200	19.377	57.358	20.519	91.034
Extracurricular arts	130.098	198.7	0	1328	129.497	203.13	130.723	194.257
Watching TV/on Internet	125.374	80.101	0	720	125.74	75.203	125	84.92
<i>Mother characteristics:</i>								
Age	36.544	4.399	25	49	36.599	4.522	36.479	4.254
Urban	0.647		0	1	0.663		0.628	
Single parent	0.158		0	1	0.157		0.159	
In good/excellent health m	0.491		0	1	0.504		0.477	
Higher education diploma	0.374		0	1	0.428***		0.312***	
Ethnically other than Russian	0.095		0	1	0.092		0.097	
Poverty	0.43		0	1	0.471***		0.382***	
Household income, in 2011 prices	34.533	27.348	1.196	259.943	34.463	25.967	34.615	28.888
Alcohol cons.> 1 time per week	0.041		0	1	0.037		0.045	
<i>Household consumption:</i>								
Vegetables/legumes	2.28	4.776	0	56	2.308	4.821	2.248	4.728
Fruit (fresh and canned)	3.488	4.453	0	80	3.236*	3.416	3.775*	5.387
Meat and poultry	3.134	3.79	0	52	3.214	3.146	3.041	4.423
Dairy	5.627	4.182	0	31.5	5.887**	4.233	5.329**	4.107
Vodka and liquors (in liters)	0.177	0.606	0	7	0.146*	0.522	0.213*	0.69
Refined sugar	1.699	5.399	0	50	1.626	5.046	1.785	5.788
Candy and high-sugar treats	1.567	1.324	0	12	1.591	1.22	1.539	1.435
Starches	8.35	7.601	0	85	7.99	7.098	8.774	8.14
Essential food items	8.885	5.98	0	72.078	8.856	5.433	8.919	6.567
Discretionary expenditures	4.868	12.894	0	224.183	4.45	9.292	5.357	16.114
Purchase of durable goods	0.365		0	1	0.329**		0.407**	
Household savings	1.426	6.541	0	86.455	1.413	7.372	1.441	5.418
Loan payments (per month)	3.713	8.414	0	162.588	3.258*	6.265	4.245*	10.359
Number of observations	1050				577		473	

Legend: stars denote p-values for t-tests on equality of means (non-binary variables)/chi-square tests on equality of proportions (for binary variables): ***- 1 % sign., **-5% sign., *- 10% sign

used the MC chose to invest it in improving housing conditions (*Russian Pension Fund, 2018*).

Descriptive statistics on the sample of second children, and on the two subgroups are provided in Table 3.1. In total, this analytical sample contains information on 1050 children, 577 of whom (54.9% were born after the MC program was introduced in January 2007. Children born before and after January 1st, 2007 had broadly similar characteristics for many health, educational and developmental outcomes. In particular, participating respondents considered that their children were in good/excellent health and had good/excellent GPA at school in around 76% and 68% of cases. However, there is also evidence that the demographics that responded to MC stimuli may not completely mirror population averages. In particular, MC eligible families reported an on average higher rate of poverty (47.1% vs. 38.2% in MC eligible and ineligible families, respectively), as well as a lower level of basic expenditures (67.8 vs. 82.6 thousand Rubles) and purchase of durable goods (32.9% vs. 40.7%). This suggests that MC subsidy may have attracted a disproportionately higher number of poorer households, who may have considered MC certificates as an opportunity to improve financial standing/ease financial strain on household budgets. In addition, MC eligible families report a higher percentage of boys compared to girls, as well as a slightly younger age of their children.

Overall, the majority of Russian children were reported to enjoy good/excellent health and have good/excellent GPA at school. After school classes, besides spending considerable time preparing homework, many children were also involved in extracurricular educational activities, such as private lessons and art classes (20 and 130 minutes per week on average, respectively). In the meantime, there is quite a significant variation in the amount of time spent on these activities.

As for mothers, the average age was around 36.5 years for both groups, with around 15.8% of families comprised of a single parent. Although on average mothers also shared similar characteristics with regards their ethnicity, urban/rural residency and self-reported alcohol consumption, MC claim eligible mothers appear to have a higher level of graduation from a higher education institution. This likely results from the fact that higher education graduation rates have been rapidly increasing in Russia in the 2010s and the fact that 2007-2009 birth cohort is observed on average later in the analytical 2011-2017 period (i.e. children born in 2007-2009 are observed in RLMS waves from 2013 to 2017, depending on the date of interview)⁷.

As far as household dietary choices are concerned, overall the most highly consumed food group were starches, which include popular crops and staples, such as bread, potato, rice, etc. (see Appendix 1 for details). Healthy food groups whose consumption is supported by dietary guidelines, such as vegetables and fruits, accounted for a significant amount

⁷However, it is worthwhile to note that a significant part of this growth in numbers was at least in part due to the pervasive problem of 'diploma mills', 25% of higher education credentials in Russia being counterfeit, according to expert unofficial estimates (*Kommersant journal, 2015*).

of food intake as well (around 5.7 kg). In the meantime, foods high in refined sugars also comprised a noticeable part of households' diet (on average, around 1.57 kg of high-sugar treats, such as candy, chocolate and cakes, and a comparable amount of refined sugar). Overall, MC eligible and ineligible families reported largely similar diets, with minor differences in fruit, dairy and alcohol consumption. Finally, while general household spending patterns were largely similar in families with different MC eligibility statuses in total spending on essential food items and discretionary expenditures, descriptive evidence suggests that MC eligible were less likely to purchase a durable good during the 1-month period prior to answering the RLMS questionnaire.

4 Empirical strategy and main results

The empirical strategy to evaluate the impact of MC relies on the fact that MC claim eligibility dependent on a clear cut-off date of the 1st January 2007. After this date, any family giving birth to a second or subsequent child was automatically given MC claim rights. For the sake of comparability, we concentrate solely on households having 2 children, who constitute by far the most common case among MC claim eligible families. The reform was announced unexpectedly for the vast majority of the Russian population. This circumstance effectively rules out any anticipatory effect at the cut-off date since Russians could not adapt their behaviors prior to the introduction of MC.

The focus of this Chapter consists in comparing outcomes of children in the immediate neighborhood of the intervention cut-off date, where theoretically any observed difference in outcome variables must be solely attributable to the policy intervention due to a near complete treatment randomization. To this end, we select children who were born 3 years around the cut-off date (i.e. between the 1st January 2004 and 31 December 2009). Subsequently, we analyze different outcomes of these children when they reach the age of 6-8 and start attending primary school. This cohort is followed along RMLS surveys 2011-2017, until they reach the age of 9 and start going to middle school⁸.

It is important to note that we focus on the MC eligibility effect rather than actual use of MC, for which data are missing for most RLMS waves. In addition, since MC certificates have no expiration date, MC funds can be claimed at any point in the mother's lifetime. Hence, the actual use of available funds is impossible to establish without data on every time period since the 2nd child birth.

In order to ensure that similar incentives are being compared, in the main specifications we only concentrate on 2nd children (effects on 1st children who did not directly cause the acquisition of MC claim rights but who could still benefit from it are presented in subsection 5.3 on robustness checks).

To estimate the effect of MC on child outcomes, we are testing regression discontinuity

⁸See Appendix 2 for descriptive statistics on different sub-samples of the analytical sample

models of the functional form:

$$\begin{aligned} Outcome_{ip} = & \alpha + \gamma_1 trend_t + \gamma_2 trend_t^2 + \gamma_3 trend_t^3 + \gamma_4 trend_t \cdot post_t \\ & + \beta post_t + \lambda' X_{ip} + \theta' Z_{ip} + \nu_p + \epsilon_{ip}, \end{aligned}$$

where for a child i born in year-month t and observed in RLMS survey wave p , $Outcome_{ip}$ is the analyzed outcome of interest, $trend_t$ is linear trend for the month of birth relative to January 2007, $trend_t^2$ and $trend_t^3$ are its quadratic and cubic terms, respectively; $post_t$ is an indicator for births occurring after the 1st January 2007, X_{ip} is a vector of child-specific controls including child's age, sex and urban (i.e. city or regional center)/rural residence⁹; Z_{ip} is a vector of mother and household-specific controls, including mother's age and a set of dummies on higher education status, health status, non-Russian ethnicity, frequent alcohol consumption and the indicator that she raises a child as a single parent; ν_p are RLMS wave dummies; ϵ_{ip} is a random error term. In all tested models error terms are clustered at the regional level. Apart from implementing a standard coefficient significance test, we also report coefficients that remain significant after Bonferroni correction for multiple comparisons.

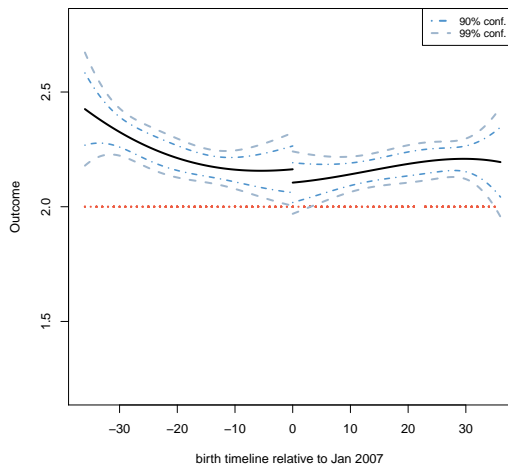
The main coefficient of interest β stands for the impact of MC on analyzed outcome variables. The advantage of the regression discontinuity design (RDD) specifications consists in the fact that theoretically they emulate a complete randomization at the immediate neighborhood of the intervention cut-off level, thus providing unbiased estimates of local average treatment effects (LATE). However, it is worth noting that RDD models critically depend on functional specifications. This consideration leads us to test for each analyzed outcome a set of models that vary in terms of functional form of the trend (ranging from linear and to 3rd degree of polynomial), birth timeline window around the cut-off period of January 2007 (models with 36, 24, 12 months' window width are tested)¹⁰. In addition, to allow for more functional flexibility, non-parametric local polynomial regressions (LOWESS) are tested as robustness checks in Section 5

4.1 Child health outcomes

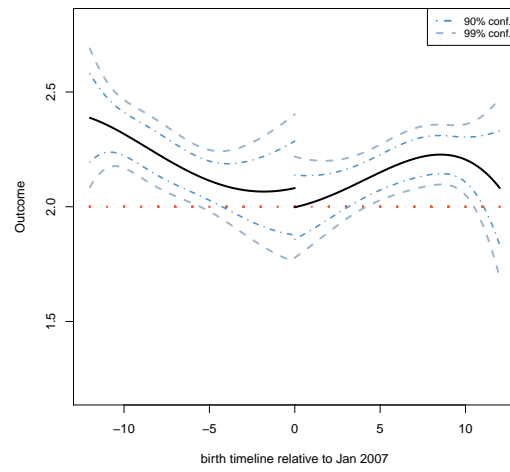
First, we evaluate the impact of MC for a range of reported child health indicators. The most straightforward of them are provided by parents when in RLMS they were explicitly asked to assess the state health of all their children on a score from 1 (best) to 5 (worst).

⁹Urban is defined as living in a city or a regional center, in accordance with RLMS statistical classification

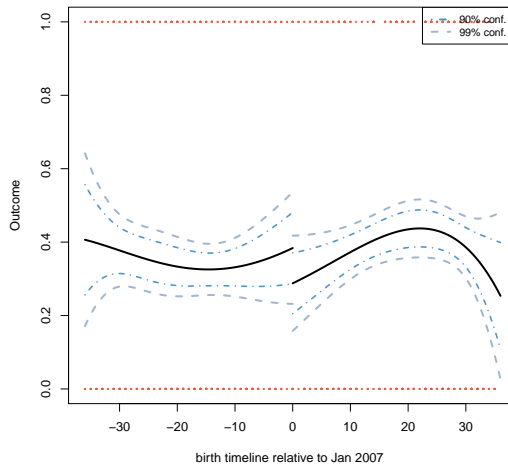
¹⁰Since RDD models estimate local treatment effects (i.e. for observations exactly at the cut-off level), testing models with larger window widths theoretically does not affect the unbiasedness of estimates, provided that the tested specification is chosen correctly. However, in practice, polynomial regression specifications can become more unstable if a larger period is considered, especially at distribution corners, since the correct functional specification is never known to empirical researchers. This makes it important to compare results from several window widths



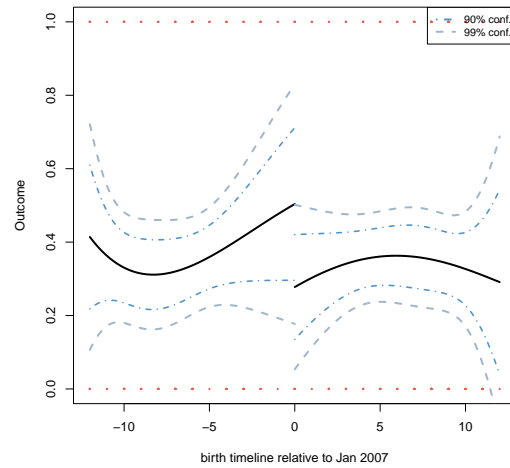
(a) Health score (36 months width)



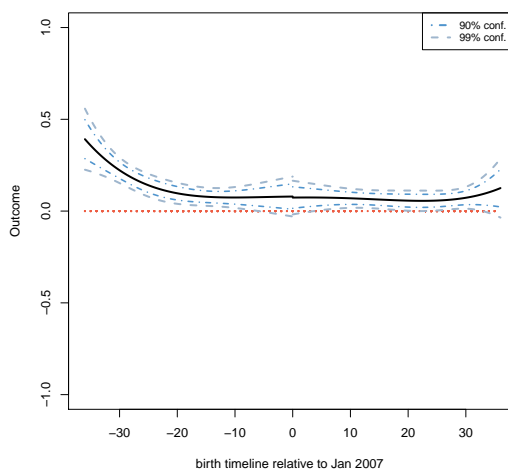
(b) Health score (12 months width)



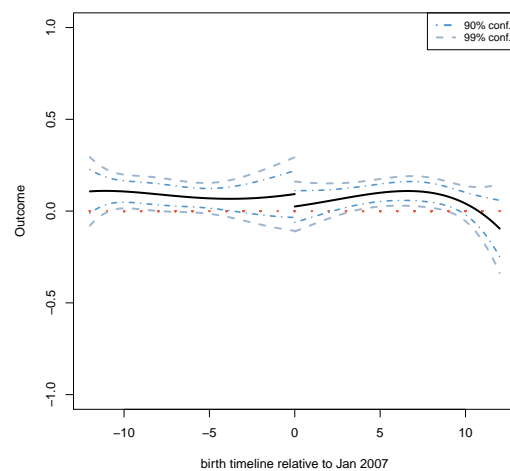
(c) Health problem in last 30days (36 months width)



(d) Health problem in last 30days (12 months width)



(e) Number of chronic conditions (36 months width)



(f) Number of chronic conditions (12 months width)

Figure .3.1 – Regression discontinuity (3rd order polynomial) estimates for Maternity Capital effect on 2nd child health outcomes

Children who received scores of 1 and 2 were deemed to have good/excellent health. Both of these measures are subjective and are liable to be affected by a host of factors, such as parental education, social background and parents' own perception of what stands for good and bad health. To partially overcome this issue, these variables are complemented with more objective health metrics that include the number of known child chronic conditions, occurrence of a health problem in last 30 days.

However, it worthwhile to note that since the RLMS survey is based entirely on answers provided by adult household members, one can still expect a higher rate inaccuracies and inconsistencies in respondents' answers compared to measurement/records obtained in a clinical setting.

Estimation results are presented in Tables 3.2 and Table 3.19 in Appendix. Estimates of RDD models with varying window width around cut-off are provided in Table 3.2. Overall, the results do not significantly change along model specifications and the window width parameter, and indicate that none of the five tested outcomes was affected by MC. It concerns all subjective health measures (good/excellent health and health score) and objective clinical measures reported by survey respondents.

Table 3.2 – Regression discontinuity estimates of impact of MC on 2nd child health outcomes (12, 24 and 36 months' window at 1st January 2007 cut-off birth date)

Variable	RDD 36m (1)	RDD 36m (2)	RDD 36m (3)	RDD 24m (4)	RDD 24m (5)	RDD 24m (6)	RDD 12m (7)
In good/excellent health	-0.0345 (0.0547)	0.0151 (0.0761)	0.0214 (0.0735)	0.0121 (0.0649)	-0.0198 (0.0903)	-0.0264 (0.0875)	-0.0153 (0.1004)
Health score (1-best, 4-worst)	0.0197 (0.0652)	-0.047 (0.0906)	-0.0544 (0.0877)	-0.04 (0.0759)	-0.0227 (0.1053)	-0.0216 (0.1019)	0.0176 (0.1077)
Num. of chronic conditions	0.0677 (0.0478)	-0.0249 (0.075)	-0.0127 (0.069)	0.0276 (0.0542)	-0.066 (0.0913)	-0.05 (0.086)	-0.0782 (0.1014)
Health problem in last 30d	0.0261 (0.0655)	-0.1153 (0.0936)	-0.0995 (0.0919)	-0.0777 (0.0806)	-0.1311 (0.1201)	-0.1133 (0.1182)	-0.0019 (0.1302)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	YES
<i>Linear trend</i>	YES	YES	YES	YES	YES	YES	YES
<i>Quadratic trend</i>	NO	YES	YES	NO	YES	YES	NO
<i>Cubic trend</i>	NO	YES	YES	NO	YES	YES	NO
<i>RLMS year controls</i>	NO	YES	YES	NO	YES	YES	NO
Num. obs.	1029	1029	1010	725	725	714	318

*Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at regional level. Subscript (b) denotes 5% signif. after Bonferroni correction*

Table 3.19 contains RDD estimates of MC impact on 2nd children with 36 months' window on both sides of the cut-off birth date of the 1 January 2007. For each outcome two models are tested, one containing only linear, quadratic and cubic birth timeline trends, and the other being complemented with child- and household level characteristics. Although these covariates do not affect the MC reform estimates in terms of bias and consistency, they can increase efficiency of estimates. Overall, the estimation results are similar to Table

3.2.

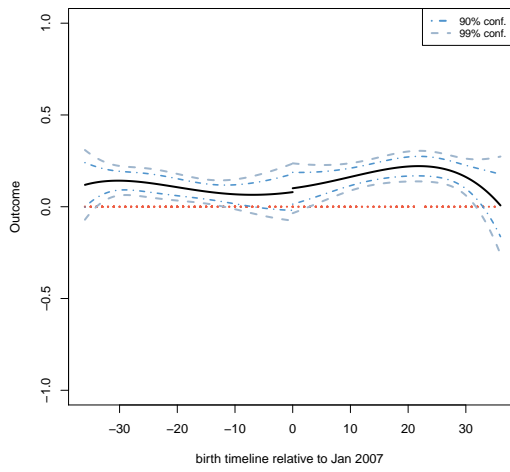
4.2 Child's education and development outcomes

In this Subsection, we evaluate the impact of the MC reform on educational and developmental outcomes. The standard and currently most common approach to measure educational achievement in education research is using students' school GPA ¹¹. Although the RLMS questionnaire does not ask parents to provide exact school performance records, it enquires about parents' perceptions of the most common grades that their children have been receiving at school. In this study, children receiving mostly grades 5 and occasionally 4, are considered to be in good/excellent GPA category (see Appendix 1 for a more detailed description). However, it is worth noting that in Russian schools there exists a high degree of heterogeneity in grading accuracy and practices among different schools. Therefore, this analysis is further complemented with additional variables reflecting the amount of time spent by children on different out-of-school activities, measured in minutes per week. These activities can be directly related to study through preparing school assignments, or reflect parental investments in child educational capital through providing opportunities for extracurricular classes (most often in the form of private tutoring for a foreign language and other school subjects, or attending an evening specialized school) or for extracurricular arts (including evening arts school, dance classes, etc.). In addition, we included a variable on the reported amount of time a child spends on watching TV/using computer for non-educational purposes (such as gaming and browsing non-educational websites). This outcome serves as a proxy of time devoted to unproductive leisure activities.

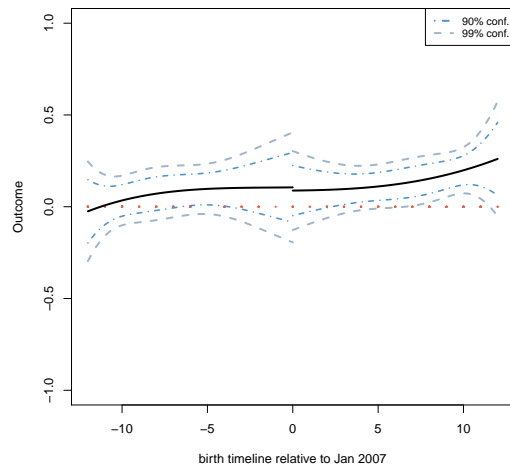
Estimation results for RDD models with a varying sets of covariates and window widths are provided in Table 3.3. Overall, both measures on student GPA show insignificant results. The marginally significant coefficient in model (2) for Good/Excellent GPA, respectively fall below the minimal 10% significance level when window width is changed in models (4)-(7) and additional covariates are included in the model. In addition, after the Bonferroni correction for multiple outcomes is applied, this coefficient loses its minimal significance.

As for variables reflecting the time spent by children, overall there is no strong evidence that patterns differed between MC claim eligible and ineligible families. Although models (1) and (6) provide weak evidence that children raised in MC claim eligible families spent on non-productive leisure (watching TV/on Internet) 22-52 mins per week was more than their MC claim ineligible counterparts, these coefficients lose their statistical significance after correcting p-values for multiple outcomes. In addition, this result is sensitive with

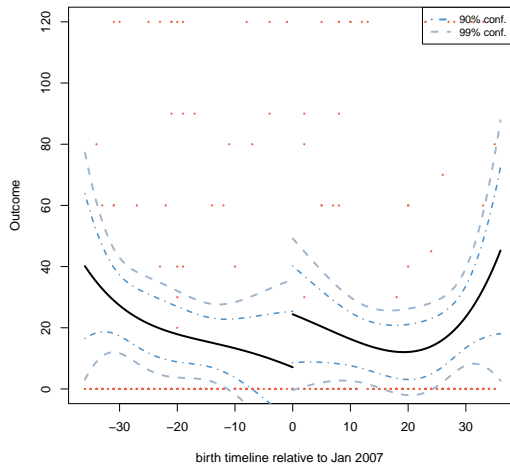
¹¹In Russian schools students are graded by 5 (excellent), 4 (good), 3 (satisfactory), and 2 (unsatisfactory). Grade 1 (unsatisfactory), while theoretically possible, is extremely rare and is usually given if rules or discipline were broken in class. As a result, it also oftentimes entails a disciplinary sanction



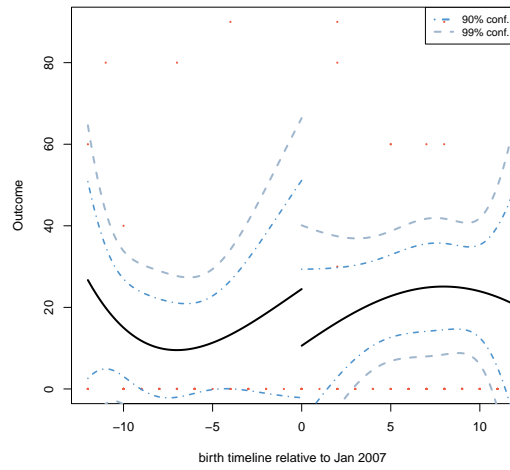
(a) GPA bad or satisfactory (36 months window)



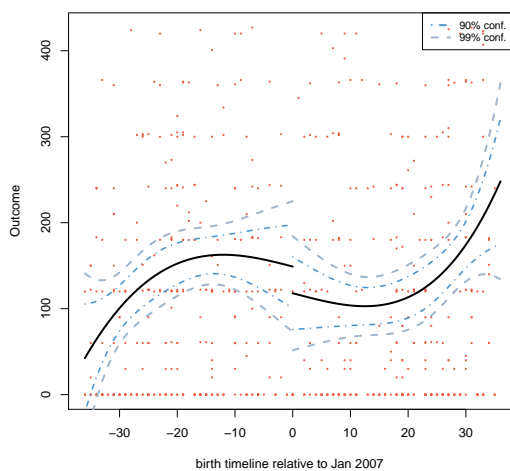
(b) GPA bad or satisfactory (12 months window)



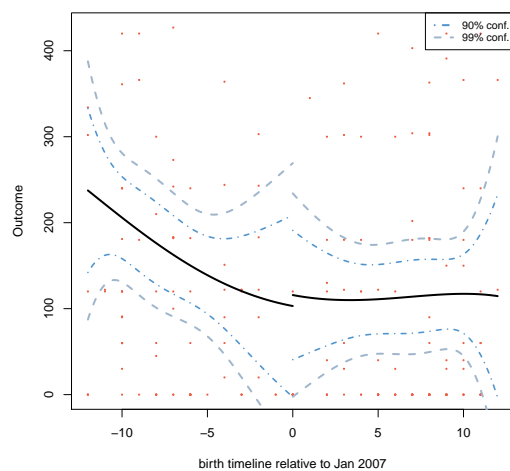
(c) Extracurricular study (36 months window)



(d) Extracurricular study (12 months window)



(e) Extracurricular arts (36 months window)



(f) Extracurricular arts (12 months window)

Figure .3.2 – Regression discontinuity (3rd order polynomial) estimates for Maternity Capital effect on 2nd child education and development outcomes

Table 3.3 – Regression discontinuity estimates of impact of MC on 2nd child education and development outcomes (12, 24 and 36 months’ window at 1st January 2007 cut-off birth date)

Variable	RDD 36m (1)	RDD 36m (2)	RDD 36m (3)	RDD 24m (4)	RDD 24m (5)	RDD 24m (6)	RDD 12m (7)
Good/Excellent GPA	-0.0137 (0.084)	-0.0356 (0.1268)*	-0.0241 (0.1224)	-0.0107 (0.1045)	-0.0145 (0.1575)	-0.0165 (0.1562)	-0.0894 (0.1644)
<i>Time spent on (in mins per week):</i>							
School homework/assignments	19.2828 (34.8)	39.8468 (49.6462)	32.496 (46.8631)	2.2194 (42.3338)	11.9547 (7.9803)	11.677 (5.861)**	7.2869 (6.8279)
Extracurricular study	9.9837 (9.8018)	15.7544 (14.3314)	15.9607 (14.8234)	8.8173 (9.0384)	7.6334 (12.5478)	10.7477 (12.2747)	5.0163 (14.1143)
Extracurricular arts	-11.3991 (3.2325)***	-16.3598 (41.8925)	-29.7815 (44.0794)	-38.2454 (39.0114)	16.8872 (50.255)	5.367 (52.1418)	39.7026 (58.7902)
Watching TV/on Internet	21.8549 (13.0915)*	12.8549 (18.2401)	18.0592 (17.7846)	15.0647 (15.3777)	17.4857 (22.5444)	52.476 (21.3784)**	34.046 (23.6267)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	YES
<i>Linear trend</i>	YES	YES	YES	YES	YES	YES	YES
<i>Quadratic trend</i>	NO	YES	YES	NO	YES	YES	NO
<i>Cubic trend</i>	NO	YES	YES	NO	YES	YES	NO
<i>RLMS year controls</i>	NO	YES	YES	NO	YES	YES	NO
Num. obs.	1029	1029	1010	725	725	714	318

*Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors s are given in parentheses under the coefficient. Error terms are clustered at regional level. No coef. are significant below 5% after Bonferroni correction*

respect to functional specifications. The evidence for an increased amount of time spent on school homework in MC eligible families is also weak, the coefficient in model (6) losing significance at the 5% level after the Bonferroni correction. As earlier, this result is sensitive to functional specifications and the choice of subsample. However, in both of these cases β for Watching TV/on Internet and Time spent on homework the coefficients are of a consistently positive sign.

4.3 Child well-being outcomes

A third set of child outcomes analyzed in this Chapter relates to the child’s physical/emotional well-being and parental effort to ensure a sufficient level of the child’s material comfort. To measure factors that are known to influence the level of children life satisfaction we include factors such as regular physical activity and interaction with parents and peers. We also included in the set of studied outcome variables the indicator variable for attending a cultural event (exhibitions, museums, galleries, cinema theaters, etc.)/going on a trip in the past 12 months, an indicator variable for a child seeing her/his friends on a regular basis at least 3 times a week, and an indicator variable that at least one parent spent the child’s school vacation with them.

RDD estimates of the MC subsidy claim eligibility on the described well-being outcomes

Table 3.4 – Regression discontinuity estimates of impact of MC on 2nd child well-being outcomes (12, 24 and 36 months’ window at 1st January 2007 cut-off birth date)

Variable	RDD 36m (1)	RDD 36m (2)	RDD 36m (3)	RDD 24m (4)	RDD 24m (5)	RDD 24m (6)	RDD 12m (7)
Has cell phone	-0.1407 (0.073)*	-0.093 (0.1036)	-0.0964 (0.1019)	-0.1408 (0.0892)	-0.1386 (0.1377)	-0.1172 (0.1345)	0.0156 (0.1438)
Vacation with parent in 1yr	-0.1359 (0.0581)**	-0.0287 (0.0849)	-0.036 (0.0831)	-0.0431 (0.0712)	0.0505 (0.1103)	0.0635 (0.1079)	0.0236 (0.1184)
Went to excursion/gallery	-0.0203 (0.0515)	-0.0331 (0.0748)	-0.0082 (0.0722)	-0.0207 (0.0636)	-0.1096 (0.0948)	-0.0781 (0.093)	-0.0799 (0.0971)
Sees friends > 2 times per week	0.0293 (0.0679)	0.0151 (0.0955)	0.0163 (0.0934)	0.0123 (0.082)	0.0353 (0.1218)	0.0483 (0.1204)	0.0801 (0.1383)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	YES
<i>Linear trend</i>	YES	YES	YES	YES	YES	YES	YES
<i>Quadratic trend</i>	NO	YES	YES	NO	YES	YES	NO
<i>Cubic trend</i>	NO	YES	YES	NO	YES	YES	NO
<i>RLMS year controls</i>	NO	YES	YES	NO	YES	YES	NO
Num. obs.	1029	1029	1010	725	725	714	318

*Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors s are given in parentheses under the coefficient. Error terms are clustered at regional level. No coef. are significant below 5% after Bonferroni correction*

are provided in Table 3.4. In line with previous findings, estimates for β are not robustly significant for any of the tested outcomes, the occasional significant coefficients in model (1) likely being a simple artefact of statistical testing and they lose their significance after the Bonferroni correction for multiple outcome testing is applied.

4.4 Household diets and spending

The last set of analyzed outcomes concentrates on household general spending patterns and household consumption of different food categories. To the best of our knowledge, the latter have not so far been examined in any previous study on child benefits/subsidies. The impact of the MC program on household dietary choices can be multifaceted. First, since MC relaxed the budget constraint for eligible families, the latter would be able to purchase higher quality nutrition. On the other hand, relaxation of budget constraints in some cases may exacerbate the already existing unhealthy food habits. Although nutritional guidelines are particularly notorious for being prone to change, there is a general consensus that consuming certain nutritional groups (notably, vegetables and fruit) is associated with better health outcomes, while others lead to the opposite (refined sugar, high-sugar treats, processed foods high in saturated fat, and alcohol). In addition, government state support and the accompanying public campaigning could have raised parental awareness with regard to child well-being and child-specific nutritional needs, in particular.

Since children are dependent on their parents for their diets and development of food

habits, the impact of parental good nutritional choices, or a lack thereof, can have a considerable and long-lasting impact on a host of health and well-being outcomes. In nutrition research, it has been shown that good dietary choices in childhood affect, among other things, the probability of developing chronic disease (including early development of metabolic syndrome), child's life satisfaction and self-esteem, and even longevity (*Grao-Cruces et al. 2014; Uauy et al. 2008; Vaiserman, 2014*).

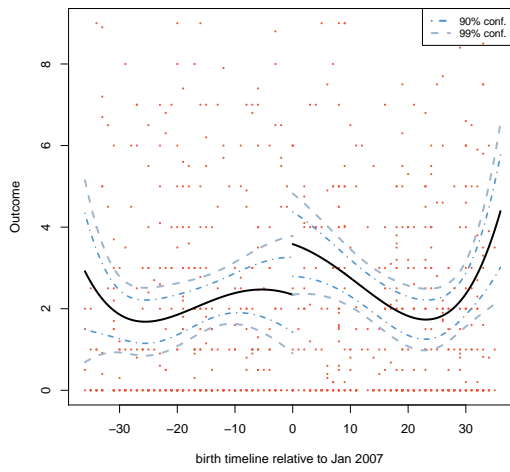
The RLMS survey provides extraordinarily detailed information on self-reported household consumption of more than product and food categories. We aggregated these variables in 9 major consumption groups: vegetables/legumes (both fresh and canned), fruit (both fresh and canned), meat and poultry, dairy (including milk, butter, cheese, etc.), refined sugar, high-sugar treats (tarts, candies, caramels, chocolate, etc.), starches (i.e. high-glycemic crops and foods, such as bread, potatoes, pasta, buckwheat). The last included group - consumption of high-alcohol beverages (e.g. vodka, rum, etc.) - was assumed to be destined only for adult consumption reflects the expenses made by heads of household for their own recreation/sustaining their habits. A more detailed description is provided in Appendix 1.

Analyzed outcomes reflecting household general spending include essential food items, discretionary (luxury) expenditures and financial outcomes (household savings and reported loan payments). These measures represent monthly reported household spending (in 2011 prices) on each of these categories. In addition, in this analysis we include an indicator variable for the purchase of a durable good by surveyed adult household members in the last 3 months preceding their RLMS interview.

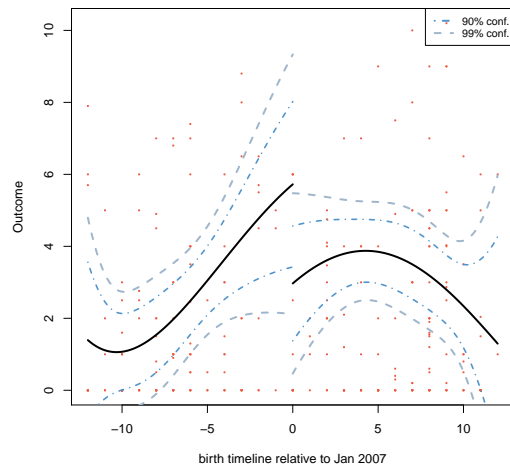
It is worth noting that while the RLMS provided information on the entire household consumption, no information is available in terms of how much was consumed by the child herself/himself. Thus, this analysis relies on the assumption that household spending and dietary habits have a spillover effect on children, although the precise magnitude of such influence is difficult to evaluate for each participating family.

The RDD estimates of MC impact on these household dietary outcomes are presented in Table 3.5. Overall, the results indicate no robustly significant difference between MC claim eligible and ineligible families for most food groups. Models with a 36 and 24 month's window are suggestive of the fact that MC eligible families may have increased consumption of starches (by around 3.5 kg) and decreased strong alcoholic beverages consumption (by around 0.1 liters). The sign of this effect remains stable across all time windows, although the significance of the estimated effects is overall marginal and goes away after the Bonferroni correction.

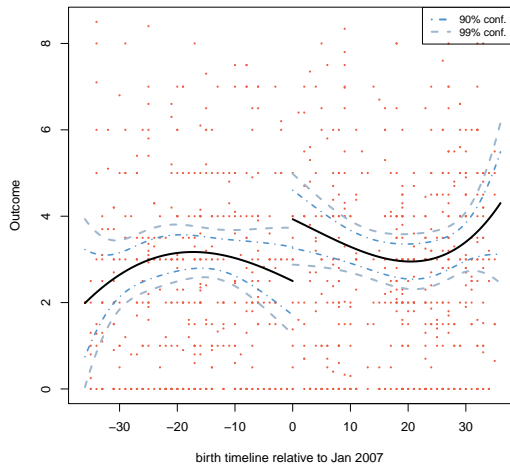
Regarding household consumption by category, models (2), (3) and (6) provide tentative and marginally statistically significant evidence in that MC eligible families may have increased their expenditure on essential food items and discretionary expenditures by around 1-3K and Rubles 1.2-3.5K Rubles per month, respectively (around \$30-90 and \$35-110 in USD equivalent), although the estimates' magnitudes for these variables tend



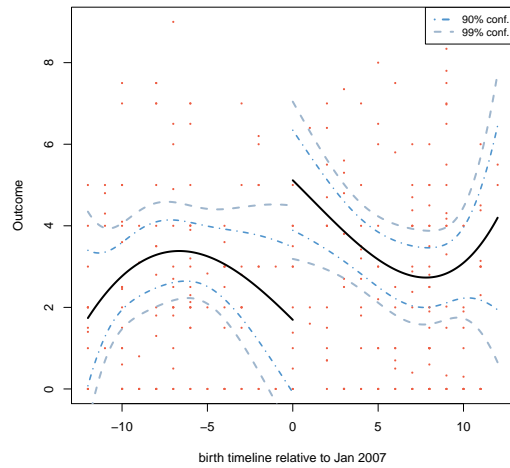
(a) Vegetables, fresh and canned (36 month width)



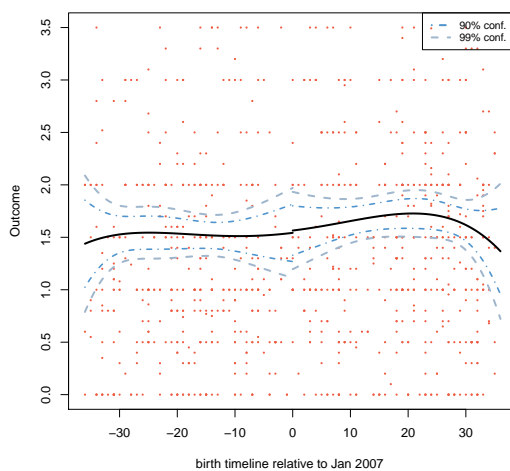
(b) Vegetables, fresh and canned (12 month width)



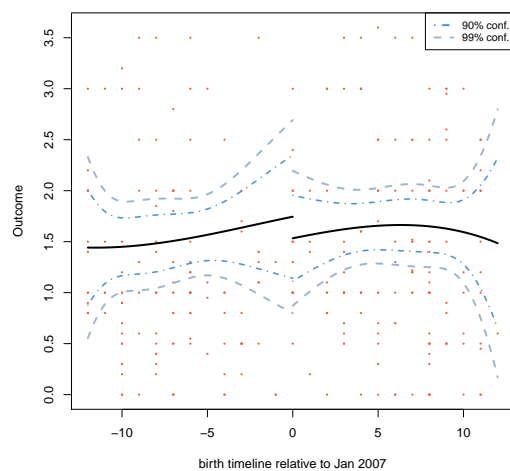
(c) Meat and poultry (36 month width)



(d) Meat and poultry (12 month width)



(e) High-sugar treats (36 month width)



(f) High-sugar treats (12 month width)

Figure 3.3 – Regression discontinuity (3rd order polynomial) estimates for Maternity Capital effect on household food consumption (in kilograms per week), by food categories

Table 3.5 – Regression discontinuity estimates of impact of MC on household diets and household consumption (12, 24 and 36 months’ window at 1st January 2007 cut-off birth date)

Variable:	RDD 36m (1)	RDD 36m (2)	RDD 36m (3)	RDD 24m (4)	RDD 24m (5)	RDD 24m (6)	RDD 12m (7)
<i>Household diet:</i>							
Vegetables/legumes	0.5579 (0.7581)	1.061 (1.1886)	0.9185 (1.178)	0.6409 (0.9742)	0.44 (1.5759)	0.3933 (1.5493)	-1.1273 (1.7373)
Fruit	-0.4223 (0.6103)	1.079 (0.7057)	0.4619 (0.7305)	-0.0888 (0.6729)	0.8489 (0.9877)	0.2777 (0.8942)	1.4565 (1.8933)
Meat and poultry	0.2627 (0.5937)	1.4942 (0.8812)*	1.3208 (0.9071)	0.957 (0.8084)	1.3715 (1.1755)	1.1326 (1.2319)	0.9894 (1.5278)
Dairy	0.6277 (0.6516)	0.4463 (0.9337)	0.353 (0.9097)	-0.1256 (0.8038)	-0.1693 (1.1351)	-0.1305 (1.1201)	-1.1891 (0.9609)
Vodka and spirits (in liters)	-0.1385 (0.0725)*	-0.1142 (0.099)	-0.1116 (0.0983)	-0.1175 (0.0805)	-0.056 (0.1367)	-0.0712 (0.14)	-0.2166 (0.1361)
Refined sugar	-0.1275 (0.4202)	0.2675 (0.5514)	0.1032 (0.5103)	-0.1749 (0.4591)	0.7721 (0.7993)	0.6119 (0.7944)	0.1143 (0.4943)
Candy and high-sugar treats	0.1286 (0.1971)	0.0848 (0.2975)	0.0306 (0.3027)	-0.0001 (0.2577)	0.0385 (0.3935)	-0.1001 (0.402)	0.183 (0.403)
Starches	3.2142 (2.058)	5.7851 (3.0654)*	4.4865 (2.8487)	4.1551 (2.363)*	3.1524 (4.8014)	3.5838 (3.6049)	-2.0776 (1.3786)
<i>Monthly consumption (by category)</i>							
Essential food items	1.4935 (0.9351)	2.3521 (1.2881)*	2.9919 ^(b) (1.0539)***	1.7792 (1.2044)	2.1383 (1.6445)	2.7378 (1.2726)**	0.938 (0.789)
Discretionary expenditures	1.2235 (1.2829)	2.6801 (1.8245)	3.1885 (1.6227)**	1.7075 (1.3253)	2.9799 (2.046)	3.4804 (1.874)*	-0.0728 (2.034)
Purchase of durable goods	-0.0479 (0.0644)	-0.0519 (0.0894)	-0.0053 (0.094)	-0.0506 (0.0792)	0.0264 (0.1131)	0.0265 (0.1192)	-0.0109 (0.1312)
Household savings	1.958 (0.8775)**	2.054 (1.1952)*	2.7825 (1.2289)**	1.7728 (1.1045)	1.9131 (1.5315)	2.4107 (1.5636)	2.0636 (1.4053)
Loan payments	0.2134 (1.0017)	1.6342 (1.1513)	2.1099 (1.186)*	1.1904 (1.0428)	3.5615 (1.3887)**	3.4238 (1.3828)**	1.8196 (1.1946)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	YES
<i>Linear trend</i>	YES	YES	YES	YES	YES	YES	YES
<i>Quadratic trend</i>	NO	YES	YES	NO	YES	YES	NO
<i>Cubic trend</i>	NO	YES	YES	NO	YES	YES	NO
<i>RLMS year controls</i>	NO	YES	YES	NO	YES	YES	NO
Num. obs.	1029	1029	1010	725	725	714	318

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at regional level. Subscript (b) denotes 5% signif. after Bonferroni correction

to be unstable across regression time windows. In addition, after the Bonferroni correction for multiple outcome testing is applied, only model (3) for essential food items remains significant at 5% . The variable for monthly household savings follows a similar pattern, with a tentative increase by around 2K per month in MC eligible families. Finally, the purchase of durable goods does not seem to have been affected by the introduction of the MC as none of the coefficients in models (1)-(7) are statistically significant, while coefficient signs and magnitudes are unstable across specifications.

5 Robustness checks

5.1 LOWESS estimation

As mentioned in Section 4, despite the fact that the RDD model can produce unbiased treatment effect estimates due to a near-complete randomization around the cut-off level, this property depends on the correctness of functional specification of the trend before and after the intervention. While the 3rd degree polynomials with a jump at the cut-off level is a very widespread approximation used in applied research, it is generally recommended to apply semi- and non-parametric methods to provide additional graphical representations allowing for more flexibility in fitting data.

Table 3.6 – LOWESS (1d and 2d kernel) regression discontinuity estimates of MC impact on the 2nd child outcomes and household consumption and diet (12, 24 and 36 months' window at 1st January 2007 cut-off birth date)

Variable	RDD 12m 2d kernel (1)	RDD 12m 1d kernel (2)	RDD 24m 2d kernel (3)	RDD 24m 1d kernel (4)	RDD 36m 2d kernel (5)	RDD 36m 1d kernel (6)
<i>Child health outcomes:</i>						
In good/excellent health	-0.0204	0.0864	0.009	0.0146	-0.011	0.0213
Health score (1-best, 4-worst)	-0.0077	-0.1253	-0.0419	-0.0554	-0.0109	-0.0491
Num. of chronic conditions	-0.0071	-0.0331	0.0091	-0.0013	0.037	0.0201
Health problem in last 30d	-0.135	-0.1302	-0.0855	-0.0933	-0.0129	-0.0375
<i>Education & development outcomes</i>						
Good/Excellent GPA	-0.0194	0.0071	-0.011	-0.0117	-0.0145	-0.0155
School homework/assignments	15.5366	12.2198	2.7102	9.4988	21.0654	21.1229
Extracurricular study	-0.3559	-6.1399	9.233	11.0502	11.3362	12.1686
Extracurricular arts	27.9579	2.2618	-29.6273	-22.6312	-9.518	-6.8308
Watching TV/on Internet	57.5203	45.7874	21.2029	40.0554*	18.3457	17.106
<i>Well-being</i>						
Extracurricular sports	29.1077	33.7861	-14.5728	84.8295	-21.7518	-23.5942
Has cell phone	0.0902	0.1433	-0.1317	-0.053	-0.1251	-0.0906
Vacation with parent in 1yr	0.0798	0.011	-0.0203	0.0299	-0.0103	-0.0094
Went to excursion/gallery	-0.0828	-0.134	-0.0337	-0.033	-0.0204	-0.0069
Sees friends > 2 times per week	0.1766	0.1806	0.1114	0.0992	0.0715	0.0799
<i>Household diets</i>						
Vegetables/legumes	-1.3307	-1.1486	0.761	0.8378	0.6543	1.0383
Fruit (fresh and canned)	-0.9121	-0.7956	-6e-04	0.0966	-0.1392	-0.1871
Meat and poultry	1.8161	2.4097*	1.016	1.0964	0.6081	0.7893
Dairy	-2.1411	-3.0258*	-0.0674	-0.0612	0.5333	0.6865
Vodka and spirits (in liters)	-0.296*	-0.2858**	-0.1035	-0.1117	-0.1322	-0.122
Refined sugar	0.3947	0.3235	-0.0199	-0.1496	-0.03	-0.2353
Candy and high-sugar treats	-0.1234	-0.1062	-0.016	-0.028	0.1031	0.0754
Starches	0.3112	0.1825	3.9885	2.2764	3.7144	3.4048
<i>Household consumption (by category):</i>						
Essential food items	0.3979	1.7866	1.78	0.7897	1.6335	2.187*
Discretionary expenditure	-3.1057	-1.4947	2.0494	-2.8628	1.6489	0.9185
Purchase of durable goods	-0.0642	-0.14	-0.0347	-0.0955	-0.0662	-0.0358
Savings	0.0805	-0.6813	1.8059	0.8756	1.7784	1.8765
Loan payments	2.3651	0.9134	2.1463	1.0671	0.9047	1.7828
Num. of observations	318	318	725	725	1029	1029

Legend: ***- 1 % sign., **-5% sign., *- 10% sign.

However, this practical convenience comes at a cost. In our context, the most important disadvantage of these methods lies in a lack of robust methods for deriving errors and confidence intervals. Neither of the two popular derivation methods – window overlap (used in this section) and bootstrap – can produce consistent standard errors without making restrictive model assumptions. Secondly, these methods cannot completely overcome the issue of choosing a functional form since point estimates in non-parametric models (such as LOWESS and kernel regression) also rely on functional assumptions. Finally, this moderate advantage of increased flexibility also comes at the cost of having to deal with the curse of dimensionality, which, in particular, manifests itself in the fact that for an increasingly high share of data/prediction points it becomes increasingly difficult to find close matches among available observations for most kernel weighing functions. As a result, it severely restricts the number of covariates that can be included in models. Thus, it is generally recommended to view semi- and non-parametric estimation as a complement rather than substitute for functional RDD. (*Lee and Lemieux, 2010*)

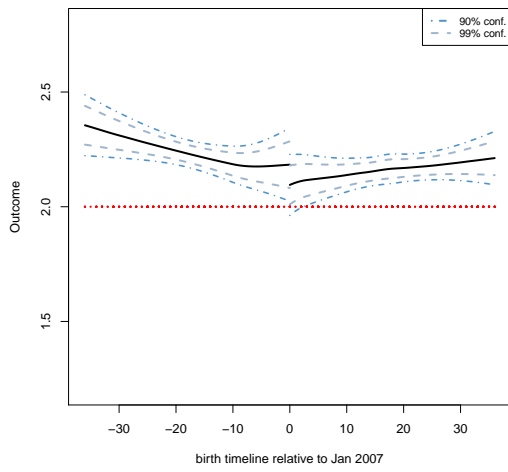
In this subsection, we test local linear regression models (LOWESS) of the form $y_i = f(\text{birthtimeline}_i + \text{birthpost2007}_i) + \epsilon_i$ wherein as explanatory variables we include a linear trend *birthtimeline* and post January 2007 birth indicator *post2007*. Observation weights, provided by a tricubic kernel function, are computed with and without taking *post2007* variable into account (“2d kernel” and “1d kernel” fit).

The results for non-parametric LOWESS estimation of MC eligibility effect on various child and household outcomes are presented in Table 3.6. Overall, they are in line with parametric RDD and indicate no robustly significant change in most analyzed outcome variables. In particular, none of LOWESS estimates was significant in health-related child outcomes. Although MC estimates are sporadically significant in some models for (watching TV/on Internet) and household consumption of dairy, meat/poultry, there appears to be no clear and robust pattern across different model specifications.

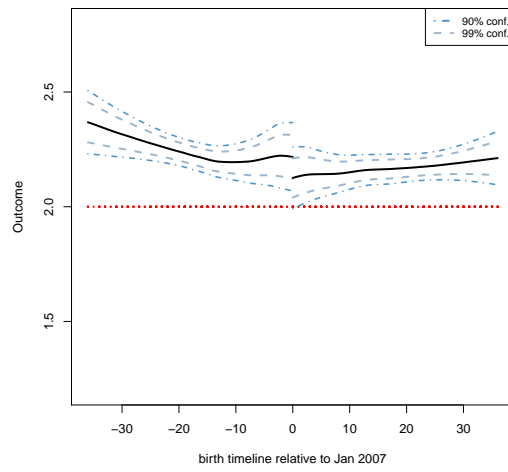
In the meantime, Models (1) and (2), coupled with a consistently negative coefficient for different time windows, provide tentative evidence for a decreased consumption of strong liquors by around 0.1-0.3 liters per week, echoing the evidence presented in *Haeck et al. (2014)* on Quebec households. In addition, LOWESS results show increased spending on essential food items (by around 1-2K Rubles/30-60 USD).

As for other household spending categories, variables for monthly household savings and loan payments show the same sign of the MC effect as in linear regression RDD models tested in Section 5. However, none of the coefficients for these variables are significant in the tested LOWESS models.

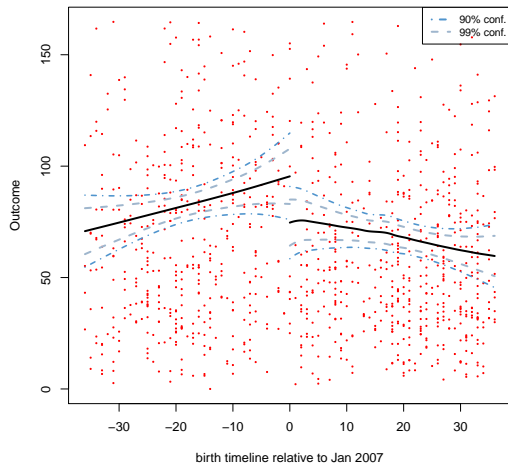
Finally, coupled with the main results described in Section 4, overall estimates for the remaining child and household-level dietary do not seem to stand the test of robustness with regards to functional specifications, the choice of model type, and the choice of window width around the cut-off birth date.



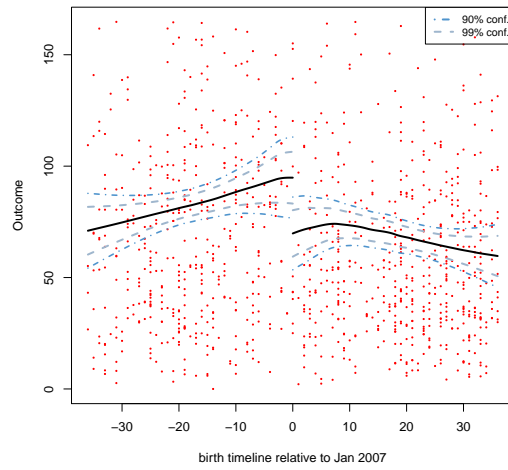
(a) Health score (2d kernel)



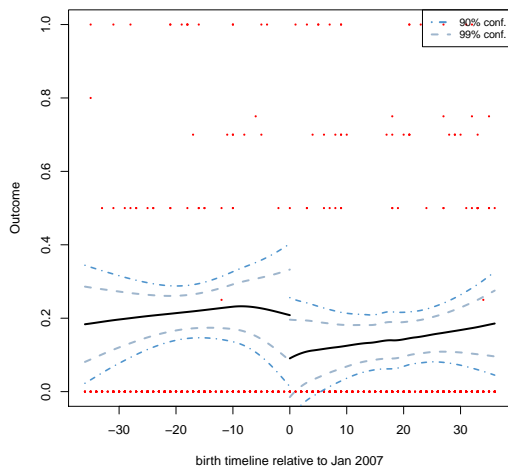
(b) Health score (1d kernel)



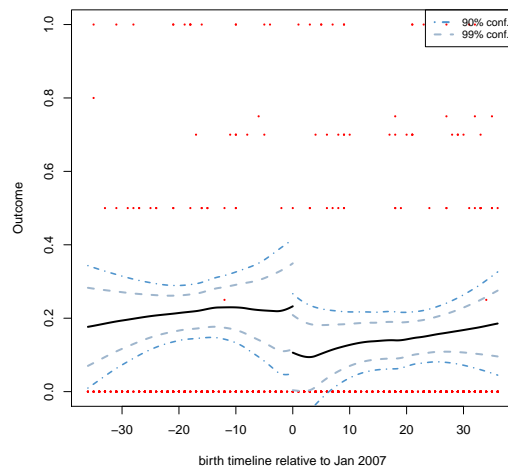
(c) Basic expenditure (2d kernel)



(d) Basic expenditure (1d kernel)



(e) Vodka and liquor (2d kernel)



(f) Vodka and liquor (1d kernel)

Figure .3.4 – LOWESS regression discontinuity estimates for Maternity Capital effect on child and household-level outcomes

5.2 Effect heterogeneity

By child gender

The MC subsidy may have impacted different sub-populations of eligible Russian families in a different fashion. The fact that investing in the MC subsidy in improving housing conditions was the most preferred use of the fund makes this reform share some of its key characteristics with such a widely known family relocation program as Moving to Opportunity (MTO) in several major U.S. cities (Baltimore, Boston, Chicago, Los Angeles, and New York City). Two influential studies on the effect of MTO by *Kling et al. (2005, 2007)* found that assigning housing vouchers to allow participating families to relocate to less poverty affected areas had a differential impact on children participating in the experiment. In particular, the authors conclude that young females benefited substantially from the MTO participation in terms of their mental health, education and engagement in risky activities. However, these positive effects for female children were offset by nearly identical adverse impacts that MTO had on their male children.

In this subsection we investigate whether MC eligibility affected male and female children differently in terms of their health, development and well-being outcomes. A summary of estimation results is provided in Table 3.7 and Appendix 3 of this Chapter.

Table 3.7 – Regression discontinuity estimates (LOWESS and 3rd order polynomial) of the MC impact on the 2nd child outcomes (36 months’ window at 1st January 2007 cut-off birth date),

Outcome	Female			Male			test $\beta_f = \beta_m$
	LOWESS 36m (1d)	LOWESS 36m (2d)	RDD 36m m3	LOWESS 36m (1d)	LOWESS 36m (2d)	RDD 36m m3	
In good/excellent health	0.007	-0.018	-0.044	0.169	0.165	0.131	0.166
Health score (1-best, 4-worst)	-0.079	-0.046	0.021	-0.21	-0.22	-0.163	0.174
Num. of chronic conditions	0.051	0.075	0.133	-0.157	-0.157	-0.25*	0.027
Health problem in last 30d	-0.006	0.04	-0.019	-0.072	-0.063	-0.132	0.493
Good/Excellent GPA	0.122	0.097	0.105	-0.025	-0.194	-0.284	0.101
School homework/assignments	-9.204	-20.78	-20.38	87.528	81.32	-52.069	0.84
Extracurricular study	6.677*	6.372	16.218	11.63	12.744	10.466	0.756
Extracurricular arts	125.878**	76.682	121.592**	-41.393	-25.65	-43.242*	0.004^(b)
Watching TV/on Internet	34.848	24.19	28.88	42.702	18.614	22.2	0.869
Has cell phone	0.034	-0.016	0.077	-0.18	-0.244	-0.03	0.443
Vacation with parent in 1yr	0.162	0.063	0.131	-0.024	-0.114	-0.18	0.091
Went to excursion/gallery	-0.144	-0.088	-0.067	0.033	0.022	-0.019	0.91
Sees friends > 2 times per week	0.224	0.213	0.255*	-0.128	-0.113	-0.151	0.055
Num of. obs		503			547		

*Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Reported RDD models use full specifications with individual controls, RLMS wave fixed effects and polynomial trends up to the 3rd degree. In linear RDD models error terms are clustered at the regional level. Subscript (b) denotes 5% signif. after Bonferroni correction. Test column shows p-values for equality of coefficients*

Overall, estimation results suggest that the MC program did not affect girls and boys differently for the majority of tested child health and well-being outcomes. However, a

few of them show a statistically significant effect for both genders. In particular, female children living in MC eligible households were estimated to have better a greater involvement in extracurricular art activities (by about 1.2-2 hours per week). Evidence presented in Table 3.7, combined with estimates in Table 3.20 of Appendix 3, point to a reportedly better level of socialization represented by a higher frequency of contacts with friends and going on vacation with at least one parent in the past 1 year. A statistically weak opposite trend is observed for boys, who may have reduced their time spend on extracurricular arts activities. However, none of the mentioned coefficients remain significant after p-values are adjusted for multiple outcome testing.

The differences in outcomes between sexes are tested in the last column in Table 3.7). The test results provide marginally statistically significant evidence of a differential impact of the MC on the probability to go on vacation with a parent and to see friends more than 2 times a week, and statistically somewhat stronger evidence (even after the Bonferroni correction) for the differences in time spent on extracurricular arts between sexes.

In the meantime, while MC eligible parents report a lower number of chronic diseases in boys compared to girls, there seems to be no consistent patterns with respect to other measures of child physical health, most parameters remaining insignificant and/or having inconsistent signs across specifications.

As expected, the LOWESS models report more stable and, in general, smaller coefficients than their linear RDD counterparts. In general, smaller samples in estimated models lead to relatively wide coefficient confidence intervals, which likely results in a smaller t-test power and, consequently, over-acceptance of the H0 hypothesis of a zero coefficient.

By rural/urban areas

The nature of the MC subsidy makes it likely that different Russian regions were impacted differentially by this reform. Insofar as the size of the MC certificate – around 10,000 USD in 2007 prices – was not adjusted with respect to local price levels, this subsidy had vastly different purchasing power across Russia (see Section 2 for more context). In particular, with regard to fertility, this point was addressed by *Sorvachev and Yakovlev (2020)*, who showed a stronger impact of the MC in regions with a greater subsidy to regional housing prices ratio. Since RMLS data used in this Chapter contains a limited number observation that does not allow us to carry out analysis at the regional level, we approximate the difference between low- and high-cost areas by examining differences by Russian rural and urban areas, respectively. RDD regressions are estimated parametrically and non-parametrically (LOWESS with a 36-month time window and 1d kernel) and are summarized in Table 3.8 and in Appendix 3.

As far as household diets and consumption patterns are concerned, the estimation results indicate that rural area residents responded relatively weakly to MC incentives, with most of the observed MC effects being driven by urban population. In particular, MC eligible households living in cities reported an increased spending on essential food items,

Table 3.8 – Regression discontinuity estimates (LOWESS and 3rd order polynomial) of the MC impact on household diet and consumption (36 months’ window at 1st January 2007 cut-off birth date), by rural/urban residence and poverty status

Outcome	Rural areas		Urban			Poor (self-declared)		Not poor (self-declared)		
	LOWESS 36m (1d)	RDD36m2	LOWESS 36m (1d)	RDD36m2	test $\beta_r = \beta_u$	LOWESS 36m (1d)	RDD36m2	LOWESS 36m (1d)	RDD36m2	test $\beta_p = \beta_n$
Vegetables/legumes	1.3745	0.8584	0.2011	1.1795	0.961	-0.0457	-0.2598	0.1347	1.4802	0.428
Fruit (fresh and canned)	-1.129	1.0505	0.6694	1.302*	0.842	-1.0076	-0.0338	1.3816	2.271***	0.476
Meat and poultry	2.6476	2.1391	0.432	0.8112	0.616	2.4104*	2.0513*	0.2343	1.0314	0.347
Dairy	-1.0779	-1.1239	0.1801	0.3064	0.489	0.9547	0.6681	-1.2042	-0.4408	0.345
Vodka and liquors (in liters)	-0.1086	0.0334	-0.1469	-0.0786	0.622	0.1434	0.3265**	-0.3964**	-0.3424**	0.002^(b)
Refined sugar	0.2373	0.4038	0.0933	0.6812	0.752	-0.3542	0.501	1.2495	0.5304	0.911
Candy and high-sugar treats	-0.6153	-0.3254	0.4091	0.5372*	0.143	-0.4971	-0.0873	0.548	0.7487**	0.167
Starches	2.437	7.1876	0.8973	7.4344*	0.295	-2.4073	12.9607*	3.0642	2.1679	0.076
Essential food items	1.3563	0.7158	1.9293	3.711*** ^(b)	0.286	0.2862	1.8105	2.8527	4.0747*** ^(b)	0.207
Discretionary expenditure	0.1732	3.401	0.5029	2.3035	0.751	1.0106	4.9025	0.6739	4.077**	0.959
Purchase of durable goods	-0.1664	-0.1779	0.0236	0.1174	0.101	-0.2442*	-0.1313	0.1847	0.1592	0.082
Savings	1.9385	1.2082	1.7902	3.941**	0.234	0.035	-1.6897	3.9808*	5.957**	0.004^(b)
Loan payments	0.2618	-0.6439	2.9655	3.04*	0.179	0.547	0.1334	3.2254	3.4514*	0.241
Num. of observations	371		679			449		596		

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Reported RDD models use full specifications with individual controls, RLMS wave fixed effects and polynomial trends up to the 3rd degree. In linear RDD models error terms are clustered at the regional level. Subscript (b) denotes 5% signif. after Bonferroni correction. Test column shows p-values for equality of coefficients

as well increased levels of saving and loan payments. While the magnitude of effects in linear regression RDD is highly variable, the LOWESS models tend to report more stable and conservative coefficient estimates due to the estimator’s reliance on point estimates. However, when a differential MC effect is tested for urban and rural populations, none of the analyzed outcomes accept this hypothesis at the minimum 10% significance threshold.

By poverty status

A heterogeneity of the MC impact can plausibly be expected for households with different levels of material wealth. The fact that in the majority of cases the MC subsidy size in itself was not sufficient to purchase a complete housing unit can render the benefits of the reform inaccessible to poorer households that cannot afford a mortgage/make a downpayment and/or do not have savings to be able to purchase real estate directly. This possibility is studied in the present Subsection, wherein the poverty status is determined by RLMS survey respondents themselves, who were asked whether they experienced difficulty providing themselves with the most essential consumer goods in the last 12 months (see Appendix 1 for more details on RMLS questionnaire). As earlier, estimation results are estimated parametrically and non-parametrically and are presented in Table 3.8 and in Appendix 3.

For a few outcomes, the results indicate significant differences in MC impact across the self-reported poverty line. Regarding household diets, there is tentative and statistically weak evidence that richer households shifted their diets toward fruit and foods rich in added sugars. In addition, poorer households, unlike families facing fewer financial difficulties,

may have increased alcohol consumption by around 0.15-0.3 liters per week (in contrast to a comparable consumption decline observed in more financially secure households).

As for more general consumption patterns, contrary to poorer households, wealthy households responded to MC incentives by increasing savings, which is suggestive of the household's past and/or future intent to use the MC to improve housing conditions. The test for the differential effect of the MC points to a different reaction of these two subgroups to MC eligibility, the test p-value remaining statistically significant at the 5% level even after the Bonferroni correction is made. These results suggest that MC program may have disproportionately benefited middle-class Russian families, who had enjoyed the opportunity to use the MC fund more freely and readily with the aim of purchasing new housing.

5.3 Effects on the first child

As mentioned in Section 2 on reform context, MC subsidy was not tied to the 2nd child born on the 1st January 2007 onward, but rather to his/her mother who could have the discretion to decide how the MC subsidy would be spent after a 3 years' wait period and, if she were to choose to use MC to ensure better education for her children, who among them would benefit from it. This feature of the MC reform provides us with the opportunity to test the impact of the MC program on 1st children, who did not directly create MC claim eligibility rights for the household, but who could still benefit from it either directly (through paid education and child care) or indirectly (through improved housing conditions or various penitential spill-over effects from other family members) when a second child is born.

As in Section 4, widow widths around the 2nd child birth date cut-offs are set to 36, 24 and 12 months to provide more robustness to the analysis. We concentrate on the same child outcomes as in Section 4. Estimates of MC impact using parametric RDD are presented in Table 3.9.

Overall, the results confirm the previously drawn conclusion on the absence of a global MC effect on a wide array of health, educational, developmental and well-being child outcomes. Even though estimated MC impacts are significant for models (2)-(4) for the number of chronic conditions, and marginally significant for Watching TV/on Internet in models (5)-(7), the variation of the functional form and/or window widths at birth date cut-off lead to a loss of results' statistical significance.

Table 3.9 – Regression discontinuity estimates of impact of MC on 1st child outcomes (12, 24 and 36 months' window at 1st January 2007 cut-off birth date)

Variable	RDD36m1 (1)	RDD36m2 (2)	RDD36m3 (3)	RDD24m1 (4)	RDD24m2 (5)	RDD24m3 (6)	RDD12m1 (7)
<i>Child health outcomes:</i>							
In good/excellent health	0.0183 (0.1488)	-0.0157 (0.1836)	-0.0217 (0.2037)	-0.0177 (0.1741)	0.0176 (0.1785)	0.0144 (0.1962)	0.0729 (0.1879)
Health score (1-best, 4-worst)	-0.1181 (0.1757)	0.2059 (0.1996)	0.3258 (0.2342)	0.1511 (0.2071)	-0.2392 (0.1971)	-0.1716 (0.2004)	-0.0457 (0.1899)
Num. of chronic conditions	0.15 (0.1435)	0.5438 (0.2498)**	0.5046 (0.2253)**	0.2567 (0.1552)*	-0.0187 (0.1192)	-0.0716 (0.1002)	-0.1105 (0.0781)
Health problem in last 30d	0.0928 (0.1949)	0.0419 (0.2612)	0.0545 (0.2617)	0.0391 (0.2267)	0.0319 (0.3164)	0.041 (0.3182)	0.0596 (0.3262)
<i>Education & development outcomes</i>							
Good/Excellent GPA	0.018 (0.2263)	0.02649 (0.2652)	0.0927 (0.2546)	-0.0946 (0.2655)	0.0382 (0.2539)	0.0319 (0.2674)	0.0467 (0.2907)
School homework/assignments	172.7832 (170.5803)	70.5778 (141.0918)	-16.5321 (145.9348)	66.8893 (184.679)	9.1494 (113.956)	-48.9428 (119.3631)	-221.08 (129.675)*
Extracurricular study	73.146 (87.6198)	105.0656 (126.4313)	139.663 (145.0118)	128.778 (117.3267)	118.9417 (152.6494)	156.0974 (172.9865)	142.9573 (166.9117)
Extracurricular arts	-27.1292 (58.6233)	21.7523 (78.6884)	27.832 (81.2304)	-11.799 (70.7734)	71.0345 (78.2661)	71.3917 (79.986)	53.2292 (64.7641)
Watching TV/on Internet	21.1873 (57.931)	-14.8781 (56.9382)	-12.1974 (45.5714)	-50.3979 (40.9458)	-175.2074 (92.0604)*	-155.4011 (85.9209)*	-44.4531 (24.1834)*
<i>Well-being:</i>							
Extracurricular sports	-39.037 (132.837)	-26.9118 (81.4377)	-29.4058 (112.4897)	-38.8977 (102.2805)	-20.3097 (103.5206)	-27.9703 (129.2387)	-12.8742 (127.8285)
Has cell phone	0.0251 (0.179)	0.0183 (0.2078)	0.0442 (0.1761)	0.025 (0.2047)	0.0915 (0.1891)	0.0296 (0.1686)	0.0548 (0.1907)
Vacation with parent in 1yr	0.0565 (0.1066)	0.0758 (0.1679)	0.0729 (0.1446)	0.0677 (0.1286)	0.0414 (0.1439)	0.0398 (0.1324)	0.0461 (0.1315)
Went to excursion/gallery	0.0244 (0.1305)	0.0331 (0.1426)	0.0367 (0.1265)	0.0249 (0.1534)	0.025 (0.1461)	0.0263 (0.123)	0.0592 (0.1575)
Sees friends > 2 times per week	-0.0554 (0.1709)	-0.0149 (0.2063)	-0.0809 (0.2124)	-0.0987 (0.2037)	0.0332 (0.2017)	0.0257 (0.2101)	0.0411 (0.2151)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	NO
<i>Linear trend</i>	YES	YES	YES	YES	YES	YES	YES
<i>Quadratic trend</i>	NO	YES	YES	NO	YES	YES	YES
<i>Cubic trend</i>	NO	YES	YES	NO	YES	YES	YES
<i>RLMS year controls</i>	NO	YES	YES	NO	YES	YES	NO
Num. of observations	965	965	965	676	676	676	337

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient. Error terms are clustered at regional level. Subscript (b) denotes 5% signif. after Bonferroni correction

6 Discussion and conclusion

This Chapter evaluates the impact of Maternity capital (MC) program child subsidy of 250,000 Rub (7,150 euros or 10,000 USD, in 2007) that was introduced on 1st January 2007 - unexpectedly to the vast majority of Russians - for giving birth to/adopting the 2nd and subsequent children. Since 2007 the reform made it possible for eligible Russian families to allocate these additional funds to improve family housing conditions, to sponsor children education, or to invest them in mother's retirement fund. The objective of this Chapter is to evaluate the impact of the MC reform on various health and well-being outcomes among children aged 6-8 (i.e. typical age of elementary school students in Russia), as well as and household-level consumption patterns. Using RLMS individual and household representative panel surveys from 2011 to 2017, we find through regression discontinuity models no significant difference in health, educational and well-being outcomes between children raised in MC claim eligible and ineligible families. However, an analysis conducted separately for female and male children points to the possibility of heterogeneous MC effects in terms of out-of-school socialization. As for the household spending patterns, MC eligible Russian families on average reported no significant shifts in household diet. In the meantime, certain general expenditure categories saw a significant increase (most notably, savings), and this effect may be heterogeneous with respect to the self-reported poverty status.

The results are robust to various polynomial functional forms, inclusion of child, mother and household-specific covariates. In addition, to relax the assumption on the functional form of the trend variable we tested non-parametric local linear estimator models (LOWESS), which produce results largely mirroring functional RDD estimation. Identical conclusions are reached with respect to 1st children, who were not at the origin of MC claim eligibility for their families, but who, according to MC conditions, could still benefit from the subsidy both directly (through paid education) or indirectly (through improved housing conditions, spill-over effects from 2nd children and other affected household members).

The contribution of this study is threefold. First, it concentrates on a middle-income transitioning country, for which very limited research with regards to pro-natalist policies is currently available. Second, it features a number of child-level outcome variables that, to the best of our knowledge, no other study had attempted to investigate. Lastly, the structure of the RLMS survey allowed us to incorporate household-level variables on the consumption of various food categories, which was never studied in the context of child subsidies.

Regression discontinuity relies on a set of restrictive assumptions, including the inability of participants to manipulate the assignment of MC near the cut-off date and the correctness of functional specification of timeline trend variables. Thanks to the fact that announcement of the reform was Sine the introduction of the MC program was completely

unexpected for the vast majority of Russians, the risk of strategic self-selection into Maternity Capital assignment near the cut-off date is negligible. To address the concerns over the correctness of functional specification, we tested RDD models with linear and polynomial trends, as well as non-parametric LOWESS models. Overall, the results produced by different specifications point to an absence of any impact of MC on analyzed child health, educational and well-being outcomes, as well as on household-level consumption patterns. However, there is weak and tentative evidence that girls may have responded better to the MC program eligibility in terms of out-of-school socialization, while poorer families did not benefit from the MC as fully as wealthier families did as far as improving housing conditions is concerned.

It should be stressed that, similarly to previous studies on MC relying on RLMS data (*Slonimczyk and Yurko (2014)*), model estimates reflect the impact of MC claim eligibility and not that of the actual use of MC subsidy by eligible households. Notably, the fact that MC subsidy could be split between many potential uses over several years and among multiple beneficiaries (i.e. children, mother herself, or household in general) makes it difficult to track the actual use of the MC subsidy. In addition, the question of MC use was only present in a few RLMS waves, making it even more problematic to employ in policy evaluation analysis.

Since RLMS data is of the survey type, wherein respondents provide their personal evaluations with regard to requested information, they are likely to be influenced by various factors related to their personality traits, past experiences and current socio-economic status. While not affecting the biasedness of RDD estimates, it likely brings additional and, plausibly, significant noise terms to outcome and covariate variables. This has the consequence of decreased efficiency of obtained estimates.

It should be noted that the fact that in some specifications we use 36 month windows around the eligibility cut-off date may result in self-selection of households based on unobserved characteristics. To correct for it, most tested models featured specifications in which the window widths were reduced to 12 and 24 months, and which did not alter the main results of the study. In addition, to alleviate this issue in models with 36 month window, we include a broad set of relevant child, mother and household covariates, which are likely to further significantly reduce the impact of this concern.

Child and household outcomes analyzed in this Chapter were likely to be affected by the MC eligibility through several channels. First, in purely financial terms, the MC subsidy provides a very sizable income supplement that can relax the household budget constraint for many years after eligibility rights were acquired. As a result, affected families can and, in theory, must respond to MC incentives by re-optimizing household spending behavior. In particular, households would need to make distributional decisions with regards to the share of the MC to be invested in children's human capital vs. family members' personal consumption and other household long- and short-term needs. Even in the latter case MC has the potential to positively impact children's well-being through wealth and

human capital spill-over effects within the household. Second, as intended by the reform, for a lot families the MC subsidy likely resulted in improved housing conditions and, arguably to a lesser extent, a better access to child education and childcare (see Section 2 for more details on the MC reform context). However, it is important to stress that, similar to previous studies on MC, due to data limitations this Chapter concentrates on the impact of the MC claim eligibility and not that of the actual use of the subsidy. MC subsidy was designed such that it could be spent on several household members, toward several eligible uses and continuously over a considerable time span. Lastly, the MC reform was accompanied by a broad public campaign aimed at encouraging Russian families to consider child well-being and development as major life priority. Apart from promoting general family-oriented values, it might have also raised public awareness of issues related to child well-being.

The conclusions of this Chapter provide an additional insight into the impact of pro-fertility reforms on an array of child outcomes and household consumption patterns. In particular, the fears that families may massively use MC subsidies to improve their financial standing with no regard to future well-being of their children does not seem to be supported by data. However, this Chapter provides suggestive evidence that the MC program may have affected male and female children differently, girls benefiting from the MC in terms of a better socialization. The likely mechanism behind it may stem from differences in psychological mechanisms boys and girls use to cope with changes in environment, which are discussed for example, in *Kling et al. (2005, 2007)*.

In addition, the fact that poorer families de facto have a more restricted set of options when deciding how to spend the MC subsidy may result in a decreased subsidy efficiency in places where this intervention is arguably most needed. In terms of policy implications, this consideration may require that MC policy design be revised such that families in straitened circumstances enjoy the same opportunity and MC spending flexibility as their more well-off peers. Plausibly in part to address this point, in 2018 in the Russian government allowed the MC subsidy to be used as a monthly allowance if the household can justify their acute need for this financial support. This can be further implemented, for example, through adjusting the subsidy amount with respect to the household income.

The results of this Chapter could be generalized to institutionally, demographically and economically close countries, most notably Belarus. Arguably to a lesser extent, these conclusions can be used while developing policy recommendations in other largely similar countries, which, however, may differ from Russia in terms of type of state institutions (most Eastern European members of EU with governments largely perceived as more democratic, such as Poland, Hungary, Bulgaria), prevalent societal norms with regards to family and religion (for instance, predominantly Muslim but economically similar Kazakhstan, Turkey and Malaysia), and in terms of the level of economic development and standards of living (for example, culturally close but, arguably, economically more challenged Ukraine). The fact the MC implementation was in various aspects similar to

Singaporean pro-fertility policies also makes the MC reform an important reference point while comparing the policy impacts in these two, otherwise different, countries. Finally, the topic of becomes even more relevant in light of an increased popularity of pro-natalist reforms introduced in several EU counties, most recently in Hungary, whose government announced in 2019 that, along with other measures of family support, it would introduce tax-waivers of up to \$36000 US for household with multiple children.

7 Appendixes

Appendix 1. RLMS questionnaire

Table 3.10 – RLMS survey question on child and household outcomes

Outcome variable:	RLMS survey question:	Values
Child health outcomes:		
Health score	How would evaluate the health of your child? Is it :	1 - Very good 2 - Good 3 - Average; neither bad nor good 4 - Bad 5 - Very bad
In good/excellent health	Idem. Note: value 1 if answers '1' or '2', 0 - otherwise	Sum of condition groups with answer 'yes'
Number of chronic conditions	Does your child have any chronic illness related to : heart, lungs, liver, kidneys, GI, spine, otorhinolaryngology, neurology, eyes, allergies or other chronic conditions	1 - Yes, 2 - No
Health problems in last 30d	In last 30 days, did your child have any health problems ? <i>Note: computed using R statistical package 'anthro' based on reported child weight, height and sex</i>	1 - almost all fives 2 - mostly fives and fours 3 - mostly fours 4 - mostly fours and threes 5 - mostly threes 6 - mostly threes and often twos
Educational achievement:		
Good excellent GPA	How would you evaluate you child's academic achievement (grades)? Note: value 1 if answers '1' or '2', 0 - otherwise	1 - almost all fives 2 - mostly fives and fours 3 - mostly fours 4 - mostly fours and threes 5 - mostly threes 6 - mostly threes and often twos
Child's time spending patterns:		
School homework/assignments	How much time does your child spend on doing school homework under parents' or adult relatives' supervision?	time spent by parents or adult relatives, in minutes per week
Extracurricular study	Before or after school, does your child study - programming and computer technology, such as learning to work with programs, programming languages and web technology - foreign language - other subjects, chosen by parent or child, to get additional knowledge in, for example, math, biology, history	time spent by child, in minutes per week
Extracurricular arts	Before or after school, does your child - practice music or drawing/painting - practice dancing, photography, theater and drama, other types of arts? - go to a young scientist club (in Russian: 'junyj tekhnik'), technical modelling club, naturalist club, practice handicraft, pottery, clay sculpting, or other type of technical and applied art activity?	Idem.
Watching TV/on Internet	Before or after school, does your child - watch TV, video, play video or computer games - browse Internet or a local web network	Idem.

Child well-being :			
Vacation with parent in 1yr	In the last 12 months, did anyone among child's parents spend a vacation with him/her?		1-Yes, 2-No
Went to excursions/gallery	In the last 12 months, did your child go to a theater, museum, exhibition, zoo, circus, or other cultural and entertainment events?		1-Yes, 2-No
Sees friends > 2 times a week	How often does your child meet his/her peers outside school or childcare facility? (Note: value 1 if answers '4' or '5', 0 - otherwise)		1: 1 to 3 times per month 2: 1 time per week 3: 2 times per week 4: 3-4 times per week 5: every day
Household consumption:	How much of the following products did your household buy in the last 7 days?		
Vegetables/legumes	Cucumbers (including pickled), tomatoes (including pickled), beets, carrots, zucchini, pumpkin, and other vegetables.		in kilograms per week
Fruit (fresh and canned)	Watermelon, melon (including dried and pickled), fruit and berry preserves, fresh berries, fresh fruit, dried fruit and berries, nuts and seeds		Idem.
Meat and poultry	Meat preserves, fresh beef and veal, fresh mutton and goat meat, pork, animal sub-products, and poultry		Idem.
Dairy	Milk preserves, powdered milk, fresh milk, sour milk products (kefir, yogurt etc.), sour cream and cream, cottage cheese, cheese		Idem.
Refined sugar	Refined sugar		Idem.
Candy and high-sugar treats	Ice cream, candy, chocolate, honey, jam, cookies, cakes, tarts, waffles, pryaniks, pastries		Idem.
Starches	White bread, rye bread, rice, pasta, potatoes, etc.		Idem.
Vodka and spirits (liquors)	Vodka, wine and liquors		in liters per week
Discretionary (luxury) expenditures	Vodka and spirits, cigarettes, private lessons and education, travel for pleasure, cinema/theater tickets, cosmetics and perfume, market stock, foreign currency purchase		In Rubles per month
Essential food items	All food items listed in RLMS, except for Vodka and spirits (liquors) and Candy and high-sugar treats		Idem.

Appendix 2. Descriptive statistics

Table 3.11 – Distribution of observations in the analytical sample, by child’s year of birth

	All sample	MC eligible	MC ineligible
birth 2004	130	0	130
birth 2005	198	0	198
birth 2006	145	0	145
birth 2007	167	167	0
birth 2008	178	178	0
birth 2009	221	221	0
birth 2010	11	11	0
Total:	1050	577	473

Table 3.12 – Distribution of observations in the analytical sample, by year of measurement

	All sample	MC eligible	MC ineligible
year 2011	114	0	114
year 2012	179	0	179
year 2013	179	50	129
year 2014	152	104	48
year 2015	178	175	3
year 2016	155	155	0
year 2017	93	93	0
Total:	1050	577	473

Table 3.13 – Distribution of observations in the analytical sample, by RLMS wave

	All sample	MC eligible	MC ineligible
RLMS wave 2011	125	0	125
RLMS wave 2012	174	0	174
RLMS wave 2013	184	54	130
RLMS wave 2014	141	100	41
RLMS wave 2015	185	182	3
RLMS wave 2016	149	149	0
RLMS wave 2017	92	92	0
Total:	1050	577	473

Table 3.14 – Descriptive statistics for whole analytical sample (RLMS 2011-2017 respondents with children born between 2004 and 2010 and aged 6-8 years, in households with two children), by period of birth

Variable	All sample				Born after Jan 1st 2007		Born before Jan 1st 2007	
	Mean	SD	Min	Max	Mean	SD	Mean	SD
<i>Child characteristics and outcomes:</i>								
Age	7.037	0.804	6	8	6.975***	0.81	7.099***	0.794
Sex (male)	0.528		0	1	0.561***		0.494***	
Second child	0.521		0	1	0.558***		0.484***	
Urban	0.643		0	1	0.676***		0.61***	
In good/excellent health	0.774		0	1	0.776		0.772	
Health score (1-best, 5-worst)	2.188	0.498	1	4	2.183	0.509	2.193	0.488
Health problem in last 30d	0.359		0	1	0.384**		0.335**	
Num. of chronic conditions	0.105	0.376	0	3	0.076***	0.306	0.135***	0.433
Good/Excellent GPA	0.669		0	1	0.645		0.685	
Went to excursion/gallery	0.819		0	1	0.831		0.807	
Sees friends > 2 times per week	0.699		0	1	0.684		0.714	
Has cell phone	0.667		0	1	0.632***		0.698***	
Vacation with parent in 1yr	0.747		0	1	0.76		0.734	
School homework/assignments	215.807	242.748	0	1500	227.839*	242.453	205.267*	242.677
Extracurricular study	21.534	74.674	0	1200	18.161*	55.937	24.579*	88.165
Extracurricular arts	133.032	206.198	0	1861	137.712	220.426	128.84	192.606
Watching TV/on Internet	122.193	75.945	0	720	121.458	75.021	122.833	76.785
<i>Mother characteristics:</i>								
Age	33.667	5.111	20	49	33.898*	5.108	33.432*	5.105
Single parent	0.135		0	1	0.127		0.143	
In good/excellent health	0.535		0	1	0.522		0.548	
Higher education diploma	0.392		0	1	0.423***		0.36***	
Ethnically other than Russian	0.101		0	1	0.092		0.11	
Household income, in 2011 prices	35.778	27.244	1.295	281.479	35.013	25.34	36.563	29.061
Alcohol cons. > 1 time per week	0.035		0	1	0.031		0.04	
<i>Household consumption:</i>								
Vegetables/legumes	2.331	4.997	0	80	2.152	4.17	2.512	5.708
Fruit (fresh and canned)	3.351	4.047	0	80	3.072***	3.276	3.626***	4.669
Meat and poultry	3.107	3.746	0	54	3.11	3.458	3.104	4.015
Dairy	5.956	4.5	0	31.5	6.024	4.422	5.888	4.577
Vodka and liquors (in liters)	0.18	0.657	0	10	0.143**	0.557	0.218**	0.743
Refined sugar	1.739	5.555	0	50	1.528*	4.657	1.952*	6.331
Candy and high-sugar treats	1.548	1.304	0	12	1.507	1.197	1.589	1.399
Starches	8.462	8.222	0	89	7.723***	7.238	9.212***	9.054
<i>Household spending:</i>								
Essential food items	8.835	5.969	0	72.078	8.38***	5.418	9.295***	6.449
Discretionary expenditures	5.353	14.205	0	224.183	5.06	12.706	5.649	15.578
Purchase of durable goods	0.386	0.487	0	1	0.345***	0.476	0.427***	0.495
Household savings	1.204	6.668	0	183.857	1.078	5.895	1.333	7.37
Loan payments	3.697	9.761	0	294.285	3.691	11.004	3.703	8.321
Number of observations	2015				1014		1001	

Legend: stars denote p-values for t-tests on equality of means (non-binary variables)/chi-square tests on equality of proportions (for binary variables): ***- 1 % sign., **-5% sign., *- 10% sign

Table 3.15 – Descriptive statistics for all sample (RLMS 2011-2017 respondents with second children born between 2004 and 2010 and aged 6-8 years, in households with two children), by urban/rural residence

Variable	All sample				Rural		Urban	
	Mean	SD	Min	Max	Mean	SD	Mean	SD
<i>Child characteristics:</i>								
Age	6.984	0.802	6	8	6.995	0.795	6.978	0.806
Sex (male)	0.521		0	1	0.566**		0.496**	
In good/excellent health	0.764		0	1	0.784		0.754	
Health score (1-best, 5-worst)	2.201	0.503	1	4	2.168	0.52	2.22	0.493
Health problem in last 30d	0.355		0	1	0.35		0.357	
Num. of chronic conditions	0.095	0.35	0	3	0.065**	0.258	0.112**	0.391
Good/Excellent GPA	0.683		0	1	0.689		0.678	
Went to excursion/gallery	0.812		0	1	0.679***		0.885***	
Sees friends >2 times per week	0.693		0	1	0.75***		0.66***	
Has cell phone	0.664		0	1	0.59***		0.71***	
Vacation with parent in 1yr	0.768		0	1	0.662***		0.825***	
School homework/assignments	200.703	223.864	0	1200	189.476	215.497	207.455	228.722
Extracurricular study	19.933	75.617	0	1200	10.242***	47.41	25.996***	88.332
Extracurricular arts	130.098	198.7	0	1328	113.37*	210.484	140.403*	190.582
Watching TV/on Internet	125.374	80.101	0	720	123.686	74.779	126.388	83.2
<i>Mother characteristics:</i>								
Age	36.544	4.399	25	49	35.499***	4.172	37.115***	4.417
Single parent	0.158		0	1	0.116***		0.181***	
In good/excellent health m	0.491		0	1	0.519		0.476	
Higher education diploma	0.374		0	1	0.21***		0.464***	
Ethnically other than Russian	0.095		0	1	0.103		0.09	
Poverty	0.43		0	1	0.488***		0.398***	
Household income, in 2011 prices	34.533	27.348	1.196	259.943	25.183***	17.794	39.549***	30.133
Alcohol cons. > 1 time per week	0.041		0	1	0.03		0.047	
<i>Household diet:</i>								
Vegetables/legumes	2.28	4.776	0	56	1.947	5.713	2.462	4.172
Fruit (fresh and canned)	3.488	4.453	0	80	2.784***	3.526	3.874***	4.848
Meat and poultry	3.134	3.79	0	52	2.735**	5.016	3.35**	2.9
Dairy	5.627	4.182	0	31.5	3.922***	3.778	6.551***	4.102
Vodka and liquors (in liters)	0.177	0.606	0	7	0.135*	0.516	0.2*	0.649
Refined sugar	1.699	5.399	0	50	2.857***	8.25	1.069***	2.623
Candy and high-sugar treats	1.567	1.324	0	12	1.684*	1.603	1.501*	1.135
Starches	9.751	20.295	0	405.5	11.852**	22.549	8.599**	18.864
<i>Household monthly spending:</i>								
Essential food items	8.885	5.98	0	72.078	8.134***	6.341	9.296***	5.736
Discretionary expenditures	4.868	12.894	0	224.183	4.611	17.578	5.008	9.411
Purchase of durable goods	0.365		0	1	0.364		0.365	
Household savings	1.426	6.541	0	86.455	0.907**	3.919	1.71**	7.589
Loan payments	3.713	8.414	0	162.588	2.488***	5.244	4.382***	9.656
Number of observations	1050				371		679	

Table 3.16 – Correlation matrix for main variables, whole sample

	Age	Sex (male)	Second child	In good/excellent health	Health score	Health problem in last 30d	Num. of chronic conditions	Good/Excellent GPA	Age m	Urban	Single parent	In good/excellent health m	Higher education diploma	Ethnically other than Russian	Household income	Alcohol cons. >1 time per week
Age	1	0.025	-0.074	0.045	-0.004	-0.032	-0.076	-0.183	0.05	0.031	0.026	-0.017	-0.017	0.004	0.046	-0.01
Sex (male)	0.025	1	-0.119	0.12	-0.086	0.042	0.103	-0.265	-0.042	-0.089	-0.003	0.088	0.008	0.143	-0.041	-0.059
Second child	-0.074	-0.119	1	-0.126	0.076	-0.004	0.007	-0.03	0.623	0.032	0.091	-0.1	-0.063	-0.027	0.04	0.077
In good/excellent health	0.045	0.12	-0.126	1	-0.906	-0.166	-0.305	0.12	-0.024	0.043	-0.005	0.323	-0.057	0.069	-0.022	-0.063
Health score	-0.004	-0.086	0.076	-0.906	1	0.166	0.285	-0.133	0.029	-0.042	0.004	-0.314	0.009	-0.047	0	0.026
Health problem in last 30d	-0.032	0.042	-0.004	-0.166	0.166	1	0.176	-0.034	0.005	0.005	0.062	-0.07	0.119	-0.006	0.014	-0.079
Num. of chronic conditions	0.103	0.007	0.007	0.285	0.285	0.176	1	-0.011	-0.025	0.008	0.065	-0.108	0.121	0.13	-0.118	-0.043
Good/Excellent GPA	-0.183	-0.265	-0.03	-0.133	-0.133	-0.034	-0.011	1	0.026	0.037	-0.095	0.044	0.195	-0.102	-0.027	0.019
Age mother	0.05	-0.042	0.623	-0.024	0.029	0.005	-0.025	0.026	1	0.162	-0.053	-0.127	0.199	-0.067	0.027	0.054
Urban	0.031	-0.089	0.032	0.043	-0.042	0.005	0.008	0.037	0.162	1	0.039	-0.006	0.157	-0.027	0.263	0.075
Single parent	0.026	-0.003	0.091	-0.005	0.004	0.062	0.065	-0.095	-0.053	0.039	1	0.04	-0.039	-0.066	-0.065	-0.048
In good/excellent health m	-0.017	0.088	-0.1	0.323	-0.314	-0.07	-0.108	0.044	-0.127	-0.006	0.04	1	-0.025	0.174	0.023	-0.108
Higher education diploma	-0.017	0.008	-0.063	-0.057	0.009	0.119	0.121	0.195	0.199	0.157	-0.039	-0.025	1	-0.121	0.138	0.008
Ethnically other than Russian	0.004	0.143	-0.027	0.069	-0.047	-0.006	0.13	-0.102	-0.067	-0.027	-0.066	0.174	-0.121	1	-0.115	-0.062
Household income	0.046	-0.041	0.04	-0.022	0	0.014	-0.118	-0.027	0.027	0.263	-0.065	0.023	0.138	-0.115	1	-0.009
Alcohol cons. >1 time per week	-0.01	-0.059	0.077	-0.063	0.026	-0.079	-0.043	0.019	0.054	0.075	-0.048	-0.108	0.008	-0.062	-0.009	1

Table 3.17 – Correlation matrix for main variables, whole sample (continued)

Age	1																																								
Sex (male)	0.025	1																																							
Second child	-0.074	-0.119	1																																						
Went to excursion/gallery	0.112	-0.009	-0.05	1																																					
Sees friends >2 times per week	0.015	0.014	0.022	0.102	1																																				
Has cell phone	0.087	-0.009	-0.089	0.102	0.032	1																																			
Vacation with parent in 1yr	-0.034	0.025	0.014	-0.116	-0.071	1																																			
School homework/assignments	0.025	0.029	0.091	0.053	0.093	0.001	1																																		
Extracurricular study	0.048	-0.002	0.03	0.093	0.093	0.008	0.058	1																																	
Extracurricular arts	-0.048	0.029	-0.091	-0.044	0.038	-0.017	0.134	0.055	1																																
Watching TV/on Internet	0.115	0.014	0.045	-0.151	-0.137	-0.004	-0.112	0.065	0.071	1																															
Age mother	0.05	-0.042	0.623	0.023	-0.028	-0.026	0.033	-0.063	0.098	0.05	1																														
Urban	0.031	-0.089	0.032	0.145	-0.108	-0.055	0.075	-0.006	0.086	-0.005	0.162	1																													
Single parent	0.026	-0.003	0.091	0.066	-0.041	-0.008	0.039	0.039	0.067	0.155	0.039	0.039	1																												
In good/excellent health m	-0.017	0.088	-0.1	-0.108	0.031	-0.012	0.06	-0.037	0.067	0.155	-0.053	0.039	0.04	1																											
Higher education diploma	-0.017	0.008	-0.063	0.16	-0.039	0.093	0.123	-0.039	0.087	-0.185	0.199	0.157	-0.039	0.04	1																										
Ethnically other than Russian	0.004	0.143	-0.027	-0.073	0.073	0.028	0.081	0.131	-0.039	-0.04	0.006	-0.027	-0.066	0.174	0.174	1																									
Household income	0.046	-0.041	0.04	0.153	-0.001	-0.067	0.036	-0.14	0.048	-0.084	0.027	0.263	-0.065	0.023	0.138	0.138	1																								
Alcohol cons. >1 time per week	-0.01	-0.059	0.077	0.059	-0.078	-0.035	-0.072	0.03	-0.033	0.095	0.038	0.054	0.075	-0.048	0.008	0.008	-0.062	1																							

Table 3.18 – Correlation matrix for main variables, whole sample (continued)

Age	1	0.025	-0.074	0.045	-0.004	0.042	-0.032	-0.076	-0.183	0.05	0.031	0.026	-0.017	-0.017	0.004	-0.041	-0.019	-0.037	-0.016	0.016	-0.003	0.001	0.041	-0.084	-0.066	
Sex (male)	0.025	1	-0.119	0.12	-0.086	0.042	0.042	0.103	-0.265	-0.042	-0.089	-0.003	0.088	0.008	0.143	-0.041	-0.059	0.065	0.021	0.04	0.016	-0.006	0.001	0.001	0.041	-0.084
Second child	-0.074	-0.119	1	-0.126	0.076	-0.004	0.007	-0.004	0.007	-0.003	0.032	0.091	-0.1	-0.063	0.027	0.04	0.077	0.019	0.058	-0.016	-0.178	0.028	-0.006	0.008	0.017	0.069
In good/excellent health	0.045	0.12	-0.126	1	-0.906	-0.004	-0.166	-0.305	0.12	-0.024	0.043	-0.005	0.323	-0.057	-0.069	-0.022	-0.063	-0.067	-0.03	0.117	-0.028	-0.026	0.061	0.02	-0.005	-0.001
Health score	-0.004	-0.086	0.076	-0.906	1	0.166	0.285	-0.133	0.029	-0.042	0.004	-0.004	-0.314	0.009	-0.047	0	0.026	0.081	0.025	-0.131	0.014	0.046	-0.118	0.088	0.06	
Health problem in last 30d	-0.032	0.042	-0.004	-0.166	0.166	1	0.176	-0.034	0.005	0.005	0.005	0.062	-0.07	0.119	-0.006	0.014	-0.079	0.08	0.035	0.023	0.074	0.03	-0.066	0.003	0.027	
Num. of chronic conditions	-0.076	0.103	0.007	-0.305	0.285	0.176	1	-0.011	-0.011	-0.025	0.008	0.065	-0.108	0.121	0.13	0.039	-0.043	0.063	0.019	0.055	0.142	0.053	-0.017	0.133	0.237	
Good/Excellent GPA	-0.183	-0.265	-0.03	0.12	-0.133	-0.034	-0.011	1	0.026	0.037	-0.095	0.044	0.195	-0.102	-0.027	0.019	0.002	0	0.118	0.008	0.03	-0.059	-0.119	0.006	-0.005	
Age mother	0.05	-0.042	0.623	0.029	0.005	0.005	0.025	0.026	1	0.162	0.037	-0.053	-0.127	0.199	-0.067	0.027	0.054	0.086	0.118	0.001	0.215	-0.014	-0.126	0.054	-0.125	
Urban	0.031	-0.089	0.032	0.043	-0.042	0.005	0.008	0.037	0.162	1	0.039	-0.006	0.157	-0.027	0.263	0.075	0.075	0.086	0.118	0.091	0.215	-0.014	-0.126	0.054	-0.125	
Single parent	0.026	-0.003	0.091	-0.005	0.004	0.062	0.065	0.095	-0.053	0.039	1	0.04	-0.039	-0.066	-0.065	-0.048	-0.039	-0.039	-0.099	-0.136	-0.115	0.048	-0.074	-0.053	-0.104	
In good/excellent health m	-0.017	0.088	-0.1	0.323	-0.314	-0.07	-0.108	0.044	-0.127	-0.006	0.04	1	-0.025	0.174	0.023	-0.108	-0.121	0.022	0.075	0.011	0.019	0.008	-0.125	-0.109	-0.109	
Higher education diploma	-0.017	0.008	-0.063	-0.057	0.009	0.119	0.121	0.195	0.199	0.157	-0.039	-0.025	1	-0.121	0.138	0.008	0.094	0.043	0.041	0.079	-0.015	-0.144	0.005	-0.013	-0.013	
Ethnically other than Russian	0.004	0.143	-0.027	0.069	-0.047	-0.006	0.13	-0.102	-0.067	-0.027	-0.066	0.174	-0.121	1	-0.115	-0.062	-0.074	-0.101	-0.078	-0.069	-0.023	0.056	-0.016	-0.055	-0.055	
Household income	0.046	-0.041	0.04	-0.022	0	0.014	-0.118	-0.027	0.027	0.263	-0.065	0.023	0.138	-0.115	1	-0.009	0.253	0.121	0.194	0.246	0.157	-0.055	-0.003	-0.035	-0.035	
Alcohol cons. >1 time per week	-0.01	-0.059	0.077	-0.063	0.026	-0.079	-0.043	0.019	0.054	0.075	-0.048	-0.108	0.008	-0.062	-0.009	1	0.021	-0.009	0.004	0.001	0.052	-0.028	0	0.007	0.007	
Vegetables/legumes	-0.019	0.065	0.019	-0.067	0.081	0.08	0.063	0.002	0.086	0.086	-0.039	-0.121	0.094	-0.074	0.253	0.021	1	0.062	0.21	0.304	0.046	0.038	0.135	0.338	0.338	
Fruit (fresh and canned)	-0.037	0.021	0.058	-0.03	0.025	0.035	0.019	0	0.118	0.118	-0.099	0.022	0.043	-0.101	0.121	-0.009	0.062	1	0.056	0.171	0.018	-0.065	0.092	0.244	0.244	
Meat and poultry	-0.016	0.04	-0.016	0.117	-0.131	0.023	0.055	0.018	0.009	0.091	-0.136	0.075	0.041	-0.078	0.194	0.004	0.21	0.056	1	0.302	0.214	0.167	0.225	0.06	0.06	
Dairy	-0.003	0.016	-0.178	-0.028	0.014	0.074	0.142	0.03	-0.056	0.215	-0.115	0.011	0.079	-0.069	-0.069	0.001	0.304	0.171	0.302	1	0.12	0.037	0.258	0.191	0.191	
Vodka and liquors (in liters)	0.001	-0.006	0.028	-0.026	0.046	0.03	0.053	-0.059	0.009	-0.014	0.048	0.019	-0.015	-0.023	0.157	0.052	0.046	0.018	0.214	0.12	1	-0.007	0.069	-0.005	-0.005	
Refined sugar	0.041	0.098	0.02	0.061	-0.118	-0.066	-0.017	-0.119	-0.005	-0.126	-0.074	0.008	-0.144	0.056	-0.055	-0.028	0.038	-0.065	0.167	0.037	-0.007	1	0.098	0.12	0.12	
Candy and high-sugar treats	-0.084	0.017	-0.005	-0.107	0.088	0.003	0.133	0.006	-0.04	0.054	-0.054	-0.125	-0.103	-0.016	-0.003	0	0.135	0.092	0.225	0.258	0.069	0.098	1	0.172	0.172	
Starches	-0.066	0.069	-0.001	-0.043	0.06	0.027	0.237	-0.07	-0.005	-0.125	-0.104	-0.109	-0.103	-0.055	-0.035	0.007	0.338	0.244	0.06	0.191	-0.005	0.12	0.172	1	1	

Appendix 3. Additional regression tables

Table 3.19 – Regression discontinuity models on 2nd child health outcomes (36 months' window at 1st January 2007 cut-off birth date)

	Good/excellent health (self-reported)	Health score (1-best 4-worst)	N. of chronic cond.	Health problem in 30d
post	0.024 (0.071)	0.0214 (0.0735)	-0.065 (0.083)	-0.054 (0.087)
Child's sex (male)	-0.051 (0.026)**	0.05 (0.03)*	0.007 (0.063)	-0.013 (0.069)
<i>Mother's characteristics:</i>				
Age	0.073 (0.04)	0.003 (0.006)	0.003 (0.003)	0.04 (0.02)**
Higher education	-0.07 (0.028)***	0.045 (0.036)	0.009 (0.02)	-0.103 (0.894)
Ethnicity other than Russian	0.028 (0.037)	-0.069 (0.051)	0.006 (0.04)	-0.099 (0.031)
In good excellent health	0.25 (0.002)***	-0.031 (0.003)***	-0.074 (0.021)***	0.042 (0.031)
<i>Household characteristics:</i>				
Single parent	-0.004 (0.038)	0.001 (0.004)	0.049 (0.037)	0.036 (0.049)
Poverty	-0.051 (0.027)*	0.044 (0.032)	0.005 (0.023)	0.066 (0.032)**
Urban	-0.027 (0.021)	0.025 (0.037)	0.067 (0.028)**	0.015 (0.052)
Intercept	0.81 (0.04)***	0.78 (0.26)***	2.16 (0.05)***	-0.074 (0.021)***
<i>Other controls</i>				
<i>Linear trend</i>	NO	YES	NO	YES
<i>Quadratic trend</i>	YES	YES	YES	YES
<i>Cubic trend</i>	YES	YES	YES	YES
<i>RLMS gear controls</i>	NO	YES	YES	YES
N. observations	NO	YES	NO	NO
R-squared	1029	1019	1021	1011
R-squared adjusted	0.018	0.13	0.014	0.064
	0.014	0.112	0.01	0.043

Legend: ***- 1% sign., **-5% sign., *- 10% sign. Coefficient std. errors s are given in parentheses under the coefficient. Error terms are clustered at regional level

Table 3.20 – Regression discontinuity estimates of impact of MC on 2nd child, by child gender (24 and 36 months' window at 1st January 2007 cut-off birth date)

outcome	Female						Male					
	RDD36m1 (1)	RDD36m2 (2)	RDD36m3 (3)	RDD24m1 (4)	RDD24m2 (5)	RDD24m3 (6)	RDD36m1 (7)	RDD36m2 (8)	RDD36m3 (9)	RDD24m1 (10)	RDD24m2 (11)	RDD24m3 (12)
In good/excellent health	-0.0265 (0.0718)	-0.0122 (0.0993)	-0.0442 (0.1032)	-0.0312 (0.0849)	0.0496 (0.1169)	-0.0283 (0.1224)	0.0232 (0.0889)	0.1237 (0.1215)	0.1319 (0.1159)	0.1859 (0.109)*	0.015 (0.1506)	0.0034 (0.1412)
Health score (1-best, 4-worst)	-0.035 (0.0899)	-0.0447 (0.1257)	0.0218 (0.1266)	-0.0322 (0.1062)	-0.1436 (0.1472)	-0.0105 (0.1472)	-0.0435 (0.0962)	-0.1759 (0.1304)	-0.1635 (0.1232)	-0.246 (0.1151)**	-0.0437 (0.1583)	-0.0205 (0.1478)
Num. of chronic conditions	0.0749 (0.0539)	0.0639 (0.0713)	0.1332 (0.0815)	0.0958 (0.0618)	0.0168 (0.0768)	0.0808 (0.0905)	-0.0578 (0.0938)	-0.2326 (0.1537)	-0.2507 (0.148)*	-0.1639 (0.1163)	-0.1983 (0.1978)	-0.1885 (0.1863)
Health problem in last 30d	0.0733 (0.0921)	-0.0213 (0.1301)	-0.0192 (0.1393)	0.0002 (0.1157)	0.0543 (0.1631)	0.0494 (0.1752)	0.0017 (0.0954)	-0.1781 (0.1367)	-0.1325 (0.1462)	-0.1241 (0.1188)	-0.0919 (0.1775)	-0.0234 (0.1941)
Good/Excellent GPA	0.0632 (0.0969)	0.0883 (0.11)	0.1055 (0.1267)	0.0706 (0.1069)	0.191 (0.1624)	0.2548 (0.1973)	-0.2054 (0.1415)	-0.1709 (0.2107)	-0.2841 (0.2025)	-0.0784 (0.1873)	0.1086 (0.2469)	-0.0374 (0.2397)
School homework/assignments	-13.9998 (48.5745)	-11.6006 (69.7878)	-20.3836 (72.158)	-49.5319 (58.427)	18.7272 (92.2992)	1.218 (89.4472)	88.9505 (26.279)	-4.8824 (92.9281)	-52.0691 (90.6138)	41.0816 (73.3342)	62.8191 (115.891)	7.5579 (113.1081)
Extracurricular study	2.2104 (11.8795)	6.3263 (14.6523)	16.218 (16.3448)	7.2476 (14.0824)	15.8468 (14.7513)	9.438 (16.297)	12.6437 (15.8947)	12.7914 (22.1921)	10.4663 (23.11959)	8.738 (21.0879)	7.9156 (21.0879)	6.8809 (23.8225)
Extracurricular arts	-15.1898 (38.5521)	100.4622 (50.1207)**	121.5929 (51.5517)**	88.9732 (45.472)*	153.2374 (59.957)**	139.2381 (60.8059)**	-35.2147 (21.8555)	-40.2545 (29.7252)	-41.7811 (33.8823)	-18.1595 (26.5631)	-47.5248 (36.8238)	-36.0247 (643.3956)
Watching TV/on Internet	17.2917 (19.7054)	15.7901 (26.9243)	28.8842 (27.4254)	14.168 (23.7587)	49.7926 (33.8843)	57.0068 (32.6266)*	18.3718 (18.8048)	-0.1085 (28.8958)	22.2032 (28.7393)	10.4961 (22.7162)*	60.0704 (32.4716)*	79.8934 (33.5191)**
Has cell phone	-0.0753 (0.1048)	0.0099 (0.149)	0.0777 (0.1646)	-0.05 (0.1294)	0.1338 (0.1883)	0.1588 (0.2084)	-0.1675 (0.1022)	-0.0351 (0.1586)	-0.0336 (0.1575)	-0.0301 (0.1243)	-0.0406 (0.2225)	-0.04256 (0.2424)
Vacation with parent in 1yr	-0.0456 (0.0813)	0.1149 (0.1133)	0.1315 (0.125)	0.0459 (0.0974)	0.2917 (0.1453)**	0.2297 (0.161)	-0.2332 (0.0854)**	-0.1651 (0.1258)	-0.1803 (0.1222)	-0.0964 (0.1057)	0.0233 (0.1631)	-0.0059 (0.1591)
Went to excursion/gallery	-0.0543 (0.0767)	-0.0755 (0.1125)	-0.0676 (0.1139)	-0.0643 (0.0981)	-0.2368 (0.1508)	-0.2201 (0.1521)	0.0239 (0.0741)	-0.0249 (0.1052)	-0.0115 (0.1038)	0.0186 (0.0929)	0.0318 (0.1324)	0.0653 (0.1303)
Sees friends > 2 times per week	0.1416 (0.093)	0.3059 (0.1239)**	0.2547 (0.135)*	0.2288 (0.1097)**	0.1984 (0.1625)	0.1612 (0.1795)	-0.1621 (0.0986)	-0.0937 (0.1428)	-0.1512 (0.1469)	-0.1077 (0.1244)	-0.1003 (0.1864)	-0.1358 (0.1981)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	YES
<i>Linear trend</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Quadratic trend</i>	NO	YES	YES	NO	YES	YES	NO	YES	NO	YES	YES	YES
<i>Cubic trend</i>	NO	YES	YES	NO	YES	YES	NO	YES	NO	YES	YES	YES
<i>RLMS year controls</i>	NO	YES	YES	NO	YES	YES	NO	YES	NO	YES	YES	YES
Num. obs.	503	503	503	347	347	347	547	547	547	360	360	360

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient

Table 3.21 – Regression discontinuity estimates of the MC impact on the 2nd child, by rural/urban residence (24 and 36 months’ window at 1st January 2007 cut-off birth date)

Outcome	Rural areas						Urban					
	RDD36m1	RDD36m2	RDD36m3	RDD24m1	RDD24m2	RDD24m3	RDD36m1	RDD36m2	RDD36m3	RDD24m1	RDD24m2	RDD24m3
In good/excellent health	-0.0316 (0.0944)	-0.0243 (0.127)	-0.0749 (0.1248)	0.017 (0.1126)	0.1598 (0.1503)	0.0386 (0.1482)	-0.0244 (0.0684)	0.0446 (0.0928)	0.0625 (0.0965)	0.0349 (0.0815)	-0.0809 (0.1139)	-0.0247 (0.1149)
Health score (1-best, 4-worst)	0.016 (0.1102)	-0.0268 (0.1524)	0.0595 (0.1498)	-0.0657 (0.1305)	-0.162 (0.1772)	0.0052 (0.1757)	-0.0123 (0.0796)	-0.0773 (0.1052)	-0.0902 (0.1078)	-0.0707 (0.0921)	0.0146 (0.1277)	-0.0399 (0.126)
Num. of chronic conditions	0.0033 (0.0593)	-0.0114 (0.0799)	-0.0014 (0.0814)	-0.006 (0.0706)	-0.0314 (0.0983)	0.0017 (0.1029)	0.0463 (0.0667)	-0.0087 (0.1022)	-0.087 (0.1092)	-0.0039 (0.0771)	-0.0542 (0.1289)	-0.1095 (0.1341)
Health problem in last 30d	0.0091 (0.1078)	-0.1348 (0.1482)	-0.0643 (0.1576)	-0.0955 (0.1336)	-0.0037 (0.2009)	0.0331 (0.2105)	0.0563 (0.0799)	-0.0883 (0.1128)	-0.1143 (0.1192)	-0.0478 (0.0978)	-0.0729 (0.1432)	-0.0854 (0.1547)
Good/Excellent GPA	-0.0787 (0.1537)	-0.0138 (0.2338)	-0.0588 (0.2447)	-0.0192 (0.1999)	-0.034 (0.2971)	-0.0375 (0.3277)	-0.021 (0.1077)	-0.068 (0.1582)	-0.0463 (0.1502)	-0.0442 (0.1338)	0.02134 (0.1864)	0.02765 (0.1778)
School homework/assignments	-3.2342 (63.121)	-22.1215 (105.68)	-23.3886 (117.8343)	8.6946 (82.8082)	-90.0761 (129.9201)	-109.2994 (142.6694)	57.2528 (44.9258)	-10.1735 (65.7653)	-52.0732 (61.8322)	-19.491 (55.2281)	119.3037 (78.661)	95.4176 (74.6461)
Extracurricular study	-6.0291 (14.7141)	22.232 (18.2611)	31.8327 (20.7252)	11.5917 (18.4146)	36.7264 (21.6554)*	34.6925 (22.9573)	11.4475 (12.8104)	2.8141 (17.4568)	6.3068 (17.3016)	2.753 (10.4663)	4.263 (15.2623)	8.2617 (16.7494)
Extracurricular arts	-87.5699 (58.5034)	20.4538 (72.7684)	89.7056 (92.217)	-9.4443 (75.9498)	89.4175 (89.0791)	127.9131 (106.6239)	104.794 (36.7646)***	-32.8838 (47.2515)	25.3873 (43.2364)	-37.0765 (43.0259)	10.9145 (56.2297)	72.7162 (52.645)
Watching TV/on Internet	30.1416 (22.401)	2.2805 (34.5073)	23.9382 (38.5549)	21.2278 (28.67)	88.5499 (42.2484)**	104.794 (47.5544)**	16.7406 (16.2413)	20.9514 (21.8146)	27.2097 (22.2183)	14.1685 (18.2618)	53.3413 (26.6148)**	42.1719 (25.3161)*
Has cell phone	-0.1165 (0.1264)	-0.0428 (0.1583)	-0.0474 (0.1799)	-0.021 (0.1551)	-0.0596 (0.2228)	-0.0599 (0.2275)	-0.1447 (0.087)*	-0.0159 (0.1315)	0.0514 (0.1408)	-0.1063 (0.1074)	0.03 (0.1717)	0.0764 (0.1847)
Vacation with parent in 1yr	-0.03686 (0.1127)	-0.0245 (0.163)	-0.1271 (0.1568)	-0.0284 (0.1384)	0.184 (0.2024)	0.1479 (0.1926)	-0.0298 (0.0651)	0.0705 (0.0952)	0.007 (0.1025)	0.1394 (0.08)*	0.1188 (0.1211)	0.0613 (0.132)
Went to excursion/gallery	0.0733 (0.1103)	0.022 (0.1537)	0.051 (0.1608)	0.0486 (0.1365)	-0.1842 (0.1979)	-0.2136 (0.1983)	-0.0508 (0.0503)	-0.0746 (0.0719)	-0.0522 (0.076)	-0.0316 (0.0621)	-0.046 (0.0932)	-0.0405 (0.101)
Sees friends \$>2\$ times per week	-0.1447 (0.1078)	0.0978 (0.1465)	0.0209 (0.1563)	-0.0348 (0.134)	0.0551 (0.2088)	-0.0295 (0.215)	0.0443 (0.0848)	0.0193 (0.1172)	-0.0153 (0.1235)	0.0735 (0.1027)	-0.1332 (0.1443)	-0.1252 (0.1491)
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	YES
<i>Linear trend</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Quadratic trend</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>Cubic trend</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>R/LMS gear controls</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES
Num. obs.	371	371	371	268	268	268	679	679	679	439	439	439

Legend: stars denote p-values for t-tests on equality of means (non-binary variables)/chi-square tests on equality of proportions (for binary variables): ***- 1 % sign., **-5% sign., *- 10% sign

Table 3.22 – Regression discontinuity estimates of the MC impact household diet and consumption, by rural/urban residence (24 and 36 months' window at 1st January 2007 cut-off birth date)

Outcome	Rural areas						Urban					
	RDD36m1	RDD36m2	RDD36m3	RDD24m1	RDD24m2	RDD24m3	RDD36m1	RDD36m2	RDD36m3	RDD24m1	RDD24m2	RDD24m3
<i>Household diet:</i>												
Vegetables/legumes	1.0254 (1.6807)	0.5419 (2.588)	0.8584 (2.8102)	0.5432 (2.2919)	-0.4025 (3.713)	0.2736 (3.632)	0.4262 (0.7744)	0.5819 (1.1076)	1.1795 (0.9935)	0.3895 (0.9628)	0.4677 (1.3889)	1.3823 (1.1439)
Fruit (fresh and canned)	-0.7192 (0.744)	0.2547 (1.095)	1.0505 (1.2114)	0.0296 (0.9746)	-1.5738 (1.4203)	-1.3153 (1.4565)	-0.0707 (0.7869)	1.3591 (0.7422)*	1.302 (0.7491)*	-0.0212 (0.7864)	1.8632 (1.2331)	2.1137 (1.3049)
Meat and poultry	0.2706 (1.3962)	1.8848 (1.8916)	2.1391 (2.099)	2.0657 (2.0397)	1.5292 (2.5251)	0.8919 (2.6359)	0.0704 (0.5389)	0.8285 (0.723)	0.8112 (0.6397)	0.2616 (0.6655)	0.8713 (0.8811)	1.191 (0.7286)
Dairy	-0.6399 (0.9024)	-1.4471 (1.2849)	-1.1239 (1.1569)	-1.6392 (1.1227)	-0.6859 (1.509)	-1.034 (1.3966)	0.9253 (0.7342)	0.8923 (1.0246)	0.3064 (1.0076)	0.2363 (0.8895)	0.8571 (1.2529)	0.7506 (1.1894)
Vodka and liquors (in liters)	-0.0605 (0.081)	-0.0084 (0.1241)	0.0334 (0.1461)	-0.0926 (0.0952)	-0.0323 (0.1729)	-0.1498 (0.1808)	-0.2001 (0.0996)**	-0.1995 (0.1301)	-0.0786 (0.1379)	-0.1426 (0.1101)	-0.1668 (0.1688)	-0.0903 (0.1672)
Refined sugar	-0.5638 (1.4975)	1.5168 (1.6422)	0.4038 (1.7425)	-0.0731 (1.5006)	1.9264 (1.9597)	-0.1021 (2.0277)	0.2575 (0.2758)	0.5552 (0.5613)	0.6812 (0.5613)	-0.0051 (0.2937)	0.1953 (0.4308)	0.2318 (0.4616)
Candy and high-sugar treats	0.2201 (0.3743)	-0.3555 (0.5209)	-0.3254 (0.5697)	-0.3526 (0.4711)	-0.7955 (0.6819)	-0.7458 (0.6697)	0.2528 (0.217)	0.4751 (0.3088)	0.5372 (0.311)*	0.2915 (0.2745)	0.5668 (0.3958)	0.6664 (0.4099)
Starches	12.2315 (5.3632)**	9.6354 (7.1911)	7.1876 (8.0586)	9.2639 (6.2857)	5.8737 (9.9018)	4.6341 (10.1397)	0.8948 (2.0833)	7.3662 (3.9385)*	7.4344 (3.9289)*	3.225 (2.0222)	6.3696 (6.1835)	6.721 (6.6966)
<i>Household consumption:</i>												
Essential food items	1.7023 (1.4234)	1.0913 (1.7952)	0.7158 (2.0288)	1.7358 (1.7105)	-0.5031 (2.2908)	-2.0859 (2.4428)	1.4147 (1.2143)	2.95 (1.6609)*	3.7107 (1.2217)***	1.7819 (1.5813)	3.0414 (2.1344)	4.5877 (1.4485)***
Discretionary expenditure	0.5187 (2.3231)	4.9141 (3.5111)	3.401 (3.2575)	2.6182 (2.5078)	4.2291 (2.8743)	1.1707 (2.9445)	1.5936 (1.4959)	0.8573 (1.803)	2.3035 (1.4162)	1.0648 (1.4228)	1.622 (2.1969)	3.2364 (2.0542)
Purchase of durable goods	-0.1488 (0.11)	-0.2957 (0.1578)*	-0.1779 (0.1607)	-0.2401 (0.1377)*	-0.1289 (0.2068)	-0.1022 (0.204)	-0.0017 (0.0794)	0.0784 (0.1092)	0.1174 (0.1158)	0.043 (0.097)	0.1178 (0.1353)	0.1028 (0.149)
Savings	1.3482 (1.1595)	1.0759 (1.6954)	1.2082 (1.8801)	0.8515 (1.5393)	1.9012 (2.4188)	2.0385 (2.3504)	2.2606 (1.1824)*	2.5589 (1.5984)	3.9405 (1.6189)**	2.2761 (1.4687)	1.9204 (1.9304)	2.9781 (1.984)
Loan payments	-0.7734 (1.1705)	-1.3067 (1.4745)	-0.6439 (1.7895)	-0.6005 (1.1918)	1.0819 (1.7097)	1.2338 (1.8327)	0.6297 (1.3789)	3.172 (1.6073)**	3.0401 (1.7251)*	2.1971 (1.4349)	5.0249 (1.9615)**	4.4111 (1.9765)**
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	YES
<i>Linear trend</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Quadratic trend</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>Cubic trend</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>RLLMS year controls</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES
Num. obs.	371	371	371	268	268	268	679	679	679	439	439	439

Legend: stars denote p-values for t-tests on equality of means (non-binary variables)/chi-square tests on equality of proportions (for binary variables): ***- 1 % sign., **-5% sign., *- 10% sign

Table 3.23 – Regression discontinuity estimates of the MC impact on the 2nd child, by poverty status (24 and 36 months’ window at 1st January 2007 cut-off birth date)

Outcome	Poor (self-reported)										Not poor (self-reported)									
	RDD36m1	RDD36m2	RDD36m3	RDD24m1	RDD24m2	RDD24m3	RDD36m1	RDD36m2	RDD36m3	RDD24m1	RDD24m2	RDD24m3								
In good/excellent health	-0.0343 (0.0858)	0.0222 (0.1183)	0.053 (0.1259)	0.0538 (0.102)	-0.0676 (0.1505)	0.0453 (0.162)	-0.0065 (0.0736)	0.0023 (0.0956)	-0.0134 (0.0958)	0.0127 (0.0893)	0.0253 (0.1155)	-0.0552 (0.1113)								
Health score (1-best, 4-worst)	0.029 (0.1022)	-0.034 (0.1437)	-0.0439 (0.1478)	-0.0736 (0.1204)	0.066 (0.1773)	-0.0501 (0.1807)	-0.0424 (0.0839)	-0.0412 (0.1128)	-0.0043 (0.1075)	-0.0669 (0.0996)	-0.0787 (0.1299)	0.0204 (0.1215)								
Num. of chronic conditions	0.1186 (0.0608)*	0.0969 (0.0719)	0.082 (0.0837)	0.0664 (0.0603)	0.0989 (0.086)	0.0375 (0.0913)	-0.0407 (0.0702)	-0.1383 (0.1158)	-0.1453 (0.117)	-0.0597 (0.085)	-0.1275 (0.1277)	-0.1267 (0.1311)								
Health problem in last 30d	0.0328 (0.101)	-0.0261 (0.1407)*	-0.022 (0.1474)	-0.154 (0.1229)	-0.0222 (0.1857)	-0.0245 (0.203)	0.0473 (0.0843)	-0.0137 (0.1181)	0.0359 (0.1256)	-0.0182 (0.1029)	0.0583 (0.1457)	0.1305 (0.1523)								
Good/Excellent GPA	-0.0311 (0.1502)	-0.0941 (0.2488)	-0.1517 (0.2693)	0.0233 (0.1926)	-0.0573 (0.3221)	-0.0502 (0.3443)	-0.0228 (0.1086)	-0.257 (0.1476)*	-0.2642 (0.1382)*	-0.1509 (0.1377)	-0.029 (0.1897)	-0.064 (0.184)								
School homework/assignments	24.2283 (49.9502)	40.6785 (84.8179)	11.781 (89.1123)	32.6136 (66.6567)	124.1933 (112.2649)	118.8139 (117.2068)	61.8987 (52.0342)	-40.3738 (71.7276)	-69.7657 (70.3357)	-7.5921 (61.7487)	9.0985 (85.5834)	-5.3638 (87.2985)								
Extracurricular study	-4.7238 (18.3998)	-13.917 (19.9741)	-5.8461 (23.3549)	-13.6437 (16.3319)	28.936 (19.6381)	21.8985 (20.4683)	17.1951 (11.2992)	30.5666 (17.0366)*	28.4639 (17.9736)	19.2537 (11.8636)	14.571 (16.6544)	10.1997 (17.2318)								
Extracurricular arts	-121.2918 (59.5767)**	13.768 (74.853)	94.8302 (77.0257)	10.7985 (26.5181)	115.8359 (128.7456)	185.3668 (85.2642)	-74.8485 (99.5638)*	3.2928 (34.7689)**	3.2928 (50.6754)	-47.0247 (51.0021)	30.1923 (64.0215)	58.6528 (61.3233)								
Watching TV/on Internet	32.0694 (23.6328)	27.5246 (36.6516)	57.1392 (37.9347)	26.5181 (27.8564)	128.7456 (49.9425)**	137.8889 (50.4668)**	7.3849 (15.3379)	-0.5042 (21.0721)	8.5968 (21.0319)	2.0015 (17.8829)	20.2867 (25.0205)	26.2692 (22.4859)								
Has cell phone	-0.0508 (0.1129)	-0.1002 (0.1887)	0.0105 (0.1987)	-0.157 (0.1427)	0.0553 (0.2425)	0.0693 (0.2701)	0.0693 (0.0951)	-0.0224 (0.1266)	-0.1821 (0.1346)	-0.1714 (0.1146)	-0.03172 (0.1634)	-0.0225 (0.1767)								
Vacation with parent in 1yr	-0.1147 (0.0948)	0.0284 (0.137)	-0.056 (0.1375)	-0.0117 (0.117)	0.2407 (0.1851)	0.1372 (0.1847)	-0.0187 (0.0745)*	-0.0187 (0.1067)	-0.067 (0.1121)	-0.0726 (0.0902)	0.0348 (0.1364)	0.0589 (0.1466)								
Went to excursion/gallery	-0.0504 (0.075)	-0.0901 (0.1036)	-0.1267 (0.1006)	-0.0526 (0.0895)	-0.1523 (0.1287)	-0.0253 (0.1398)	-0.0132 (0.0705)	-0.0361 (0.1026)	-0.0361 (0.1016)	-0.0068 (0.0878)	-0.0845 (0.1348)	0.0198 (0.1357)								
Sees friends > 2 times per week	-0.1191 (0.1013)	0.1796 (0.1411)	0.163 (0.1366)	0.0424 (0.1202)	0.0993 (0.191)	0.0667 (0.1885)	0.0374 (0.0919)	-0.0136 (0.1292)	-0.074 (0.1385)	0.0447 (0.1128)	-0.1088 (0.1633)	-0.0573 (0.1757)								
Other controls	NO	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	YES								
Linear trend	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES								
Quadratic trend	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES								
Cubic trend	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES								
RLMs year controls	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES								
Num. obs.	449	449	449	283	283	283	596	596	596	420	420	420								

Legend: stars denote p-values for t-tests on equality of means (non-binary variables)/chi-square tests on equality of proportions (for binary variables): ***- 1% sign., **-5% sign., *- 10% sign

Table 3.24 – Regression discontinuity estimates of the MC impact on household diet and consumption, by self-reported poverty status (24 and 36 months’ window at 1st January 2007 cut-off birth date)

outcome	Poor (self-reported)						Not poor (self-reported)					
	RDD36m1 (1)	RDD36m2 (2)	RDD36m3 (3)	RDD24m1 (4)	RDD24m2 (5)	RDD24m3 (6)	RDD36m1 (7)	RDD36m2 (8)	RDD36m3 (9)	RDD24m1 (10)	RDD24m2 (11)	RDD24m3 (12)
<i>Household diet:</i>												
Vegetables/legumes	0.657 (1.4992)	-1.0604 (2.3549)	-0.2598 (2.2261)	-0.5866 (2.0376)	-2.487 (3.5012)	-1.1585 (3.2991)	-0.019 (0.7328)	0.9158 (0.9504)	1.4802 (1.0721)	0.4451 (0.7522)	0.9902 (1.2189)	1.8653 (1.3023)
Fruit (fresh and canned)	-1.2255 (0.752)	-0.2371 (1.1637)	-0.0338 (1.1445)	-0.8404 (0.9659)	-1.0667 (1.4906)	-1.3547 (1.4644)	0.6935 (0.8162)	1.636 (0.7739)**	2.2707 (0.8262)**	0.5857 (0.8403)	2.1343 (1.124)*	3.1636 (1.4593)**
Meat and poultry	0.7884 (1.1102)	1.3586 (1.2884)	2.0513 (1.1867)*	1.4636 (1.4624)	2.6851 (1.5813)*	3.0029 (1.5075)**	-0.4663 (0.5056)	0.9543 (0.8495)	1.0314 (0.8843)	0.2595 (0.7228)	0.2507 (0.9109)	0.17 (0.8716)
Dairy	0.4963 (1.0184)	1.0105 (1.5109)	0.6681 (1.2168)	0.1012 (1.2811)	2.7076 (1.9018)	1.9991 (1.418)	0.2312 (0.7618)	-0.7916 (1.0431)	-0.4408 (1.0212)	-0.9799 (0.9199)	-0.8883 (1.2767)	-0.7986 (1.2204)
Vodka and liquors (in liters)	0.0391 (0.0895)	0.2901 (0.1043)**	0.3265 (0.1294)**	0.1753 (0.0765)**	0.2908 (0.1509)*	0.3085 (0.1821)*	-0.2917 (0.1026)**	-0.4254 (0.142)**	-0.3424 (0.1523)**	-0.3486 (0.1201)**	-0.3912 (0.1785)**	-0.3616 (0.1839)*
Refined sugar	-0.5923 (0.4744)	0.2395 (0.8993)	0.501 (0.8115)	-0.312 (0.4335)	-0.5602 (0.7024)	-0.2284 (0.769)	0.8117 (0.8714)	1.1296 (1.0804)	0.5304 (1.0338)	0.3259 (0.9596)	1.263 (1.184)	-0.1667 (1.2376)
Candy and high-sugar treats	0.1327 (0.3609)	-0.3143 (0.5486)	-0.0873 (0.5675)	-0.4831 (0.4913)	-0.3259 (0.7793)	-0.0428 (0.7903)	0.4557 (0.2224)**	0.6692 (0.304)**	0.7487 (0.3212)**	0.4658 (0.2544)*	0.6004 (0.3006)*	0.7674 (0.3825)**
Starches	2.1376 (4.5853)	9.8426 (7.0685)	12.9607 (7.3801)*	5.4426 (4.4054)	14.9038 (10.9911)	18.4187 (10.906)*	5.5127 (1.9393)**	4.6784 (2.5295)*	2.1679 (1.96)	3.5764 (2.5268)	-2.5011 (4.5036)	-4.5469 (3.6774)
<i>Household consumption (per month):</i>												
Essential food items	-0.261 (1.8801)	-0.5241 (2.5843)	1.8105 (1.6294)	-1.0775 (2.5222)	1.4681 (3.4532)	3.406 (2.0821)	2.4581 (0.9873)**	4.2615 (1.3792)**	4.0747 (1.3898)**	3.393 (1.1816)**	3.037 (1.7116)*	2.7614 (1.6136)*
Discretionary expenditure	0.2049 (1.7088)	5.1471 (3.3688)	4.9025 (3.6686)	2.0398 (1.8355)	6.4413 (4.2166)	6.4069 (5.6911)	2.083 (2.0125)	2.3762 (2.413)	4.077 (1.8822)**	3.2249 (1.9977)	2.3001 (2.621)	4.0524 (2.5386)
Purchase of durable goods	-0.2101 (0.0954)**	-0.1719 (0.1328)	-0.1313 (0.1389)	-0.2552 (0.1179)**	-0.1058 (0.1727)	-0.0692 (0.1863)	0.0813 (0.089)	0.0803 (0.1265)	0.1592 (0.1301)	0.1344 (0.1096)	0.2024 (0.1613)	0.2052 (0.1676)
Savings	0.1298 (0.5549)	-1.1622 (0.9077)	-1.6897 (1.1164)	-1.01 (0.9656)	0.1187 (1.3219)	-1.0335 (1.3471)	4.038 (1.6017)**	4.8929 (2.3248)**	5.957 (2.3814)**	4.6161 (1.966)**	3.9118 (2.9212)	4.5721 (2.9917)
Loan payments	-0.8878 (1.6796)	1.0654 (1.6832)	0.1334 (1.8483)	-0.1136 (1.4366)	3.1783 (2.7951)	-0.5487 (2.0788)	0.8416 (1.3897)	2.3739 (1.7665)	3.4514 (1.7632)*	2.3924 (1.592)	4.1858 (1.9473)**	4.8858 (2.0483)**
<i>Other controls</i>	NO	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	YES
<i>Linear trend</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>Quadratic trend</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>Cubic trend</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES
<i>RLMS year controls</i>	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES
Num. obs.	449	449	449	283	283	283	596	596	596	420	420	420

Legend: ***- 1 % sign., **-5% sign., *- 10% sign. Coefficient std. errors are given in parentheses under the coefficient

General Conclusion

Main Results and Contributions

The studies presented in this PhD thesis provided insights into the challenges encountered by public policy makers who, in particular, make difficult distributional decisions and face the need to foresee unintended consequences of their decisions. The 3 chapters presented in this dissertation cover widely different institutional settings in terms of economic development, political culture and regulatory environment – namely, France, Ontario (Canada) and Russia. The analysed reforms implemented in these countries generally aimed at improving the efficiency of resource allocation in order to achieve various final goals, ranging from better health outcomes (as in chapter 2 on orthopedic care reform in Ontario) to correcting for inequities in budget distributions (as in chapter 1 on the DRG refinement in France) and improving fertility in a country facing long-term demographic challenges (chapter 3 on the Maternity Capital program in Russia). In the meantime, research questions investigated in this PhD thesis placed a particular emphasis on unintended side effects of these policy interventions.

In chapter 1 we investigated the impact of the obstetric DRG refinement and ensuing tariff incentives on the choice between scheduled C-sections and normal deliveries in France. This policy shock occurred in 2012 in an attempt to reduce financial risk for obstetric care providers by introducing additional parameters into the price formula, such as multiparity, multiple birth, gestational age and new comorbidities, bringing the payment system closer to a cost-plus approach. These hospital-level tariff changes created an exogenous source of variation, which could potentially affect the clinical practice of French obstetricians and midwives. We adopted both probabilistic and deterministic approaches to measuring DRG price incentives and tested them in difference-in-difference models. Our results suggested that the 2012 DRG reform did not have a significant impact on healthcare providers' obstetrical practice. Models featuring alternative formulations for measures of financial incentives led to the same conclusion. Moreover, there was no evidence that effects varied across years and for different types of care institutions (for-profit or non-profit).

Chapter 1 filled a relative lack in the literature on the impact of changes in financial incentives resulting from DRG refinements. To the best of our knowledge, the consequences of DRG refinements have never been studied specifically in the realm of obstetrics. In terms of public policy, this Chapter highlighted the fact large scale interventions produce

significant byproduct incentives oftentimes overlooked by policy makers. Whether or not these incentives will play significant role depends, in particular, on the strength of institutional safeguards and regulations. In Chapter 1, more specifically, we drew the conclusion that legal protections enjoyed by French medical professionals made it more problematic to transfer hospital-level incentives to individual care providers. From a methodological standpoint, our contribution consisted in developing a novel approach to modeling patient-level changes in DRG financial incentives. More precisely, we constructed measures of tariff change measures based on the difference of probability-weighted DRG tariffs at the decision tree node where an obstetrician is faced the choice between a scheduled C-section and vaginal delivery. This approach made it possible to reflect tariff variations more fully and, thus, mitigate a potential source of attenuation bias.

In chapter 2 we analyzed the effects of a healthcare reform in Ontario, whereby the Ontario Ministry of Health intended to replace global hospital budgets by a prospective funding scheme with a pay-for-performance component. Thus, since 2012 unilateral hip and knee joint replacements started to be funded through Quality-Based Procedures and the Health Based Allocation Model, with bilateral replacements joining the list of eligible procedures in 2014. Since the reform was only implemented in Ontario, we used two other Canadian provinces (namely, Alberta and British Columbia) as control groups in our difference-in-difference analysis, complemented with matching estimators. The main results indicated that the QBP component of the reform led to a significant reduction in the acute length of stay. However, evidence for quality improvement in orthopedic surgery— the main objective of the reform – was overall weak and insufficient. From a policy standpoint, the 2012 Ontario reform represents an important example of a so-called “policy-drift”, where initial policy objectives were considerably revised during the implementation phase of the reform. This Phd thesis provides evidence that poorly communicated policy goals and resulting information asymmetries are likely to translate into a poorer performance, even in a setting with otherwise robust institutions.

In Chapter 3 we analyzed the impact of the Maternity Capital (MC) program (i.e a child subsidy of around 10,000 USD, in 2007 prices) introduced in Russia in 2007. The MC made it possible for Russian families who gave birth to/adopted a 2nd or subsequent children to use this fund towards improving housing conditions, funding child education, or investing in the mother’s retirement fund. Chapter 3 attempted to evaluate the impact of the MC reform on various child health and well-being outcomes, as well as household-level consumption patterns. Testing regression discontinuity design model on RLMS data from 2011 to 2017, we found that differences between MC eligible and ineligible households in terms of analysed health, educational and well-being outcomes were overall insignificant. However, we also found evidence that the MC effects can be heterogeneous. In particular, female children demonstrated a better performance in out-of-school socialization compared to male children. In addition, there were signs that self-reportedly poor households benefited less fully from the MC program since they could not afford to invest the

MC in buying real estate.

In terms of public policy, this dissertation Chapter highlighted the importance and inevitability of the agency dilemma in most public policy contexts, wherein principals, agents and state funders have inherently different utility functions and, thus, interests and behaviors. It can manifest itself in the moral hazard problem, post-contractual opportunism and, as a result, inefficient allocation of resources. Partially as a result of it, the MC reform designers put in place a number of and stringent use restrictions to minimize the risk of the MC subsidy being spend inefficiently. In particular, this chapter emphasizes the importance of taking into account known behavioral traits and cognitive biases while making such policy decisions as the choosing between one-time transfer vs. a flow of smaller transfers, introducing mandatory wait times before activation of payment rights, or imposing limits on whom a subsidy can be spent.

Research prospects

In chapter 1, an interesting avenue for future research would be to study the effect of price incentives from the point of view of empirical industrial organization. In France there has been a trend towards merging or closing small maternity wards, leading to larger institutions belonging to different financial groups. The way in which these industrial changes interacted with changes in price incentives remains an unanswered question.

In chapter 2, as a further line of research one may attempt to look into the impact of hospital concentration and competition on care parameters. Insofar as pay-for-performance systems introduce stimuli for hospitals to specialize in services in which they have comparative advantages (and to drop those for which they are less efficient in terms of cost and quality of care), one would expect a decrease in the diversity of procedures provided in Ontario hospitals as a result the QBP reform. Moreover, this effect can hypothetically be observed only in areas where patients have a choice among multiple providers (i.e. where concentration of/competition among providers is the greatest). As in the present study, Alberta and British Columbia could be used as control groups.

Second, distributional comparisons of care parameters can be made in the spirit of *Contoyannis and Wildman (2007)*. This would shed more light as to the QBP/HBAM-induced variation of care pathways, technology adoption, uniformity of treatment practices, comorbidity distributions, etc.

Third, changes in the structure of costs and their relation to care efficiency could be analysed provide an additional data module is made available d by CIHI. In particular, one could investigate what component of costs (material, administrative, labor etc.) is likely to yield maximum increases in affected care parameters. In addition, a general cost-benefit analysis could also be performed if the amounts of QBP/HBAM payments are released to the public in the future.

Finally, the impact of the QBP/HBAM policies could be analyzed with respect to physi-

cian characteristics and practice styles. This would allow assessing the impact of characteristics such as gender, age, school of graduation, which may have an impact on practice styles.

Chapter 3 can be complemented in a number of ways. First, additional outcomes related to family relationships can be investigated in more depth. For example, since MC subsidy is tied to the mother, additional funds brought by this subsidy can provide a higher degree of mothers' financial independence. Thus, such outcomes as the probability of divorce, spending on personal items and the patterns of spending leisure time by mothers can potentially be affected by MC stimuli. Other household members can also be affected by these changes through, for example, the channel of a relaxed budget constraint or through spill-over effects.

Second, the analysis of the household-level reaction to MC stimuli could be deepened by analysing various aspects of financial behavior after MC claim rights were acquired. This would be particularly relevant since MC subsidy is likely to dynamically affect future income flows, as well as such household financial decisions as borrowing, lending, investing and reliance on insurance. The advantage of RLMS data consists in providing very detailed information on the source/recipients of borrowed/lent funds (e.g. from/to a friend, family member, financial institution), as well as on the type of insurance contracts signed by survey respondents.

Third, after additional RLMS waves become available, this study may be extended by examining various outcomes of MC on adolescents and even adults. Indeed, the structure of RLMS surveys makes it possible to retrospectively reconstitute a range of biographical information, including birth dates, birth order and the fact of living in MC claim eligible households.

Finally, conditional on RLMS survey designers consistently collecting the information of the MC subsidy use, it would be possible to investigate the share of households using their MC over time, as well as the share of MC subsidy spent on eligible uses. In particular, one may inquire about the share of MC that directly benefited children versus other uses that were aimed at improving household conditions in general.

This dissertation covered 3 different institutional environments (Ontario, France and Russia), and was centered around several key stakeholders: (1) policy-makers (or social planners), who put forward policies aimed at achieving a better distributional efficiency; (2) hospitals and individual healthcare providers who act in accordance with the proposed policies while having their individual utility functions (3) patients and families, who are on the receiving end of the public policy reforms and whose welfare largely depends on the cooperation between the two former stakeholders.

This PhD presents 3 main empirical findings. First, as evidenced in Chapter 1, even significant variations in hospital tariffs do not always lead to noticeable changes in the clinical practice, which can be explained by institutional safeguards entrenched in healthcare systems and bounded rationality of agents. Second, in Chapter 2 we provided evidence

that weak incentives aiming to improve healthcare quality are unlikely to be effective in practice. Finally, the last Chapter suggests that the effect of sizable child subsidies on a broad range of intra-family distributional decisions can be greatly affected by the subsidy design. In particular, the absence of a statistically significant MC effect on the majority of analyzed outcomes highlighted the importance of such distinctions as one-time vs. regular payments, unrestricted vs. limited use of transfers, and the universal coverage vs. means-tested eligibility conditions.

This dissertation provides only so many pieces of evidence to the immensely huge public policy debate on the distributional efficiency and the mechanisms that policy makers have at their disposal to achieve it. Additional research is undoubtedly needed to shed more light on these fundamental economics issues. We hope that the evidence provided in this dissertation stimulates an increased interest in readers and enhances awareness of the challenges that policy makers are facing when implementing large-scale public policy reforms.

Bibliography

- [1] Abadie, A., Imbens, G. Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, Vol.74, pp.235-267. (2006).
- [2] Abadie, A., Imbens, G. Bias-Corrected Matching Estimators for Average Treatment Effects. *Journal of Business and Economic Statistics*, Vol.74, pp.235-267. (2011).
- [3] Abadie, A., Imbens, G. On the Failure of the Bootstrap for Matching Estimators. *Econometrica*, Vol.76. (2008), pp. 1537-1557.
- [4] Abankina I., Filatova L. Accessibility of pre-school education in Russia. *Nauka i Obrazovanie*, 2018. (in Russian)
- [5] Ajefu J. Parental income, child labour, and human capital accumulation: evidence from trade liberalisation in India. *Development in Practice*, Vol. 28(8), 2018, pp. 1071–1082.
- [6] Allin, S., Baker, M., Isabelle, M., Stabile, M. (2015). Physician Incentives and the Rise in C-sections: Evidence from Canada. *National Bureau of Economic Research*.
- [7] Barlet, Muriel and Coldefy, Magali and Collin, Clémentine and Lucas-Gabrielli, Véronique (2012). Local Potential Accessibility (LPA): A New Measure of Accessibility to Private General Practitioners. *Questions d' économie de la santé*, 174, p1-8, <http://www.irdes.fr/EspaceAnglais/Publications/IrdesPublications/QES174.pdf>
- [8] An L., Bluhm J., Foldes S., Alesci N., Klatt C., Center B., Nersesian W., Larson M., Ahluwalia J., Manley M. A Randomized Trial of a Pay-for-Performance Program Targeting Clinician Referral to a State Tobacco Quitline. *Archives of internal medicine*. Vol.168., (2008).
- [9] Attanasio, O. The determinants of human capital formation during the early years of life: theory, measurement, and policies. *Journal of the European Economic Association*, Vol.13, 2015, pp. 949-997.
- [10] Baker L. Acquisition Of MRI Equipment By Doctors Drives Up Imaging Use And Spending *Health Affairs*. 2010; Vol. 29, No. 12.

- [11] Baughman R., Stacy Dickert-Conlin S. Did Expanding the EITC Promote Motherhood? *The American Economic Review*, Vol. 93(2), 2003.
- [12] Baughman R. and Duchovny N. State Earned Income Tax Credits and the Production of Child Health: Insurance Coverage, Utilization, and Health Status. *National Tax Journal*, Vol. 69 (1), 2016, pp. 103–132
- [13] Born K., Dhalla I. Ontario hospital funding: confusion for 2012/2013?; (2012); <https://healthydebate.ca/2012/02/topic/cost-of-care/ontario-hospital-funding-confusion>
- [14] Brown, H. (1996). Physician demand for leisure: implications for cesarean section rates. *Journal of Health Economics*, 15(2), pp.233-242.
- [15] Busse R. et al. Diagnosis-related Groups in Europe: Moving Towards Transparency, Efficiency and Quality in Hospitals, 2011.
- [16] CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada, 4th Edition. 2017. https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf
- [17] Canadian Institute for Health Information. Patients receiving treatment within the benchmark for Hip Replacement (accessed on 18 November 2019) <http://waittimes.cihi.ca/procedure/hip#year>
- [18] Canadian Joint Replacement Registry Annual Report. Hip and Knee Replacements in Canada, 2017–2018. *Canadian Institute for Health Information*; 2019. https://www.cihi.ca/sites/default/files/document/cjrr-annual-report-2019-en-web_0.pdf
- [19] Cameron D., Ubels J., Norström F. On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: a systematic review. *Global Health Action*. Vol. 11(1), 2018.
- [20] Case A., Lubotsky D., Paxson C. Economic Status and Health in Childhood: The Origins of the Gradient. *American Economic Review*, Vol. 92 (5), 2002, pp. 1308-1334.
- [21] Carmo de Souza Cruz R., Barroso Azevedo de Moura L., José Soares Neto J. Conditional cash transfers and the creation of equal opportunities of health for children in low and middle-income countries: a literature review. *International Journal for Equity in Health*. Vol. 16. (2017).

-
- [22] Cartier-Bechu C., Gherardi A., Sivignon M., Pignata M., Petitjean A., Monnier R., Roze S. Is There a Threshold in France?: First Exhaustive Review of Published Health-Economic Appraisals by the Haute Autorite De Sante (HAS), (French National Authority for Health). *Value in Health*, Vol. 19, Issue 7, 2016
- [23] Christianson, J., Leatherman, S., Sutherland, K. Lessons From Evaluations of Purchaser Pay-for-Performance Programs A Review of the Evidence. *Medical care research and review*, (2008).
- [24] CIAN. How did apartment prices evolve in Moscow? (in Russian, page last cosulted on September 16, 2020) <https://www.cian.ru/stati-kak-izmenilis-tseny-na-kvartiry-v-moskve-218051/>
- [25] Clemens, J., Gottlieb J. (2014). Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health? *American Economic review*, 104 (4), pp. 1320-49.
- [26] Cohen A., Dehejia R., and Romanov D. Financial incentives and fertility *The Review of Economics and Statistics*, Vol. 95(1), 2013, pp. 1–20.
- [27] Conti G., Heckman J. The economics of child well-being. *NBER Working paper series*, 2012.
- [28] Contoyannis P., Wildman J. Using relative distributions to investigate the body mass index in England and Canada. *Health economics*. Vol.16, (2007).
- [29] Cooper K., Stewart K. Does money affect children's outcomes? *JRF*, 2018. <https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/money-children-outcomes-full.pdf>
- [30] Coudin, E., Pla, A., Samson, A. L. (2015). GP responses to price regulation: evidence from a French nationwide reform. *Health economics*, 24(9), pp.1118-1130.
- [31] Currie J. Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, Vol. 47, No. 1, 2009, pp. 87-122.
- [32] Currie J.W., MacLeod B. (2008) First Do No Harm? Tort Reform and Birth Outcomes. *The Quarterly Journal of Economics*, 123, (2), pp. 795–830.
- [33] Czeisler M., Lane R., Petrosky E., Wiley J., Christensen A., Njai R., Weaver M., Robbins R., Elise R., Facer-Childs E., Barger L., Czeisler C., Howard M., Rajaratnam S. Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic — United States, June 24–30. *Morbidity and Mortality Weekly Report* Vol. 69(32) pp. 1049–1057, 2020. <https://www.cdc.gov/mmwr/volumes/69/wr/mm6932a1.htm>

-
- [34] Dahl G., Lochner L. The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *The American Economic Review*, Vol. 102(5), 2012, pp. 1927-1956.
- [35] Dafny, L.(2005). How Do Hospitals Respond to Price Changes? *The American Economic Review*, 95(5), pp. 1525-1547.
- [36] Delattre, E., Dormont, B. (2003). Fixed fees and physician-induced demand: A panel data study on French physicians. *Health Economics*, 12(9), pp.741-754
- [37] Delattre, E., Dormont, B. (2005). La régulation de la médecine ambulatoire en France: quel effet sur le comportement des médecins libéraux?. *Dossiers solidarité et santé*, (1), pp.135-161.
- [38] Deutscher N., Breunig R. Baby Bonuses: Natural Experiments in Cash Transfers, Birth Timing and Child Outcomes. *Economic Record*, Vol. 94(304), 2018, pp.1–24.
- [39] Dubay, L., Kaestner, R., Waidmann. T. (1999). The Impact of Malpractice Fears on Caesarean Section Rates. *Journal of Health Economics*, 18 (4) pp.491-522.
- [40] Duncan G, Morris P., Rodrigues C. Does Money Really Matter? Estimating Impacts of Family Income on Young Children’s Achievement With Data From Random-Assignment Experiments. *Developmental Psychology*. Vol. 47(5). (2011).
- [41] Eggleston K. Multitasking and mixed systems for provider payment. *Journal of Health Economics*. Volume 24, Issue 1, (2005), pp. 211-223.
- [42] Ellis R. Creaming, skimping and dumping: provider competition on the intensive and extensive margins, *Journal of Health Economics*, Vol. 17, pp. 537—555. (1998)
- [43] Ellis R., McGuire T. (1996) Hospital response to prospective payment: Moral hazard, selection, and practice-style effects, *Journal of Health Economics*, Vol. 15, pp. 257–277.
- [44] Emmert M., Eijkenaar F., Kemter H., Esslinger A., Schöffski O. . Economic evaluation of pay-for-performance in health care: A systematic review. *The European journal of health economics*. Vol. 13., (2011).
- [45] Ferman B., Pinto C. (2019) Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity. *The Review of Economics and Statistics*, 101:3, pp. 452-467.
- [46] Gaitz J., Schurer S. Bonus Skills: Examining the Effect of an Unconditional Cash Transfer on Child Human Capital Formation. *IZA Discussion papers*, 2017.

-
- [47] Galizzi M., Tammi T., Godager G., Linnosmaa I., Wiesen D. (2015) Provider altruism in health economics. National Institute for Health and Welfare (THL). Discussion paper 4/2015. Helsinki, Finland.
- [48] Geissler, A. , Scheller-Kreinsen, D. , Quentin, W. on behalf of the EuroDRG group. Do Diagnosis-Related Groups Appropriately Explain Variations In Costs And Length Of Stay Of Hip Replacement? A Comparative Assessment Of DRG Systems Across 10 European Countries. *Health Econ* (2012) 21: 103-115. Doi:10.1002/Hec.2848
- [49] Gilman B. (2000) Hospital response to DRG refinements : the impact of multiple reimbursement incentives on inpatient length of stay. *Health Econ Policy Law*, Vol. 1, Issue 1, pp.79-90.
- [50] Giacomo M., Piacenza M., Siciliani L., Turati G. (2017) Do public hospitals respond to changes in DRG price regulation? The case of birth deliveries in the Italian NHS. *Health Economics* , Vol.26, Issue S2, pp. 23-37.
- [51] Goddard M. , Hauck K., Smith P.C. (2006) Priority setting in health - a political economy perspective *Health Economics*, online version
- [52] González L. The Effect of a Universal Child Benefit on Conceptions Abortions, and Early Maternal Labor Supply. *American Economic Journal: Economic Policy*, 2013, 5(3), pp. 160–188.
- [53] Government of Ontario. Moving To Patient-Based Funding Will Improve Care, 2012. <https://news.ontario.ca/mohltc/en/2012/03/moving-to-patient-based-funding-will-improve-care.html>.
- [54] Grant, D. (2009) Physician financial incentives and cesarean delivery: new conclusions from the healthcare cost and utilization project. *Journal of health economics*, 28(1), pp.244-250.
- [55] Grao-Cruces A.; Fernández-Martínez A., Nuviala A. Association of Fitness With Life Satisfaction, Health Risk Behaviors, and Adherence to the Mediterranean Diet in Spanish Adolescents. *Journal of Strength and Conditioning Research*, Vol. 28(8), 2014, p.2164-2172.
- [56] Gruber J., Owings M. (1996) Physician financial incentives and cesarean section delivery, *RAND Journal of Economics*, 27 (1), pp. 99-123.
- [57] Gruber, J., Kim J., and Mayzlin D. (1999). Physician Fees and Procedure Intensity: The Case of Cesarean Delivery *Journal of Health Economics*, 18 (4), pp. 473-490.

-
- [58] Grytten, J., Monkerud, L., Sorensen, R. (2012). Adoption of diagnostic technology and variation in caesarean section rates: a test of the practice style hypothesis in Norway. *Health services research*, 47(6), pp.2169-2189.
- [59] Grytten, J., Skau I., Sorensen R. (2011). Do Expert Patients Get Better Treatment than Others? Agency Discrimination and Statistical Discrimination in Obstetrics. *Journal of Health Economics*, 30 (1), pp. 163-180.
- [60] Grytten, J., Skau, I., Sorensen, R. (2013). Do mothers decide? The impact of preferences in healthcare. *Journal of Human Resources*, 48(1), pp. 142-168.
- [61] Hafsteinsdottir E., Siciliani L. (2010). DRG prospective payment systems: refine or not refine? *Health Economics*, 19, pp. 1226-1239
- [62] Health Canada. Final Report of The Federal Advisor on Wait Times. (2005) https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/hcs-sss/alt_formats/hpb-dgps/pdf/pubs/2006-wait-attente/index-eng.pdf
- [63] Herck P., De Smedt D., Annemans, L., Remmen, R., Rosenthal, M., Sermeus, W. Systematic Review: Effects, Design Choices, and Context of Pay-For-Performance in Health Care. *BMC health services research*, Vol.10, (2010).
- [64] Hoddinott J., Lawrence H. , Does Female Income Share Influence Household Expenditures? Evidence from Cote d'Ivoire, *Oxford Bulletin of Economics and Statistics*. Vol. 57(1), 1995, pp. 77-96.
- [65] Holmstrom B., and Milgrom P. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization* Vol. 7, Special Issue: [Papers from the Conference on the New Science of Organization, January 1991] (1991), pp. 24-52.
- [66] Janulevicuite J., Askildsen J., Kaarboe O., Siciliani L.(2016) How Do Hospitals Respond To Price Changes? Evidence from Norway. *Health Economics*, 25, pp. 620–636.
- [67] Johnson E., Rehavi M. (2016) Physicians Treating Physicians: Information and Incentives in Childbirth. *American Economic Journal: Economic Policy* (2016), 8(1), pp. 115-141.
- [68] Kaarboe, O. Siciliani, L. (2008). Multitasking, quality and pay for performance. *Health Economics*, Vol.20, Issue2, (2011), pp. 225-238.
- [69] Kalemli-Ozcan S. A stochastic model of mortality, fertility, and human capital investment. *Journal of Development Economics*, Vol. 70, 2003, pp. 103–118.
- [70] Kling, J., Liebman J., Katz L. Experimental Analysis of Neighborhood Effects. *Econometrica*. Vol.75(1), 2007, pp. 83-119.

-
- [71] Kling, J., Ludwig J., Katz L. “Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment.” *Quarterly Journal of Economics*. Vol.120(1), (2005), pp.87-130.
- [72] Kochanek K., Anderson R., Arias E. Changes in Life Expectancy at Birth, 2010–2018. *National Center for Health Statistics*, 2020.
- [73] Kommersant journal. Counterfeit diploma holders, 2015 <https://www.kommersant.ru/doc/2724365> (in Russian)
- [74] Kuhn, M. Siciliani, L. . Performance Indicators for Quality with Costly Falsification. *Journal of Economics Management Strategy*. Vol. 18., (2009), pp.1137-1154.
- [75] Laibson, D. Hyperbolic Discount Functions, Undersaving, and Savings Plans,” NBER Working Paper 5635, 1996.
- [76] Lebihan L., Takongmo C. The impact of universal child benefits on family health and behaviours. *Research in Economics*. Vol. 72, 2018, pp. 415–427.
- [77] Lee D., Lemieux T. Regression Discontinuity Designs in Economics *Journal of Economic Literature*, Vol.48, 2010, pp.281–355.
- [78] Lefèvre M. (2014) Physician induced demand for C-sections: does the convenience incentive matter?,” Health, Econometrics and Data Group (HEDG) Working Papers 14/08, HEDG, c/o Department of Economics. University of York.
- [79] Levada Center. Trust in public institutions (Survey), 2014. <https://www.levada.ru/2014/11/13/doverie-institutam-vlasti-3/> (in Russian)
- [80] Levada Center. Religion (Survey), 2014. <https://www.levada.ru/2017/07/18/religioznost/> (in Russian)
- [81] Lindenauer P., Remus D., Roman S., Rothberg M., Benjamin E., Ma A., Bratzler D.. Public Reporting and Pay for Performance in Hospital Quality Improvement. *New England Journal of Medicine*. Vol.356(5), (2007), pp.486–96.
- [82] Lo J. (2003) Patients‘ attitudes vs. physicians‘ determination: implications for cesarean sections. *Social Science and Medicine*, 57(1), pp.91-96.
- [83] Lo J. (2008) Financial incentives do not always work - an example of cesarean sections in Taiwan. *Health policy*, 88(1), pp.121-129.
- [84] Makris M., Siciliani L. Optimal Incentive Schemes for Altruistic Providers. *Journal of Public Economic theory*. Vol. 15, Issue5, (2013), pp. 675-699.

- [85] McEwen A., Stewart J. The Relationship between Income and Children's Outcomes: A Synthesis of Canadian Evidence. *Canadian Public Policy / Analyse de Politiques*. Vol. 40 (1), 2014, pp. 99-109.
- [86] Milcent C. (2016) Upcoding and Heterogeneity in Hospitals' Response: a Natural Experiment. *PSE Working Paper*. URL: <https://www.parisschoolofeconomics.eu/docs/milcent-carine/upcoding.pdf>
- [87] Milcent C., Rochut J. (2009) Tarification hospitalière et pratique médicale La pratique de la césarienne en France *Revue économique* , 60, pp. 489-506.
- [88] MacKinnon J., Webb M. (2020) Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*. <https://www.sciencedirect.com/science/article/abs/pii/S0304407620301445?via%3Dihub>
- [89] Milligan K. Subsidizing the Stork: New Evidence on Tax Incentives and Fertility. *The Review of Economics and Statistics*, Vol. 87(3), 2005, pp. 539-555.
- [90] Milligan K., Stabile M. Do Child Tax Benefits Affect the Well-being of Children? Evidence from Canadian Child Benefit Expansions. *American Economic Journal: Economic Policy*, 2011, pp. 175–205.
- [91] Moreira L. Health literacy for people centred care Where do OECD countries stand? *OECD Health Working Paper* No. 107, 2018.
- [92] Mullen K. Frank R., Rosenthal M. Can you get what you pay for? Pay-for-performance and the quality of healthcare providers *The RAND Journal of Economics*. Volume 41, Issue 1, (2010), pp. 64-91.
- [93] Nahra T., Reiter K., Hirth R., Shermer J., Wheeler, J. Cost-Effectiveness of Hospital Pay-for-Performance Incentives. *Medical care research and review*. Vol.63., (2006).
- [94] National Health Service. Hip replacement (Overview, How it is performed, Risks, Recovery, Metal implant advice) (2019a) <https://www.nhs.uk/conditions/hip-replacement/>
- [95] National Health Service. Knee replacement (Overview, How it is performed, Risks, Recovery) (2019b) <https://www.nhs.uk/conditions/knee-replacement/>
- [96] Niskanen, W.A. (1971) *Bureaucracy and Representative Government* (Chicago: Aldine, Atherton).
- [97] Niskanen, W.A. (1975) Bureaucrats and Politicians *Journal of Law and Economics*, Vol. 18, Issue 3, pp.617-43.

-
- [98] Norton C. Incentive regulation of nursing homes. *Journal of Health Economics*, Vol. 11, Issue 2, 1992, pp. 105-128.
- [99] OECD (2019), Health spending (indicator). doi: 10.1787/8643de7e-en (Accessed on 18 November 2019)
- [100] OECD. Health at a Glance 2019: OECD Indicators. Caesarean sections (2019) [https://www.oecd-ilibrary.org/sites/fa1f7281-en/index.html?itemId=/content/component/fa1f7281-en#:~:text=Across%20OECD%20countries%2C%2028%25%20of%20five%20years%20\(Figure%209.17\).](https://www.oecd-ilibrary.org/sites/fa1f7281-en/index.html?itemId=/content/component/fa1f7281-en#:~:text=Across%20OECD%20countries%2C%2028%25%20of%20five%20years%20(Figure%209.17).)
- [101] Ontario Hospital Association. HSFR Resources, 2019 (accessed on 18 November 2019) <https://www.oha.com/health-system-transformation/health-system-funding-reform/hsfr-resources>
- [102] Palmer, K.S., Brown, A.D., Evans, J.M. et al. Standardising costs or standardising care? Qualitative evaluation of the implementation and impact of a hospital funding reform in Ontario, Canada. *Health Research Policy Systems*. Vol. 16(74), (2018a).
- [103] Palmer, K. S., Brown, A. D., Evans, J. M., Marani, H., Russell, K. K., Martin, D., Ivers, N. M. Qualitative analysis of the dynamics of policy design and implementation in hospital funding reform. *PloS one*, Vol.13(1), (2018b).
- [104] Papanicolas I., McGuire A. (2015). Do financial incentives trump clinical guidance? Hip Replacement in England and Scotland, *Journal of Health Economics*, 44 (2015), pp.25-36
- [105] Phillips, R., Thornton J., Gleicher, N. (1982), Physician bias in cesarean sections, *Journal of the American Medical Association*, 18 (2), pp. 1082-1084.
- [106] Quality-Based Procedures Clinical Handbook for Primary Hip and Knee Replacement. *Health Quality Ontario & Ministry of Health and Long-Term Care* (2013) http://www.health.gov.on.ca/en/pro/programs/ecfa/funding/hs_funding_qbp.aspx
- [107] Richards T., Green G. Environmental Choices and Hyperbolic Discounting: An Experimental Analysis. *Environmental and Resource Economics volume*. Vol. 62, (2015).
- [108] Rochaix, L., Jacobzone, S. (1997). L'hypothèse de demande induite: un bilan économique. *Economie prévision*, 129(3-4), pp.25-36.
- [109] Rosrealt. Apartment prices in Russian cities in October 2014 (in Russian, page last consulted on September 16, 2020) <https://rosrealt.ru/cena/kvartira/159>

- [110] ROSSTAT. Consumption and income statistics of Russian households: how do we assess changes? (in Russian) (2020) https://rosstat.gov.ru/free_doc/new_site/rosstat/smi/konferenz/ovcharova_prez.pdf
- [111] Russian Pension Fund. Press release of January 22th, 2018. http://www.pfrf.ru/press_center/~2018/01/22/151428, in Russian
- [112] Ryan A.. The Effects of the Premier Hospital Quality Incentive Demonstration on Mortality and Hospital Costs in Medicare. *Health services research*. Vol.44., (2009), pp.821-42.
- [113] Sarma S., Mehta N., Devlin R., Kpelitse K., Li L. Family physician remuneration schemes and specialist referrals: Quasi-experimental evidence from Ontario, Canada. *Health Economics*. Vol.27(10), (2018).
- [114] Schultz, T. P. , Testing the Neoclassical Model of Family Labor Supply and Fertility, *Journal of Human Resources*. Vol. 25, 4, pp. 599-634, (1990).
- [115] Shin E. Hospital responses to price shocks under the prospective payment system *Health Economics*. 2018; 28: 245– 260. <https://doi.org/10.1002/hec.3839>
- [116] Shigeoka H., Fushimib K. Supplier-induced demand for newborn treatment: Evidence from Japan *Journal of Health Economics*. 2014; 35: 162–178
- [117] Shurtz I. (2013) The impact of medical errors on physician behavior: Evidence from malpractice litigation. *Journal of Health Economics*, 32 (2), pp. 331-340.
- [118] Siciliani L. Paying for performance and motivation crowding out. *Economics Letters*. Vol.103, (2009), pp. 68–71.
- [119] Siciliani, L.. (2006) Selection of treatment under prospective payment systems in the hospital sector, *Journal of Health Economics*, 25, pp.479-499.
- [120] Slonimczyk F., Yurko, A. Assessing the impact of the maternity capital policy in Russia. *Labour Economics*, Vol.30, 2014, pp. 265-281.
- [121] Sorvachev I., Yakovlev E. Could a child subsidy increase long-run fertility and stability of families? Could it have equilibrium effects? Evidence from the Maternity Capital program in Russia. *SSRN Working papers*. (2020). [file:///C:/Users/alksp/Downloads/SSRN-id3416509%20\(1\).pdf](file:///C:/Users/alksp/Downloads/SSRN-id3416509%20(1).pdf)
- [122] Stender S. Industrially produced trans fat in popular foods in 15 countries of the former Soviet Union from 2015 to 2016: a market basket investigation. *BMJ Open*, 2019. <https://bmjopen.bmj.com/content/bmjopen/9/2/e023184.full.pdf>

-
- [123] Sutherland, J. M., Hellsten, E., Yu, K. Bundles: An opportunity to align incentives for continuing care in Canada? *Health Policy*, 107(2-3), pp.209–217., (2012)
- [124] Schwartz M., Tartter P. Decreased Length of Stay for Patients with Colorectal Cancer: Implications of DRG Use. *Journal for healthcare quality*, (2011).
- [125] Tamura R. Human capital and economic development *Journal of Development Economics*, Vol. 79, 2006, pp. 26– 72.
- [126] Teptsova T., Bezdenezhnyh T., Fedyaeva V., Musina N., Hachatryan G., Tarasov V. Determination of a willingness-to-pay threshold and decision-making in financing the healthcare technologies. *Farkakoekonomika* (in Russian). Vol 11, No 3, 2018.
- [127] Uauy R., Kain J., Mericq V., Rojas J., Corvalán C. Nutrition, child growth, and chronic disease prevention. *Annals of Medicine*. 2008; Vol. 40(1), pp. 11-20.
- [128] Vaiserman M. Early-life nutritional programming of longevity. *Journal of Developmental Origins of Health and Disease*, Vol. 5(5), 2014. pp. 325-338.
- [129] Vanlerenberghe J.-M. (2015) Rapport d’information fait au nom de la commission des affaires sociales sur l’enquête de la Cour des comptes relative aux maternités. No. 243. Sénat.
- [130] Verzulli R., Fiorentini G., Lippi Bruni M., Ugolini C. (2016) Price changes in regulated healthcare markets: do public hospitals respond and how? *Health Economics*, online version
- [131] Viscusi W., Masterman C. Income Elasticities and Global Values of a Statistical Life. *Journal of Cost Benefit Analysis*. Vol.8 (2), 2017, pp. 226-250.
- [132] Wettstein M., Palmer K., Kulkarni G, Paterson J., Ling V., Lapointe-Shaw, L., Li, A., Brown, A., Taljaard M., Ivers, N. Management Strategies and Patient Selection After a Hospital Funding Reform for Prostate Cancer Surgery in Canada. *JAMA Network Open*. Vol.2., (2019).
- [133] WHO statement on caesarean section rates (2015). https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/cs-statement/en/
- [134] Yang Y., Mello M., Subramanian S., Studdert D. (2009) Relationship Between Malpractice Litigation Pressure and Rates of Cesarean Section and Vaginal Birth After Cesarean Section. *Medical Care*. 47(2), pp. 234-242.

Essays in Health Economics

Abstract:

This dissertation sheds light on the challenges policy makers are facing while implementing large-scale healthcare and welfare reforms. This dissertation highlights the necessity to take into account unintended and oftentimes overlooked distributional consequences of policy interventions, as well as the broad institutional framework. In Chapter 1, we test the effects of a considerable DRG tariff refinement that occurred in 2012, designed to reduce financial risks of French maternity wards. After controlling for multiple patient, hospital and regional characteristics and allowing for hospital and year effects, we found that introducing new severity levels and clinical factors into the reimbursement algorithm had no significant effect on the probability of a scheduled C-section being performed. In Chapter 2, we evaluate the effects of a hospital funding reform, whereby hip and knee replacements were reimbursed based on a price by volume formula, with the expectation that payments would be subsequently adjusted with respect to hospital performance on quality indicators. We found a significant decrease in acute length of stay associated with Quality based payments (QBPs), although it did not result in an improved quality of care. Chapter 3 looks into the effect a maternity subsidy – called Maternity Capital (MC) – introduced in Russia in 2007. We conclude that the MC did not significantly affect the vast majority of analyzed child and household outcomes. In the meantime, there is tentative evidence that the MC may have resulted in improved socialization of female children. In addition, wealthier households may have benefited more fully from the MC by accumulating more savings that could be plausibly used for improving housing conditions.

KEYWORDS: DRG; tariff refinement; Hospital funding; Activity-based funding; Quality Based Procedures, Health Based Allocation Model; Child subsidies.

Résumé:

Cette thèse met en lumière les défis auxquels sont confrontés les décideurs politiques lors de la mise en œuvre de réformes à grande échelle. Cette thèse souligne la nécessité de prendre en compte les conséquences distributives non intentionnelles et souvent négligées des interventions politiques, ainsi que le cadre institutionnel général. Dans le chapitre 1, nous testons les effets d'un important raffinement tarifaire DRG intervenu en 2012, destiné à réduire les risques financiers des maternités françaises. Après avoir contrôlé les multiples caractéristiques au niveau du patient, de l'hôpital et de la région et tenu compte des effets de l'hôpital et de l'année, nous avons constaté que l'introduction de nouveaux niveaux de gravité et de nouveaux facteurs cliniques dans l'algorithme de remboursement n'avait aucun effet significatif sur la probabilité d'une césarienne programmée. Au chapitre 2, nous évaluons les effets d'une réforme du financement des hôpitaux, selon laquelle les arthroplasties de la hanche et du genou étaient remboursées sur la base d'une formule "prix par volume", dans l'espoir que les paiements seraient ultérieurement ajustés en fonction de la performance de l'hôpital sur les indicateurs de qualité. Nous avons constaté une diminution significative de la durée de séjour aiguë associée aux paiements fondés sur la qualité (QBP), bien que cela n'ait pas entraîné une amélioration de la qualité des soins. Le chapitre 3 examine l'effet d'une allocation de maternité - appelée Maternity Capital (MC) - introduite en Russie en 2007. Nous concluons que la MC n'a pas eu d'incidence significative sur la grande majorité des résultats analysés pour les enfants et les ménages. En même temps, la MC pourrait avoir abouti à une meilleure socialisation des filles. En outre, les ménages les plus aisés ont peut-être plus pleinement bénéficié de la MC en accumulant davantage d'épargne qui pourrait être vraisemblablement utilisée pour améliorer les conditions de logement.

MOTS-CLEFS: DRG; raffinement tarifaire; financement hospitalier; financement à l'activité; Quality Based Procedures, Health Based Allocation Model; allocation de maternité.