CCHE CCES
Canadian Centre for Health Economics
Centre canadien en économie de la santé

Working Paper Series
Document de travail de la série

IMPACT OF QUALITY-BASED PROCEDURES ON ORTHOPEDIC CARE QUANTITY AND QUALITY IN ONTARIO HOSPITALS

Alex Proshin, Adrian Rohit Dass, Lise Rochaix, Audrey Laporte

Working Paper No: 200002

www.canadiancentreforhealtheconomics.ca

June, 2020

Canadian Centre for Health Economics
Centre canadien en économie de la santé
155 College Street
Toronto, Ontario

CCHE/CCES Working Paper No. 200002
June, 2020

Impact of Quality-Based Procedures on Orthopedic Care Quantity and Quality in Ontario Hospitals

Alex Proshin[1,2,3], Adrian Rohit Dass[4,5], Lise Rochaix[1,2,3], Audrey Laporte[4,5]

[1] *PSE - Paris School of Economics*
[2] *PJSE - Paris Jourdan Sciences Economiques*
[3] *Hospinomics*
[4] *Institute of Health Policy, Management and Evaluation, University of Toronto*
[5] *Canadian Centre for Health Economics, Toronto, Canada*

Abstract

In 2012 the Ontario Ministry of Health introduced Quality-Based Procedures (QBPs), 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. From the onset, unilateral hip and knee replacements were included in QBPs, whereas bilateral hip and knee replacements were added in 2014. 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 Canadian Discharge Abstract Database (DAD), we evaluate 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 to 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 is 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 our study is to comprehensively evaluate the impact of a funding reform introduced in Ontario aiming to replace global budgets 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 paper 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 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 our study. The main results of the paper are presented and interpreted in Section 6. Section 7 presents robustness checks, Section 8 discusses the significance and limitations of the results. Section 9 concludes and suggests avenues for future research.

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

2.1 Prospective Payment Systems (PPS)

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 paper 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 by 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 most 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, 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 paper 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 Minessota, *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 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).

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 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.

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

Table 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

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, 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 \$463and \$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 fist 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 patent-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 conceptors 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 structured were considered too different from those of larger providers.

[1] For example, a hospital preforming 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 $150 \cdot 5214 = 782100$\$.

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 decease. 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 budget[2], 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 pre- and post-reform as of November 2019. *(Wettstein et al., 2019).*

[2]However, 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

Hip and knee replacements are one 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 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). 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).

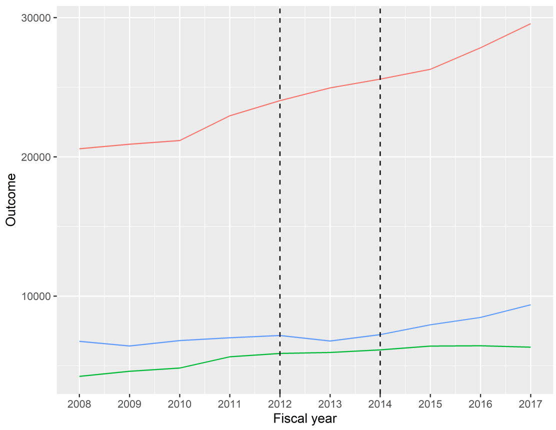
4 Research questions

The purpose of this study is to evaluate how incentives resulting from QBPs 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 with 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.

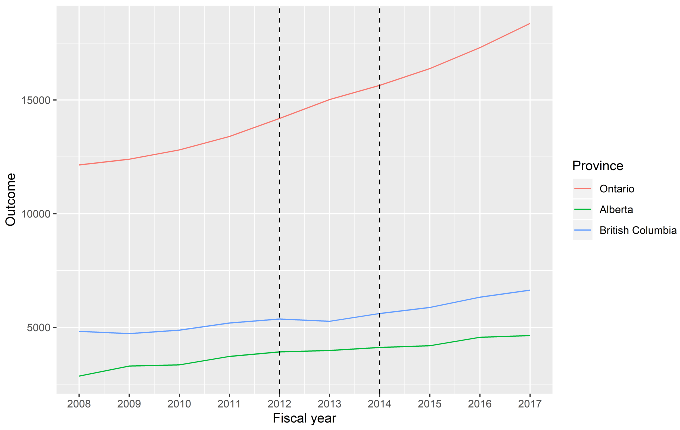
In 2012 QBP and HBAM payments were simultaneously introduced for unilateral hip and knee replacements, making it difficult to separate the effects of the two without comparing pre- and post-reform outcomes to similar populations of patients who did not experience the same policy changes. In difference-in-difference 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.

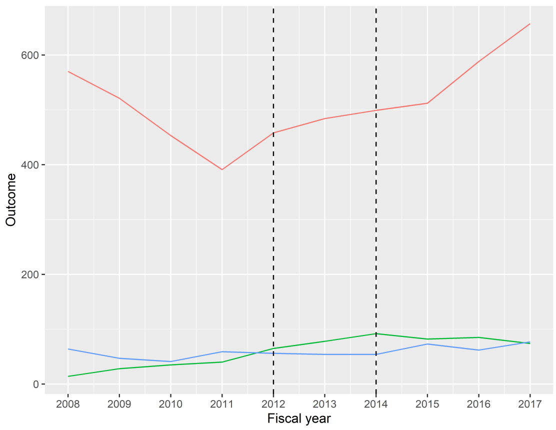
In our models, as control populations we use comparable patients residing in the Canadian provinces of Alberta (AB) and British Columbia (BC). This choice is dictated by two main reasons. First, their funding mechanics are similar to Ontario and remained



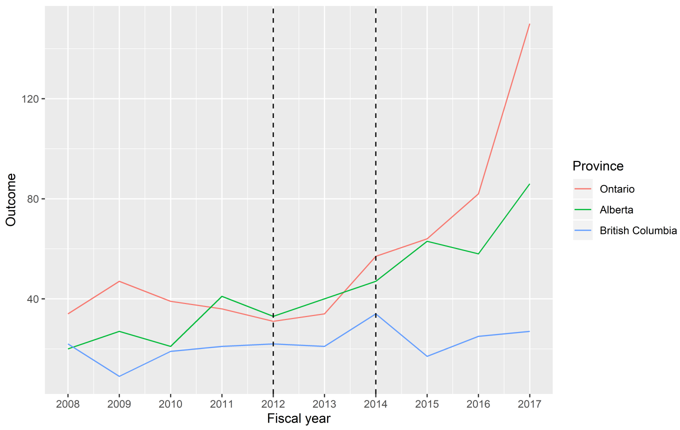
(a) Unilateral knee replacements



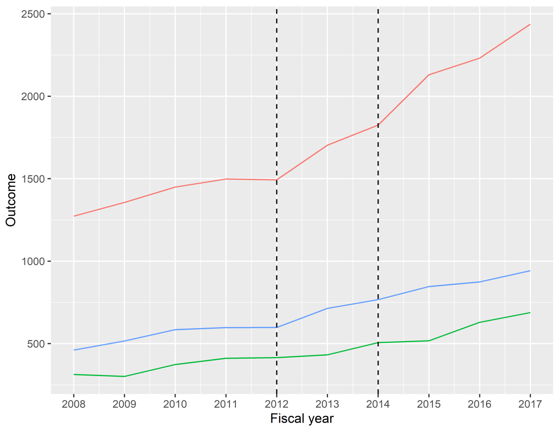
(b) Unilateral hip replacements



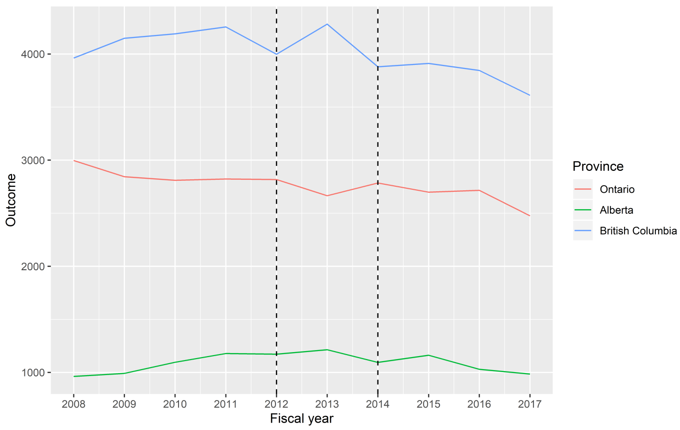
(c) Bilateral knee replacements



(d) Bilateral hip replacements



(e) Other joint replacements



(f) Surgery unrelated to arthroplasty

Figure 1: Total number of joint surgeries by type in Ontario, Alberta and British Columbia between 2008-2017

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 forth 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.

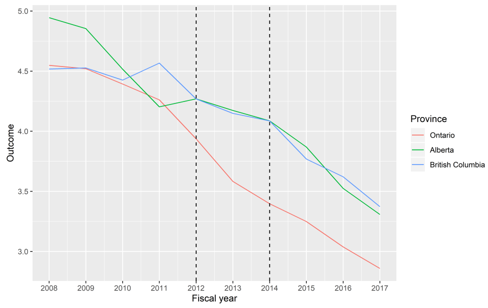
Secondly, we study the question whether stimuli that arose in unilateral and bilateral hip and knee replacement surgeries affected quality, process outcomes and appropriateness of other types of closely-related joint replacement surgeries (e.g. ankle and shoulder replacements). Theoretical justification 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 unincetivized 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 replacements included in QBPs.

5 Data

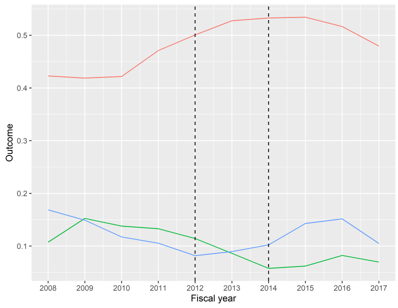
The Discharge Abstract Database (DAD) constitutes the main source of data in our study. It is a data set which if comprised of patient-level characteristics of every hospital stay in all Canadian provinces, except for Quebec. Data are submitted to the Canadian Institute of Health Information (CIHI) by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

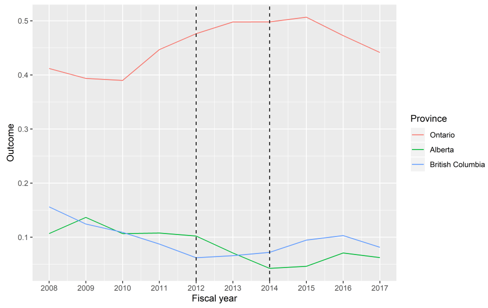
(a) Acute LOS (hip unilateral)



(b) Acute LOS (knee unilateral)

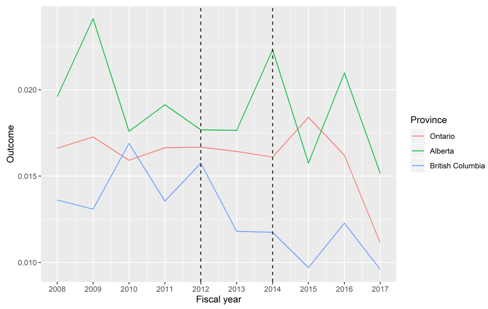


(c) Discharged home & support (hip unilateral)



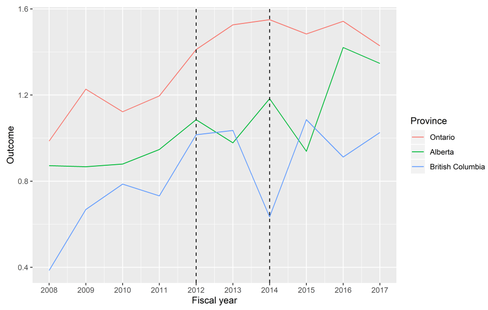
(d) Discharged home & support (knee unilateral)

(e) Revision in 180 days (hip unilateral)



(f) Revision in 180 days (knee unilateral)

(g) Elixhauser post-admission (hip unilateral)



(h) Elixhauser post-admission (knee unilateral)

Figure 2: Outcome trends unilateral knee/hip replacements in Ontario, Alberta and British Columbia between 2008-2017

- 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 by 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 excluded[3]

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.

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,

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

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.

Table 2: Descriptive statistics for whole sample and by province

					By province:					
Variables	*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
Length of stay (days):										
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
Responsibility for payment:										
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
Stay leads to:										
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
By replacement :										
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
Discharged to:										
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
Diagnoses (observed pre- and post-admission):										
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
Comorbidity indexes (pre- and post-admission)										
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:

$$y_{ipht} = \alpha + \beta_1' postQBP_{ipht} + \beta_2' X_{ipht} + \beta_3' Z_{ipht} + \psi_h + \gamma_t + \omega_i + \epsilon_{ipht}, \quad (1)$$

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 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 by 0.05 points and a strongly significant decrease by 0.15 index points for hip and knee unilateral replacements, respectively. However, as for 180-day revision and 180-day readmission rates, both there was no meaningful or statistically strong change after 2012 nor 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 care outcomes, likely due to unobserved patient characteristics such as official language ability/familiarity with local healthcare systems and ability to effectively navigate them.

Results for bilateral hip and knee replacement are presented in Table 4. 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

Table 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)	Elixhauser index	N.of interv.	Acute LOS	Revision(180d)	Readm.(180d)	N. diag.(post admit)	Elixhauser index	N.of interv.
Dependent var.	1	2	3	4	5	6	7	8	9	10	11	12
(Intercept)	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)***
$post2012 \cdot QBP$	-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)***
$post2014 \cdot QBP$	-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)***
age	-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)***
age^2	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)***
Male	-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:												
Worker insurance	-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)
Other	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												
Lowest	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)
Medium-low	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)
Medium	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)
Medium-high	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)
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Diagnosis FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	167322	167322	167322	167322	167322	167322	265620	265620	265620	265620	265620	265620
R2	0.36	0.1	0.011	0.21	0.22	0.41	0.25	0.15	0.01	0.31	0.13	0.26
AdjR2	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*

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.

Table 4: Difference-in-difference estimates by post-reform year for bilateral knee and hip replacements (quality measures)

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

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:

$$y_{ipht} = \alpha + \beta_1' postQBP_{ipht} + \beta_2 \cdot QBP_{iph} + \beta_3' X_{ipht} + \beta_4' Z_{ipht} + \psi_h + \gamma_t + \epsilon_{ipht}, \quad (2)$$

where variables and effects are denoted identically to equation (1), 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 5 and 6 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 6, 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 5 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 unincetivized 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 6 and Table 19 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 6), 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, we tentatively tried to estimate the impact of QBP and HBAM on the likelihood of different modes of patient discharge. Tables 18 -20 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 21 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.

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.

Table 5: Hospital-level difference-in-difference estimates for unilateral hip replacements

	Appropriateness		Process		Quality			
Outcome	Antibiotic	Pain	Acute LOS	Transfer	Readm. (180d)	Revision (180d)	Charlson post	Elixhauser post
Dependent variable:								
(Intercept)	-0.023	0.179	1.868	-1.047	-0.006	-0.148	1.481	-1.504
	(0.051)	(0.079)**	(6.063)	(0.312)***	(0.054)	(0.069)**	(0.496)***	(5.183)
post2012 QBP	-0.002	0.002	-0.286	< 0.001	-0.002	-0.001	-0.003	-0.06
	(0.001)*	(0.002)	(0.165)**	(0.013)	(0.002)	(0.002)	(0.012)	(0.13)
post2014 QBP	0.001	-0.001	-0.265	0.013	0.001	0.001	-0.002	-0.001
	(0.001)	(0.002)	(0.154)**	(0.009)	(0.002)	(0.002)	(0.011)	(0.126)
age	0.001	0.006	0.195	0.031	< 0.001	0.006	-0.04	0.049
	(0.002)	(0.003)**	(0.204)	(0.01)***	(0.002)	(0.002)***	(0.016)**	(0.172)
age2	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	(< 0.001)	(< 0.001)**	(0.002)	(< 0.001)***	(< 0.001)	(< 0.001)**	(< 0.001)**	(0.001)
female	0.003	-0.018	-0.193	-0.073	-0.018	-0.007	0.126	1.027
	(0.006)	(0.01)*	(0.769)	(0.032)**	(0.007)***	(0.009)	(0.056)**	(0.589)*
urban	0.001	0.001	0.666	-0.01	0.001	-0.001	-0.007	0.221
	(0.001)	(0.001)	(0.11)***	(0.005)*	(0.001)	(0.001)	(0.008)	(0.087)**
QBP flag dummy	-0.002	-0.002	-0.442	-0.043	-0.002	0.002	0.009	0.302
	(0.001)*	(0.001)	(0.121)***	(0.009)***	(0.001)	(0.001)	(0.009)	(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*

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

Table 6: 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							
Acute LOS	0.226	(0.529)	-0.558	(0.503)	554	0.194	0.167
Imaging procedures	0.052	(0.046)	-0.001	(0.054)	265	0.13	0.066
N.of post-admit diag	-0.127	(0.089)	-0.039	(0.085)	557	0.09	0.059
N.of interventions	0.019	(0.139)	0.069	(0.101)	557	0.126	0.097
Pain symptoms	-0.003	(0.013)	0.005	(0.012)	557	0.028	-0.005
Share bilateral repl.	-0.004	(0.028)	-0.011	(0.027)	557	0.475	0.458
Total LOS	0.314	(0.538)	-0.666	(0.543)	556	0.172	0.144
Transferred	-0.069	(0.055)	0.029	(0.056)	557	0.205	0.179
Charlson post	-0.061	(0.214)	-0.121	(0.216)	557	0.168	0.14
Elixhauser post	0.203	(0.448)	-0.441	(0.396)	329	0.077	0.023
Hip bilateral							
Acute LOS	1.03	(1.348)	-1.938	(1.308)	263	0.254	0.199
Imaging procedures	0.052	(0.046)	-0.001	(0.054)	265	0.13	0.066
N.of post-admit diag	0.404	(0.23)*	-0.401	(0.219)*	265	0.149	0.087
N.of interventions	0.092	(0.15)	-0.089	(0.128)	265	0.084	0.017
Pain symptoms	-0.072	(0.049)	0.049	(0.042)	265	0.021	0.019
Share bilateral repl.	-0.005	(0.005)	0.004	(0.007)	265	0.06	-0.008
Total LOS	2.14	(1.587)	-3.033	(1.594)*	265	0.268	0.214
Elixhauser post	0.159	(0.372)	-0.337	(0.363)	265	0.51	0.475
Knee unilateral							
Acute LOS	-0.304	(0.117)***	-0.18	(0.107)*	1119	0.678	0.646
Antibiotic use	-0.001	(0.002)	0.001	(0.001)	1152	0.593	0.555
Inhospital death	0	(0)	0	(0)	1152	0.435	0.382
Diagnostic procedures	-0.002	(0.002)	0.003	(0.003)	1152	0.287	0.219
Imaging procedures	-0.004	(0.004)	-0.002	(0.005)	1152	0.374	0.315
N.of post-admit diag	-0.007	(0.015)	0.001	(0.015)	1152	0.457	0.406
N.of interventions	-0.116	(0.052)**	0.17	(0.043)***	1152	0.511	0.465
Pain symptoms	0	(0.001)	0.001	(0.001)	1152	0.054	0.047
Readmission (180d)	0.002	(0.001)	0	(0.001)	1152	0.349	0.288
Revision (180d)	0.002	(0.001)	0	(0.001)	1152	0.749	0.726
Share unilateral repl.	0.002	(0.005)	0	(0.005)	1152	0.221	0.148
Total LOS	-0.355	(0.122)***	-0.197	(0.111)*	1140	0.673	0.642
Transferred	0.005	(0.009)	0.01	(0.006)	1152	0.45	0.398
Charlson post	0.005	(0.011)	-0.018	(0.011)*	1094	0.153	0.069
Elixhauser post	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* 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) $\tau = E[Y_i(1) - Y_i(0)]$ and average treatment effects on the treated (ATT) $\tau^{tr} = E[Y_i(1) - Y_i(0) | W_i = 1]$ are estimated as:

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

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

[5]Yearly 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

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

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 $\widehat{Y}(S_i)$ 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 7. It reflects the difference in outcome attributable to QBP and HBAM, as long as - 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 7 and 8 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,

Table 7: Estimates of nearest-neighbor covariate matching for unilateral hip replacements

Variable	Effect	Estimate 2008-11 [1]	SE	Estimate 2012-13 [2]	SE	Estimate 2014-17 [3]	SE	Effect 2012-13 [2] − [1]	SE	Effect 2014-17 [3] − [1]	SE
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.*

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 by 15-25 percentage points).

These results go 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 associ-

Table 8: Estimates of nearest-neighbor covariate matching for unilateral knee and other replacements

		Knee unilateral				Other replacements			
Variable	Effect	Effect 2012-13	SE	Effect 2014-17	SE	Effect 2012-13	SE	Effect 2014-17	SE
Acute LOS	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
Antibiotic use	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
Discharged home	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
Discharged home with support	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**
Discharged support service	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**
Inhospital death	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
Diagnostic procedures	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
Imaging procedures	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**
N.of post-admit diag	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
N.of interventions	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
Pain symptoms	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*
Readmission (180d)	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
Revision (180d)	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
Share unilateral repl.	ATE	0.004	0.005	0.005	0.004				
	ATT	0.003	0.007	0.006	0.006				
Total LOS	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
Transferred	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
Charlson post	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
Elixhauser post	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.*

ated 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 22 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 8, 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 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-weighed PSM the predicted outcome values in unobserved counterfactual state are a weighed average of all hospitals from the opposite (treatment or control) group[6]

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

$$\widehat{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0 \\ \dfrac{1}{\sum\limits_{W_m \neq W_i} w_m k_m} \sum\limits_{W_m \neq W_i} w_m k_m Y_m & \text{if } W_i = 1 \end{cases}$$

$$\widehat{Y}_i(1) = \begin{cases} \dfrac{1}{\sum\limits_{W_m \neq W_i} w_m k_m} \sum\limits_{W_m \neq W_i} w_m k_m Y_m & \text{if } W_i = 0 \\ Y_i & \text{if } W_i = 1, \end{cases}$$

where Y_i and w_i follow the notation provided in Subsection 7.1. Kernel weights k_i are obtained by applying the Epannechnikov kernel function K to estimated propensity scores, such that $k_i = \dfrac{1}{n_c h} K\left(\dfrac{x_c - x_t}{h}\right)$, 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 Epannechnikov kernel is used with the plug-in bandwidth parameter $h = 1.06\hat{\sigma}n^{-1/5}$.

[6] Thus, NNM is a particular, akin to degenerate, case of sample weight whose functional form is $1(X = \text{argmin}\ \{M_j\})$

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

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 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 9, while results for the remaining procedure types and for all replacements types combined can be found in Table 25 of Appendix G. Obtained coefficients go 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-weighed 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 Epachennikov kernel weights used in the aforementioned subsection

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

Table 9: Propensity score matching estimates for unilateral hip, unilateral knee and other replacements

Variable	Effect	Hip unilateral Effect 2012-13	SE	Effect 2014-17	SE	Knee unilateral Effect 2012-13	SE	Effect 2014-17	SE	Other replacement Effect 2012-13	SE	Effect 2014-17	SE
Acute LOS	ATE	-0.151	0.426	-0.404	0.426	-0.147	0.188	-0.243	0.143*	0.117	0.345	-0.084	0.271
	ATT	-0.185	0.58	-0.398	0.58	-0.188	0.225	-0.236	0.164	-0.38	0.434	-0.059	0.338
Antibiotic use	ATE	-0.002	0.004	-0.002	0.004	0.001	0.003	0.003	0.002	-0.005	0.002**	-0.005	0.002**
	ATT	-0.001	0.002	-0.001	0.002	0.001	0.004	0.003	0.002	-0.001	0.001	-0.001	0.001
Discharged home	ATE	0.097	0.027***	0.093	0.027***	0.1	0.025***	0.1	0.023***	-0.007	0.018	0.013	0.01
	ATT	0.087	0.03***	0.095	0.03***	0.098	0.024***	0.1	0.022***	-0.003	0.018	0.001	0.013
Discharged home with support	ATE	0.123	0.051**	0.137	0.051**	0.102	0.064	0.11	0.052**	0.043	0.048	0.094	0.042**
	ATT	0.145	0.055***	0.165	0.055***	0.122	0.062*	0.104	0.049**	0.11	0.045**	0.144	0.037***
Discharged support service	ATE	-0.101	0.018***	-0.098	0.018***	-0.103	0.017***	-0.112	0.017***	0.017	0.014	-0.007	0.007
	ATT	-0.099	0.021***	-0.105	0.021***	-0.106	0.017***	-0.113	0.016***	0.006	0.011	-0.005	0.009
Inhosptal death	ATE	0.001	0.006	0.001	0.006	0	0.001	0	0	-0.001	0.002	-0.001	0.001
	ATT	0.001	0.001	0.001	0.001	0	0.002	0	0	-0.001	0.003	-0.001	0.001
Diagnostic procedures	ATE	-0.003	0.004	0.007	0.004	-0.001	0.004	0.004	0.004	-0.004	0.013	0.007	0.008
	ATT	-0.002	0.004	0.008	0.004	-0.001	0.006	0.003	0.005	-0.016	0.014	0.005	0.008
Imaging procedures	ATE	-0.005	0.01	-0.011	0.01	-0.006	0.005	-0.006	0.006	-0.023	0.018	-0.021	0.009**
	ATT	-0.009	0.008	-0.011	0.008	-0.006	0.005	-0.006	0.004	-0.025	0.03	-0.021	0.011*
N.of post-admit diag	ATE	-0.018	0.036	0.004	0.036	-0.014	0.027	-0.009	0.019	-0.001	0.041	0.014	0.026
	ATT	-0.015	0.034	0.008	0.034	-0.001	0.031	-0.005	0.021	0.019	0.053	0.026	0.028
N.of interventions	ATE	-0.005	0.051	0.029	0.051	-0.048	0.075	0.122	0.062*	0.148	0.146	0.065	0.093
	ATT	-0.019	0.054	0.039	0.054	-0.064	0.081	0.115	0.067*	0.096	0.12	0.075	0.086
Pain symptoms	ATE	0.011	0.019	-0.001	0.019	-0.002	0.007	-0.003	0.006	-0.022	0.044	-0.016	0.035
	ATT	0.01	0.02	-0.004	0.02	-0.002	0.008	-0.005	0.008	0.034	0.05	-0.012	0.034
Readmission (180d)	ATE	-0.002	0.002	-0.001	0.002	0.003	0.002	0.003	0.001**	-0.003	0.007	-0.005	0.004
	ATT	0	0.002	0	0.002	0.003	0.003	0.003	0.001**	-0.004	0.009	-0.002	0.004
Revision (180d)	ATE	-0.002	0.005	-0.002	0.005	0.003	0.004	0.004	0.003	-0.001	0.008	-0.004	0.006
	ATT	0.001	0.005	0.001	0.005	0.003	0.004	0.005	0.003*	0	0.013	-0.004	0.007
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**	-0.075	0.37	-0.508	0.318
	ATT	-0.442	1.309	-0.365	1.309	-0.263	0.292	-0.442	0.186**	-0.404	0.5	-0.479	0.455
Transferred	ATE	0.011	0.018	0.011	0.018	0.008	0.013	0.016	0.01	-0.005	0.011	-0.001	0.008
	ATT	0.019	0.021	0.015	0.021	0.011	0.016	0.016	0.013	0.002	0.015	0.01	0.01
Charlson post	ATE	-0.004	0.016	-0.011	0.016	0.003	0.012	-0.017	0.011	-0.038	0.043	-0.012	0.04
	ATT	0.001	0.017	-0.006	0.017	0.002	0.012	-0.02	0.013	-0.019	0.048	-0.019	0.034
Elixhauser post	ATE	-0.021	0.136	-0.105	0.136	-0.065	0.156	0.118	0.136	0.536	0.588	0.04	0.385
	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

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

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 10, while tables with treatment effect estimates for other procedures can be found in Appendix 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 10: Difference-in-difference estimates with synthetic kernel-weighed control group, for unilateral hip and unilateral knee replacements

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

There are several results in this paper that may have policy implications. First, this study 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 a 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, it can be argued that exist several concerns 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 comorbidites 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 concern capable of putting at risk, to a largely unknown degree, the internal validity of the results may come from 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, albeit never null, 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 of this study can– with due diligence– be generalized to most Canadian jurisdictions and to countries having a similar institutional setting. To name a few and, arguably, the most important 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.

Further lines of research in which our research agenda can be expanded and elaborated are suggested in Section 9.

9 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 to 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.

Our study can be continued and complemented in a number of directions. First, 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 get specialized in services in which they have comparative advantages and to drop services at which they are less efficient in terms of cost and quality of care, one would expect a decrease in 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 can 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 can be looked into on the condition that an additional module of data is requested and provided by CIHI. In particular, one can 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 can be also performed if the amounts of QBP/HBAM payments are released to the public in the future

Finally, the impact of the QBP/HBAM policies can be analyzed with respect to physician characteristics and practice styles. This would allow for assessment of the impact of such characteristics as gender, age, school of graduation, which may have an impact on practice styles.

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:

<u>DAD – All Levels of Care</u>

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.

<u>NACRS – All Levels of Care</u>

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
2nd 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 (1st 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 (1st 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

DIAGNOSIS CODES OF INTEREST					
ICD-10-CA DIAGNOSIS DESCRIPTION	ICD-10-CA DIAGNOSIS CODES				NOTES
	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 &autog r	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	1VG53LAPPN	-	-	-	No change
Implant tri comp prosth knee OA &comb tis	1VG53LAPPQ	-	-	-	No change
Implant cement spacer knee OA	1VG53LASLN	-	-	-	No change
Implant internal device, patella	1VP53	-	-	-	1VP53LAPN added in v2012 1VP53LAPNN added in v2012
Implant dev patella OA &prosthesis synth mater	1VP53LAPMN	-	-	-	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 metatatarsophalangeal 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

<u>CMG/CACS Codes</u>

CASE MIX GROUPING (CMG) / Comprehensive Ambulatory Classification System (CACS)					
CMG/CACS DESCRIPTION	**CMG/CACS CODES**				**NOTES**
	v2006	**v2009**	**v2012**	**v2015**	
CMG CODES					
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.
CACS CODES					
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

Ontario

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 Financement 2nd temps
% des patients EDS dénutris pour lesquels une action correctrice pré opératoire a été mise en place	Processus		
% 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 *Nb d'ETP dédié à la coordination des épisodes de soins / Nb d'épisodes de soins*	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 ≥ 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 o ver 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 recommendations.

– 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 D. 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**
Added QBP procedures	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
QBPs intended funding in %	0	6	15
QBPs actual funding in %	0	6	12
HBAM intended funding in %	1,5	40	40
HBAM actual funding in %	0	34	34
Global budget intended funding in %	98,5	54	45
Global budget actual funding in %	100	60	54

Funding Components	**2014/15**	**2015/16**	**2016/17**
Added QBP procedures	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
QBPs intended funding in %	30	30	30
QBPs actual funding in %	13	14	15
HBAM intended funding in %	40	40	40
HBAM actual funding in %	33	32	32
Global budget intended funding in %	30	30	30
Global budget actual funding in %	54	54	54

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

Table 11: 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
Intercept	6.24	4.621	2.545	-0.004	-0.017	-0.01	0.001	-0.01	0.001
	(0.37)***	(0.479)***	(0.578)***	(0.007)	(0.007)*	(0.016)	(0.004)	(0.005)*	(0.013)
$post2012 \cdot QBP$	-0.3	-0.266	-0.273	-0.002	-0.003	-0.001	-0.003	-0.002	< 0.001
	(0.084)***	(0.086)**	(0.101)**	(0.002)	(0.002)	(0.002)	(0.001)*	(0.001)*	(0.001)
$post2014 \cdot QBP$	-0.359	-0.35	-0.132	0.001	< 0.001	-0.003	0.001	0.001	-0.002
	(0.079)***	(0.08)***	(0.114)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
age	-0.08	-0.076	-0.045	0.001	0.002	0.002	< 0.001	< 0.001	< 0.001
	(0.011)***	(0.011)***	(0.013)***	(< 0.001)***	(< 0.001)***	(< 0.001)***	(< 0.001)**	(< 0.001)***	(< 0.001)***
age^2	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)***
male	-0.273	-0.271	-0.231	< 0.001	< 0.001	-0.001	< 0.001	< 0.001	< 0.001
	(0.03)***	(0.03)***	(0.035)***	(0.001)	(0.001)	(0.001)	(< 0.001)	(< 0.001)	(< 0.001)
QBP flag	-0.418			0.002			-0.001		
	(0.052)***			(0.001)*			(0.001)*		
Resp. for payment:									
Worker insurance			-0.386			-0.001			0.005
			(0.288)			(0.009)			(0.006)
Other			0.586			-0.018			-0.009
			(0.791)			(0.003)***			(0.001)***
Income quintile:									
Lowest			0.515			-0.01			-0.012
			(0.229)*			(0.014)			(0.012)
Medium-low			0.305			-0.011			-0.013
			(0.226)			(0.014)			(0.012)
Middle			0.311			-0.011			-0.013
			(0.228)			(0.014)			(0.012)
Medium-high			0.249			-0.012			-0.013
			(0.225)			(0.014)			(0.012)
Highest			0.121			-0.012			-0.013
			(0.225)			(0.014)			(0.012)
Other controls	NO	NO	YES	NO	NO	YES	NO	NO	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Diagnosis FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Hospital FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
N	240954	240954	167322	240971	240971	167322	240971	240971	167322
R2	0.35	0.36	0.36	0.08	0.09	0.1	0.007	0.011	0.011
AdjR2	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 12: Difference-in-difference estimates for unilateral hip replacements (quality measures, continued)

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.
Dep. var./Model	1	2	3	4	5	6	7	8	9
Intercept	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)***
$post2012 \cdot QBP$	-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)
$post2014 \cdot QBP$	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)*
age	-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)***
age^2	< 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)***
male	-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)
QBP flag	0.037 (0.005)***			0.136 (0.019)***			0.068 (0.005)***		
Resp. for payment:									
Worker insurance			-0.031 (0.025)			-0.082 (0.098)			-0.073 (0.029)*
other			0.178 (0.148)			-0.122 (0.271)			-0.059 (0.08)
Income quintile:									
Lowest			0.013 (0.048)			0.194 (0.099)*			0.041 (0.034)
Medium-low			-0.002 (0.048)			0.235 (0.098)*			0.03 (0.034)
Middle			-0.002 (0.048)			0.235 (0.098)*			0.029 (0.034)
Medium-high			-0.007 (0.048)			0.288 (0.098)**			0.023 (0.034)
Highest			-0.023 (0.048)			0.313 (0.098)**			0.013 (0.034)
Other controls	NO	NO	YES	NO	NO	YES	NO	NO	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Diagnosis FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Hospital FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
N	240971	240971	167322	240971	240971	167322	240971	240971	167322
R2	0.2	0.21	0.21	0.22	0.22	0.22	0.37	0.4	0.41
AdjR2	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 13: 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
Intercept	8.203	7.562	7.318	0.062	0.066	0.048	0.062	0.065	0.042
	(0.336)***	(0.344)***	(0.373)***	(0.009)***	(0.01)***	(0.012)***	(0.007)***	(0.008)***	(0.01)***
$post2012 \cdot QBP$	-0.307	-0.271	-0.326	0.002	0.002	0.002	0.002	0.003	0.002
	(0.039)***	(0.04)***	(0.043)***	(0.001).	(0.001).	(0.001)	(0.001)*	(0.001)**	(0.001).
$post2014 \cdot QBP$	-0.206	-0.183	-0.042	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	(0.035)***	(0.035)***	(0.043)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
age	-0.158	-0.149	-0.138	< 0.001	< 0.001	< 0.001	-0.001	-0.001	-0.001
	(0.01)***	(0.01)***	(0.011)***	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)***	(< 0.001)***	(< 0.001)**
age^2	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)**
male	-0.237	-0.228	-0.225	0.004	0.004	0.004	0.004	0.004	0.004
	(0.013)***	(0.013)***	(0.014)***	(< 0.001)***	(< 0.001)***	(< 0.001)***	(< 0.001)***	(< 0.001)***	(< 0.001)***
QBP flag	-0.089			-0.001			-0.005		
	(0.025)***			(0.001).			(0.001)***		
Respons. for payment:									
Worker insurance			0.104			0.002			0.001
			(0.05)*			(0.003)			(0.002)
other			-0.087			-0.012			-0.007
			(0.172)			(0.006).			(0.003)*
Income quintile:									
Lowest			0.311			0.011			0.001
			(0.103)**			(0.003)***			(0.005)
Medium-low			0.166			0.011			0.001
			(0.103)			(0.003)***			(0.005)
Middle			0.136			0.011			0.001
			(0.103)			(0.003)***			(0.005)
Medium-high			0.107			0.011			0.001
			(0.102)			(0.003)***			(0.005)
High			0.009			0.011			< 0.001
			(0.103)			(0.003)***			(0.005)
Other controls	NO	NO	YES	NO	NO	YES	NO	NO	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Diagnosis FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Hospital FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
N	374165	374165	265620	374317	374317	265620	374317	374317	265620
R2	0.24	0.26	0.25	0.14	0.14	0.15	0.01	0.02	0.01
AdjR2	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 14: Difference-in-difference estimates for unilateral knee replacements (quality measures, continued)

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.
Dependent var.	1	2	3	4	5	6	7	8	9
Intercept	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)***
$post2012 \cdot QBP$	-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)***
$post2014 \cdot QBP$	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)***
age	-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)***
age^2	< 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)***
male	-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)***
QBP flag	0.023 (0.003)***			0.113 (0.013)***			0.071 (0.003)***		
Resp. for payment:									
Worker insurance			-0.037 (0.033)			0.044 (0.043)			0.021 (0.014)
other			-0.184 (0.077)*			0.145 (0.094)			-0.019 (0.038)
Income quintile:									
Lowest			0.328 (0.101)**			-0.269 (0.157).			0.045 (0.035)
Medium-low			0.266 (0.101)**			-0.24 (0.157)			0.04 (0.035)
Middle			0.224 (0.101)*			-0.218 (0.157)			0.045 (0.035)
Medium-high			0.195 (0.101).			-0.2 (0.157)			0.053 (0.035)
Highest			0.157 (0.101)			-0.158 (0.157)			0.07 (0.035)*
Other controls	NO	NO	YES	NO	NO	YES	NO	NO	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Diagnosis FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Hospital FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
N	374317	374317	265620	374316	374316	265620	374317	374317	265620
R2	0.09	0.1	0.31	0.11	0.13	0.13	0.26	0.36	0.26
AdjR2	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 15: Difference-in-difference estimates by post-reform year for unilateral replacements (quality measures)

	Unilateral Hip						Unilateral knee					
Outcome	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
$qbp \cdot year2012$	-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)
$qbp \cdot year2013$	-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)***
$qbp \cdot year2014$	-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)***
$qbp \cdot year2015$	-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)***
$qbp \cdot year2016$	-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)***
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Diagnosis FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	167322	167322	167322	167322	167322	167322	265620	265620	265620	265620	265620	265620
R2	0.36	0.01	0.1	0.21	0.22	0.41	0.27	0.01	0.15	0.11	0.13	0.38
AdjR2	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 16: Difference-in-difference estimates by post-reform year for bilateral knee and hip replacements (quality measures)

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

Table 17: Hospital-level difference-in-difference estimates for unilateral hip and unilateral knee replacements, by year pre- and post-reform years

	QBP year 2008	Sd	QBP year 2009	Sd	QBP year 2010	Sd	QBP year 2012	Sd	QBP year 2013	Sd	QBP year 2014	Sd	QBP year 2015	Sd	QBP year 2016	Sd	QBP year 2017	Sd	N	R^2	$AdjR^2$
Hip unilateral:																					
Acute LOS	-0.024	(0.296)	0	(0.285)	7	(0.286)	-0.133	(0.274)	-0.513	(0.287)*	-0.654	(0.275)**	-0.718	(0.274)***	-0.581	(0.272)**	-0.416	(0.279)	1128	0.795	0.773
Antibiotic use	-0.001	(0.002)	-0.001	(0.002)	-0.001	(0.002)	-0.002	(0.002)	-0.004	(0.002)**	-0.003	(0.002)	-0.003	(0.002)	-0.001	(0.003)	-0.002	(0.002)	1138	0.627	0.586
Discharged home	-0.08	(0.041)**	-0.079	(0.039)**	-0.075	(0.036)**	0.049	(0.031)	0.047	(0.028)*	0.05	(0.026)*	0.053	(0.026)**	0.036	(0.025)	0.039	(0.025)	1138	0.556	0.508
Discharged home with support	-0.064	(0.074)	-0.126	(0.069)*	-0.046	(0.069)	0.037	(0.067)	0.07	(0.066)	0.131	(0.063)**	0.106	(0.069)	0.063	(0.072)	0.103	(0.069)	1138	0.518	0.466
Discharged support service	0.073	(0.036)**	0.071	(0.034)**	0.05	(0.029)*	-0.044	(0.022)**	-0.061	(0.021)***	-0.064	(0.021)***	-0.07	(0.021)***	-0.058	(0.021)***	-0.063	(0.021)***	1138	0.564	0.517
Inhospital death	0	(0.002)	0	(0.001)	0.001	(0.001)	0.002	(0.001)	0	(0.001)	0	(0.001)	0	(0.001)	0.002	(0.001)	0.001	(0.001)	1138	0.675	0.64
Diagnostic procedures	-0.003	(0.005)	0	(0.005)	-0.006	(0.005)	-0.004	(0.005)	-0.008	(0.005)	-0.003	(0.005)	0.002	(0.005)	0.007	(0.007)	0.005	(0.005)	1138	0.409	0.345
Imaging procedures	0.004	(0.01)	-0.003	(0.01)	0.003	(0.01)	-0.005	(0.008)	-0.012	(0.009)	-0.016	(0.009).	-0.007	(0.011)	-0.006	(0.008)	-0.013	(0.008)	1138	0.539	0.489
N.of post-admit diag	-0.02	(0.036)	-0.02	(0.037)	-0.02	(0.039)	-0.028	(0.038)	-0.035	(0.034)	-0.029	(0.034)	-0.015	(0.036)	0.001	(0.033)	0.008	(0.033)	1138	0.482	0.426
N.of interventions	0.043	(0.058)	0.05	(0.054)	0.046	(0.059)	-0.005	(0.054)	-0.04	(0.054)	-0.012	(0.06)	0.06	(0.052)	0.071	(0.051)	0.03	(0.053)	1138	0.587	0.542
Pain symptoms	-0.003	(0.004)	-0.001	(0.003)	0.001	(0.004)	0.001	(0.003)	0.003	(0.004)	0	(0.003)	0.001	(0.003)	0	(0.003)	0.003	(0.003)	1138	0.042	0.037
Readmission (180d)	-0.005	(0.003)*	-0.004	(0.002)*	-0.001	(0.002)	-0.005	(0.003)*	-0.004	(0.003)	-0.003	(0.002)	-0.007	(0.003)**	-0.002	(0.002)	0	(0.002)	1138	0.361	0.292
Revision (180d)	0	(0.003)	-0.002	(0.003)	0.002	(0.003)	-0.003	(0.003)	0.001	(0.003)	0.003	(0.003)	-0.003	(0.003)	0.002	(0.003)	0	(0.003)	1138	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)	0.001	(0.002)	0.002	(0.003)	-0.001	(0.002)	-0.001	(0.003)	1138	0.39	0.324
Total LOS	0.26	(0.336)	0.139	(0.337)	0.009	(0.348)	0.035	(0.321)	-0.469	(0.35)	-0.599	(0.321)*	-0.947	(0.34)***	-0.639	(0.335)*	-0.6	(0.36)*	1134	0.654	0.616
Transferred	-0.005	(0.024)	0.004	(0.022)	0.02	(0.023)	-0.003	(0.022)	0.014	(0.02)	0.014	(0.018)	0.016	(0.017)	0.019	(0.017)	0.026	(0.017)	1138	0.533	0.483
Charlson post	-0.022	(0.02)	0.033	(0.019)*	-0.003	(0.019)	-0.001	(0.018)	-0.002	(0.017)	0	(0.019)	0.003	(0.018)	-0.007	(0.017)	-0.01	(0.018)	1106	0.216	0.128
Elixhauser post	-0.08	(0.229)	-0.07	(0.233)	-0.098	(0.224)	-0.152	(0.223)	-0.088	(0.212)	-0.069	(0.21)	0.1	(0.231)	-0.257	(0.215)	-0.247	(0.271)	1106	0.208	0.12
Knee unilateral:																					
Acute LOS	0.089	(0.198)	0.137	(0.185)	0.216	(0.167)	-0.057	(0.157)	-0.349	(0.189).	-0.514	(0.172)***	-0.393	(0.151)***	-0.326	(0.149)**	-0.305	(0.159)*	1119	0.68	0.647
Antibiotic use	0.001	(0.002)	-0.003	(0.004)	-0.001	(0.002)	-0.002	(0.002)	-0.001	(0.002)	-0.001	(0.002)	0	(0.002)	-0.001	(0.002)	-0.001	(0.002)	1152	0.594	0.553
Discharged home	-0.097	(0.039)**	-0.076	(0.036)**	-0.042	(0.029)	0.055	(0.024)**	0.067	(0.024)***	0.062	(0.023)***	0.069	(0.023)***	0.063	(0.023)***	0.059	(0.022)***	1152	0.51	0.46
Discharged home with support	-0.07	(0.082)	-0.081	(0.079)	-0.048	(0.079)	0.099	(0.078)	0.112	(0.077)	0.144	(0.08)*	0.107	(0.081)	0.061	(0.081)	0.089	(0.079)	1152	0.421	0.363
Discharged support service	0.085	(0.036)**	0.068	(0.033)**	0.032	(0.025)	-0.064	(0.019)***	-0.074	(0.02)***	-0.075	(0.019)***	-0.087	(0.019)***	-0.083	(0.019)***	-0.08	(0.019)***	1152	0.519	0.471
Inhosptal death	0	(0.001)	0	(0.001)	-0.001	(0.001)*	0	(0)	0	(0)	-0.001	(0.001)	0	(0)	0	(0.001)	0	(0)	1152	0.437	0.38
Diagnostic procedures	-0.005	(0.004)	-0.006	(0.004)	-0.002	(0.003)	-0.004	(0.003)	-0.006	(0.004)*	-0.004	(0.003)	-0.004	(0.005)	0.001	(0.005)	-0.001	(0.005)	1152	0.288	0.216
Imaging procedures	0.009	(0.007)	0.002	(0.006)	0	(0.007)	0.006	(0.005)	-0.008	(0.006)	-0.008	(0.005)	-0.003	(0.006)	0	(0.008)	-0.003	(0.008)	1152	0.377	0.313
N.of post-admit diag	-0.019	(0.028)	-0.013	(0.028)	-0.005	(0.027)	-0.012	(0.027)	-0.019	(0.025)	-0.02	(0.026)	-0.016	(0.027)	-0.015	(0.025)	-0.007	(0.025)	1152	0.458	0.403
N.of interventions	-0.002	(0.08)	0.024	(0.092)	-0.016	(0.09)	-0.173	(0.086)**	-0.06	(0.077)	-0.033	(0.079)	0.032	(0.071)	0.106	(0.07)	0.115	(0.069)*	1152	0.517	0.468
Pain symptoms	0	(0.002)	0	(0.002)	0.001	(0.002)	0	(0.002)	0	(0.002)	0	(0.002)	0.001	(0.002)	0	(0.002)	0.002	(0.002)	1152	0.039	0.037
Readmission (180d)	0	(0.002)	0	(0.002)	-0.001	(0.002)	0.002	(0.002)	0.001	(0.002)	0	(0.002)	0.002	(0.002)	0.001	(0.002)	0.002	(0.002)	1152	0.351	0.285
Revision (180d)	0.001	(0.002)	0	(0.002)	-0.001	(0.002)	0.002	(0.002)	0.002	(0.002)	0.001	(0.002)	0.003	(0.002)*	0.001	(0.002)	0.002	(0.002)	1152	0.75	0.725
Share unilateral repl.	-0.008	(0.009)	-0.006	(0.008)	-0.004	(0.008)	-0.004	(0.008)	0	(0.008)	0.001	(0.009)	-0.001	(0.009)	-0.003	(0.008)	-0.004	(0.008)	1152	0.222	0.143
Total LOS	0.086	(0.206)	0.108	(0.194)	0.234	(0.175)	-0.097	(0.165)	-0.414	(0.197)**	-0.557	(0.179)***	-0.454	(0.158)***	-0.415	(0.157)***	-0.395	(0.168)**	1140	0.675	0.642
Transferred	0.003	(0.021)	0.004	(0.02)	0.007	(0.017)	0.009	(0.016)	0.007	(0.016)	0.013	(0.015)	0.016	(0.015)	0.02	(0.015)	0.022	(0.014)	1152	0.451	0.395
Charlson post	0.003	(0.018)	0	(0.019)	0.015	(0.017)	0.01	(0.015)	0.011	(0.017)	0.001	(0.021)	-0.006	(0.015)	-0.01	(0.016)	-0.016	(0.016)	1094	0.155	0.064
Elixhauser post	-0.124	(0.203)	0.14	(0.196)	0.114	(0.206)	0.021	(0.214)	0.127	(0.233)	0.352	(0.235)	0.324	(0.236)	0.213	(0.241)	-0.035	(0.217)	1094	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 18: 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:							
Discharged home	0.029	(0.109)	-0.008	(0.116)	557	0.304	0.281
Discharged home with support	0.121	(0.065)*	-0.015	(0.064)	557	0.089	0.059
Discharged support service	0.042	(0.111)	-0.022	(0.117)	557	0.365	0.343
Transferred	-0.069	(0.055)	0.029	(0.056)	557	0.205	0.179
Knee unilateral:							
Discharged home	0.112	(0.016)***	0.003	(0.011)	1152	0.502	0.455
Discharged home with support	0.152	(0.048)***	-0.005	(0.047)	1152	0.42	0.365
Discharged support service	-0.113	(0.015)***	-0.013	(0.009)	1152	0.513	0.467
Transferred	0.005	(0.009)	0.01	(0.006)	1152	0.45	0.398
Hip bilateral:							
Discharged home	0.464	(0.151)***	-0.226	(0.146)	265	0.269	0.216
Discharged home with support	0.209	(0.136)	0.057	(0.143)	265	0.304	0.253
Discharged support service	-0.37	(0.142)***	0.252	(0.137)*	265	0.319	0.269
Transferred	-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 19: Hospital-level difference-in-difference estimates for other replacements

	Variable (SE)				Model statistics:		
	post2012 QBP	Sd	post2014 QBP	Sd	N	R2	AdjR2
Other replacement:							
Acute LOS	0.082	(0.208)	-0.056	(0.201)	1011	0.362	0.337
Total LOS	-0.009	(0.309)	-0.394	(0.324)	1143	0.201	0.173
Discharged home	-0.003	(0.009)	0.01	(0.008)	1201	0.196	0.169
Discharged home with support	0.029	(0.038)	0.046	(0.038)	1201	0.274	0.249
Discharged support service	0.003	(0.008)	-0.012	(0.006)*	1201	0.243	0.218
Transferred	0.005	(0.006)	0.002	(0.005)	1201	0.145	0.116
N.of post-admit diag.	0.01	(0.027)	0	(0.025)	1201	0.224	0.198
N.of interventions	0.046	(0.073)	-0.034	(0.072)	1201	0.427	0.408
Pain symptoms	0	(0.009)	0.005	(0.007)	1201	0.047	0.039
Readmission (180d)	0.001	(0.003)	-0.002	(0.003)	1201	0.119	0.089
Revision (180d)	0.007	(0.005)	-0.005	(0.005)	1201	0.235	0.21
Elixhauser post	0.272	(0.461)	-0.326	(0.448)	690	0.082	0.027
Diagnostic procedures	-0.004	(0.008)	0.004	(0.008)	1201	0.122	0.092
Imaging procedures	-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 20: 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			
(Intercept)	0.906	-2.519	0.499
	(0.733)	(1.621)	(0.674)
post2012 QBP	0.104	0.11	-0.099
	(0.02)***	(0.043)**	(0.015)***
post2014 QBP	-0.004	0.047	-0.011
	(0.014)	(0.042)	(0.011)
age	0.003	0.046	-0.016
	(0.024)	(0.054)	(0.021)
age2	0	0	0
	(0)	(0)	(0)
female	-0.043	-0.219	0.122
	(0.071)	(0.198)	(0.065)*
urban	-0.048	-0.018	0.062
	(0.009)***	(0.029)	(0.008)***
QBP flag dummy	-0.144	0.208	0.181
	(0.017)***	(0.032)***	(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 21: Hospital-level difference-in-difference estimates for all replacements

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

Table 22: Estimates of nearest-neighbor covariate matching for bilateral hip and bilateral knee replacements

		Hip bilateral				Knee bilateral			
variable	Effect	Effect 2012-13	SE	Effect 2014-17	SE	Effect 2012-13	SE	Effect 2014-17	SE
Acute LOS	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
Antibiotic use	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
Discharged home	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
Discharged home with support	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
Discharged supportservice	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
Inhosptal death	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
Diagnostic procedures	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
Imaging procedures	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
N.of post-admit diag	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***
N.of interventions	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
Pain symptoms	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
Readmission (180d)	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
Revision (180d)	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
Share bilateralreplacements	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
Total LOS	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
Transferred	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*
Charlson post	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
Elixhauser post	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 23: Estimates of nearest-neighbor covariate matching for all joint replacements

Variable	Effect	Effect 2012-13	SE	p-val	Effect 2014-17	SE	p-val
Acute LOS	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
Antibiotic use	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
Discharged home	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
Discharged home with support	ATE	0.114	0.048**	0.018	0.143	0.037***	0
	ATT	0.223	0.065***	0.001	0.232	0.052***	0
Discharged support service	ATE	-0.06	0.016***	0	-0.071	0.015***	0
	ATT	-0.087	0.022***	0	-0.095	0.02***	0
Elbow replacement	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
Share bilateral hip	ATE	-0.001	0	0.183	-0.001	0*	0.096
	ATT	0	0.001	0.609	-0.001	0.001*	0.083
Share unilateral hip	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
Inhosptal death	ATE	-0.002	0.001	0.102	-0.002	0.001*	0.056
	ATT	0	0.001	0.99	0	0.001	0.638
Diagnostic procedures	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
Imaging procedures	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
Share bilateral knee	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
Share unilateral knee	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
N.of post-admit diag	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
N.of interventions	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
Share other replacements	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
Pain symptoms	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
Readmission (180d)	ATE	0.003	0.002*	0.083	0.004	0.001***	0
	ATT	0.004	0.002**	0.048	0.005	0.001***	0
Revision (180d)	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
Share shoulder replacement	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
Total LOS	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
Transferred	ATE	0.019	0.01**	0.045	0.037	0.008***	0
	ATT	0.04	0.014***	0.004	0.05	0.012***	0
Charlson post	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
Elixhauser post	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 24: Propensity score matching estimates with kernel weights (bilateral hip and bilateral knee replacements)

		Hip bilateral				Knee bilateral			
Variable	Effect	Effect 2012-13	SE	Effect 2014-17	SE	Effect 2012-13	SE	Effect 2014-17	SE
Acute LOS	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
Antibiotic use	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
Discharged home	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
Discharged home with support	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
Discharged support service	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
Diagnostic procedures	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
Imaging procedures	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**
N.of post-admit diag	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
N.of interventions	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
Pain symptoms	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
Readmission (180d)	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
Revision (180d)	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
Share bilateral repl.	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
Total LOS	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
Transferred	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 25: Propensity score matching estimates with kernel weights (all joint replacements)

Variable	Effect	Effect 2012-13	SE	p-val	Effect 2014-17	SE	p-val
Acute LOS	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
Antibiotic use	ATE	-0.001	0.002	0.655	0.001	0.001	0.48
	ATT	-0.001	0.003	0.724	0	0.002	1
Day surgery	ATE	0	0.008	1	0	0.004	1
	ATT	-0.001	0.013	0.939	-0.001	0.008	0.896
Discharged home	ATE	0.105	0.024***	0	0.101	0.02***	0
	ATT	0.102	0.027***	0	0.097	0.021***	0
Discharged home with support	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
Discharged support service	ATE	-0.101	0.019***	0	-0.109	0.017***	0
	ATT	-0.102	0.019***	0	-0.105	0.017***	0
Elbow replacement	ATE	-0.001	0.001	0.48	-0.001	0.001	0.48
	ATT	-0.001	0.001	0.48	0	0.001	1
Share bilateral hip	ATE	0	0.001	1	-0.001	0.001	0.317
	ATT	0	0.001	1	-0.001	0.001	0.317
Share unilateral hip	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
Diagnostic procedures	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
Imaging procedures	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
Share bilateral knee	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
Share unilateral knee	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
N.of post-admit diag	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
N.of interventions	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
Share other replacements	ATE	-0.001	0.006	0.864	0	0.004	1
	ATT	-0.002	0.009	0.832	-0.001	0.006	0.864
Pain symptoms	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
Readmission (180d)	ATE	0	0.002	1	0.001	0.001	0.48
	ATT	0	0.002	1	0.001	0.001	0.48
Revision (180d)	ATE	0	0.003	1	0.002	0.002	0.371
	ATT	-0.001	0.004	0.782	0.002	0.003	0.48
Share shoulder replacement	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
Total LOS	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
Transferred	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
Charlson post	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
Elixhauser post	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 H. Synthetic control group estimators (additional tables)

Table 26: Difference-in-difference estimates with synthetic kernel-weighed control group, for bilateral hip and bilateral knee replacements

	Variable (SE)				Model statistics:		
	post2012 QBP	Sd	post2014 QBP	Sd	N	R2	AdjR2
Hip bilateral:							
Acute LOS	2.498	(1.619)	-2.415	(1.323)*	242	0.398	0.35
Discharged home	0.16	(0.159)	0.133	(0.152)	242	0.127	0.056
Discharged support service	-0.092	(0.148)	-0.035	(0.136)	242	0.214	0.15
Imaging procedures	0.02	(0.044)	0.033	(0.052)	242	0.116	0.045
N.of post-admit diag	0.723	(0.333)**	-0.728	(0.297)**	242	0.335	0.282
N.of interventions	0.362	(0.272)	-0.266	(0.193)	242	0.138	0.068
Pain symptoms	-0.151	(0.084)*	0.156	(0.069)**	242	0.024	0.019
Share bilateral repl.	-0.003	(0.003)	0.003	(0.003)	242	0.159	0.091
Total LOS	2.703	(1.67)	-3.748	(1.645)**	242	0.366	0.315
Transferred	-0.056	(0.055)	-0.114	(0.065)*	242	0.229	0.167
Elixhauser post	0.019	(0.111)	0.025	(0.111)	242	0.328	0.277
Bilateral knee :							
Acute LOS	0.341	(0.953)	-0.92	(0.996)	538	0.114	0.083
Discharged home	0.177	(0.12)	-0.1	(0.123)	538	0.098	0.066
Discharged support service	-0.17	(0.128)	0.122	(0.131)	538	0.082	0.05
Diagnostic procedures	-0.024	-0.033	0.023	(0.025)	538	0.062	0.029
Imaging procedures	-0.039	(0.034)	0.013	(0.047)	538	0.113	0.082
N.of post-admit diag	-0.14	(0.115)	-0.021	(0.118)	538	0.134	0.104
N.of interventions	0.031	(0.223)	0.464	(0.209)**	538	0.131	0.101
Pain symptoms	-0.013	(0.019)	0.023	(0.033)	538	0.038	0.025
Share bilateral repl.	0.004	(0.008)	-0.006	(0.008)	538	0.082	0.051
Total LOS	0.105	(0.965)	-0.776	(1.084)	538	0.129	0.099
Transferred	-0.005	(0.036)	-0.017	(0.036)	538	0.086	0.054
Elixhauser post	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 27: Difference-in-difference estimates with synthetic kernel-weighed control group, all joint replacements

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

References

[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] 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).

[5] Baker L. Acquisition Of MRI Equipment By Doctors Drives Up Imaging Use And Spending *Health Affairs*. 2010; Vol. 29, No. 12.

[6] 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

[7] Busse R. et al. Diagnosis-related Groups in Europe: Moving Towards Transparency, Efficiency and Quality in Hospitals, 2011.

[8] 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

[9] 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

[10] 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).

[11] Contoyannis P., Wildman J. . Using relative distributions to investigate the body mass index in England and Canada. *Health economics*. Vol.16, (2007).

[12] Dafny, L.(2005). How Do Hospitals Respond to Price Changes? *The American Economic Review*, 95(5), pp. 1525-1547.

[13] Eggleston K. Multitasking and mixed systems for provider payment. *Journal of Health Economics*. Volume 24, Issue 1, (2005), pp. 211-223.

[14] Ellis R. Creaming, skimping and dumping: provider competition on the intensive and extensive margins, *Journal of Health Economics*, Vol. 17, pp. 537—555. (1998)

[15] 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).

[16] 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

[17] 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.

[18] 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

[19] 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).

[20] 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.

[21] 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.

[22] Johnson E., Rehavi M. (2016) Physicians Treating Physicians: Information and Incentives in Childbirth. *American Economic Journal: Economic Policy* (2016), 8(1), pp. 115-141.

[23] Kaarboe, O. Siciliani, L. (2008). Multitasking, quality and pay for performance. *Health Econonomics*, Vol.20, Issue2, (2011), pp. 225-238.

[24] Kuhn, M. Siciliani, L. . Performance Indicators for Quality with Costly Falsification. Journal of Economics Management Strategy. Vol. 18., (2009), pp.1137-1154.

[25] 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.

[26] Makris M., Siciliani L. Optimal Incentive Schemes for Altruistic Providers. Journal of Public Economic theory. Vol. 15, Issue5, (2013), pp. 675-699.

[27] 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, Issue1, (2010), pp. 64-91.

[28] 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).

[29] National Health Service. Hip replacement (Overview, How it is performed, Risks, Recovery, Metal implant advice) (2019a) https://www.nhs.uk/conditions/hip-replacement/

[30] National Health Service. Knee replacement (Overview, How it is performed, Risks, Recovery) (2019b) https://www.nhs.uk/conditions/knee-replacement/

[31] Norton C. Incentive regulation of nursing homes. *Journal of Health Economics*, Vol. 11, Issue 2, 1992, pp. 105-128.

[32] OECD (2019), Health spending (indicator). doi: 10.1787/8643de7e-en (Accessed on 18 November 2019)

[33] Ontario Hospital Association. HSFR Resources, 2019 (accessed on 18 November 2019) https://www.oha.com/health-system-transformation/health-system-funding-reform/hsfr-resources

[34] 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).

[35] 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).

[36] 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

[37] 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

[38] 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.

[39] 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).

[40] 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

[41] Shigeoka H., Fushimib K. Supplier-induced demand for newborn treatment: Evidence from Japan *Journal of Health Economics*. 2014; 35: 162–178

[42] Siciliani L. Paying for performance and motivation crowding out. *Economics Letters*. Vol.103, (2009), pp. 68–71.

[43] 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)

[44] Schwartz M., Tartter P. Decreased Length of Stay for Patients with Colorectal Cancer: Implications of DRG Use. *Journal for healthcare quality*, (2011).

[45] 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).