Trois essais sur l'impact des incitatifs financiers sur la productivité du système de la santé au Québec

Thèse

Nizar Ghali

Doctorat en économique
Philosophiæ doctor (Ph.D.)

Québec, Canada

Trois essais sur l'impact des incitatifs financiers sur la productivité du système de la santé au Québec

Thèse

Nizar Ghali

Sous la direction de:

Bernard Fortin, directeur de recherche
Guy Lacroix, codirecteur de recherche

Résumé

Les incitatifs financiers dans le réseau de la santé sont reconnus avoir un effet sur le comportment des différents acteurs : médecins, patients, gestionnaires d'établissements. Ils peuvent prendre la forme de modèles de financement comme le financement à l'activité ou le financement selon la meilleure pratique ou bien des récompenses/pénalités financières offertes/appliquées selon le niveau des résultats atteints par rapport aux cibles fixées. Ces incitatifs financiers peuvent être liés directement au financement de l'établissement de santé ou à la rémunération médicale ou les deux simultanément.

L'impact de ces leviers a été très peu testé dans le contexte québécois. L'objectif de cette thèse est de contribuer à la littérature d'évaluation de cet impact sur différents aspects de la productivité du système de santé au Québec. Ce dernier étant un système public caractérisé par une gouvernance coordonnée par le ministère de la santé et des services sociaux (MSSS). Les modèles de financement du réseau étant basé sur le modèle historique alors que la rémunération des professionnels de la santé étant majoritairement basée sur l'acte médical. Ces deux enveloppes budgétaires sont distinctes et le médecin prend le statut de travailleur autonome.

Dans le premier chapitre, nous étudions l'effet causal d'un programme de financement à l'activité, le programme d'accès à la chirurgie, sur l'accès aux services et la qualité des soins. En utilisant des données administratives du Québec et un groupe de contrôle (données similaires de la Colombie britannique) et en se plaçant dans le cadre d'une approche de différence en différence, nous montrons que ce programme instauré depuis 2004 pour le secteur de la chirurgie et pour l'ensemble du réseau de la santé a permis de baisser les délais d'attente moyens aisni que les durées de séjour à l'hôpital, notamment pour les chirurgies de hanche et de genou, sans qu'il y ait détérioration dans certains indicateurs de la qualité des soins. En plus, l'effet de ce financement est non seulement positif mais aussi croît avec le niveau de financement.

Dans le deuxième chapitre, mon analysons l'effet causal de l'introduction d'un programme de dépistage du cancer colorectal pour certains établissements pilotes. En utilisant les données des établissements non traités comme groupe de contrôle et à l'aide d'un modèle de transition multi-états, nous montrons l'effet positif de ce programme sur la qualité des soins ainsi que sur la santé de la

population. Étant introduit au début en tant que stratégie clinique et combiné par la suite avec un financement récurrent selon la performance, ce programme a contribué à la baisse des durées de séjour pour un retour à domicile et à la diminution des coûts de traitement pour une chirurgie colorectale. Contrairement aux résultats du premier chapitre, cette analyse n'a pas permis de démontrer un effet positif du financement sur la baisse des durées de séjour. Ceci peut être dû à la courte durée de notre échantillon à partir de la date de l'annonce du financement additionnel pour les établissements pilotes. De l'autre côté, ce financement à la performance a contribué à une utilisation accrue d'une approche de traitement moins invasive.

Dans le dernier chapitre, nous réalisons une analyse coût-efficience de ce programme de qualité afin de juger de la pertinence de poursuivre l'application de cette stratégie basée sur les protocoles cliniques ainsi que la pertinence de poursuivre le financement additionnel. Nous démontrons que le ratio bénéfice-coût de la stratégie clinique (tests de dépistage et protocole clinique) est non seulement supérieur à l'unité mais aussi supérieur à celui du programme incluant le financement additionnel. Ces résultats suggère une revue de la stratégie des incitatifs financiers en lien avec ce programme.

Dans cette thèse, nous montrons comment les incitatifs financiers peuvent contribuer au changement du comportement et améliorer certains aspects de la productivité du système de la santé. Les leviers financiers ont été capables d'agir sur le comportement des médecins, dans la majorité des situations, malgré qu'ils ne sont pas directement liés à la rémunération médicale. Ceci témoigne d'une façon d'agir de la part des médecins qui n'est pas encore complètement documentée mais qui n'est certes pas détachée du contexte financier de l'établissement de santé.

Cependant, ces leviers financiers doivent être utilisés dans le cadre général d'une stratégie clinique offrant un certain seuil minimal de conditions de réussite et d'atteinte des objectifs. Ils ne peuvent agir seuls dans le sens de l'objectif à atteindre mais certes en cohérence avec toute stratégie clinique basée sur un partenariat clinico-administratif.

Abstract

Financial incentives in health network system are supposed to have an effect on the behaviour of the different healthcare stakeholders : physicians, patients, facilities managers. They may be proposed as funding models such as activity based funding or best practice paiement or rewards / financial penalties offered / applied according to outcome indicators achieved and compared to targets. Financial incentives could be linked directly to the healthcare facility funding or to medical payments or both.

The impact of these levers has been little tested in the Quebec context. The objective of this thesis is to contribute to the literature of financial incentives impact assessment on various aspects of healthcare system productivity in Quebec. The latter being a public system characterized by a governance coordinated by the Ministry of Health and Social Services (MSSS). Network funding models being based on global budget while the healthcare professionnels paiement is mainly based on fees for services. Both budgets are distinct and the physician takes self-employed status.

In the first chapter, we assess for the causal effect of the first Quebec activity based funding program, access to surgery program, on access to services and healthcare quality. Using Quebec administrative data, a control group (similar data from British Columbia) and a difference in difference approach, we show that this program, introduced in 2004 for the surgery sector in all facilities, reduced waiting times and hospital lengths of stay, especially for hip and knee replacements, without deterioration in some healthcare quality indicators. In addition, the effect of this funding is not only positive but also increases with funding level.

In the second chapter, we estimate the causal effect of the introduction of a colorectal cancer screening program for some pilot facilities. Using data from untreated hospitals as control group and a multistate model, we show the positive impact of this program on the healthcare quality and population health. Introduced at the beginning as a clinical strategy and combined later with recurrent performance payments, the program has contributed to the decrease of hospital lengths of stay with home discharge and also lower treatment costs for colorectal surgeries. Contrary to the results of the first chapter, this analysis did not demonstrate a positive effect of financial incentives on lower lengths of stay. This may be due to the short duration of our sample since the date of additional funding announcement. On the other side, financial incentives contributed to increased use of a less invasive

treatment approach.

In the last chapter, we perform a cost-effectiveness analysis of this healthcare quality program to identify the relevance of continuing implementing the strategy based on clinical protocols and additional funding. We demonstrate that the benefit-cost ratio of clinical strategy (screening tests and clinical protocol) is not only greater than unity but also higher than the program including the additional funding ratio. These results suggest a review of the financial incentives strategy for this program.

In this thesis , we show how financial incentives can help behaviour move and improve certain productivity aspects in the healthcare system. The financial levers have been able to influence the physicians behaviour, in most situations, although they are not directly related to their payments. This reflects a way of behaving for physicians that is not yet fully known but is certainly not disconnected from facilities financial context.

Finally, these financial levers must be used in the general framework of a clinical strategy providing a minimum level of success conditions and achievement of objectives. They can not act alone in the direction of the goal but certainly they should be consistent with any clinical strategy especially when based on clinical-administrative partnership.

Table des matières

Liste des tableaux

Liste des figures

Remerciements

En premier lieu, je remercie Dieu qui m'a facilité la réalisation de ce travail tout au long de ces dernières années. Je remercie les membres du jury de ma thèse pour leur disponibilité et pour l'intérêt qu'ils ont porté à mon travail. Je remercie également M. Bernard Fortin et M. Guy Lacroix qui ont encadré mes recherches et qui se sont toujours soucié de m'offrir les meilleurs conditions de travail possibles. Je les remercie également pour leur disponibilité, leur support tout au long de mes travaux ainsi que pour la confiance qu'ils m'ont accordée.

Je remercie le département d'économie de l'université Laval ainsi que la chaire Industrielle Alliance de m'avoir accordé leut soutien financier, académique et moral lors de mes travaux.

Je tiens aussi à remercier toutes les personnes du ministère de la santé du Québec qui m'ont aidé et plus particuièrement M. Normand Lantagne pour son soutien et sa qualité scientifique. Je cite aussi Mme Le boeuf, M. Jackson, Mme Katsahian et M. Deschênes pour leurs commentaires forts précieux.

Je voudrais également exprimer toute ma reconnaissance à mes parents et à ma femme pour leur aide précieuse, leur patience et leur soutien moral surtout lors des périodes plus difficiles.

Introduction

The current economic context of the Quebec and the budget problems could make more and more pressure on government spending. Health system is one of the most targeted sectors because of its significant share in the budget. To optimize these expenses and improve the efficiency of the Quebec health system, system performance evaluation is one of the first essential steps. In many cases, improving system productivity requires the use of financial incentives for physicians or facilities. These financial incentives may be a part of physicians remuneration or activity based funding for hospitals but also may be offered as payments by result. These levers are supposed to act on certain important parameters of productivity such as behaviour change, supply level, organization of healthcare services.

Our thesis deals with this topic. Specifically, we studied the impact of financial incentives on the evolution of the Quebec healthcare system productivity. This performance is defined by some outcome indicators such as mortality, readmission, length of stay, wait times, etc.

The Quebec healthcare system is public system that is characterized by global budget funding for facilities (based on historical spending) which is independent of the physicians remuneration. The system is based on a set of health facilities (hospitals, CLSCs [1], youth centers, etc.) which are funded directly by the Ministry of Health and Social Services (MSSS).

Usually, financial incentive is considered as a lever for behaviour move. Since the change involves many people (physicians, nurses) or entities (facilities, medical staff), it would be appropriate to separate the effect of the lever on each actor. This concern becomes even more important in our Quebec context. Indeed, physicians remuneration and facilitiesl funding are respectively under the responsibility of the MSSS and the RAMQ [2]. Thus, the two budgets are separate. We can therefore have two main actors : physicians and hospital staff. By implementing a financial incentive that affects one or the other of these two actors, it would be appropriate to ensure that the interests for the change are not divergent. If this is the case, this may compromise the lever effect and reduce the level of achievement of the expected objective.

1. Centre local de soins courants.
2. Régie de l'assurance maladie du Québec.

To analyze this question, we consider in our thesis two Quebec study cases with financial incentive experience. The first chapter deals with the impact of financial incentives on waiting times and average length of stay in the surgery sector. We take the example of hip and knee replacements to analyze the activity based funding impact (access to surgery program) on access to services and healthcare quality.

In the second chapter, we discuss the relationship between financial incentives and the healthcare quality. We assessed for the effect of the implementation of clinical guidelines, coupled with financial incentives, on healthcare quality indicators and populational health.

In the last chapter we perform a cost-effectiveness analysis of the latter gourvenmental program. The objective is to determine the relevance of the implementation of such a program by comparing costs and expected benefits. .

Finally, we hope with this study contribute to the effort to install a new culture of analysis and evaluation of public policies in the Quebec healthcare system. Indeed, this change is desired to provide the necessary information and useful evidence to support any moving behaviour strategies.

Chapitre 1

Impact of financial incentives on access to services and healthcare quality : The surgery access program

1.1 Introduction

The health networks in Quebec and Canada have for some time given rise to various discussions relating to performance and productivity of health systems. This issue has been driven by an economic environment characterized by budget problems. Increased funding is not the miracle solution in front of high debt ratio in the province, pressures are increasingly rising to make spending cuts, especially in the health sector. These cuts require a preliminary calculation of system performance and, possibly, the introduction of measures to improve productivity of Quebec institutions.

The current design of the Quebec health system in terms of financial efficiency, access and quality of care can be improved. Besides, the problem of access has always been one of the points to be improved in this network. Unfortunately, the desire to provide a better universal access in a public system is confronted with a desire of the government to control spending in the health system. As a consequence, supply of services does not fully meet the demand which brings an imbalance and therefore waiting lists, especially for examinations and targeted surgeries.

Waiting lists are seen as a mechanism for the rationing of care, which replaces rationing based on prices to ensure balance between supply and demand. Otherwise, the supply would be tempted to respond to increased growth in demand. Thus, it would be better to see the problem as resources optimization instead of supply control to get better health results from the money being invested.

It is important to note that waiting lists and waiting times are two separate concepts. While a waiting list is a direct result of excess demand, a waiting period often reflects a problem of organization of services. However, a link definitely exists between the two.

The waiting time for some types of surgeries is a variable often used to assess the level of access to health services. But waiting time should be compared to waiting list for better healthcare access evaluation. Currently in Quebec [1], several types of surgeries are experiencing waiting times that exceed the medically reasonable time. This has consequences not only on the quality of services provided to the population, but also leads to unnecessary costs for the health system and the economy in general.

To this end, several studies have addressed the sources of variation in waiting times and its resulting effects. Indeed, the optimal waiting times are never zero. Short lists of patients waiting for elective surgery can be cost effective to the extent that short waiting times have no serious health consequences and to the extent that queues enable optimization hospital capacities. The solution to this optimization problem requires action on demand or supply. The latter seems to be valuable to the population as it ensures more services out of public funds.

Financial incentives to institutions form an alternative to act on offer in this market. This alternative has been explored in several experiments and gave encouraging results concerning the evolution of wait times.

It is in this context that our study falls. Indeed, we seek to measure the impact of financial incentive policy for health institutions on their waiting times for surgeries. This will validate the theoretical expectations of these models and suggest necessary improvements for the proper functioning of these procedures.

This policy is known as the surgery access program (SAP) [2] introduced in Quebec in recent years, as an activity based funding experience. The purpose of this program is to improve access of the population to a large set of surgeries encouraging hospitals to produce more surgeries and to have shorter waiting time and better organization of services.

In the next section, we present a review of the literature on the importance of costs of long waiting times, on variables which act on the waiting time, on solutions which have been adopted to reduce them and on some approaches used to evaluate these programs. Then, we introduce the surgery access

1. SIMASS(Système d'information sur les mécanismes d'accès aux services spécialisés) data on waiting time, 2011-2012, Quebec Health Ministry.
2. In french : Programme d'accès à la chirurgie.

program and explain its theoretical foundations. Section 4 will focus on the data used in our study. Section 5 explains the empirical model used to evaluate the reform. Finally, we present the results of the empirical model and we conclude on the results obtained.

1.2 Literature Review

In recent years, several studies have highlighted the exorbitant economic costs of long waiting times, especially for surgeries such as hip and knee replacements as well as cataract surgeries [3]. Globerman [78] and Esmail [63] attempted to estimate the economic cost of waiting times through the opportunity cost caused by these patients (waiting) while their period of inactivity and the reduction of their contribution to the labour market. Cullis and Jones [26] and Propper [8] have estimated the opportunity cost of the price paid by the private sector to be treated more quickly.

Others, like Lindsay and Feigenbaum [10], Martin and Smith [79], Martin and al. [58], considered the issue in a partial equilibrium perspective of the market and have studied the effect of waiting times on the demand and supply of services. They found the existence of a link between waiting times and supply of services : a more intensive use of resources and thus a higher supply level will lead to reductions in waiting times. However, the elasticity of demand with respect to waiting time is quite low.

Following these unanimous conclusions on the costs of this type of rationing, several researchers, policy makers and stakeholders of the network have questioned themselves on the best strategy to reduce these waiting times and limit these costs. It is clear that action on the supply of services is much more socially acceptable than action on the demand side, for example by imposing a ceiling.

Financial incentives are among the means that tend to proliferate around the world [4]. This implies the introduction of programs and government policies aiming to increase hospitals' yield and reduce unit costs, which improves productivity. The actual consequences of this type of intervention remain unknown despite lots of initiatives in several countries. Indeed, very few studies have addressed the problem of evaluating the effect of financial incentives on the evolution of waiting times in hospitals.

In Canada, for example, several funding policies were seen as beneficial in terms of increasing volume. For example, in Alberta, an investment of 18 million $ in a project of hip and knee replacement

3. The economic cost of wait times in Canada (2008) : The Centre for Spatial Economics.

4. Norway, United Kingdom, Australia, France and United States are examples of countries which have introduced some financial incentives over the years through reward programs for better performing hospitals as well as policies of activity based funding.

brought a real reduction in waiting times [5]. Elsewhere in the world, a similar experience was started in Sweden with "OrthoChoice" program for hip and knee replacement surgeries. After a first year of application, the program is deemed to have improved the quality of service, reduced waiting times and increased production [6].

All these conclusions mentioned above are general and lack solid empirical support. Besides, the study of the evolution of waiting times due to such policies can be based on several methodological aspects. Some like Lofvendahl and al. [53] and Coyte and al. [11] have taken the path of qualitative research through questionnaires sent to patients in order to have more information about waiting times.

A range of studies conclude that waiting times are shorter after the introduction of financial incentives. However, given the bias brought by qualitative studies, applying a quantitative measure in this kind of topic can bring a valuable contribution to the discussion. Thus, other researchers have tried to calculate an estimator of waiting times. Propper and al. [69] uses difference in difference methodology to show that the funding strategy or performance objectives that have been set up in England led to a reduction in waiting times without harm to other aspects of health services, such as quality of care. It concludes that this kind of mechanism is likely to work better in systems that have progression margins in terms of their productivity and performance. For their part, Dimakou and al. [13] have tried to estimate the impact of government policy implementation in order to reduce waiting times. They estimated the instantaneous probability of being admitted (duration model) for each patient who is on the waiting list at time t. The study concludes that these rates vary randomly each time the preset target deadlines change.

Mervin and Jackson [7], Hurst and Siciliani [39] analyzed the factors that influence waiting times. Using multiple regression models, they showed that waiting times are inversely related to the number of beds and medical staff (number of specialists, number of nurses). Iacone and al. [34] showed that a greater proportion of elderly causes longer waiting times.

On the other hand, the link between financial incentives for hospitals and the quality of care has been the subject of many studies. Several econometric techniques have been tried such as instrumental variables estimator (Heckman and al. [31], Earle [15], Hadley and al. [27], Town and al. [74]), difference in difference (Farrar and al. [21], Rocha and al. [71], Schreyogg [81]) and regression discontinuity design. Often outcome indicators (readmission rates, hospital mortality rates, length of stay, etc) are used to measure healthcare quality. The main results of these studies are that, in the majority of the financial

5. One-stop shops for assessment and treatment : Alberta Hip and Knee Replacement Project gets results, Susan Usher and Cy Frank

6. Swedish Waiting Times for Health Care in an International Perspective, The Swedish Association of Local Authorities and Regions

incentives experiments, there is no decrease in the quality care level. Sometimes, the policy effect is minimal or not significant.

In our analysis, we use a survival model, like Dimakou [13], by introducing a control group in order to make the model results more robust. We assess the effects of introduction of financial incentives on both waiting time and length of stay of patients treated in Quebec hospitals(average treatment effects on the treated). In addition, we seek to estimate if the hospital manager behaviour, with respect to the funding amount and surgeries tariffs level, has effect on outcome variables.

1.3 Surgery Access Program (SAP)

The Surgical Access Program is a first experiment of an activity-based funding in the Quebec health-care system. Although it is much less aggressive than other international examples of financial policy, the PAC is a first shift in Quebec in improving the performance of the health system. This program was implemented in Quebec on April 1st, 2004. It defines a new funding policy for institutions that have an additional volume of surgeries compared to a starting threshold.

The selected threshold is the level of production in 2002-2003. The idea is then to offer to any hospital that succeeds in producing a volume of surgeries greater than that of 2002-2003, additional funding to its historical funding. This additional funding is the product of the additional production, according to data from Med-Echo [7] , and a fixed tariff for each type of surgery. This additional funding encourages hospitals to increase their surgical production and reduce their waiting list. Several types of surgery are covered by this program. However, in this test, and for methodological reasons, we will choose only targeted surgeries : hip and knee replacement(tariffs for these surgeries are 10 800$ for knee replacement and 11 000$ for hip replacement). Indeed, these two types of surgery appear to be similar across various databases surgeries in others jurisdictions. In addition, they are medical acts which have not been subjected to major technological advances in recent years. This makes them comparable through time.

The SAP program has been able to bring a marked increase in the volume of surgeries, particularly targeted surgeries. However, few rigorous analyses have sought to evaluate the program and the level of achievement of objectives especially at the level of access to services and healthcare quality. One reason of this is the absence of reliable data on waiting time before the development of SIMASS database in 2008.

Before the introduction of the SAP program in 2004, the historical funding of institutions did not

7. Med-Echo : administrative database for inpatients and day surgeries cases.

encourage the increase of production since this funding mechanism is based on historical expenditure and not on the evolution of production volumes. Following the introduction of the activity based funding (SAP), additional funding received by the hospital encourages it to increase its surgery production. This will have a direct effect on the waiting lists and the net effect will depend on the rate of growth of demand.

Theoretically, this stimulus on the supply (new equipment, new medical staff, new operating rooms) will bring the institution to reflect on the improvement of its performance. The institution will put in place strategies for assessing and improving its productivity : better arrangement of its waiting lists, better organization and planning of services, introduction of mechanisms and standard of quality, etc. As well, not only the waiting lists will shrink but also the waiting times will be optimized. Also, unnecessary days of hospitalization will be avoided and lengths of stay will be shortened. However, the challenge to keep a minimum level of healthcare quality should remain at the heart of this process. Arbitration should be traced between these three performance dimensions : financial optimization, accessibility to services and the quality of care offered.

The next section presents the data used in this analysis in order to validate this theoretical expectation of the introduction of the PAC.

1.4 Data description

In order to apply our difference in difference approach, we use data over eight years, three years before the reform and five years after. In this study we use data from two provinces : Quebec as treatment group and British Columbia as a control group. The administrative files contain detailed information on individual characteristics : patient age and gender, diagnostic code, treatment code, surgery date, visit date, Resources Intensity Weights(RIW), discharge destination, length of stay, Diagnostic-Related-Group (DRG) of the case and other informations on institutions : hospital code, number of beds, physician staff, locality and distance from home. The Quebec data come from three sources :

— RAMQ data on medical fees. These data contain medical acts of specialists in Quebec hospitals from 1 April 2001 to 31 March 2009, for hip and knee replacements.

— Outpatient RAMQ data. These data contain external visits to specialist physicians in Quebec from the 1 April 1999 to 31 March 2009.

— Med-Echo data on hospitalizations. These data cover the period from 1 April 2001 to 31 March 2009, for the same kind of surgeries.

Quebec did not have an official database for waiting times before 2008 [8]. To calculate waiting times and lengths of stay for the period covered in this study, we reconstructed the episode of care for each patient operated for a hip or knee replacement : from the date of the first visit until the date of the hospital discharge after surgery. For this, we merged the RAMQ data on surgery (hip and knee) for the year in question with specialized outpatient visits for three years : the surgery year and two previous years. The reason to dig into data of the previous two years is to make sure we obtain the maximum number of visits likely to be candidates for the visit prescribing the operation. However, and according to predetermined criteria, there should not be any surgeries that are more than two years from the date of the visit to the orthopedist [9] .

The merge of Med-Echo and RAMQ database was not perfect because there is no common patient ID. However, the choice of knee and hip surgeries in our analysis made this merger easier than other surgery cases [10].

The control group data are those of British Columbia for the same type of surgeries. This choice is motivated by some reasons : first the two health systems are comparable in terms of healthcare organization, funding models and system governance. Also, population in the two provinces are similar. Finally, as we will see later, treated patients in Quebec and British Columbia for hip and knee replacement have similar individual characteristics like age, clinical weights and gender. These data come from the University of British Columbia (BC Popdata) and are derived primarily from two sources :

— Hospital Separations File data on hospitalizations for hip and knee surgeries for the same period as Quebec.

— Medical Services Plan data on physicians' fees : these data include all visits to specialists for the same period of this study.

Since BC has no official waiting time database prior to 2008, we applied the same process as Quebec to merge the data and reconstruct the episode of care of the surgery since specialist consultation. This merge is made with an anonymous merger key. The control group of British Columbia was chosen for several reasons :

— Quebec and British Columbia compare well in terms of their health system, the evolution of medical practice and the technological advancement in the field of health.

— British Columbia has sufficient volumes of this type of surgeries to be compared to Quebec.

8. Since this date, SIMASS is the official database used for surgeries waiting time in Quebec.
9. See appendix A for more details about data treatment.
10. We used others variables to merge these databases like institution code, age, gender, diagnosis.

The choice of the interval of eight years is dictated by empirical considerations. The period should be long enough to provide robust estimates. The preparation of the data resulted in 71 258 observations in all and includes 125 hospitals (88 in Quebec and 37 in BC). For knee surgeries, the sample include 21 025 patients with a mean age of 69.9 years for Quebec (the treatment groupe) against 19 892 patients with a mean age 69.8 years in BC(the control group). For hip surgeries, we have 14 918 patients with a mean age of 66 years for Quebec and 15 423 patients with mean age of 65.9 years for BC. For the five years after the introduction of the PAC, the average funding received by Quebec hospitals was 1.1 M$ for knee surgeries and 1.25 M$ for hip surgeries. Two groups are similar with respect to age and gender. On the other hand, waiting times for knee replacements are shorter in BC and length of stay are longer in Quebec for both hip and knee replacements. The descriptive analysis for the two groups data (Quebec and British Columbia) is given by tables 1.1 and 1.2 .

TABLE 1.1 – Descriptive analysis - Quebec

Variable	**n**	**Mean**	**Min.**	**Max.**	**Standard deviation**	**Q1**	**Mediane**	**Q3**
Knee*(censored 0.7%)*	156							
Waiting time	21 025	173.4	28	730	138.8	74	129	223
Length of stay	21 025	8.63	1	124	5.09	6	7	10
Age	21 025	69.9	15	95	9.3	62	69	75
Gender	21 025	1.61	1	2	0.49	1	2	2
Funding	21 025	1.10	0	4.63	1.21	0	0.75	1.71
Hip*(censored 0.5%)*	81							
Waiting time	14 918	141.2	28	730	117.8	68	118	203
Length of stay	14 918	7.99	1	270	5.90	5	7	9
Age	14 918	66	13	94	12.1	58	67	75
Gender	14 918	1.51	1	2	0.49	1	2	2
Funding	14 918	1.25	0	4.63	1.34	0	0.85	2.19

Waiting times calculated for each patient confirm our theoretical expectations. Figures 1.1 and 1.2 shows the evolution of the average waiting time for each type of surgery.

The waiting time for being operated for hip or knee replacement, saw a sharp reduction from the year following the implementation of access surgery program, 2004-2005. Changes during the first year, is less clear. This is due, possibly, to an adaptation period for institutions to the new demand characterized by an increase in the capacity of waiting lists.

In figures 1.3 and 1.4, the evolution of lengths of stay shows a sustained decline since 2001-2002. In fact, several hospitals have introduced better modes of organization for this type of orthopaedic surgery. This has helped to shorten the lengths of stay for this type of surgery and to discharge patients earlier. It will be interesting to test, in the empirical framework, if financial incentives contributed to

TABLE 1.2 – Descriptive analysis - British Columbia

Variable	n	Mean	Min.	Max.	Standard deviation	Q1	Mediane	Q3
Knee*(censored 1.3%)*	277							
Waiting time	19 892	154.6	28	730	128.95	60	112	206
Length of stay	19 892	5.2	1	357	6.46	3	4	6
Age	19 892	69.8	20	95	9.70	63	71	77
Gender	19 892	1.58	1	2	0.49	1	2	2
Funding	19 892	0	0	0	0	0	0	0
Hip*(censored 1.0%)*	155							
Waiting time	15 423	141.4	28	730	118.00	57	102	184
Length of stay	15 423	5.9	1	155	6.54	3	5	6
Age	15 423	65.9	13	94	12.02	60	70	77
Gender	15 423	1.55	1	2	0.49	1	2	2
Funding	15 423	0	0	0	0	0	0	0

Source : author's computations using Quebec and BC data

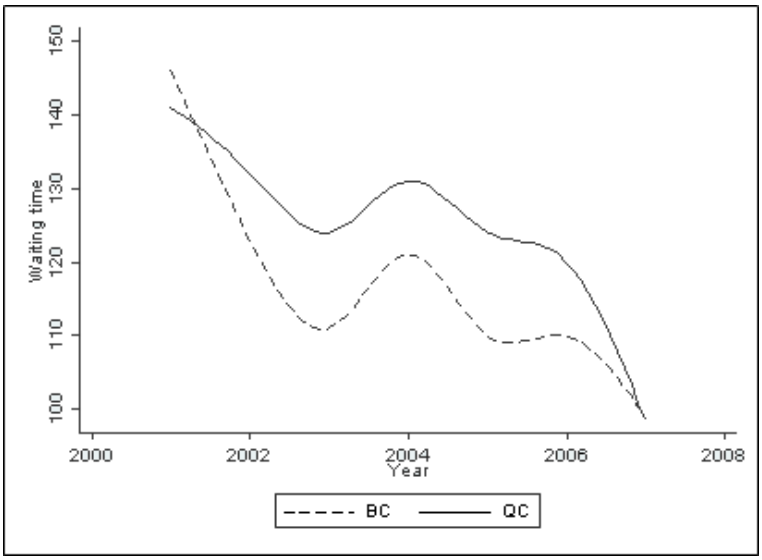


FIGURE 1.1 – Waiting times (days) in Qc and BC (Knee replacement)

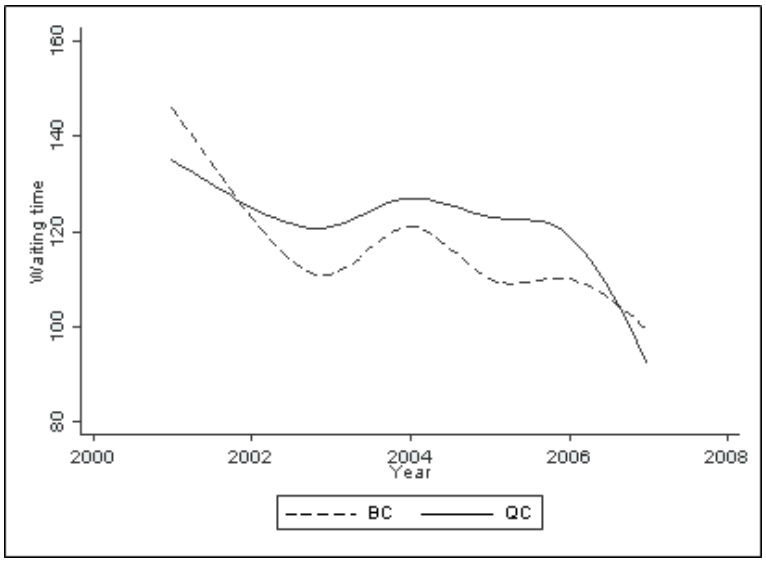


FIGURE 1.2 – Waiting times (days) in Qc and BC (Hip replacement)

this increase and that's cutting the length of stay is a part of the strategy of doing more surgeries

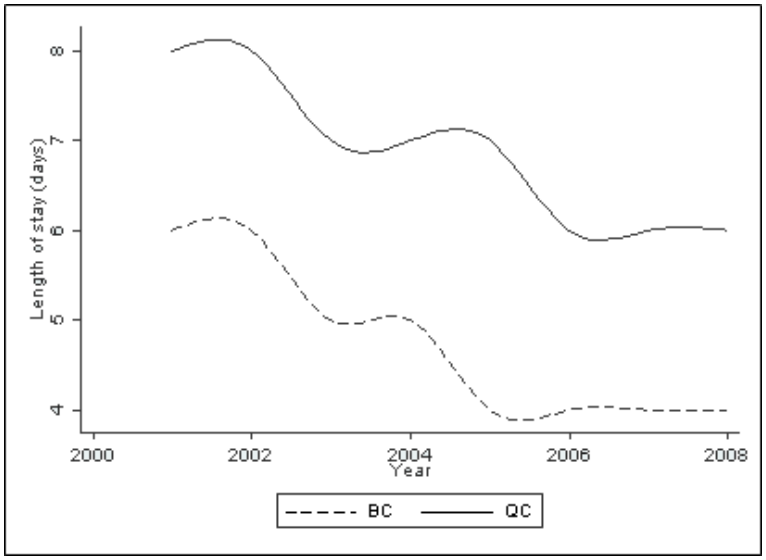


FIGURE 1.3 – Lengh of stay in Qc and BC (Knee replacement)

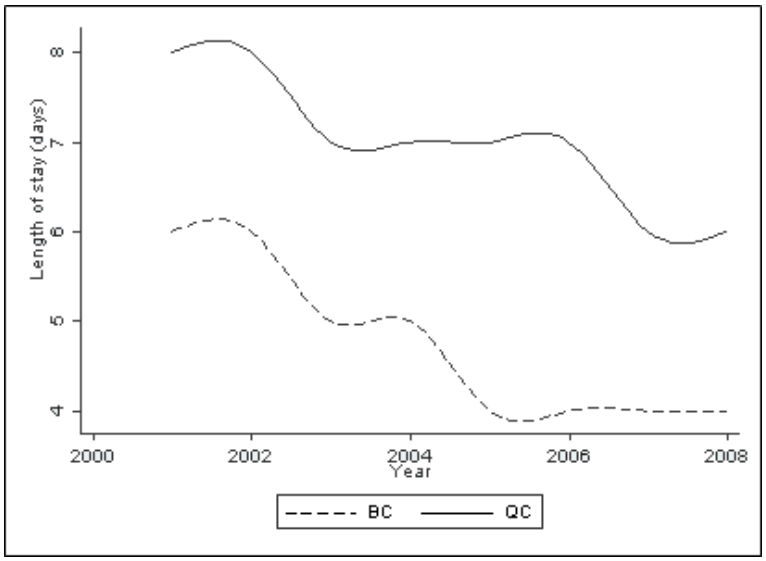


FIGURE 1.4 – Lengh of stay in Qc and BC (Hip replacement)

Estimation of the survival function (Kaplan-Meier analysis)

We estimate the survival function of waiting time model with the Kaplan-Meier method. The advantage of this method is that it takes into account the censored data. We plot pre-2004 and post-2004 curves for hip and knee replacement (together) in Quebec data and we compare this survival curve before and after the reform.

Figure 1.5 clearly show that the exit rates of the treatment group declined significantly after the 2004 reform. The median waiting time, decreased from 125 to 115 days. Log-rank and Wilcoxon statistics test the null hypothesis that the survivor functions are the same. Both are chi-square statistics with a single degree of freedom. They differ only insofar as the Wilcoxon statistics weights early times more heavily than later times.

In the next section, we introduce the empirical model analyzed in order to support these descriptive statistics.

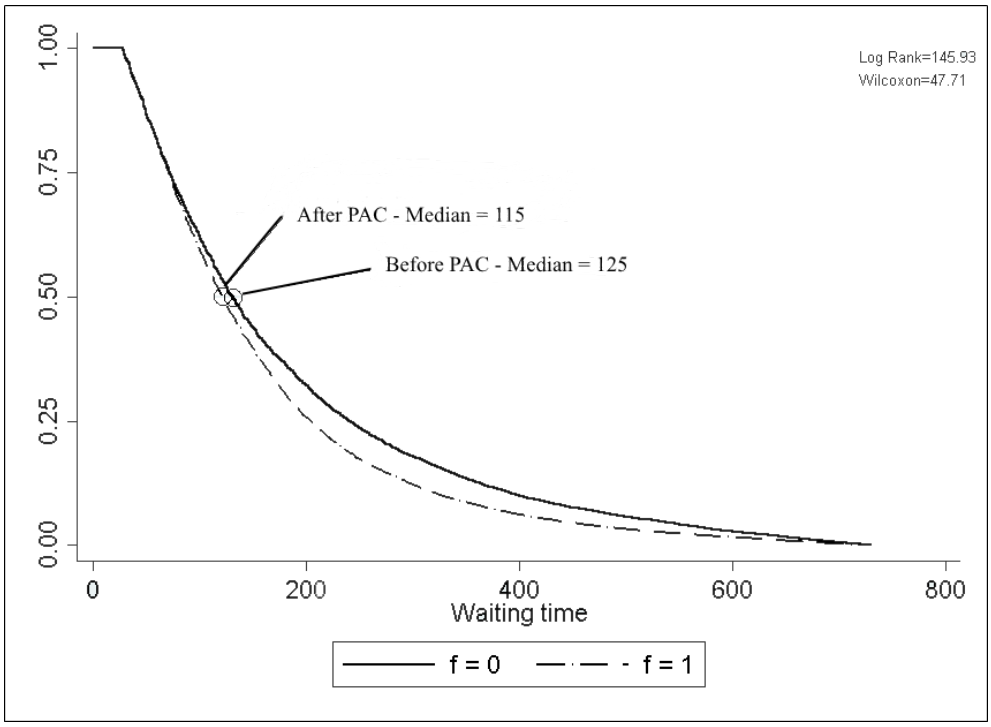


FIGURE 1.5 – Kaplan-Meier daily survival rates, waiting time

1.5 Empirical Framework

This empirical analysis focuses on the impact of the introduction of the PAC on waiting times and lengths of stay. Our data sample consists of two groups : a treatment group (Quebec) and a control group (British Columbia). The episode of care for hip or knee surgery(figure 1.6) is defined according to three important dates : the start date of the waiting for the surgery(or the surgery decision date), the surgery date and finally the discharge date.

We use a Mixed Proportional Hazard (MPH) model in which in each hospital j, each patient i is faced with three possible states s :

— At time t_1, either remain on the waiting list (s = 1) or be operated (s = 2) [11]. Thus the waiting time, equal to the duration between the start date and the end date of the state 1, is a random variable.

— At time t_2, either remain hospitalized (s = 2) or leave the hospital (s = 3). Thus, the length of stay, equal to the duration between the date of the surgery and the date of discharge from the hospital, is also a random variable. We will see later, in some model specification, that the third state should be two seperate states : death or other.

Let T be this positive random variable indicating the time that separates the date t=0 (date of entry to the state s= 1) from the date at which occurred the transition to the state s=2. The hazard function $\gamma^s(t)$ is a key function that measures, for a patient at time t, the instantaneous probability to be operated/to leave the hospital given that the patient is still on the waiting list/hospitalized until this date t. The hazard function of the MPH model in its simple form, as defined by Han and Hausman[28] and Meyer[3] , is given by

$$\gamma^s(t/z^s) = \exp(z^s\beta)\gamma_0^s(t) \qquad s = 1,2,3 \tag{1.1}$$

The covariates z^s, appearing in the first term of the right hand side of equation 3.1, are of two types : the variables which vary according to hospital j and others which vary according to the patient i. The β vector is the set of covariate parameters that we will have to estimate. $\gamma_0^s(t)$ is the baseline hazard function which is common to all patients. It corresponds to the instantaneous probability of transition in a model without unobserved or observed heterogeneity, and therefore must be positive. This function is modeled as a constant hazard by interval, thus a "piecewise" function, which depends only on time. We assume that the baseline hazard is not influenced by the reform nor by hospital characteristics.

This kind of model presents two advantages. First, we do not need to determine a functional form to the baseline hazard function. Indeed, we are more interested in the hazard ratio which is independent

11. We make the assuption that surgery date is the first day of hospitalization.

of time (that is the basic assumption for this model).

Next, the advantage of MPH model is that it can take into account the censored data. In our case, it was a censorship to the right because patients who have had a medical visit but have not yet been operated as at 31 March 2009 will be lost from view and our estimates may be biased. The MPH model allows us to include them in the estimate but as censored data. We assume that the censoring is non-informative. In our case it would be important to mention that the cause of the censorship is independent of the event of interest, either the waiting time up to the operation date (the stay in the hospital until the date of discharge).

This basic form in equation 3.1 does not take into account the unobserved heterogeneity which may create an estimation bias. To include this aspect, we added the term ϑ_{ij}^{s} in equation 2.2. The individual hazard function for each state s becomes

$$\gamma^{s}(t/z_{i}^{s}(t),\vartheta_{ij}^{s}) = \exp(z^{s}\beta)\gamma_{0}^{s}(t)\vartheta_{ij}^{s} \qquad s=1,2,3 \tag{1.2}$$

The term ϑ_{ij}^{s} reflects unobserved heterogeneity. The heterogeneity is found between hospitals and not between patients. If this term was equal to 1, it would be in the particular case of the COX model.

In the absence of unobserved heterogeneity but there is only measurement errors, this term captures the over-dispersion. The ratio of the hazard functions for two patients will be determined by the observed and unobserved heterogeneity and does not depend on time. We note that time does not appear in the term which is located to the right side of the equality sign in the following equation

$$\frac{\gamma_{1}(t/z_{1}^{s},\vartheta_{1j}^{s})}{\gamma_{2}(t/z_{2}^{s},\vartheta_{2j}^{s})} = \frac{\exp(z_{1}^{s}\beta)\vartheta_{1j}^{s}}{\exp(z_{2}^{s}\beta)\vartheta_{2j}^{s}} \tag{1.3}$$

Certainly, the choice of the distribution of this term of heterogeneity inluences the results (Heckman and Singer[38], Hougaard and al. [33], Keiding and al.[43]). However, the literature puts forth few arguments in favor of a distribution rather than another.

In our analysis, the random variable capturing the unobserved heterogeneity is assumed distributed according to a gamma density function whose mean is equal to 1 and variance to θ. This choice is based on two points. First, Abbring and Van den Berg[37] have shown that the unobserved heterogeneity distribution converges to a gamma distribution. In addition, the Monte Carlo simulations of Baker and Melino[54] have shown that the use of the non-parametric approach for this random term leads to a

non-converging estimator if too many mass points.

By introducing a dimension of time, as defined by Drolet and al.[25], and knowing that we have two groups (a treatment group and a control group), the logarithm of the hazard function to leave the state 1 for a patient i at time t, knowing that he is still on the waiting list until this date t is given by

$$\delta_i^T(t/z_i) = \delta_0^T(t) + (\alpha_P + \alpha_R)Post_R + z_i(t)\alpha + \varepsilon_i^T \tag{1.4}$$

For the treatment group and

$$\delta_i^C(t/z_i) = \delta_0^C(t) + (\beta_P + \beta_R)Post_R + z_i(t)\beta + \varepsilon_i^C \tag{1.5}$$

For the control group.

$Post_R$ is a dummy variable equal to zero for the period before the introduction of the PAC and to 1 for the period following the establishment of the PAC. The parameters α_R and β_R measure the impact (in percentage) of the PAC on the hazard rate in the treatment group and the control group respectively.

The parameters α_P and β_P control an additional effect of the specific time which can change the hazard ratio after the introduction of the PAC on a permanent basis. $\delta_0^T(t)$ and $\delta_0^C(t)$ are the baseline hazard function (log) for the treatment group and the control group respectively. It represents the rate of exit common to all individuals of the same group at a time t before the reform (when $Post_R = 0$) and for covariates null ($z_i(t) = 0$).

The error terms, random ε_i^T and ε_i^C, are i.i.d with mean equal to zero and reflect the unobserved heterogeneity of permanent patients. The parameters α and β are the parameters to estimate. Only $(\alpha_P + \alpha_R)$ and $(\beta_P + \beta_R)$ are separately identified and without imposing additional restrictions, α_R and β_R can not be estimated.

The two estimators we use in the empirical analysis are the special cases of this general framework. They have been obtained by imposing some identification restrictions. First, the BA estimator is obtained by assuming that :

— there is no specific effect at times which have a permanent impact on the exit rate of the treatment group after the reform ($\alpha_P = 0$),

— the control group is not affected (directly or indirectly) by the reform ($\beta_R = 0$)

— and the covariates are orthogonal to the error term.

However, this estimator BA will be biased in the presence of variables not observable which will permanently affect the intercept of the hazard of the treatment group after the reform.

An alternative approach is the difference in difference (DID) estimator. It corrects a deficiency in the BA estimator in authorizing, after the reform, a change in the hazard rate which is common to both groups, while imposing that the reform will have no effect on the control group.

A simple way to compare this DID estimator the BA estimator is to combine the equations 1.4 and 1.5, while imposing $\beta_R = 0$:

$$\delta_i(t) = \delta_0^C(t) + Treat_{it}\hat{\delta}_0^T(t) + \beta_P Post_R + (\alpha_P + \alpha_R - \beta_P)Treat_{it}Post_R + z_i(t)\beta + Treat_{it}z_i(t)\hat{\alpha} + \varphi_i \tag{1.6}$$

$Treat_{it}$ a dummy variable that takes the value one when the patient is in the treatment group and zero otherwise. $\hat{\delta}_0^T(t)$ is the gap of the baseline hazard of the two groups. $\hat{\alpha}$ is the difference of the two parameters relating to the covariates in the two groups. φ_i is the error term which is equal to ε_i^T for the treatment group and ε_i^C for the control group.

In equation 1.6, we obtain the DID estimator by equating the specific effect of time in the two groups ($\alpha_P = \beta_P$). Thus, the effect of the reform is measured by the parameter α_R. This parameter can be interpreted as a change in the hazard rate of the treatment group after the reform beyond any other change which is common to both groups (by controlling, of course, for the covariates included in $z_i(t)$).

On the other side, the BA estimator is obtained by nullifying the time specific effect in the treatment group ($\alpha_P = 0$). Since the condition $\alpha_P = \beta_P$, was already in the equation (6) , the effect of the treatment will therefore be given by $\beta_P + \alpha_R - \beta_R = \alpha_R$.

We should ensure if it is right to assume that the theoretical model studied is perfectly identical for the two provinces or on the contrary if it exists specific characteristics of each province. In other words, we want to ensure if health sector and especially hip and knee surgeries are characterized by different seasonal cycles that may exert a significant impact on the coefficients estimates. Failure to meet this assumption can lead to biased estimates of coefficients and especially biased estimates of the standard errors.

Six annual dummies (three for Quebec and three for BC, for the three years before treatment) are

introduced to take into account unobservable factors such as technical and demographic changes that vary over time and which are assumed at first not common to Quebec and British Columbia. We performed wald test and likelihood ratio test under the null assumption that dummies of the same year are equal in the two groups. We cannot reject the null hypothesis of equal coefficients at the 5% level. So in next simualtions, we adopt common trend dummy variables for the two provinces.

Proportionality assumption test

The MPH model rests on the proportionality assumption. In our case, it implies that if Quebec had not been subjected to the SAP treatment, both groupes (QC and BC) would have experienced the same time trends conditional on covariates.

In cases where this assumption would be violated, the statistical tests and the thresholds observed are disabled and the risk reports can no more be interpreted correctly. It would therefore be important to validate this hypothesis in the case of the variable of interest in this model (treatment variable).

We note F^s the treatment variable and λ the parameter to estimate for this variable. We begin first by testing whether the effect of the funding variable varies through time. We then add an interaction term between F^s and the log of the time

$$\gamma^s(t/z_i^s(t), \vartheta_{ij}^s) = \exp(z^s\beta + F^s\lambda^1 + log(t)F^s\lambda^2)\gamma_0^s(t)\vartheta_{ij}^s, \quad s = 1,2,3 \tag{1.7}$$

In equation 1.7, the effect of the variable F^s is $[\lambda^1 + log(t)\lambda^2]$. It is understood that if λ^2 is different from zero, the effect of the treatment variable increases or fades with time. Lee and Wang [19] suggested to test the significance between the interaction between the covariate and the log time. If the null hypothesis $\lambda^2 = 0$ is not rejected then the proportionality condition is satisfied.

Then, we look at the graphs of standardized Schonefeld residuals (Schoonefeld [12]) as a function of the log of the time. In our case, the Schonefeld residuals to the date t for a patient i is the difference between the level of funding corresponding to this individual operated on the date t and the average funding levels of individuals waiting at the date t. It is standardized since it will be divided by its variance.

If the hypothesis is verified then these residuals must be distributed in the same manner as a function of time. The graph of residuals as a function of the log of time will look like a horizontal straight line in this case.

All previous analyses have helped to validate the proportionality assumption for the financing factor. Figure 1.7 and 1.8 show the graph of the Schonefeld residuals as a function of the log of the time in the form of a horizontal straight line.

Test of the null hypothesis of parameters β

To investigate the robustness of our findings, we test the coefficients of the covariates under the null hypothesis by performing three tests : the Likelihood Ratio test, the Score test also called Rao test and the Wald test. All of them reject the null hypothesis : $(H0) : \beta = 0$, so the estimates of our model are different from zero.

Also, we simulated the placebo effect (treatment isn't real) on Quebec data introduced in 2003 (dummy variable equal to 0 for 2001 and 2002 years and equal to 1 for the other years). The estimated coefficient was not statistically significant.

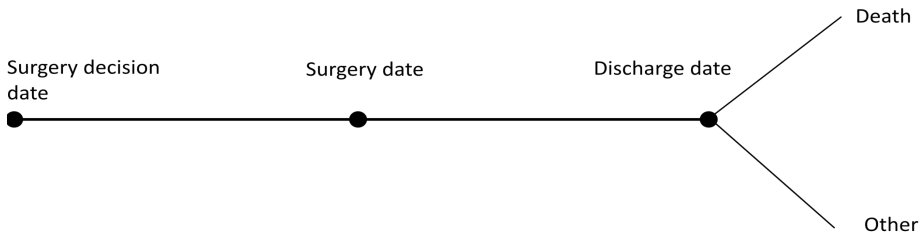


FIGURE 1.6 – Episode care for hip and knee surgeries

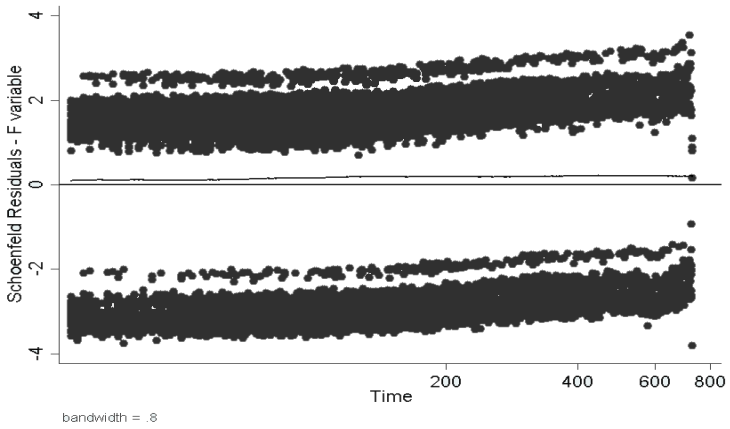


FIGURE 1.7 – Plot of Schonefeld residuals and the log of time (logt) Treatment group-COX model

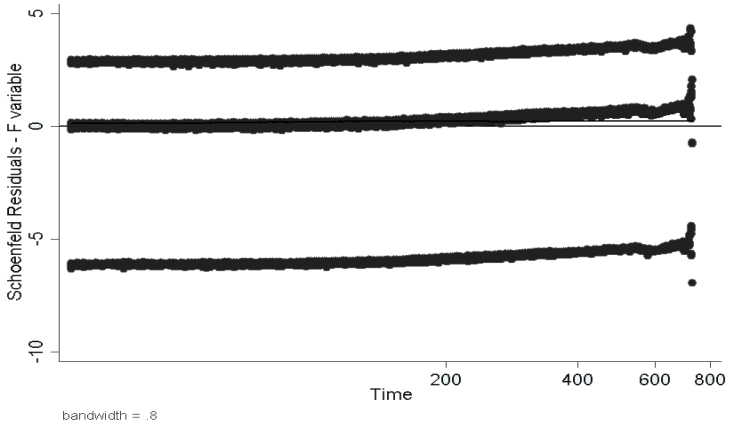


FIGURE 1.8 – Plot of Schonefeld residuals and the log of time (logt) Two groups - COX model

1.6 Results and discussion

Our specification strategy proceeds in four steps. In the first specification (model 1), we calculate the BA estimator in the context of Cox model (only with Quebec data). Then in model 2, we add a term of unobserved heterogeneity. In the third specification (model 3), we resume to the Cox model (proportional hazard with no unoservable heterogeneity) but we introduce the control group in order to calculate the DID estimator. Also, we add a common trend dummies for Quebec and BC [12]. Finally, in model 4, we add the term of unobserved heterogeneity for MPH model with a control group.

In the first specification, we tested two treatment variables : first, we use the amount of funding per year and per institution for knee and hip surgeries (equal to zero before 2004) as our treatment variable to estimate the impact of the reform on waiting times. Second, we modelize the treatment as a binary variable (0 or 1). We assume that this discrete variable is exogenous and we assess for the causal reform effect on treated hospitals. In the first case, we want to test if there is tariff motivation for hospitals to produce more hip and knee surgeries or if that production increases is only due to the willingness of reducing the waiting list. More precisely, we want to test if the activity based funding level has an impact on the outcome variable. However, we are likely to encounter endogeneity issues.

In order to assess the endogeneity problem, we instrument the continuous treatment variable, using the proportion of the elders aged 65 or above (in each locality) and the hospital status (dummy variable) in three different region levels : mega, medium and remote regions. This analysis was conducted by hospital observation, not by patient. Hausman test showed that funding variable is endogenous and that the tratement effect should be biaised if funding level is treated as exogenous variable. So, if taken as continuous, the treatment variable is indeed endogenous and the measured effect will be biased.

We then tested the two basic requirements of instrumental variable model : the instrument is correlated with the treatment variable but not with waiting time variable. First we obtain that elderly proportion is correlated with the funding regressor but not the hospital status dummy. So we keep just elderly proportion as instrument. Also F-statistic [13] for the first stage model is greater than 10. So we conclude that elderly proportion is sufficiently strong instrument for the funding variable and the model is exactly identified.

Table B.1 shows ordinary least squares (OLS) and two stage least squares (2SLS) results for treated hospitals. In colomns 1 and 3, treatment estimates are significantly negative for knee but not for hip. In

12. Common trend is defined by dummy variable for each year. We tested for this common trend in the two provinces just before treatment year. We reject the hypothesis that coefficients are different between the two groups.

13. F-statistic test for the model significance. It is function of R^2 : a bigger R^2 lead to high values of F. Litterature suggested instrument to be weak is F-statistic is less than 10.

colomns 2 and 4, treatment estimates are significantly negative on log of waiting time : -0.61 for knee and -0.70 for hip. In other words, each additionnal funding of 0.1 M$ offered to hospital decreased its average waiting time by 12.2 days for knee replacements and by 12.9 days for hip replacements. Treatment estimate of 2SLS model is greater than the estimate of OLS model. Funding amount has no fixed effect on the access to services and each additional funding has a negative marginal effect on the waiting time. So, this parameter can not be ignored and some hospitals decide to increase their surgeries production because tariffs are generous. Indeed, the positive difference between tariff and cost encourage these hospitals to manage their surgical activity in the aim to do more hip and knee replacements.

To estimate the reform effect on length of stay, we use the same four specifications as waiting time analysis. Also, we test for a competing risks model (model 5) with two possible exit state : alive (interest failure) or death(competing failure). So we compare hospital death risk with other hospital exit risk (home, other institution, etc). We want to test if the use of one failure survival model is appropriate in our case or wether we should consider competing risks.

In model 1 (BA analysis), we tried the covariates age, gender, number of beds, distance between home and hospital and the proportion of 65 years older. The treatment variable (discrete variable) as well as all other covariates (except gender and physician staff for hip and distance for knee) are statistically significant. The impact of the SAP program is positive on the waiting times. An additional funding increased the hazard to be operated by an average of 0.15 for knee surgery and 0.10 for hip surgery. In other words, this reform has helped to reduce the waiting time by 19.7 days [14] (173.20 to 153.48 days) for knee replacement and 13.1 days (159.09 to 145.99 days) for hip replacement. For the other control variables, as expected, an increase in the age (1 more year) increases the probability to be operated by 0.9 per cent as well as a higher proportion (1% more) of people older than 65 years multiplied the risk of be operated by 1.585 in the case of knee replacement and 0.65 in the case of hip replacement. However the supply variables (number of beds, number of physicians) seem to have less impact on the outcome variable.

In model 2, we have incorporated the gamma unobserved heterogeneity with mean normalized to one and variance equal to θ. Interestingly, we find that the variance θ is statistically significant according to the LR test [15]. We obtain estimated coefficient θ equal to 0.049 , and given the standard error of θ and likelihood-ratio test of $H0 : \theta = 0$, we find a significant frailty effect, meaning that the correlation within institution cannot be ignored. Parameters estimates of the covariates were almost identical to those of model 1. So we conclude to the presence of unobserved heterogeneity. Reform increases

14. These durations were calculated from the survival functions estimated by the model for an average patient (with average values for model covariates) and for two periods : F=0 before treatment and F=1 after treatment.

15. LR test is a boundary test and the null hypothesis is an equal mixture of a chi-squared (degree of freedom=o) variate and chi-squared (degree of freedom = 1).

surgeries hazard by 0.07 for knee replacement and 0.05 for hip replacement. So SAP reform reduces waiting time by 8.7 days (122.4 to 113.7 days) for knee replacement and by 6.3 days (115.3 to 109 days) for hip replacement. The detailed model 1 and 2 results are included in table C.1.

In model 3, we eliminate the gamma term and we introduce the control group. The reform still has a significantly negative impact on the exit rate. The unit funding has increased the hazard rate of 0.18 (decrease of 22.2 days on waiting time) for the knee surgery and 0.12 (decrease of 13.9 days on waiting time) for hip surgery. So, the BA estimator in model 1 probably underestimates the impact of the reform : part of the decrease in waiting times observed after 2004 is due to common factors affecting both the control and the treatment groups. We remark that since 2003, the common trend estimates was growing and these parameters are statistically significant. This specification suggests that we cannot ignore common trend. These parameters reflect some inobservable common variation between QC and BC. Age and gender variables are statistically significants too.

In model 4, we introduce an unobserved heterogeneity and calculate the DID estimator. As for model 2, the results show the presence of unobserved heterogeneity : the variance θ is statistically significant and equal to 0.05 and 0.041 for knee and hip replacement, respectively. The SAP increases the hazard by 0.12 for knee replacement and by 0.08 for hip replacement. In other words, the treatment decreases waiting time by 13.6 days (114.2 to 100.6 days) for knee surgeries and by 8.6 (105.8 to 97.2 days)for hip surgeries. Age and gender variables are statistically significants but not province dummy. Trend dummies parameters are positive and increasing with time. This model offer a better treatment estimate after controlling for observable and unobservable variables and also with respect to common trends in the two provinces. Table D.1 shows detailed results for models 3 and 4.

On the other side, when we use length of stay as outcome variable in the first specification(model 1), all estimates are statistically significant at the level of 1%. The treatment has increased the hazard rate to leave the hospital by 0.35 following a knee surgery and 0.40 following hip surgery. This is reflected by a decrease in lengths of stay of 1.05 days (8.62 to 7.57 days) for knee surgery and 1.15 days (8.88 to 7.73 days) for hip surgery. As expected, the NIRRU [16] variable seems to be the most important explanatory variable for the length of stay. Indeed, this variable has a negative effect on the discharge rate (estimates of -1.25 for knee and -1.10 for hip). When this variable is omitted, we obtain quitely the same reform parameter estimate. The age and the gender covariates have decreased the hazard rate of discharge (increase in the length of stay).

By introducing unobserved heterogeneity in model 2, the results are similar to those for the waiting

16. NIRRU(Niveau d'intensité relative des ressources utilisées) is an indicator of the intensity resources used in healthcare. In BC, this indicator is called : Resources Intensity Weight (RIW).

times : the frailty parameter is statistically significant at the level of 1% and we obtain θ equal to 0.04 and 0.03, respectively for knee and hip replacements. This unobserved heterogeneity term is irrelevant and the treatment estimates falls to 0.09 and 0.13, respectively for knee and hip surgeries. All estimates in model 2 are smaller than those of the first specification. Results for models 1 and 2 are in table E.1.

In model 3, the addition of control group shows that treatment has increased the hazard ratio to leave the hospital by 0.28 for knee replacement (length of stay shortened by 0.93 days) and 0.32 for hip replacement (lengths of stay shortened by 1.11 days). Common trend estimates has significant positive impact on length of stay. This impact is increasing over years. The province estimate is interestingly statistically significant at the level of 1% : a patient in Quebec group is more likely to exit more rapidly relatively to BC patient.

When we performed model 4, the reform impact is no longer important as in the model 3 and the reform estimate is now equal to 0.09 (-0.5 days) for knee replacements and 0.12 (-0.7 days) for hip replacements. Also, the unobserved heterogeneity estimate is statistically significant. This result may be explained by differences within hospitals in their organization care model. All the other estimates are statistically significant except province variable. The results of DID estimators (models 3 and 4) are detailed in Table F.1.

Table G.1 provides the parameter estimates for model 5 (competing risks model). In this specification, treatment coefficient is positive and statistically significant at the level of 1% : SAP program increases the hazard for patients returning to home by 0.28 for knee replacement and 0.32 for hip replacement. These coefficients are similar to those in the cox model(model 3). We conclude that competing risks model is equivalent to the single risk model. This is due to the few number of competing risk cases (hospital death). Also, we tested probit model to estimate treatment effect on the probability of hospital mortality and we obtained statistically unsignificant estimate. So we conclude that SAP treatment has no effect on the mortality rate at hospital after knee and hip replacements. The improvement in length of stay did not cause deterioration in post surgery mortality rate. Consequently, the SAP program has not led to lower levels of healthcare quality for hip and knee surgeries, as measured by these two indicators.

These results show that the introduction of the SAP has contributed to reduce waiting times for knee and hip surgeries as well as length of stay for hospitalized patients. After controlling for common trend, observable and unobservable covariates, and comparatively to BC data, the average treatment effect is positive on the hazard rate. These results show the positive impact of additional funding in the health network. This financial incentive encourages institutions to better manage their waiting lists

for surgeries. A better management of waiting lists as well as a better organization of services (surgical schedule, preparing the patient for the episode care, optimization of rehabilitation services, etc.) brings better results in terms of waiting times. Length of stay will be subsequently shorter and closer to the average reasonable durations.

1.7 Conclusion

This paper aims at analysing the effect of SAP program, introduced in Quebec since 2004 as an activity based funding. Using patient-level data for Quebec hospitals as treatment group and BC data as control group, this paper assess the impact of the reform upon waiting times and length of stay for hip and knee replacements. Based on the estimates of three-states transition model with a difference in difference approach, we estimate the hazard to exit from a state to another. First, SAP program has contributed significantly to reducing waiting times for these surgeries. It seems that a permanent positive financial incentives offered to hospitals (as is the case with SAP) brings positive results more sustainable on waiting times. However, this strategy seems to work less on the long term when the reduced time is accompanied by an increase in demand [17]. Given the slight difference between the waiting time and waiting list, it should be mentioned that the solution to reduce waiting times must take into consideration that this reduction should not be accompanied by a deterioration of the surgical access level. This analysis did not examine whether the reduction of waiting times is the result of a reduction in waiting lists by moving resources (therefore an increase in waiting lists in other places) and therefore put constraints to admission to a waiting list. However, when we modelized treatment as endogenous variable (funding amount), we show that each additionel funding encourages hospital to increase the hip and knee surgeries production because of their generous tariffs. This increase may be associated to a decrease in others orthopedic surgeries production.

Second, SAP program has helped to reduce length of stay for hospitalizations following knee and hip replacement. The increase in surgical production encouraged hospitals to discharge patients more quickly, as long as this medical act does not decline the patient's health by more than a critical level. In parallel, there is no evidence that length of stay improvement was associated with healthcare quality deterioration. The length of stay reflects a dimension of the healthcare quality but can't be considered as an outcome indicator. However, the absence of negative impact on hospital mortality should reassure us about the healthcare quality evolution.

Improving waiting times and shortening lengths of stay while treating a larger volume, is certainly an improvement in the hospital productivity. However, we must ensure that this progress doesn't lead to a deterioration of other quality of care aspects. Other heathcare quality indicators (such as readmission

17. More experiences in the litterature reports this result. Even for SAP program, in last years, we recorded a waiting time deterioration especially for hip and knee surgeries. This paper has not analysed this issue.

rates, complication rates) can then be analyzed to conclude in relation to this matter. These indicators should be relevant for this kind of surgeries, especially in terms of cases number.

Chapitre 2

Do clinical guidelines affect healthcare quality and populational health : Quebec colorectal cancer screening program

2.1 Introduction

Cancers figure among the leading causes of mortality in Quebec. Because of aging and the increase in population, it is expected that the number of cancer deaths will rise in the coming years. In order to ensure rapid access with better healthcare quality, the Quebec government introduced several strategies to fight against many kinds of cancer. Thus a screening strategy for colorectal cancer, PQDCCR [1], has been initiated with some institutions in the Quebec health network in recent years. Essentially, this strategy is based on clinical protocol for colorectal cancer screening tests. Persons between 50 and 74 years old are offered a test by mail to detect traces of occult blood in the stool.

The objective of this policy is to decrease cancer mortality, to improve health status and to reassure people who have negative tests. Indeed, colorectal cancer symptoms can occur several years before being diagnosed. So screening aims to detect and treat colorectal cancer before the appearance of these symptoms. Recovery possibilities are in consequently better in this case.

Targeted individuals should visit their general practitioner (GP) to obtain explanation about the process. The physician determines if this test is appropriate to the patient's situation and, if so, he explains how to use it [2]. In the case of a negative test, the person is asked to repeat the test two years later. In the case that the test is positive, the GP prescribes a colonoscopy to confirm or refute suspected cancerous lesions.A colonoscopy is the most reliable way to diagnose a polyp or cancer, but is proposed usually

1. In french : Programme Québécois de dépistage du cancer colorectal.
2. Physicians can prescribe a test for their patients aged 50-74 years without an invitation letter.

to persons at high or very high risk, with the presence of disease symptoms.

Therefore, it is clear that the appropriateness of a colonoscopy exam is important in this process. Variability can be observed in the way to prescribe such exams between different healthcare professionals. Standardization of medical practice would be preferred in order to improve healthcare quality and optimize activity costs.

The objective of improving the appropriateness and safety of patient care has led to many investments in the development of clinical protocols. In recent years, there has been a growing interest in the use of these clinical protocols to reduce inappropriate care and control for geographical variations in medical practice and a more efficient use of resources. However, the spread and the introduction of these clinical standards remains a difficult task and information systems have been proposed as an effecient way to implement them.

Clinical guidelines were developed to allow standardization of medical practice which may differ from one physician to another. The principle is to describe the process of best practice and thus simplify the care trajectory. These clinical standards not only allow the opportunity to set goals and to share responsibilities but also aim to improve coordination mechanisms and communication. Performance indicators are subsequently used to provide a measure of the objectives achieved in terms of compliance to clinical standards and improving the quality of care.

Usually, financial incentives also improve these clinical models and encourage behavior that moves towards the best practice. These incentives may be applied to hospitals or physicians. Among these models, we include physician remuneration based on performance targets, pay for performance based on best practices and hospital funding based on a normative pricing.

In our study, we target pay for performance based on best practices. Every participating hospital receives additional funding compared to other facilities and based on the performance indicators. Indeed, in the case of historical funding, there is no payment for additional activity and there is no incentive to adopt the best practices. However, tariffs based on the best practices encourage health care providers to follow the proposed clinical recommendations.

The implementation of guidelines can lead to prioritize some colonoscopy cases over others. In addition, this strategy is expected to induce the physician behavioral change and to comply with these standards in order to improve services access and quality of care. Is this prioritization process beneficial for the entire population ? Is there any negative health effect on patients that have had their exams

delayed or even canceled ? Is this mechanism capable of achieving the desired objectives, particularly in the context of the Quebec health system ?

It would be interesting to seperate the clinical approach effect from the financial incentives effect : is a clinical strategy better to achieve its goals without financial incentives ? Does the financial incentive tend to accentuate the causal effect of the reform or does it go in the opposite direction ?

Therefore, it's important to evaluate the level of improvement induced by the new policy. This assessment would not only judge the extent of the expected change but also detect areas of dysfunction and offer some solutions. The objective of this research is to contribute to efforts that measure the impact of medical practice policy standardization on healthcare quality and population health.

In the next section, we present a literature review about the impact of clinical guidelines with or without financial incentives on healthcare quality and health system efficiency. Then, we introduce the Quebec colorectal cancer screening program and we explain the theoretical foundations. Section 4 covers data used in our analysis. Section 5 explains the empirical model used to estimate the reform causal effect. Finally, we present results of the empirical model with a discussion, some suggestions and a conclusion following the results that were obtained.

2.2 Literature Review

In recent years, the evaluation of the impact of clinical protocols on outcome indicators has been the subject of several studies in health economics literature. Some, like Eijkenaar and al. [17], concluded that no evidence on the subject, nor interpretation, presents a complex challenge to extract the key factors for this kind of strategy.

Other studies showed significant effects of clinical guidelines. For example, Patkar and al. [65], revealed that the number of medical errors is lower in activities with clinical protocols than those without these protocols. While Mullet and al. [46] demonstrated the positive effects of these standards on the efficiency and healthcare quality : fewer days when prescribing specific drugs than what is recommended despite an increase in the number of patients receiving the same drug. Van Wijk and al. [37] has also shown a reduction by 20 % of the administering of blood tests, after the introduction of clinical guidelines.

Based on the before-after approach, Roberts and al. [76] analyzed the impact of a clinical standard for hip surgery on healthcare quality in teaching hospitals. The authors concluded that there was an

increase in the average length of stay and a decrease in readmission and discharge rates, to a long-term facility. In another study , Kim and al. [44], the authors show a positive impact of the introduction of the clinical standard in the cost of hospitalization, complications and the health status of the patient after their knee or hip replacement.

Despite these findings, there is currently little conclusive scientific evidence on the effects of the implementation of clinical guidelines, with or without funding mechanisms, on outcome indicators. The programs are not all similar and are located within different health systems, at different times and with very different goals. Also the choice of model (the link between financial incentives, healthcare quality and the outcome indicator target) may explain this variability in findings of the various analyses (Doran [14], Rosenthal [59], [60], Sempowski [36]).

In addition, financial incentives in these models are often implemented in complementarity with other elements, such as benchmarking, professional training programs and initiatives related to patient experience. So it becomes more difficult to attribute changes in a single factor due to a selection bias problem.

Given this bias, few rigorous analyses were used to assess the causal effect of these clinical policies on the healthcare quality and population health. Hence there is a need and challenge at the same time to define a credible counterfactual for the treated group. The assessment of the causal effect of public policies in the presence of this selection bias was the subject of growing literature in recent years, especially in health economics.

Randomization seems to be the preferred methodology for estimating the average treatment effect on the treated especially in health economy and epidemiology studies (Mullahy and al. [62], Jackson and al. [42]). Based on the principle of the random assignment of individuals between a treatment group and a control group, this methodology estimates the effect of treatment by comparing the average results of the two groups. So on average, people treated and untreated are similar.

When randomization is not possible, individuals can choose to be treated or not. This induces a problem of endogeneity due to a correlation between the error term and the regressors. The researchers then used panel data to estimate the causal effect of the treatment. Other econometric models were used in this research issue : instrumental variable estimators (Heckman and al. [32], Earle [16], Hadley [27], Town and al. [75]), difference in difference (Farrar and al. [22], Rocha and al. [72], Scheyrog [40]) and regression discontinuity (Eibicha [64], Bor and al. [6], Dague [48]).

Other researchers have used survival models to analyze the duration of some events of interest. For example, Kucher and al. [47] used a survival approach (hazard model) to assess the effect of the introduction of standards on infection rates. He found that there has been a decrease in the probability of major complications following infection by 41%. Also, Cavallini and al. [61] use a Proportional Cox Hazard model to show the positive effect of the introduction of clinical guidelines. It demonstrates an improvement in mortality risk following the introduction of the standard. Chang and al. [9] used outpatient data and logistic regression to conclude that preventive prescription for children with asthma was more prevalent (40% higher) in the case of physicians having greater financial incentives.

The study of the financial incentives effect on behaviour is hugely important for the Quebec healthcare system. Indeed, when the financial incentive is attributed directly to the physician, there are some positive results in terms of improving healthcare quality and medical practice (Berchi and col. [4], Rosenthal and col. [59], Doran and col. [14]). On the other hand, when it is offered to the institution, it also often has positive implications (Grossbart [83], Lindenaeur [52]). In this last case, usually physician remuneration depends directly on funding offered to facilities since the physician is employed and payed by the institution. Thus, the physician behavioral change towards better practice would seem to be supported by financial incentives.

However, in the Quebec context, the situation seems to be different : financial incentive is not available to the physician, being self-employed, but it is offered to the hospital. Hospital funding in the Quebec healthcare system is essentially a historical funding. This type of budgeting guarantees a certain financial stability for the hospital. The additional funding for participants to program will help them to cover additional production costs and others costs incurred by the PQDCCR. So, it would be interesting to assess the effectiveness of such a lever on physician behavior and subsequently on the evolution of the quality of care, since the budget is offered to the institutions and has no direct impact on physician revenue. Moreover, the leverage can be counterproductive if it is offered in an unfair manner or is insufficient to cover all costs (Kingma [45]).

Our contribution in this analysis is to show the impact of clinical guidelines on healthcare quality and population health but in a specific context : payments of health professional is independent of the healthcare institution funding. In other words, physicians are self-employed and their remuneration is not included in the hospital 's operating budget. They are expected to contribute to this process not only offered to them without additional remuneration but with a real risk of decline in medical compensation due to the lower number of procedures performed. Does the financial incentive act in the same direction and with the same efficiency when offered to facilities ? Is the physician's behavior influenced by a parameter that is not directly related to his payment ? We also seek to separate the effect of clinical strategy of financial incentive : does the financial incentive support the change as postulated theoretically ? What would be the causal effect of the clinical standard with or without this

leverage ?

To solve this problem and give some reponses to all of these questions, we assess the effects of the introduction of the PQDCCR program on both length of stay (LOS) in hospital and the mortality risk of patients treated in participating facilities (average treatment effects on the treated). In the next section we introduce the healthcare Quebec reform called PQDCCR.

2.3 Quebec Program for Colorectal Cancer Screening (PQDCCR)

Although the volume of colonoscopies has increased significantly in recent years, the gap between supply and demand remains a concern for policymakers. This gap implies an increase in waiting lists and costs and a deterioration in the healthcare quality. In the absence of official Quebec data [3], some data [4] mentionned the variability of medical practice for this activity. Beyond the question of appropriateness, these events showed a variation in medical practices, and therefore also in the quality of care.

These findings show inefficiency in the Quebec health system caused by the inappropriateness of a considerable number of colonoscopies. In order to fix this problem, the Quebec Health Ministry(MSSS) introduced in November 2010 the PQDCCR program [5] and announced its gradual implementation.

Before implementing the program at the provincial level, pilot facilities had the mandate to ensure the quality and access to colonoscopy for all persons requiring this exam based on clinical and organizational standards. The selection process of these facilities extended between May and November 2010 and was based on predefined criteria : institutional commitment, regional health services diversity, minimum requirements on the populational diversity. The selection process was conducted by an anonymous committee. Therefore, these selected hospitals committed to the project requirements in terms of data availability criteria to monitor their performance and to improve access to services and

3. In SIMASS (Information System access to specialized services mechanisms) which is the official database for waiting lists for specialized services, colonoscopies exams are not followed. Indeed, SIMASS is a system connected to the operating room while colonoscopies exams are mostly made in endoscopy units

4. We cite mainly the 2008 National Public Health Institute of Quebec (INSPQ) report which states "the heterogeneity of procedures for including the appointment and monitoring, the lack of institutional mechanisms to prioritize requests colonoscopy and the lack of adequate documentation in many services colonoscopy". Others reports have been published by the College of Physicians ("Normes d'exercice en matière de coloscopie". Le Collège des médecins, March 2010, 50(2) : p.15.) and the Quebec Order of Nurses ("Lignes directrices sur les soins infirmiers en coloscopie chez l'adulte". Ordre des infirmières et des infirmiers du Québec, September 2011.)

5. We choose this date as a date of the program introduction because it is the date of announcement of the selected participating institutions. However, the effective date of guidelines implementation varies from one site to another. This timing delay could explain later variability in some results.

quality of care. The complete list of participating institutions is given by table 2.1.

TABLE 2.1 – Participant institutions

Region	**Facilty**
Bas St-Laurent	CSSS de Rivière-du-Loup
Capitale Nationale	Centre Hospitalier Universitaire de Québec
Capitale Nationale	Centre hospitalier affilié universitaire de Québec
Mauricie and Est du Québec	CSSS d'Arthabaska-Érable
Montréal	CSSS Sud-Ouest-Verdun
Montréal	Centre universitaire de santé McGill
Montréal	Hôpital Maisonneuve-Rosemont
Chaudières-Appalaches	CSSS Alphonse Desjardins
Montérégie	CSSS Pierre Boucher

This program first proposed clinical protocols for participating hospitals. This protocol covered the episode of care between the first visit to a GP up to the colonoscopy exam. Specifically, the organizational standards for optimal pathways of individuals requiring colonoscopies and clinical standards of screening colonoscopies were developed by a provincial expert committee. Healthcare quality mechanisms have also been proposed with the aim to assess for the achievement of program objectives. This mechanism is based on some quality indicators and other performance reports presented by each participating institution. The relevant medical associations have adopted and promoted these clinical standards.

At the same time, the Quebec health ministry (MSSS) offered to these hospitals a budget essentially to implement a new clinical software in the endoscopy units, to control for healthcare quality and to support this policy move. This software is necessary to assess the practice management information and make data retrieval efficient and easy. This revalidation process measures current pratices against those defined by clinical guidelines.

In April 2012, an additional financial incentive, an activity based funding, was offered to these facilities [6]. This funding is calculated every year based on additional colonoscopies volume (volume of the year minus average of the volume of two years : 2009 and 2010) and a unit price. The latter represents 50% of the average colonoscopy unit direct cost. This additional funding is offered conditionally to some performance criteria measured with outcome indicators.

This additional budget supported necessary efforts to temporarily reduce waiting lists by eliminating

6. The funding was not announced during the facilities selection process. So participant decision to enter the program could not be based on the funding level.

inappropriate cases. So the appropriateness of caseloads were revised and if they were still required, these requests were prioritized. All in all, the program has two main components : implementation of clinical guidelines in November 2010 and additional financial incentive introduced in April 2012. When we estimate program effects in the next sections, we analyze these two components separately and globally.

2.4 Data description

2.4.1 Descriptive Analysis

The present study is based on two sets of data : inpatient data for colorectal cancer surgeries in Med-Écho database and the mortality database. This sample covers a period of eight years (2006-2013) [7]. For each patient, we have clinical information regarding the hospital episode of care : age, gender, locality, DRG (Diagnosis Related Group), length of stay, treatment code, discharge destination, level of clinical severity, mortality risk level, etc. We follow each patient until the date of his death or censoring (March 31th, 2014). More information is available on the cases of death : date, place, main cause, secondary causes, etc. [8] We have to account for right-censoring in our data but no left-censoring. Indeed, some patients were still hospitalized on March 31, 2014 the end of our observation window.

Hence, our sample allows us to calculate the complete duration in and out-of hospital for each patient over this period. We distinguish between the length of stay (LOS) and the times spent after discharge until readmission, if at all.

We have 35 451 patients treated for this type of diagnosis with colorectal surgery during this period in over 84 facilities. We only considered inpatient surgeries because it is the main treatment for colorectal cancer. Preoperative radiotherapy reduces the risk of local recurrence. Adjuvant chemotherapy completes surgery, reduces the risk of recurrence and improves survival. These three treatments are often complementary. The literature indicates that the proportion of surgery cases for colorectal cancer treatment is over 90%. For more details on the data preparation, please refer to appendix H.

All the patients will be either in the treatment group (participating hospitals) or the control group (non-participating institutions). The control group provides a conterfactual to the treatment group. Tables 2.2 and 2.3 summarizes descriptive statistics over the sample period. We observed 9 282 admissions in the treatment group and 29 010 admissions in the control group. The facilities in the two groups are comparable in terms of the average age of patients, the average gender, average resource

7. Fiscal year beginning April 1st and ending the following March 31.

8. Since mortality database is structured by the fiscal year, we have the latest information for December 31, 2013. Therefore, it would lack the data on deaths from 2013 to 2014 in the last quarter compared to data from hospital admissions

use indicator (NIRRU), the average mortality index and the level of the average clinical severity. However, the average length of stay is slightly higher in the control group while the treatment group institutions have more beds on average. Both groups have an average age close to 68 years for men and women. According to appendix I, the evolution on NIRRU is similar in both groups and the average values are quite comparable from one year to another.

The total number of deaths is higher for the control group. Deaths are decreasing in both groups (Table 2.4) since 2011. However, a higher decrease rate was observed in the case of the treatment group. In total, we have 84 hospitals (9 in the treatment group and 75 in the control group).

TABLE 2.2 – Descriptive analysis(Treatment group)

Variable	Mean	Std. Dev.	Min.	Max.
Age	67.76	12.24	15	96
Gender	0.44	0.49	0	1
NIRRU	2.59	2.06	0	47.63
Clinical severity	1.89	0.81	1	4
Number of beds	373.04	114.70	118	549
Length of stay	12.44	15.53	1	296
Mortality indicator	1.76	0.79	1	4
N		9 282		
Censored		18		

TABLE 2.3 – Descriptive analysis(Control group)

Variable	Mean	Std. Dev.	Min.	Max.
Age	68.44	11.87	7	98
Gender	0.44	0.49	0	1
NIRRU	2.60	1.92	0	52.22
Clinical severity	1.91	0.83	1	4
Number of beds	289.89	146.34	16	571
Length of stay	13.48	15.35	1	447
Mortality indicator	1.77	0.81	1	4
N		29 010		
Censored		149		

At discharge, and depending on his health status, the patient may transit towards : home, death, other hospitals, CLSC (centre local de services communautaires), rehabilitation center, others.

Table 2.5 shows the number and the proportion of hospital episodes of care by destination at discharge in both groups. Destinations "home (21)" and "home with CLSC (17)" are the most frequent in both groups and represent close to 90% for all facilities in both groups. However, the proportion of discharges to the home is higher in the treatment group than in the control group while the proportion

TABLE 2.4 – Number of deaths in the two groups

Year	Control group	Treatment group
2006	1857	607
2007	1870	576
2008	1850	592
2009	1919	555
2010	1996	608
2011	1984	591
2012	1795	530
2013	1270	361

Only 2013 data are preliminary but data for all others years are final.

TABLE 2.5 – States frequency

Type de destination	Control group		Treatment group		Total	
	No.	%	No.	%	No.	%
Hospital center	785	2.7	250	2.7	1 035	2.7
Residential and long-term care center	488	1.6	176	1.9	664	1.7
Hospital center out of Quebec province	31	0.1	5	0.0	36	0.0
Rehabilitation center	14	0.0	9	0.1	23	0.1
Local community service center	8 199	28.3	2 440	26.3	10 639	27.8
Home	18 046	62.2	5 930	64.0	23 976	62.6
Death	1 115	3.9	315	3.4	1 430	3.7
Discharge without instructions	22	0.0	9	0.1	31	0.0
Day medecine	110	0.4	46	0.5	156	0.4
non-institutional facilities	200	0.6	102	1.1	302	0.7
Total	29 010	100.0	9 282	100.0	38 292	100.0

Agregated states are : in-hospital(1), home(2), hospital death(3), home with community care (4), others(5) and out-of-hospital death(6).

of discharges to home with CLSC services is lower in the latter. All other proportions are comparable between the two groups. For the remainder of our analysis, we consider the following five aggregated states : in-hospital (state 1), home (state 2), hospital deaths (state 3), home with CLSC services (state 4) and others (State 5). The last one includes all types of destinations other than destinations 2, 3 and 4.

Figure 2.1 shows that colonoscopies growth rates decreased since 2011 for the two groups. This decrease is more marked in the treatment group than in the control group.

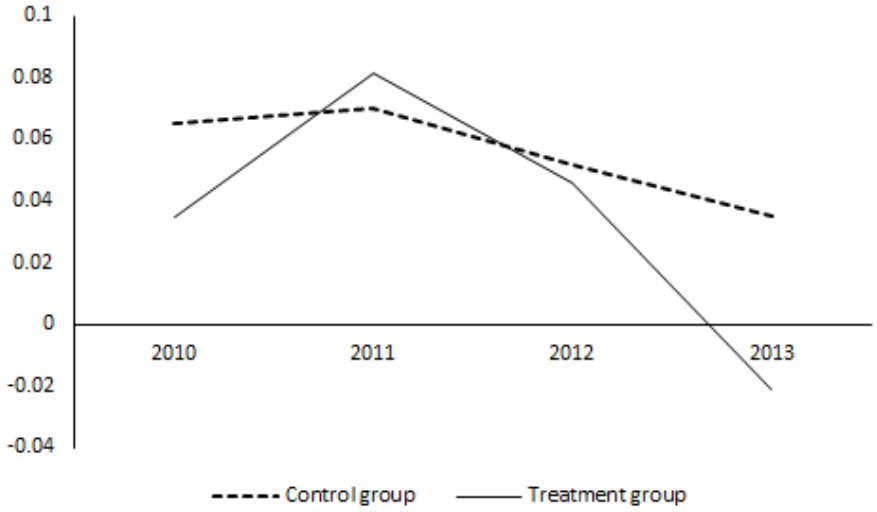
We use a probit model to ensure that the decision to participate in the program is not biased. We thus predict the probability of belonging to the treatment or the control group according to age, gender,

number of beds and year. According to the results of the table 2.6, estimates of the variables sex, year and the constant are not statistically significant while the estimates of the age and number of beds variables are statistically significant, respectively at the rate of 5% and 1%. These results show that hospitals in the treatment group have slightly younger patients and are larger (number of beds) than those in the control group.

TABLE 2.6 – Probit Results

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	12.2050	9.3794	1.30	0.1932
Age	-0.0013	0.0007	-2.03	0.0419
Gender	0.0106	0.0160	0.66	0.5102
Number of beds	0.0025	0.0001	43.81	0.0000
Year	-0.0067	0.0047	-1.44	0.1493

FIGURE 2.1 – Colonosopies volume evolution in the two groups



These ratios are annual growth rate for given year relatively to the prior year.

2.5 Empirical Model

In this section we present a multi-state transition model in continuous time to identify the causal effect of the reform on hospital length of stay, readmission and in-hospital mortality. We use a difference in difference approach where institutions that do not participate in the PQDCCR determine control group. We estimated the transition hazards between two states for each patient. This is the general case of the usual two states transition model (e.g Cox model). In the last years, multi-states models were used especially in epidemiology and health economics (Harbarth and al. [29], Price and al. [68], Andersen and al. [67], Berlin and al. [5], Rainer [86])). Our approach extends the model developed by Fiocco and al. [23].

Let i denote a patient hospitalized for colorectal surgery in hospital j. Z is a vector of covariates (some are specific to a patient and others to a hospital). We define S={1,2,3,4,5,6} as finite space of transition states : 1(in-hospital), 2(home), 3(in-hospital death), 4(home and CLSC services), 5(other destinations) and 6(out-of-hospital death). States 3 and 6 are absorbing states. We assume that when a patient leaves a state, he can go back as long as this is not one of absorbing or censored states. We distinguish between state at home with or without CLSC services because the presence of these services just after discharge may argue for a better patient autonomy and so a better health status.

Let's α_{sr} be the transition intensity for each pair of states s and r. This represents the instantaneous risk of moving from state s to state r at time t. In other words, it is the conditional probability that a patient with covariates Z and being in state s at the time t moves to state r in the interval [t, t+Δt].

$$\alpha_{sr}(t,Z) = \lim_{\Delta t \to 0} \frac{P_{sr}(t < T < t+\Delta t | T > t, Z)}{\Delta t} \tag{2.1}$$

In the case of a single initial state (state 1) and only one possible transition, we have the special case of a competing risks model, and failure must be due to only one of the four possible destinations at discharge : death, home, home and CLSC services and others.

In the more general case, we have a multi-states and multi-episodes transition model. This would allow for better evaluation of the reform through the interpretation of several transition states. Beginning with state 1, we observe at time t , one of the four states 2, 3, 4 or 5 and return to initial hospitalization state is allowed. Each patient is followed up to either entering in one of sixth absorbing states (out-of-hospital death) or to be censored. In the next section, we present the competing risks model.

2.5.1 Competing risks model

We consider a competing risk approach to deal with destination after discharge. We assess for PQDCCR causal effect on the transition hazards (average treatment on the treated) between hospitalization state and several states (different destination types) as described by figure 2.2. We conclude then about the length of stay evolution.

The choice of competing risks model in our case is motivated by the different transition hazards for different competing events. So, if we ignore competing risks or we suppose them as censored, we violate the independent censoring assumption of standard survival analytical methods such as Cox model and Kaplan-Meier and the estimators will be biased.

In our model, we control for covariates specific to patients and others specific to the institutions. We suppose that competing risks are independent. As we know from the literature, there is no test for this assumption and only a few analyses(Honoré and al. [2]) relax this hypothesis.

We use non treated institutions as the control group and we model the outcome variable, length of stay (LOS) as being determined by the following specification.

The main identifiable quantity in this model is the cause-specific hazard function $\alpha_j(t/Z)$. At each time t, only one state will be observed for patient i. The hazard function in equation 2.1 can be modeled as a proportional hazard with unobserved heterogeneity. So, hazard α_j as in equation 2.2 will be defined as the instantaneous risk (probability) of transition from state 1 to state j at time t conditional on Γ_{ij}^s, including covariates vector Z, treatment variables (equation 2.3) and also unobserved heterogeneity ϑ_h term assumed to be independent of Z. This frailty term follows a gamma distribution with unit mean and variance θ

$$\alpha_j(t/\Gamma_{ij}^s, \vartheta_h) = \vartheta_h \alpha_0(t) \exp(\Gamma_{ij}^s) \qquad j = 2,3,4,5 \tag{2.2}$$

with

$$\Gamma_{ij}^s = \gamma_1 + \gamma_2 Ind_{it} + \gamma_3 Post_{it} + \gamma_4 Post2_{it} + \sum_{k=2}^{T} \delta_k I_t + \beta_1 Ind_{it} Post_{it} + \beta_2 Ind_{it} Post2_{it} + Z_{it}\phi + \varepsilon_{it} \tag{2.3}$$

where $Post_{it}$ is a dummy for the post treatment period, $Post2_{it}$ is a dummy for the post funding period and Ind_{it} is a dummy for the treatment group. Z_{it} is a vector of covariables, I_t is dummy for the yearly trend common to the two groups and ε_{it} is an i.i.d. error term. We will discuss later the treatment specification in our approach. ϑ_h represent unobserved differences between facilities. Also, it will take into account the heterogeneity in the data, associated with unobserved covariates. β_1 is the treatment effect, β_2 is the funding effect and ϕ is the vector estimates of covariates.

Let α_0 denote the baseline hazard function. β_1, β_2 and ϕ are the parameters to estimate by maximizing likelihood function under model assumptions. Equation 2.4 defined likelihood function as the all ratios multiplication of each individual hazard by the all patients hazard, for each state j

$$L(\beta) = \prod_h \prod_i \prod_j \vartheta_h \frac{\exp \Gamma^s_{ij}}{\sum_{j \in C_i} \exp \Gamma^s_{ij}} \tag{2.4}$$

where y_j is the observed time (including censoring cases) and C_i is the set of indices j, with $y_j \geq t_i$ (those at risk at time t_i).

The cumulative incidence function gives the proportion of patients at time t who have discharged to destination j accounting for the fact that patients can discharge to other destinations. Figure 2.3 plots cumulative incidence curve (like survival function in the usual Cox model) for states 2 and 4 (home is the event of interest in our model) in treatment and control groups. For a mean patient (mean values of age, gender and clinical severity) and mean hospital(mean number of beds), cumulative hazard vary differently before and after the implementation of PQDCCR (period 1 and period 2). However, additional funding effect seems to be absent (period 2 and period 3).

On one side, the program increased the incidence to exit at home. On the other side, additional funding has no effect on this risk. Total effect (announcing program participants and introduction of addional funding) is positive and we had an increase in the incidence of discharge to home compared to other competing events. In other words, the PQDCCR program contributed to a decrease in hospital length of stay after colorectal surgery. For the long term, gaps between three curves became constant.

For the specification of the treatment effect, we then introduced a trend variable to capture year-

specific effect that should be different in the two groups. This variable is defined as dummies for each year and is specific to each group (control and treatment). We test also for other specifications (see Results section).

2.5.2 Multi-states transition model

A multi-state approach is particularly appropriate for our data, since it allows for many transitions between different states. Also, it does not allow patients to enter two states at the same time. Figure 2.4 show possible transitions between these states.

In our analysis, we focus on some events of interest in the aim to conclude about healthcare quality and populational health : in and out-of hospital mortality, readmission, discharge at home. Table 2.7 summarizes total number of transitions between each of the two states. Most frequent transitions are : discharge at home (with or without CLSC services), out-of-hospital death and in-hospital death.

TABLE 2.7 – Number of transition between states

	1	2	3	4	5	6
1	0	33616	10595	13824	4886	0
2	9396	0	0	0	0	3513
4	3587	0	0	0	0	1360
5	753	0	0	0	0	1895

1 : in-hospital ; 2 : home ; 3 : in-hospital death ; 4 : home and community care ; 5 : others ; 6 : out-of-hospital death.

Let T denote the continuous random variable that represent the duration of time units that a patient i spends in some state s in time t before experimenting with another event and let H_s- be the observation transition history before entering into state s at this time t. The hazard of transition between two states s and r, as seen in equation 2.5, will depend on this history and the covariates vector Z.

$$P_{sr}(k,t,H_s-) = P_{sr}(Z_t = r|Z_k = s,H_s-), for \;\; k < t \tag{2.5}$$

Transition intensity, derived from equation 2.4, represents the instantaneous hazard for moving to state r conditionally of being in the s state at time t. Hazard function α_{sr} in equation 2.6 can be used to estimate the average duration and the number of patients in each state at a time t.

$$\alpha_{sr}(k,t,H_s-) = \lim_{\Delta t \to 0} \frac{P_{sr}(t,t+\Delta t, Z_t = r | Z_k = s, H_s-)}{\Delta t} \tag{2.6}$$

To estimate transition intensities, we use the hazard proportional regression model including a baseline hazard (piecewise). We assume a Markov process : at time t, future transitions depend only on present state and not on transitions history. We test for this assumption by including some variables correlated with transitions history. This assumption is plausible since it is used in many analysis including health state evolution for cancer patients where age variable is dominant (Kay [73], Ladabaum [49], Ross [77]). The Markov process may be homogeneous if transitions intensities do not vary with time. Otherwise, we have a non-homogeneous process. We assume for this analysis a continuous homogeneous Markov process.

Note that even in the homogeneous Markov models, transition probabilities can depend on time. In our context, there is some evidence about this hypothesis : technology change and clinical practice variation, as well as health network shocks. Transition intensity for patient i is given by equation 2.7

$$\alpha_{sr}(t/Z_i) = \alpha^0_{sr}(t) \exp\Big(\sum_{i \in C_{sk}} \beta_{sr}(t) Z_{hi}(t)\Big) \qquad 1 \leq s \leq r \leq 6 \tag{2.7}$$

with $\alpha^0_{sr}(t)$ defined as a baseline hazard at time t and $\beta_{sr}(t)$ regression estimate for transition from s to r.

The likelihood function in equation 2.8 is then obtained by replacing transition intensity in the general likelihood statement. The last observation for each patient will be censored if the event is different from death. So we consider this right-censoring in every patient contribution to the likelihood function.

$$L = \prod_{sr} \prod_{k} \prod_{i=1}^{\eta^k_{sr}} \alpha^0_{sr}(t) \exp\Big(\sum_{i \in C_{sk}} \beta_{sr}(t) Z_{hi}(t)\Big) \qquad 1 \leq s \leq r \leq 6 \tag{2.8}$$

where the risk set C_{sk} is the set of patients i in state s that are at risk of their k^{th} transition just before time t and η^k_{sr} the number of uncensored durations in which the k^{th} episode started in state s and moved to state r

FIGURE 2.2 – Possible competing risks at the discharge

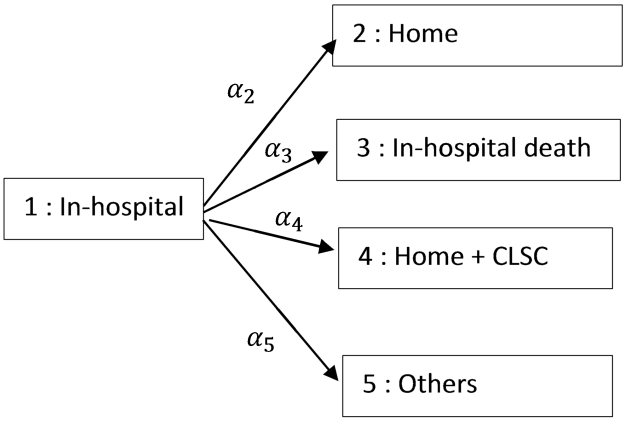
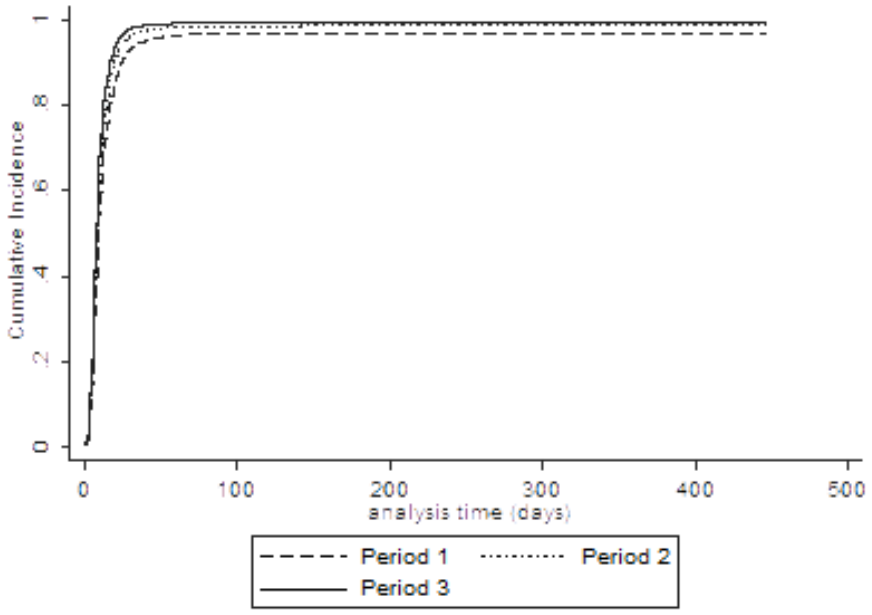
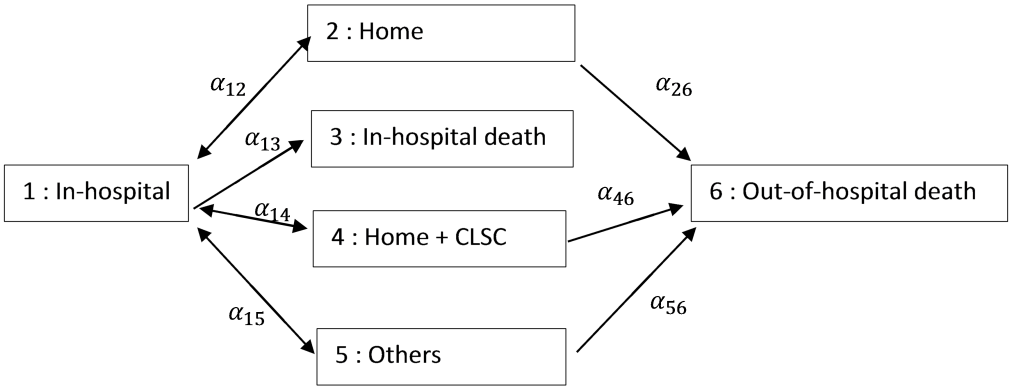


FIGURE 2.3 – Cumulative incidence function for the home discharge event



Period 1 : before november 20th, 2010 (Treatment variable=0 and funding variable=0) ; Period 2 : between november 20th, 2010 and march 31th, 2012 (Treatment variable =1 and funding variable=0) ; Period 3 : After march 31th, 2012 (Treatment variable =1 and funding variable =1).

FIGURE 2.4 – Possible transitions in the model



2.6 Results and discussions

In the first part of this section, we estimate a general mixed proportional hazard (MPH) model with and without unobserved heterogeneity (frailty). We also compare the MPH model with the semi-parametric Cox model. After that we conducted the competing risks model, analog to a difference in difference approach. We focused on the hazard for discharge at home compared to other types of destination (model 1). To check robustness and validate the choice of competing risks, different specifications of the model have been estimated and the results have been very similar. So our specification strategy is built on four steps as suggested by Mora and Regio [70]. In the first step, we test for parallel paths assumption. The treatment effect is identified by the interaction between the treated group indicator and the post-treatment period indicator.

In the second specification (model 2), we test for a common trends assumption by including common year indicators for the two groups. In the third specification (model 3), we add a linear post-treatment trend only for the treatment group. In the last specification (model 4), we introduce a fully flexible dynamics by considering common trends for two groups and specific dynamics for the treatment group in post-treatment periods.

In the second part, we adopt a multi-states model in the general case (model 5) and we estimate the transition probabilities between different states as well as the average length of stay in each of these states.

The MPH model was performed with parametric exponential assumption and shared gamma frailty (mean 1 and variance θ). Results in table J.1 show that treatment effect is positive and statistically significant at the level of 1% in the general case of the proportional hazard model (0.062) and also in the specific case of the Cox model (0.079). The funding effect is not significant for either models. Adding the frailty term in the MPH model, results in the second column of the same table show a significant treatment effect (0.059) quite similar to the estimate in the model without unobserved heterogeneity . The frailty variance estimate θ is equal to 0.055 and the likelihood ratio test for the presence of heterogeneity is significant. So there is unobserved heterogeneity between hospitals in our model. This frailty resumes unobserved effects like change in standards for data entry over the years. In the next analysis, we adopt the Cox model and competing risks model without unobserved heterogeneity effect.

Looking at other covariates, as expected, a greater age, clinical severity or atypical level case have significant negative impacts on the exit rates (increase in LOS). Also, this risk of discharge is greater in the case of females compared to males. Post-treatment estimates (Post and Post2) are positive and all of these estimates are statistically significant.

To check the common robustness of parallel trend assumption for the MPH model, we only looked at pre-treatment data and we moved placebo treatment to 2008. Results in table J.2 show that placebo treatment is not significant. So over the last years before treatment, there has been no major reform that explicitly discriminates between the two groups.

We performed the competing risks model and we compared results with the Cox model. In the competing risks specification, the reform effect is significantly positive at the rate of 1%. In other words, the reform increased the specific hazard of discharge at home comparatively to other destinations by 10% (column 1 of table K.1). This means that the length of stay (LOS) was shortened for every treated average patient who discharged at home. The earlier screening of the disease, in the case of treated facilities, lead to better patient health conditions at the admission and quicker recovery after the surgery. As seen in the second column of the same table, this effect is equal to 7.9% in the case of the Cox model. The reform effect is higher in the case of the competing risks model. So, the Cox model underestimates reform impact because we suppose that all destinations at discharge have the same hazard. This hypothesis is not plausible in our case. We performed the Gray(1988) test comparing cause-specific cumulative incidence curves. The null hypothesis of equality of failures risks is rejected. So, we should consider competing risks events otherwise reform estimates will be biased.

Considering the funding effect, the estimate is not significant in all specifications. In other words, announcing the heathcare quality reform shortened LOS whereas additional funding has no effect on LOS. All in all, the total reform effect is a decrease in hospital LOS for patients who discharged at home. This specification suggests that the funding variable could be ignored in our model. No significant funding effect on LOS may be explained by the short period of data after the implementation (only two years). Also, this financial incentive was annouced to the institutions in February 2013, in other words only 13 months before the final date in our sample.

We tested specifications other than the parallel paths, for the competing risks model. Column 2 of table K.2 shows results of our second specification (model 2). Common trend effects are growing over the time and statistically significant at the rate of 1% (except those of 2008 and 2009). So specific year shocks (common for two groups) shortened hospital LOS [9]. These shocks may be caused by greater supply for rehabilitation institutions or changes in data entry, especially for the destination at discharge data entry. The reform total effect is statistically positive and equal to 11%.

In model 3, we added a linear trend variable specific to each group. Results in column 3 of table K.2 show that the linear common trend is significant at the rate of 1%. However, reform estimates do not

9. We tested for quarterly dummies and we obtained similar results for treatment and funding variables as those in the model with annual dummies.

change greatly compared to the previous model and are positive (11% for total effect). Specific linear trend (interaction between the time dummy and the treated group indicator) is not significant in this specification. Including the linear trend variable, we controlled for a specific effect for two groups with parametric function (linear). We tested also for other polynomial forms (T^2 , T^3 and T^4) and we found that the effects of these variables are no longer statistically significant. So, introducing higher power levels in time did not change the reform parameter estimates in any significant way.

In our last specification, we replaced the linear parametric trend with fully flexible dynamics for the treatment effect in the post-treatment period [10]. This model, introduced by Mora and Regio [70], is more flexible than the linear trend model because we do not suppose parametric specific group time evolution. Results in column 4 of table K.2 show that the treatment effect is not only different for each post-treatment year but also is growing over time : 8% for 2011, 18% for 2012 and 21% for 2013. All of these estimates are statistically significant. We tested to see if the treatment effect is really different for the three post-treatment years. We used likelihood ratio test to test null hypothesis that model with three post-treatment dummies is nested in the model with only one post-treatment dummy. In other words we test if the last specification results in a statistically significant improvement in model fit compared to the third specification. The test does not reject the null hypothesis so the two models are similar. The third specification seems to be the best one and it suggests that the reform total effect is positive equal to 11%.

In this second part, we consider a multi-states and multi-episodes model to test for the reform effect (we consider only reform annoucement variable but not funding variable). We estimate transition intensities between states of model 5 using msm package with R statistical software [11]. We assume a homogenous Markovian transition process (transition intensity function is not affected by time). However, transition probability between two different states will depend on time *t*. For the context of inpatient data, several factors suggest that this probability would not be constant over the time : change in technology and medical practice, variation in hospital environment and health network and other specific period shocks.

Results in column 1 of tables L.1 and L.2 show that the reform increases transition intensity from in-hospital state to home state by 1% (4.6% for treated relative to 3.6% for controls). So discharge at home for patients treated by colorectal surgeries in the treated facilities is faster after reform. This result is consistent with the competing risks model result discussed earlier in the first part of this section. Hazard for in-hospital death decreased by 0.3% after the reform implementation. So faster

10. Fully flexible dynamics in the pre-treatment period are not significant so we kept only dummies for the post-treatment period.

11. This package employs maximum-likelihood estimation for general multi-state Markov. Main results are intensities transition matrix and maximized likelihood value. All details about this package are available at the following address : http ://cran.r-project.org/web/packages/msm/index.html

discharge has not been supported by deterioration in hospital mortality. At the same time, transition intensity from in-hospital state to home with community care services increased after the reform from 2% to 2.7%. This improvement points out a greater patient autonomy and better health conditions at discharge. There is no significant reform effect on the probability of out-of-hospital death between the two groups. This can be due to the short period targeted by our sample to see an effect on this variable .

Sojourn times, as seen in the first column in tables L.3 and L.4, show that the reform shortened average duration for hospital stay by 3.4 days (from 14.53 to 11.13 days), for state 2 (from 308.15 to 98.07 days), for state 4 (from 319 to 118.01 days) and for state 5 (from 256.21 to 74.94 days). Decrease in durations for states 2, 4 and 5 after the reform is consistent with results in tables L.1 and L.2 : treatment effect is negative in all transitions to in-hospital states from different states (home, home and CLSC, others). In other words, the reform has contributed to a faster discharge from hospital but also a faster return. However, this faster return to hospital does not argue a deterioration in readmission rates because the average LOS for states 2, 4 and 5 are respectively greater than 98, 18 and 74 days. Or, usually, readmissions are based on patient hospital return within 30 days. So there is no deterioration in the readmission rate after the reform even if average return to hospital is faster. All in all, the shortened average duration in states shows an improvement in healthcare quality and patient satisfaction.

Given the Markov assumption, transition probabilities between states depend on the time t only. We then plot the evolution of these transition probabilities on a 30-days interval for mean patient in the case of hospital discharge and re-hospitalization and on a ten-years interval in the case of out-of-hospital mortality. Figures M.1 and M.2 show that transition probabilities from state in-hospital to home state (with or without CLSC services) are increasing over time for treated and non treated groups. However, these probabilities are greater in the case of the treatment group. After the reform, LOS has decreased on average and this decrease is more important over time. Also, the transition probability is increasing from in-hospital state to other destination types, except death (figure M.3).

As seen in figure M.4, in-hospital death probability increases over time for both groups and is lower in the treatment group. The gap between the transition probabilities in the two groups is growing from the first day of the episode. This reflects an improvement in the survival rate in hospital after colorectal cancer surgery.

Figures M.5, M.6 and M.7 show that the probability of death after discharge is increasing over time in both groups and is higher in the treatment group. This may be due to the short study period after PQDCCR implementation. Also, mortality data for 2013 are preliminary and not complete for the fiscal year. We can see the same effect in figure N.1 and N.2 that shows a higher proportion of out-of-

hospital deaths in the treatment group. Figures M.8, M.9 and M.10 show that hazard of readmission, that is higher in the treatment group, is growing over time.

Mortality rates, incidence rates and the proportion of patients detected at earlier cancer stages are the best indicators for the population health (Stratmann and Leive [84], Young and Womeldorph [88]). Since these indicators are not available in the context of our analysis, [12] we chose two other indicators : the hospitalization rate per capita and the evolution of the treatment approach. The first indicator measures the change in the number of hospitalized patients for colorectal surgery versus other types of treatment such as chemotherapy or radiotherapy. The second indicator determines whether the cancer treatment is less invasive after the program implementation.

We performed a difference in difference analysis to estimate reform effects on hospitalization rates [13]. The first column of the table O.1 shows that the effect of the reform announcement in November 2010 was not significant. Furthermore, the effect of additional funding is significantly negative at the rate of 1 % : implementing additional funding for treated facilities helped them to reduce the hospitalization rate by 8 %. On the other hand, a higher 1% rate of colonoscopies increased the rate of hospitalization by 6 %. The negative causal effect of financial incentive reflect better health of the population targeted by the PQDCCR. There was a substitution between treatment approaches and the decline in hospitalization rates should have resulted in increased levels of other disease treatments such as chemotherapy or radiotherapy. Then we looked at the variation of treatment approach for colorectal surgeries after PQDCCR introduction.

We analyzed the evolution of the treatment approach. We used probit analysis to estimate the effect of treatment and financial incentives on the approach category. First we looked at the presence of surgery in the episode of care for each patient [14]. Results in the first column of table O.2 show that reform is not significant on surgery. Second, we classified all episodes of care in two possible approaches for colorectal surgery : open surgery approach (laparotomy) and second less invasive approach called laparoscopy. The results of column 2 of the same table show that the reform (PQDCCR announcement and additional funding) contributed to a greater use of the laparoscopic approach (total effect of 54%). These variables are statistically significant at the rate of 1%. This result reflects a healthier population since a less invasive approach is usually employed at earlier cancer stages. However, other factors may explain this shift like the physician's experience and available equipment in the facility.

12. This is because the short time sample and also we have some missing data to calculate these indicators
13. This ratio is calculated by 1 000 residents and is weighted by NIRRU of each patient. Indeed, gross number does not reflect for example if we treated more patients but with less clinical severity or health conditions.
14. To perform this analysis at patient level, we used all inpatient data treated for colorectal cancer and not only surgical cases. The period of data is between 2009 and 2013

All in all, our results are consistent with our expectations according to which patients treated in reformed institutions spend less days in hospital because of a better health condition caused by earlier screening cases. Empirically, this effect is reflected by a decrease in the hospitalization spells but no deterioration in the risk of in-hospital death or re-hospitalisation after few days of discharge. Our results raise an important issue regarding the measure of healthcare quality for colorectal surgeries.

Financial incentives are present to enhance the implementation of healthcare quality programs and should operate in conjunction with other systems. Theoretically, they are intended to reinforce behaviour change towards the targeted goal. Interestingly, in the case of our reform, funding has no effect on LOS according to our sample data. It can be explained by the short period of post-treatment analysis. Also, this effect may be explained by the fact that financial incentives were not delivered to the level of the clinical department (oncology unit care) but at the level of the endoscopy department. Finally, to have better results, we should be sure that the objective of financial strategy is engendering improvement across all facilities rather than just rewarding institutions/services that are already performing well.

Other factors may influence the level of achievement of financial incentives objectives such as physician age, experience or qualifications, the location and the kind of medical practice, disease severity. This analysis did not address the impact of these factors in the case of the PQDCCR program. Also this study does not focus on other risks associated with this strategy implementation like deterioration of access to services, healthcare integration, conflicts of interest between physicians and patients.

2.7 Conclusion

In November 2010, Quebec introduced the colorectal cancer screening program. Based on clinical guidelines and offering, since THE April 2012 additional financial incentive, the program has the objective to standardize the use of these exams by serving as a framework for clinical decisions and supporting best practices. Consequently, it should increase the quality of care, reduce medical risks and improve the healthcare efficicency : patients will get the right care, at the right place and at the right time.

This analysis provides some answers about the level of achievement of these objectives. Our results show that clinical guidelines helped to reduce hospital length of stay on the treated facilities at the average, for any type of discharge except death. This improvement did not cause any deterioration in hospital mortality or readmission rate. In other words, the reform improved healthcare quality and the satisfaction of patients after the episode of care.

On the other side, additional funding did not contribute to this improvement on hospital length of stay. However, total reform effect is positive : the reform, including guidelines and financial incentives, is expected to increase the probability of discharge.

Our multi states model suggests that the mean duration in all non absorbing states decreased after the implentation of the PQDCCR program. So patients spend on average less time in hospital and in home. Because the mean duration in each state is greater than 30 days (time usually used for calculation of readmission cases), there is no evidence of deterioration in readmission rate even with faster return rate to hospital. However, this result suggests to implement clinical standards for hospital episode of care. Indeed, guidelines only for colonoscopies exams may be insufficient to improve healthcare quality of the entire episode of care beginning at the first visit to a GP up to discharge after hospital episode.

We show also in this analysis that the program including financial incentives contributed to avoid the use of more invasive treatment by the use of laparoscopic surgery instead of laparotomy. Besides that, only additional funding helped to decrease the weighted hospitalization rate for colorectal surgeries. These two results shows an improvement in populational health and patient satisfaction.

Finally, our analysis shows that even when physician renmuneration and hospital funding are independent, it is possible to modify physician behaviour and improve hospital performance : discharge decision, treatment choice, admission decision, etc. Implementation of clinical guidelines and pay for performance mechanisms are relatively new concepts in the Quebec health network and we should give different stakeholders a further adaptation period to this new competitive reality.

Furthermore, our study has some limitations. First the short time sample after the implementation of PQDCCR (three years) is not suitable for studies dealing with survival or death rates at 5 or 10 years. Or such indicators are relevant for analyzing the impact of the program. Also, with the short time after the policy funding annoucement (only 13 months), we cannot conlcude about financial funding effect. Finally, a cost-effectiveness analysis may suggest an evaluation of the anticipated benefits of this screening strategy for every dollar invested. Such analysis, expected in the next steps, will be helpful for deciding to continue with this program in the future.

Chapitre 3

A cost-effectiveness analysis of the Quebec colorectal cancer screening program

3.1 Introduction

Colorectal cancer, which is often fatal, represents for Canada in general and Quebec in particular a heavy economic burden. Indeed, statistics show mortality rates of 0.22 and 0.14 per 1 000 residents and incidence rates of 0.6 and 0.4 per 1 000 residents, respectively for men and women in Canada. In addition, the relative survival rate after 5 years based on the estimate of 2006 to 2008 is about 65%. These incidence rates have declined in recent years, especially because of a rise in screening tests. [1]

The development of public policies in defeating cancer, especially those based on screening tests, can significantly reduce the economic costs of this disease by reducing the costs of treatment and increasing survival rates. This objective becomes achievable especially when policies are well implemented and strategies well communicated. The screening tests are usually offered to target populations with a high risk of the disease.

The PQDCCR program, one of these many policies based specifically on screening tests, was introduced in nine Quebec facilities since November 2010 and was funded by the Quebec Health Ministry (MSSS). It targeted the population aged between 50 and 74 years [2]. This program generated benefits but also required supporting implementation and monitoring costs. Our analysis in the previous chapter showed that this program shortened length of stay (LOS) after colorectal surgery, decreased the in-hospital mortality rate and improved the use of less invasive treatment approaches for the surgery

1. Estimated Canadian Colorectal Cancer Statistics 2015 according to data from the Canadian Cancer Society.
2. See chapter 2 section 3 for more details about this program.

episode. However, we obtained mitigated conclusions about survival rates after discharge.

Several analyses (Lieberman and al. [51], Imperiale and al. [35], Hardcastle [30], Siegel [82] and Mandel [56]) showed that the survival rate should increase through early detection and treatment of colorectal cancer. Even Stratman and al. [85]) showed that the effect of some screening strategies on the mortality rate is clearer for the age groups targeted by these strategies.

However, little or no analysis calculated the costs and benefits of this program since its implementation in Quebec. This is necessary to recommend an extension or enhancement program across the province.

The PQDCCR program offers clinical guidelines before the use of colonoscopy exams. These procedures, one way of colorectal cancer screening, are known to be efficient for the identification of adenomas and cancer and also help to remove polyps during exams. We should be careful about deaths that occur soon after the initiation of screening and that may result in biased estimation of the policy effect(Hanley [41]).

However, colonoscopies have some issues such as the risk of complications, high costs and limits to access for patients. Therefore, it would be interesting to analyze the effectiveness of such screening policy. This goal is more legitimate given the presence of other screening modalities for colorectal cancer : FOBT (Guaiac-based fecal occult blood test), FIT (Fecal Immunochemical test), Stool DNA test and virtual colonoscopy. In addition, these screening policies often target a healthy population with different levels of risk making the choice between these different alternatives not easy.

That's why some authors tried to evaluate the relevance of these different strategies. Interestingly, a recent study(Ioannidis and al. [80]), showed that screening healthy individuals does not save lives. In this sense, the development of better treatments may be a better strategy and can mitigate the advantages of earlier screening. Or because high costs of these kinds of strategies and the scarcity of resources, it will be useful to analyze the costs and benefits of each alternative, including colonoscopies, and to compare between them to know which one is the most efficient. It would therefore be interesting to assess the costs and benefits and then compare them.

The cost-effectiveness study aims firstly to estimate the benefits of this program for the Quebec population and then calculate the difference between the costs and benefits in order to get a clearer idea about the relevance of the program.

The objective of this analysis is to estimate the benefits of this program in reducing treatment costs

and compare these benefits to the program implementation costs, especially as colorectal cancer is acknowledged as being curable if diagnosed and treated early (Farmer and al. [20]). It also allows us to compare the costs and benefits for treated versus those who are not. So we compare between two main scenarios : the treatment group and the untreated group (reference scenario). This analysis presents essentially two treatment scenarios : first keeping the clinical reform based on clinical guidelines and second adding financial incentives for additional colonoscopies volumes .

In the next section we discuss a short literature review of the cost-effectiveness analysis of strategies defeating colorectal cancer. Then we present the estimation of program costs and the expected benefits. Finally we discuss the results of the analysis and conclude about the PQDCCR efficiency and relevance.

3.2 Literature review

When we want to compare different policies, one way should be analyzing costs and benefits of each one. But there may be some gaps between these values for two strategies. In general, this depends greatly on the population targeted by the reform. Public policies for high-risk populations are usually the most successful. In addition, in our case of screening policies, frequency of screening tests is the determinant key on the effectiveness of the reform. Also the choice of cancer type, the sector activity and other parameters can play an important role in explaining variation in results.

Several methodologies allow evaluating benefits and costs. The cost-effectiveness analysis in health economics is one of these methods based on the simple economic principle : maximizing the benefit of each unit cost per treatment. Often the cost estimate is easier to evaluate while the benefits may differ depending on the assumptions. The advantage of the cost-effectiveness method is to guide not only the choice between several alternative treatments given the precariousness of resources, but also to judge the appropriateness of beginning or continuing any program. Indeed, policymakers often prefer to finance reforms for which net benefits are positive relative to other alternative strategies or reforms.

Many studies used the cost-effectiveness analysis of various alternatives for colorectal cancer screening (Sarfaty and Feng [55], Lansdorp and al. [50], Schneider and Häck [40]). This is in order to choose the alternative that would provide a better cost-benefit ratio. These analyses have concluded that the majority of screening alternatives bring benefits in terms of cost savings of treatment, saving on labor productivity and life-years gain with implementation costs that are not only acceptable in developed countries but higher than situation with no screening. However, few conclusions were drawn as to the relative effectiveness of a given strategy or if an alternative is preferred to another(Elmunzer [18]).

Other economic efficiency evaluation methodologies are available and the choice of one of them depends on the context and objectives of the evaluation. More specifically, the known methods at this level are : cost-minimization analysis, cost-effectiveness analysis, cost-utility analysis, cost-benefit analysis, economic impact analysis, fiscal impact analysis and Social Return on Investment (SROI) analysis.

While the cost minimization analysis is simply to consider treatment with a lower cost with identical results, the cost-effectiveness analysis takes into consideration the benefits of treatment in clinical terms, not monetary terms. The cost-utility analysis is able to translate the expected benefits in terms of QALY (Quality Adjusted Life Year)[3], whereas the cost-benefit analysis estimates the monetary value of these benefits. The latter method may be more difficult especially when it is not easy to obtain this financial estimate.

For this analysis, we choose the cost-effectiveness methodology for two reasons : first we have information to assess the monetary value of the benefits. Also, we assess the costs and benefits from the perspective of the government and Quebec health system. So monetary evaluation is important in this case.

Often the information of the cost-effectiveness analysis is synthesized by a cost-benefit ratio. It represents the additional benefits (compared to statu quo) generated by the policy compared to additional costs (compared to staut quo). In the next section we explain PQDCCR cost estimation.

3.3 Costs estimation

The colorectal cancer screening program was introduced in selected Quebec facilities in November 2010. This new program implied additional costs including essentially clinical software system costs, waiting lists management costs and financial incentives offered to participating institutions based on additional colonoscopies volumes (compared to the 2009 and 2010 average volume) and conditionally to performance improvement.

In the cost identification, two categories are considered : fixed costs and variable costs. Fixed costs may be recurring or non-recurring. Variables costs depend on the colonoscopies volume every year.

3. The QALY is a measurement for each year of healthy life. It reflects two dimensions : quantity (life expectancy) and healthcare quality (health status).

When we calculated patient treatment costs, we included only direct costs [4]. Direct costs include drugs, nursing, laboratories. Indirect costs include administrative and support costs. On the other hand, non-medical costs including travel costs, opportunity costs, etc, are ignored in this analysis.

To estimate the program costs, we included those costs for the post treatment period (2011, 2012 and 2013) [5] and we considered all patients treated in participating facilitites between 2011 and 2013 (4 021 patients) for colorectal cancer surgery. All costs were inflated to 2011 Canadian dollars using the health and personal care component of the consumer price index for Quebec. Based on a normalized index of 100 for 2011, these discount rates are 102.3 for 2012 and 102.8 for 2013 [6]. Other costs (intangible costs) are not observable and therefore can not be estimated : pain, discomfort, etc. Below, we present different PQDCCR costs included in our analysis and how they are estimated.

3.3.1 Costs of clincial software

In the objective to assist physicians with the best practice defined by guidelines and improve health-care quality, a clinical software system was installed in the institutions participating in the experiment. First, we have purchasing costs of this system usually estimated by the real price paid by hospitals. Also, the software package implementation has required the development of endoscopy units. These fixed costs vary between 50 000$ and 100 000$ for every facility. Since we did not have the real costs of implementation, we have calculated an average of 75 000$ per facility. So, the total cost for the pilot facilities is equal to 675 000$.

3.3.2 Costs of managing waiting lists

At the beginning of this program, facilities performed preliminary work to prioritize some cases. The objective is to ensure that we consider appropriate cases with respect to clinical standards. This action required the use of an assistant nurse to prioritize relevant cases. The salary of this resource can be calculated using the top-down approach of the Quebec Health Ministry (MSSS). The cost of this work was variable from one site to another. All in all, these fixed costs are equal to 3 197 220$ [7].

3.3.3 Additional funding

Additional funding was provided to pilot institutions in order to encourage them to adopt best practice. This funding is based on the additional volume of colonoscopies produced each fiscal year, compared

4. In some scenarios of our sensitivity analysis, we include indirect costs.
5. The choice of time horizon (three years after treatment) was driven by the available data period for treated patients.
6. These rates are available on the Statistics Canada site at : www.statcan.gc.ca/tables-tableaux/sum-som/I01/cst01/econ161f-eng.htm
7. This information was available from the Quebec Health Ministry

to the threshold for each facility (the average volume of 2009 and 2010) and a tariff rate calculated from the direct average cost of production at the pilot sites. This unit rate is 50% of the direct average cost for all pilot facilities. The total funding for the two years 2012-2013 and 2013-2014 is equal to 1 088 289$.

3.3.4 Others costs

There are also other recurring costs such as administrative costs, patients costs such us travel costs or opportunity costs (absence from work, etc). This analysis does not include these costs. It also does not include intangible costs such as social or psychological effects such as stress, loss of quality life of the patient and his entourage (pain, suffering, emotional impact, etc.). All of these costs are real but rarely taken into account because of the extreme difficulty of their economic valuation. Costs due to complications (rarely observed in our sample) following screening are not considered either. Table 3.1 presents all the PQDCCR costs on period program implementation in participating facilities.

TABLE 3.1 – PQDCCR costs evaluation

Costs			
	Amount	**Treated patients**	**Per patient**
Fixed costs	3 872 220 $	4 021	963 $
Additional funding	1 089 449 $	2 342	465 $
Total program costs	4 960 509 $	4 021	1 428 $

Source : Authors' computation using Quebec Health Ministry data about PQDCCR costs. All of these costs (total costs for three years of treated patients) are deflated based on the 2011 Canadian dollar. Total treated patients are 4 021 in 2013, 1 181 in 2012 and 1 637 in 2011 and 2010 (since November).

3.4 Benefits estimation

The PQDCCR benefits may be reflected on the treated patients, on the population, on the workers and the whole health system. These benefits include potential gains in labor productivity, life-years gain, gains in terms of healthcare quality and benefits in terms of the disease treatment costs savings. Like costs, we deflated all of benefits based on 2011 consumer price index.

We demonstrated in chapter 2 some of these benefits after program implementation : LOS shortening, less invasisve treatment, lower hospitalization rate and life-years gain. We present in the next sections how we estimated these benefits based on the results of chapter 2.

3.4.1 LOS shortening

The implementation of PQDCCR helped to shorten lengths of stay for hospital episodes following colorectal surgery (see results in chapter 2) on average by 3.4 days. We estimated the cost savings from this reduction in length of stay. The considered treatment costs are only nursing costs [8].

First, we calculated the daily hospitalization cost for each year based in considered costs. For this step, we considered the 2011 average cost of the province (278.05$) [9]. Next, we multiplied the daily cost by the average number of days shortened due to the program implementation for each patient. Finally, we multiplied this average saving by the number of patients treated in the pilot facilities after the reform for the three years. Total benefits due to LOS shortening is equal to 3 745 255 $.

3.4.2 Benefits of less invasive treatment approach

PQDCCR contributed to a greater use of a less invasive approach for the surgical treatment of colorectal cancer : laparoscopic approach. This treatment substitution (between laparotomy and laparoscopy) generated benefits in terms of lower nursing costs as well as benefits for better patient satisfaction. Indeed, despite that this approach requires a longer operation time (thus higher costs in the operating room), it often allows a shorter hospital stay, less anelgesics use and a quicker recovery and return to daily activities. It also produces a more interesting aesthetic results and reduces post-operative pain which is of greater satisfaction to the patient.

To estimate net benefits for this conversion between two approaches, we estimate first benefits in treatment savings as suggested by Alkhamesi and al. [1]. Based on this analysis, we calculated net benefit

8. In some scenarios of our sensitivity analysis, we included other direct costs like drugs, laboratories, radiology, etc.
9. This information is available in the Quebec Health Ministry financial database.

as cost savings in ward and unit care minus additional costs in the operating room (see appendix P for more details about this estimation).

The laparoscopic approach has other benefits in terms of gained days in-hospital and faster patient recovery to normal life. Some studies (such as the CHUQ colorectal surgery guide, 2005) mention a return to office work after 3 to 6 weeks in the case of a laparoscopy and between 4 to 8 weeks in the case of laparotomy. While the return to a physical work is between 4 to 8 weeks in the case of laparoscopic, this is 6 to 8 weeks in the case of laparotomy. We do not estimate these benefits because we concentrated only on benefits for the health system and the government and not for the patient. That's why we also ignored in this study all costs for the patient.

3.4.3 Benefits of smaller hospitalization rate

PQDCCR analysis in chapter 2 showed a decrease in hospitalization rates following the introduction of this reform : only additional funding helped to reduce hospitalization rates by 8% and the clinical strategy variable has no effect on this rate. Benefits are generated by lower treatment costs and also faster return to the labor market. Warren and al. [87] examined trends in the initial phase of cancer treatment and suggested a significant increases in the proportion of chemotherapy treatment. However, we suppose a 50-50 proportion of treatment subtitution beween chimiotherapy and radiotherapy.

To estimate this benefit, we define the number of hospitalizations avoided following the implementation of the program. Then multiply that number by the difference between the average cost of a short hospital stay and the average cost of alternative treatment. Savings for each avoided hospitalization is equal to 7 100.15$ (see appendix Q for more details).

In our case, we had 2 342 hospitalizations for treated patients after additional funding implementation and an average decrease of 8% in this rate. So, all in all, we had 187 saved hospitalizations for the treated facilities. Total savings for two years are equal to 1 297 127$.

3.4.4 Life-Years gain

We concluded in chapter 2 that the PQDCCR program helped to improve the mortality ratio in hospital. However, survival rate out-of-hospital was declined and this result may be explained by short time data after reform. We calculated 132 in-hospital death cases in treated institutions after the reform implementation. As presented in chapter 2, the reform reduced in-hospital mortality rate by 0.3% on average. In other words, the reform avoided no more than on case of death. Because of this little value

effect, in-hospital mortality benefits will be ignored in this analysis.

3.4.5 Other benefits

Other benefits of the program would be difficult to estimate in monetary terms : better aesthetics, lower pain or discomfort, improved quality of life, etc. Benefits such as shorter waiting time to see a doctor or fewer days to be operated could be considered. Often, these benefits may be measured by the willingness of these patients to pay, to avoid the waiting times. These benefits are not included in this analysis because first we do not have information about this willigness to pay for treated patients and second we made the choice of considering costs and benefits for the health system and not population. Also, we think that these benefits will be balanced by intangible costs of the program. So we will consider finally only monetary costs and benefits and ignore other costs and benefits that have no monetary values in this analysis. Table 3.2 presents all the PQDCCR estimated benefits on period program implementation in participating facilities.

TABLE 3.2 – PQDCCR benefits estimation

Benefits		
	LOS (days)	**Per patient**
Shortened length of stay	3 745 255 $	945 $
Hospitalization rate decrease	1 297 127 $	554 $
Less invasive treatment	3 842 810 $	846 $
Total benefits	8 885 191 $	2 345 $

Source : All of these benefits (total benefits for three years of treated patients) are deflated based on the 2011 Canadian dollar. In total we have 4 021 treated patients after clinical reform including 2 342 treated patients after additional funding implementation.

3.5 Results and discussions

We now use program costs and estimated benefits to perform cost-effectiveness analysis of different cancer screening strategies. Basically, we compare costs and the estimated benefits program over a period of time (three years) and compare this with the no screening situation. Usually this information is synthesized in the benefit-cost ratio. We adopt three strategies to be compared : no change(no additional costs or benefits) as reference strategy, the only clinical strategy and the whole reform including additional funding policy. Each strategy has different costs and benefits. Table 3.3 synthesizes benefits and costs for each of these targeted strategies.

TABLE 3.3 – PQDCCR cost-effectiveness evaluation

Costs			
	Treatment	**Funding**	**Total**
Fixed costs (3 years)	3 872 220 $		3 872 220 $
Additional funding (3 years)		1 089 449 $	1 089 449 $
Treated patients (3 years)	4 021	2 342	
Treated patients (8 years)	10 625	10 625	
Total costs	3 872 220 $	1 089 449 $	4 961 669 $
Cost per patient	963 $	465 $	1 428 $
Benefits			
	Treatment	**Funding**	**Total**
Shortened length of stay	3 745 255 $		3 745 255 $
Hospitalization rate decrease		1 297 127 $	1 297 127 $
Less invasive treatment	4 458 109 $	-615 300 $	3 842 810 $
Treated patients (3 years)	4 021	2 342	
Treated patients (8 years)	10 625	10 625	
Total benefits	8 203 364 $	681 827 $	8 885 191 $
Benefit per patient	2 040 $	291 $	2 331 $

Source : Authors' computation using Quebec Health Ministry data about PQDCCR costs.

The benefit-cost ratio represents estimated benefit for each $ invested to implement a given strategy. Considering two strategies A and R, the benefit-cost ratio $T_{Benefit-cost}$ of a given strategy A compared to reference strategy R will be given by equation 3.1

$$T_{Benefit-cost} = (Benefit_A - Benefit_R) / (Cost_A - Cost_R) \tag{3.1}$$

TABLE 3.4 – Cost-benefit ratios of different PQDCCR strategies

Strategies	Costs($) per patient	Benefits($) per patient	Benefit-cost ratio
Reference strategy	0	0	-
Only clinical strategy	963 $	2 054 $	2.13
Clinical and funding strategy	1 428 $	2 345 $	1.64

Different strategies are based on different combinations of costs and estimated benefits as described in the beginning of this section.

Strategy 1 is defined as reference strategy (no reform). Strategy 3 is continuing with the PQDCCR program including additional funding while strategy 2 consists of eliminating this financial incentives and continuing only with the clinical guidelines policy. As seen in table 3.4, strategies 2 and 3 are cost effective since benefit-cost ratio is greater than 1 : 2.13 for strategy 2 and 1.64 for strategy 3. For this last one, for every 1$ invested in this program, including funding budgets, we save 1.64 $ on average. However, with strategy 2, for every 1 $ invested in this program, we save 2.13 $ on average.

Since benefit-cost ratio is greater for strategy two, we conclude that for the study period, policy including funding is no more cost effective than the one with only clinical guidelines. This result is true because funding effect is not significant on LOS as seen in chapter 2 and also funding effect on approach treatment is negative.

3.6 Sensitivity analysis

The sensitivity analysis is a good tool to test the validity of the model. Particularly, we could conclude about the likelihood of model assumptions and consequently the robustness of our results. Indeed, this kind of analysis is used to give some credibility the cost-benefit ratio of our model. In other words, with sensitivity anlysis we may assess the impact that changes in a parameter will have on the model's conclusions.

This consists of varying uncertain parameters or those that are subject to change over time or depending on context. The main candidates for sensitivity analysis are the discount rate, the value of QALY, the horizon of the evaluation of the strategy, unmeasured values, etc.

In our case, we chose to vary three parameters : the discount rate, daily hospitalization cost value and the horizon evaluation. This is because we think that these variables are the most influential on the

final results of our model. Table 3.5 shows results of these variations in these three parameters (see appendix R for more details about this estimation)

TABLE 3.5 – Sensitivity analysis results

	Strategy 2			Strategy 3		
	Costs	**Benefits**	**Ratio**	**Costs**	**Benefits**	**Ratio**
Basic discount rate	3 872 220 $	8 203 364 $	2.13	4 989 200 $	8 885 191 $	1.64
Discount rate 1	3 872 220 $	8 111 281 $	2.09	4 940 823 $	8 780 066 $	1.62
Discount rate 2	3 872 220 $	8 209 199 $	2.14	4 983 344 $	8 994 498 $	1.64
Basic daily cost	3 872 220 $	8 203 364 $	2.13	4 989 200 $	8 885 191 $	1.64
Daily cost 1	3 872 220 $	8 768 418 $	2.26	4 989 200 $	9 425 433 $	1.72
Daily cost 2	3 872 220 $	9 926 814 $	2.56	4 989 200 $	10 634 440 $	1.94
Basic Horizon	3 872 220 $	8 203 364 $	2.13	4 989 200 $	8 885 191 $	1.64
Horizon 1	2 966 856 $	42 523 893 $	14.33	11 525 046 $	47 881 522 $	4.15
Horizon 2	17 627 366 $	73 542 703 $	4.17	57 589 531 $	82 808 426 $	1.44

Source : Authors' computation using Quebec Health Ministry data about PQDCCR costs. Discount rates 1 are equal to 4.3% (2012) and 4.8% (2013) and discount rates 2 are equal to 0.3% (2012) and 0.8%(2013). Daily hospitalization cost 1 is equal to 320$ and daily hospitalization cost 2 is equal to 406$. Horizon evaluation 1 is equal to 20 years and horizon evaluation 2 is equal to 50 years.

Excluding different horizon analysis, cost estimation varies between 3.8 M$ and 4.9 M$ and benefits may vary between 8.1 M$ and 10.6 M$. All in all, the cost-benefit ratio of strategy 2 varies between 2.09 and 2.56 and those of strategy 3 varies between 1.62 and 1.94. Our sensitivity analysis indicated that in all cases, strategy 2 is more cost effective than strategy 3. In this sample,the funding effect provides too little benefit to the health system relatively to the budget spent to cover these financial incentives.

When we extend the analysis to a long horizon, our conclusion is the same : strategy 2 is more effective than strategy 3. Indeed, with a finite analysis, we have only lower values of costs and benefits. The greater ratios in the infinite horizon analysis show that benefits increase quicker than costs in the future and the program will be more efficient.

This sensitivity analysis shows the robusteness of our model evaluation. Parameter changes do not alter the conclusion when comparing strategy 2 with strategy 3.

3.7 Conclusion and limitations

The PQDCCR is a clinical initiative introduced in nine health facilities in November 2010 with the aim to reduce inappropriate colonoscopic exams. It is mainly based on screening tests for colorectal cancer (other than colonoscopy) and rules for clinical practice of colonoscopy exams. A pay-for-performance methodology was introduced in the ensuing months for the participating facilities to encourage behavioural change.

Although colonoscopy exam remains an effective means of detecting this type of cancer, this type of screening strategy has certain limitations such as clinical risk (perforation, hemorrhage, adverse effects of sedation) and perception of the population for this exam.

The implementation of the program required investments but also brought benefits. Certainly estimating the costs and benefits of a screening cancer strategy is relevant to policymakers looking at new options for screening guidelines and treatments. We conducted a cost-effectiveness analysis to estimate the effectiveness of this policy. We compared mainly two scenarios : continuing the PQDCCR program including the pay-for-performance incentive or eliminating this additional funding and continuing only with clinical strategy.

This analysis showed that the benefits of PQDCCR exceed the related costs, even if we do not consider all the possible benefits of the quality program. From a governmental perspective, the PQDCCR screening program is cost-effective compared to no clinical standardization situation. Indeed, the results of this study suggest that the PQDCCR program implemented in selected Quebec health facilities were a cost-beneficial investment of public money : for every dollar invested in the program, society would gain 1.62 $ in the worst case scenario. Our results show a positive net effect. This suggests the program to be continued and even to be expanded in the whole Quebec health network.

Our results are consistent with previous studies (Pignone and al. [66], Flanagan and al. [24]) showing that the cost-effectiveness ratios for new screening strategies compared to no screening were greater than one.

The accuracy of the results of a cost-effectiveness analysis depends on the accuracy with which the estimation of costs and benefits has been made. Thus the plausibility of assumptions depending on the context of the analysis is a basic element in the estimate of net profit. Sensitivity analysis was used to validate the results of this study. It shows that strategy without financial incentives is more effective that the one including additional funding for selected facilities, even with a long horizon analysis.

The generalization of the results of the cost-effectiveness analysis of selected facilities to other facilities can be made for several reasons : similarity in screening costs, similar resources are used in these capacities and similar population-based preferences in relation to these diagnostic services. However, benefits with life-years gain should be reevaluated and included in the cost-effectiveness analysis when sample data will be greater after the reform.

Some key elements to the success of the screening approach in this program are : choosing the best moment of implantation considering several factors and contextual elements. Also the differentiation of the implementation of the strategy by the targeted age group (50 to 74 years old). Finally, it was a clinical-administrative partnership that is very important to help change towards best practice.

This analysis has some limitations. First, for simplifying our calculations, we have not considered inflation in costs of medical equipment, which can be an additional cost to the program. However, this effect can be absorbed by the impact of future income due to reduced treatment costs that we have not adjusted for inflation.

In addition, this analysis does not consider the monetary costs of intangibles. Also, this cost-benefit analysis of the government's perspective may be different in terms of the results of a cost-benefit analysis for society. The ratio calculated in our study is not necessarily the one that maximizes the social surplus which can benefit the population. Finally, this study did not consider other societal effects of colorectal cancer screening on tax contribution and others costs for disabilty programs or social security programs.

The emergence of new screening strategies, the move towards personalized medicine and the sustained increase in treatment costs in relation to screening tests (this can inflate the savings on treatment costs from one year to another) are the main issues for the colorectal cancer screening tests during the next years.

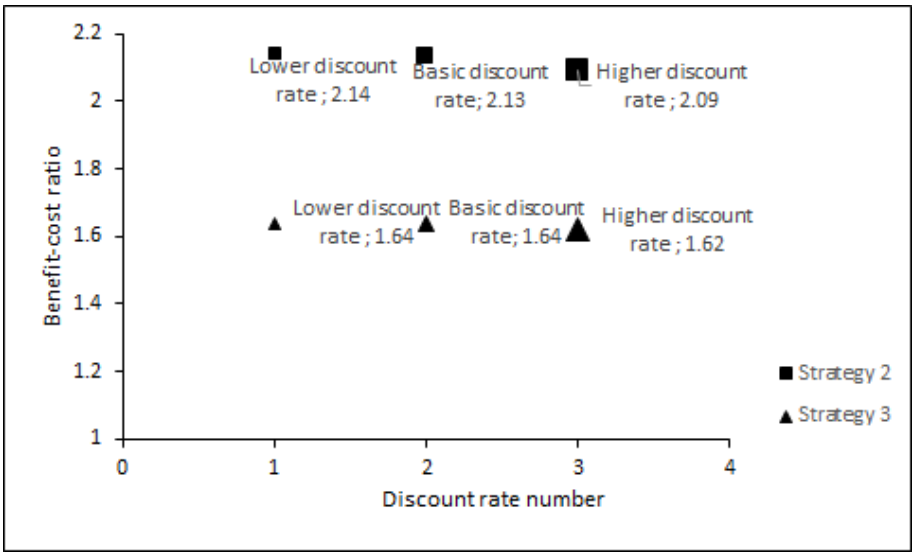
In the case of our sample data, the pay-for-performance scheme included in the PQDCCR program was associated with little or no improvement in outcome indicators (LOS, mortality and treatment costs) despite substantial expenditures. Policymakers should consider other funding strategies for improving healthcare quality and population health and enhance the positive move due to clinical guidelines.

Also cultural differences related to the health system may explain this result. Indeed, in North America and Australia, clinical protocols are established by health professionals without considering constraints and financial incentives. While in Europe, these decisions are highly dependent upon pay-

ment capabilities and financial incentives while bearing in mind the different screening alternatives.

Finally, policy makers should develop new strategies especially those based on clinical protocols of pre and postoperative surgical episodes and designed to standardize and optimize these treatment steps.

FIGURE 3.1 – Sensitivity analysis scenarios



Conclusion

The Quebec health system is experiencing some concerns with facilities performance. An improvement in the latter needs a set of consistent and continuous actions. The funding model or financial incentives in general are one of the strategies for change. Indeed, financial incentives are a lever for performance improvement through a behavioural change.

Based on this theory, we analyze some case studies in the Quebec health system with the aim to assess the real impact of this lever. The Quebec health system is a publicly funded healthcare system where hospital funding and physicians' payments are two separate budgets. On one side, the first one is basically a global budget system and it is managed by the health ministry and the second one is basically a fee-for-services payment and it is managed by RAMQ. So, an increase on the volume activity seems to be more attractive for physicians because the direct relationship the payment and the volume. We analyzed Quebec physicians' behaviour when financial incentive is introduced in the health network. More specifically, we want to study the evolution of physicians behaviour when hospital funding is altered but not their own payments. This is because physicians behaviour is important on the performance system improvement process.

Our results suggest that this lever can play a role in the Quebec health system improvement conditional on certain factors. Indeed , the collaboration of health professionals and the commitment of administrative staff are some of the sucess factors for the implementation of such strategies.

In the first case study of our thesis, we assess for the positive causal effect of funding on waiting times and lengths ofstay in the case of hip and knee replacements. This improvement has not been accompanied by a deterioration in the healthcare quality for these surgeries. However, it would be pertienent to validate whether this improvement is not just a production substitution effect with other types of surgeries especially in the same specialty. On the other hand, we showed that this improvement is increasing with the funding intensity.

In our second case study, the Quebec colorectal cancer screening program, we showed that clinical

guidelines improve the health system productivity (lower hospital length of stay and cost of treatment) and some aspects of population health. When financial incentives were introduced, there was not more improvement in the outcome indicators. This may be due to the short data period after the financial program annoucement but also may reflect the clinical relevance of supporting change through financial leverage.

In the third chapter, we performed a cost- effectiveness analysis of this program. We showed that the continuity and generalization of the program are recommended without an additional funding strategy. An adjustment to the funding methodology could bring better results in the future.

We think that our analysis can be a good beginning for more analysis and studies about physicians behaviour in the Quebec health system. In our thesis we analyzed outcomes such as length of stay, readmissions, in hospital mortality and waiting times. In the next studies, it will be interesting to introduce other indicators such us waiting lists, out-of-hospital mortality, best practice change, etc.

Annexe A

Data treatment

The definition of a waiting time period needs some assumptions. We should choose the start date (decision of the orthopedic treatment, date of 1th consultation of the family physician, the date of access to the waiting list) and the point of the end (the operation date, the discharge date) in order to calculate this duration. The following figure explains the mechanism of access [1] for these surgeries in Quebec :

We have calculated a waiting time for each patient operated for hip or knee replacement. The start date was when the specialist physician decided an episode of treatment for the patient with a surgery. This ignores of course all the waiting times which arising before meeting with a specialist or even before having access to a family physician. The waiting times may differ from one province to another. However the standard for this type of surgery is developed by "Wait Time Alliance for Timely Access to Health Care". These standard are the maximum waiting time medically required :a waiting time of three months to consult an orthopedist and twenty six weeks to be operated.

The link to be made between the surgery date and the date of the visit appears to be complicated, since it may have several visits for the same beneficiary. When the date of the visit comes after the surgery, these visits have been eliminated :these are the control visits post-surgery. Our main assumption is to consider the date of the last visit to the orthopedist in order to calculate the waiting time. However, the decision of the operation may be taken during a previous visit. In order to circumvent this problem,we have retained just the visits with diagnostic related to the knee or hip surgeries. Also, only the elective cases are considered.Thus, a patient who will consult the orthopedist for a carpal tunnel while 3 days after it is was operated for knee replacement following a fracture or an accident is eliminated from our sample.

We have also eliminated the cases of short deadlines (less than 28 days). These waiting times may

1. Source : MSSS, General Access Methodology.

be explained by a monitoring visit before the operation. These cases are not may because we have already eliminated non elective surgeries. Finally, we have eliminated all the waiting time exceeding two years. We have not found explanations for these too long duration. They have been eliminated because their deadlines are non-medically reasonable.

For the BC data, we merged the different database (hospitals, physicians payments and CIHI [2] resources intensities) using a common "studyid" variable. We used the same methodology than Quebec to calculate the waiting time. Following the adjustments mentioned above, the final sample consists of 35 943 observations for Quebec and 35 315 observations for BC.

For length of stay, there is some complication : for BC data, we have common studyid for hospital data and physicians fees data, so we have the length of stay for every surgery. But in Quebec data, we do not have the same studyid for the two database (Med-écho and RAMQ). So, we merged these two database with others variables : year, institution, locality, age, gender and diagnostic.

2. Canadian Institute for Health Information

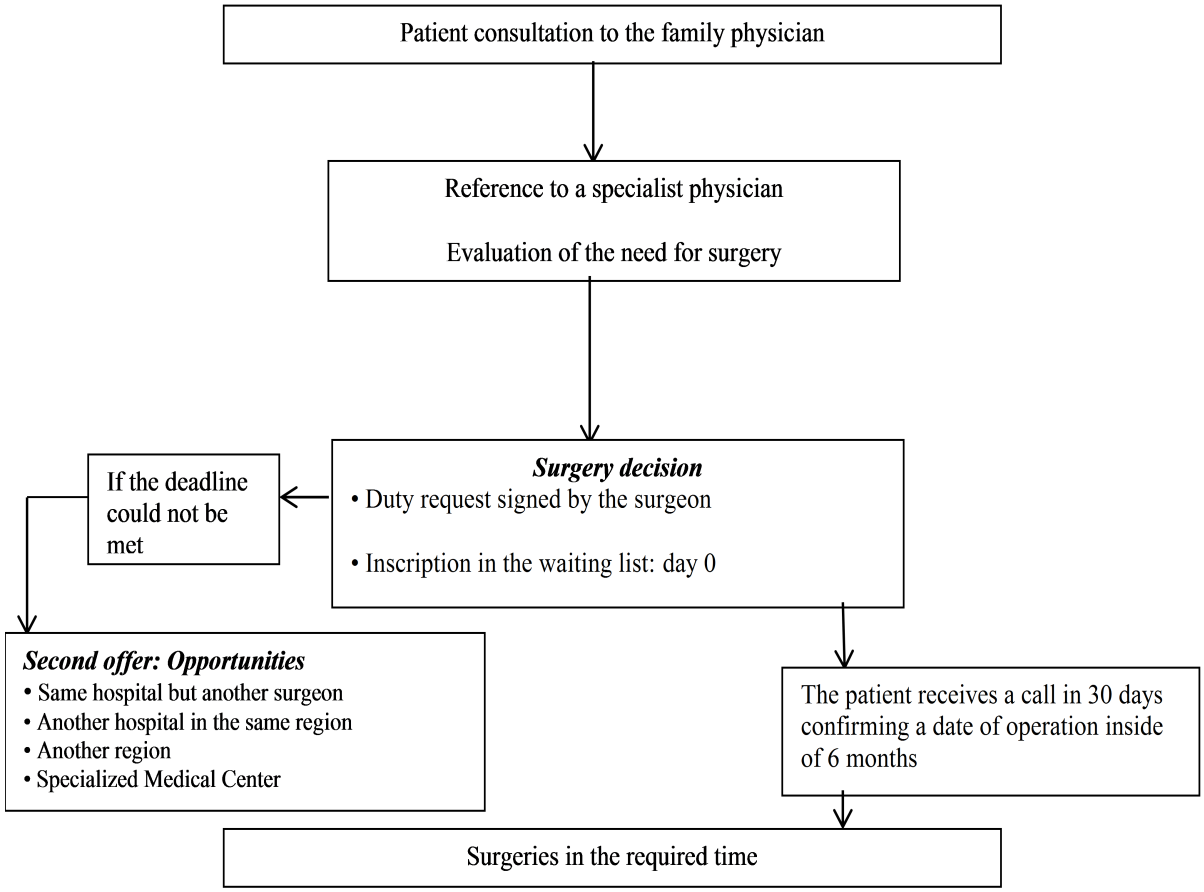


FIGURE A.1 – Access Mechanism to elective surgeries

TABLE A.1 – Variables description

Variable name	Description	Values
Nobenef	Patient ID	Continuous variable
Annee	Surgery year	2001 to 2008
Corporation	Institution code	Continuous variable
Territoire	Institution locality	Continuous variable
Age	Age of the patient	Continuous variable
Sexenum	Gender variable	0 (male) ou 1 (female)
Lits	Number of the hospital's beds	Continuous variable
Pro_65	Proportion of 65 years older persons in the corresponding territory	From 0 to 1
Death	Post-surgery mortality indicator	1 death, 0 else
Distance	Average Distance travelled between home and hospital	Continuous variable
Pro	Province	1 (Quebec), 0 (BC)
NIRRU	Resource Intensity weight	Continuous variable
Con_ETP_ch	Proportion of orthopedic physicians in the hospital	Continuous variable
Catcout	Surgery type	17 (knee) ou 18 (hip)
F	Treatment variable	1 (PAC), 0 (no PAC)
Fina	Additional funding ammount for hip and knee replacements in each hospital every year(M$)	Continuous variable

Annexe B

Endogeneity analysis

TABLE B.1 – Endogeneity analysis with IV estimator

	(1) OLS - Knee	(2) 2SLS - Knee	(3) OLS - Hip	(4) 2SLS - Hip
Age	-0.007	-0.009	-0.008**	-0.010*
	(0.004)	(0.006)	(0.003)	(0.004)
Gender	-0.109	-0.060	0.124	0.071
	(0.093)	(0.125)	(0.078)	(0.114)
Bed's number	0.000***	0.001***	0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Elder proportion	-1.963***		-2.579***	
	(0.519)		(0.583)	
Physician staff	-0.022	-0.001	-0.023	0.007
	(0.017)	(0.023)	(0.019)	(0.027)
Distance from home	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Funding variable	-0.084***	-0.612***	-0.048	-0.700***
	(0.025)	(0.184)	(0.027)	(0.207)
N	558	558	549	549

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source : Authors' computation using hospital health record database. Notes : Standard errors in parentheses. Model 1 is COX model, model 2 is OLS model and model 3 is 2SLS model with Instrumental Variable. In models 2 and 3, we use log(waiting time) as dependent variable. Elder proportion is the IV in columns 2 and 4 estimations.

Annexe C

Before-after waiting time analysis

TABLE C.1 – Before-after waiting time analysis

	(1)	(2)	(3)	(4)
	Model 1 - Knee	Model 1 - Hip	Model 2 - Knee	Model 2 - Hip
Age	0.009***	0.008***	0.007***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
Gender	0.031*	0.010	0.032*	0.017
	(0.014)	(0.017)	(0.015)	(0.017)
Beds' number	-0.000***	-0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Elder proportion	1.585***	0.650**	5.166***	3.165***
	(0.191)	(0.217)	(0.811)	(0.732)
Physicians staff	0.029***	-0.002	0.094***	0.047*
	(0.007)	(0.008)	(0.023)	(0.022)
Distance frome home	-0.000	-0.001***	0.001*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment	0.155***	0.107***	0.073***	0.056**
	(0.015)	(0.018)	(0.017)	(0.019)
Theta			0.049	0.027
LR-test theta			0.000	0.000
Log L	-183509	-125628	-26065	-18710
Number of observations	21025	14918	21025	14918

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source : Authors' computation using hospital health record database. Notes : Standard errors in parentheses. Model 1 is COX model and model 2 is MPH model with Unobserved heterogeneity.

Annexe D

Difference in difference waiting time analysis

Source : Authors' computation using hospital health record database. Notes : Standard errors in parentheses. Model 3 is COX model and Model 4 is MPH model with unobserved heterogeneity.

TABLE D.1 – Difference in difference waiting time analysis

	(1)	(2)	(3)	(4)
	Model 3 - Knee	Model 3 - Hip	Model 4 - Knee	Model 4 - Hip
Age	0.005***	0.006***	0.004***	0.004***
	(0.001)	(0.000)	(0.001)	(0.001)
Gender	0.025*	0.035**	0.026*	0.034**
	(0.010)	(0.012)	(0.010)	(0.012)
Treatment	0.182***	0.121***	0.127***	0.085***
	(0.015)	(0.017)	(0.015)	(0.018)
Province	0.055	0.022	0.044	-0.005
	(0.031)	(0.032)	(0.058)	(0.054)
Trend_2002	0.061	0.086*	0.096*	0.093*
	(0.039)	(0.039)	(0.039)	(0.039)
Trend_2003	0.171***	0.122**	0.180***	0.101**
	(0.038)	(0.039)	(0.038)	(0.039)
Trend_2004	0.124***	0.087*	0.142***	0.109**
	(0.036)	(0.038)	(0.036)	(0.038)
Trend_2005	0.289***	0.182***	0.276***	0.181***
	(0.034)	(0.036)	(0.035)	(0.037)
Trend_2006	0.348***	0.263***	0.294***	0.187***
	(0.033)	(0.035)	(0.033)	(0.036)
Trend_2007	0.485***	0.421***	0.404***	0.328***
	(0.032)	(0.035)	(0.033)	(0.036)
Trend_2008	0.553***	0.513***	0.467***	0.392***
	(0.033)	(0.036)	(0.033)	(0.037)
Theta			0.053	0.040
LR-test theta			0.000	0.000
Log L	-393070	-282453	-51965	-38561
Number of observations	40917	30341	40917	30341

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Annexe E

Before-After length of stay analysis

TABLE E.1 – Before-After length of stay analysis

	(1)	(2)	(3)	(4)
	Model 1 - Knee	Model 2 - Hip	Model 2 - Knee	Model 2 - Hip
Age	-0.014***	-0.015***	-0.007***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
Gender	-0.111***	-0.195***	-0.062***	-0.099***
	(0.011)	(0.015)	(0.011)	(0.015)
Beds' number	0.001***	0.001***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Ressources intensity weight	-1.257***	-1.107***	-0.490***	-0.481***
	(0.017)	(0.022)	(0.013)	(0.016)
Treatment	0.352***	0.406***	0.090***	0.136***
	(0.012)	(0.018)	(0.014)	(0.021)
Theta			0.046	0.034
LR-test theta			0.000	0.000
Log L	-339503	-165993	-38400	-20282
Number of observations	21025	14918	21025	14918

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source : Authors' computation using hospital health record database. Notes : Standard errors in parentheses. Model 1 is COX model and model 2 is MPH model with Unobserved heterogeneity

Annexe F

Difference in difference length of stay analysis

Source : Authors' computation using hospital health record database. Notes : Standard errors in parentheses. Model 3 is COX model and Model 4 is MPH model with unobserved heterogeneity

TABLE F.1 – Difference in difference length of stay analysis

	(1)	(2)	(3)	(4)
	Model 3 - Knee	Model 3 - Hip	Model 4 - Knee	Model 4 - Hip
Age	-0.021***	-0.024***	-0.011***	-0.014***
	(0.000)	(0.000)	(0.000)	(0.000)
Ressources intensity weight	-0.288***	-0.142***	-0.223***	-0.160***
	(0.007)	(0.007)	(0.007)	(0.007)
Treatment	0.286***	0.325***	0.097***	0.129***
	(0.012)	(0.018)	(0.014)	(0.021)
Province	-0.140***	-0.228***	0.037	-0.015
	(0.032)	(0.035)	(0.049)	(0.051)
Trend_2002	0.167***	0.145***	0.117**	0.093*
	(0.039)	(0.039)	(0.039)	(0.039)
Trend_2003	0.335***	0.231***	0.219***	0.130***
	(0.038)	(0.039)	(0.038)	(0.039)
Trend_2004	0.443***	0.346***	0.231***	0.186***
	(0.036)	(0.038)	(0.036)	(0.038)
Trend_2005	0.556***	0.505***	0.313***	0.300***
	(0.034)	(0.036)	(0.035)	(0.037)
Trend_2006	0.780***	0.793***	0.403***	0.385***
	(0.033)	(0.035)	(0.033)	(0.036)
Trend_2007	0.847***	0.789***	0.385***	0.364***
	(0.032)	(0.035)	(0.033)	(0.036)
Trend_2008	0.929***	0.846***	0.438***	0.385***
	(0.033)	(0.036)	(0.033)	(0.037)
Theta			0.031	0.028
LR-test theta			0.000	0.000
Log Ll	-558323	-325891	-61821	-38302
Number of observations	40917	30341	40917	30341

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Annexe G

Competing risks analysis

TABLE G.1 – Competing risks model for post surgery mortality

	(1)	(2)	(3)	(4)
	Model 3 - Knee	Model 3 - Hip	Model 5 - Knee	Model 5 - Hip
Age	-0.021***	-0.024***	-0.021***	-0.024***
	(0.000)	(0.000)	(0.001)	(0.001)
Ressources intensity weight	-0.288***	-0.142***	-0.299***	-0.145***
	(0.007)	(0.007)	(0.010)	(0.008)
Treatment	0.286***	0.325***	0.286***	0.330***
	(0.012)	(0.018)	(0.008)	(0.012)
Province	-0.140***	-0.228***	-0.132***	-0.242***
	(0.032)	(0.035)	(0.035)	(0.028)
Trend_2002	0.167***	0.145***	0.166***	0.145***
	(0.039)	(0.039)	(0.044)	(0.034)
Trend_2003	0.335***	0.231***	0.334***	0.232***
	(0.038)	(0.039)	(0.043)	(0.034)
Trend_2004	0.443***	0.346***	0.446***	0.351***
	(0.036)	(0.038)	(0.044)	(0.036)
Trend_2005	0.556***	0.505***	0.557***	0.506***
	(0.034)	(0.036)	(0.042)	(0.037)
Trend_2006	0.780***	0.793***	0.779***	0.793***
	(0.033)	(0.035)	(0.042)	(0.038)
Trend_2007	0.847***	0.789***	0.851***	0.793***
	(0.032)	(0.035)	(0.045)	(0.039)
Trend_2008	0.929***	0.846***	0.930***	0.849***
	(0.033)	(0.036)	(0.044)	(0.041)
Log L	-557915	-325623	-558323	-325891
Number of observations	40917	30341	40917	30341

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source : Authors' computation using hospital health record database. Notes : Standard errors in parentheses. Model 3 is COX model with one risk and Model 5 is COX model with competing risks.

Annexe H

Data treatment

We used admissions data in Med-Écho database froms 2006 to 2013. We retained only ICD-10(International Classification of Diseases version 10) codes for colorectal cancer diagnostic : C18, C19, C20, C21, C26 [1].

First, we merged these data with mortality database (2006 to 2013) and we obtained 59 926 observations, including 37 516 death cases.

We eliminated all episodes of care with null length of stay or null Resources intensity weight (NIRRU). These cases are usually episodes of mental health or long term care in hospitals. We eliminated too observations for patients treated in psychiatric institutions or rehabilitation centers or other institutions not offering colonoscopies exams. So, we have only cases for hospitals in group 1 to 4 (see below). We have now 59 135 observations.

01 : CH de courte durée 100 lits and moins ;
02 : CH de courte durée 101 à 200 lits ;
03 : CH de courte durée 201 lits and plus ;
04 : CH universitaires ;

Finally, we selected only episodes with inpatient surgeries (eliminated one day surgeries and medicine cases) and eliminated diagnostic interventions (eliminated all episodes with main treatment code beggining with 2). In our final database we have 38 292 observations (colorectal surgeries) equivalent to 35 451 patients treated in 84 institutions for all years included in our sample. This our database used

1. We retained every complete ICD code beggining by one of these short code. This list was defined by many clinical organizations specialized in cancer analysis like Canadian Cancer Society and Cancer Care Ontario. We also validated these codes with Quebec clinicians.

for competing risks model.

In the other hand, the use of msm package for multistate model required specific data preparation. First, we consider both medical and surgical cases. Indeed, many death cases are linked to medical cases so we should not ignore these cases. After that, we prepare data in long format : each patient has a number of observations equal to his transition states. Three important variables are added : date of the observation (state), the observed state and the ID. Observations must be ordered by time within subjects.

Institution variables

Variable	**Description**	**Values**
Annee	Year of the episode	2006 to 2013
Code	Hospital code	Continuous variable
Etab	Institution code	Continuous variable
Territoire	Institution territory CLSC	Continuous variable
Cap_Lit	Number of beds	Continuous variable
Post	Dummy for post clinical treatment	0 or 1
Post2	Dummy for post financial treatment	0 or 1
Ind	Dummy if patient in treated institution	0 or 1
Dxxxx	Dummy for common yearly trend	0 or 1
DTxxxx	Dummy for yearly trend in treatment group	0 or 1
Trend	Dummy if patient in treated institution	Continuous variable
Treat	Clinical treatment variable (Post*Ind)	0 or 1
Fina	Financial treatment variable (Post2*Ind)	0 or 1

	Variable	Description	Values
Patient variables	Dossier	Patient ID	Continuous Variable
	Age_adm	Age of patient when admitted	Continuous Variable
	cod_sexe	Gender of patient	0 (Male) ou 1 (Female)
	DRG	Diagnostic Related Group of patient	Continuous Variable
	NIRRU	Resources intensity weight	Continuous Variable
	SEJTOTDRG	Hospital length of stay	Continuous Variable
	Gravi	Level of clinical severity	1 : Low level 2 : Medium level 3 : High level 4 : Very high level
	Traitor1	CCI code of surgical treatment	CCI : Canadian Classification of Health Interventions
	Typ_dest	Destination code at discharge	01 : centre hospitalier de soins généraux and spécialisés ou centre hospitalier de soins psychiatriques 03 : centre d'hébergement and de soins de longue durée 09 : centre hospitalier de soins de courte durée hors province 13 : centre de réadaptation 17 : CLSC (incluant les dispensaires) 21 : domicile 30 : maison funéraire 31 : départ sans autorisation 33 : médecine de jour 40 : ressources non institutionnelle
	Dt_deces	Date of death	Date
	Li_deces	Death place	Continuous Variable
	Caus_deces	Principal cause of death	Continuous Variable

Annexe I

Resources weights evolution

Yearly NIRRU evolution in the treatment and control groups

Year	Control groupe	Treatment group
2006	2.365	2.316
2007	2.618	2.701
2008	2.681	2.626
2009	2.690	2.620
2010	2.652	2.507
2011	2.547	2.492
2012	2.690	2.554
2013	2.463	2.465
Total	2.602	2.538

Quaterly NIRRU evolution in the treatment and control groups

	Control group				Treatment group			
Quarter	1	2	3	4	1	2	3	4
2006	2.504	2.535	2.348	2.085	2.323	2.471	2.308	2.213
2007	2.666	2.587	2.554	2.650	2.695	2.969	2.512	2.665
2008	2.685	2.690	2.726	2.637	2.835	2.438	2.427	2.761
2009	2.766	2.795	2.645	2.587	2.672	2.569	2.430	2.769
2010	2.624	2.782	2.622	2.605	2.607	2.394	2.532	2.484
2011	2.570	2.575	2.470	2.572	2.463	2.417	2.413	2.614
2012	2.622	2.746	2.762	2.645	2.410	2.534	2.535	2.716
2013	2.470	2.489	2.472	2.428	2.567	2.367	2.484	2.429
Total	2.613	2.660	2.593	2.556	2.562	2.502	2.466	2.601

Annexe J

Mixed proportional model

TABLE J.1 – Mixed proportional model

	(1) Proportional hazard model	(2) Mixed proportional hazard model	(3) Cox model
Age	-0.010***	-0.008***	-0.011***
	(0.000)	(0.000)	(0.000)
Gender	-0.037***	-0.026*	-0.030**
	(0.010)	(0.010)	(0.010)
IND	0.054***	0.078	0.125***
	(0.016)	(0.067)	(0.016)
Clinical severity	-0.535***	-0.526***	-0.850***
	(0.006)	(0.006)	(0.008)
POST	0.103***	0.116***	0.218***
	(0.016)	(0.016)	(0.016)
POST2	0.116***	0.111***	0.217***
	(0.019)	(0.019)	(0.019)
Treatment	0.062**	0.059**	0.079*
	(0.032)	(0.033)	(0.032)
Funding	0.050	0.041	0.037
	(0.037)	(0.038)	(0.037)
Number of beds	-0.000***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)
Typ_prov	-0.027***	-0.015***	-0.007***
	(0.001)	(0.002)	(0.002)
Region	-0.005***	-0.030***	-0.002*
	(0.001)	(0.005)	(0.001)
Codex	-0.131***	-0.126***	-0.213***
	(0.003)	(0.003)	(0.003)
Theta		0.055***	
		(0.013)	
N	38 292	38 292	38 292

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source : Authors' computation using hospital health record database. Notes : Post is dummy for post treatment period and post2 is dummy for post funding period. Ind is an indicator for treatment group. Typ_prov is the facility group where patient is coming and codex is indicator for atypical cases. A higher codex value means a higher atypical case level. Theta is the frailty variance.

TABLE J.2 – Proportional hazard model with frailty - Placebo treatment

	(1) Model estimates
Age	0.010***
	(0.000)
Gender	0.014
	(0.009)
Clinical severity	0.537***
	(0.007)
POST_PLACEBO	-0.023*
	(0.010)
TREAT_PLACEBO	0.007
	(0.015)
IND	-0.181***
	(0.011)
Number of beds	0.000***
	(0.000)
Constant	0.696***
	(0.031)
ln(Theta)	-1.042***
	(0.057)
N	23 638
ll	-20 675

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

POST_PLACEBO is a dummy for post treatment period and IND is dummy for treatement group. So TREAT_PLACEBO is interaction of the last two dummies for the placebo treatment.

Annexe K

Competing risks model

TABLE K.1 – Cox model and competing risks model

	Competing risks model	Cox model
Age	-0.017***	-0.011***
	(0.000)	(0.000)
Gender	-0.046***	-0.030***
	(0.011)	(0.010)
IND	-0.008	0.125***
	(0.015)	(0.016)
Clinical severity	-0.776***	-0.850***
	(0.009)	(0.008)
POST	0.177***	0.218***
	(0.016)	(0.016)
POST2	0.134***	0.217***
	(0.021)	(0.019)
Treatment	0.105**	0.079*
	(0.036)	(0.032)
Funding	0.035	0.037
	(0.049)	(0.037)
N	38 292	38 292
ll	-331 884	-331 838

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE K.2 – Specifications of the treatment effect in competing risks model

	Parallel paths	Common trend	Linear trend	Flexible trend
Age	-0.017***	-0.017***	-0.017***	-0.017***
	(0.000)	(0.000)	(0.000)	(0.000)
Gender	-0.046***	-0.048***	-0.048***	-0.048***
	(0.011)	(0.011)	(0.011)	(0.011)
IND	-0.008	-0.004	-0.043	0.002
	(0.015)	(0.015)	(0.028)	(0.015)
Clinical severity	-0.776***	-0.781***	-0.781***	-0.781***
	(0.009)	(0.009)	(0.009)	(0.009)
POST	0.177***			
	(0.016)			
POST2	0.134***			
	(0.021)			
Treatment	0.105**	0.112**	0.110**	
	(0.036)	(0.034)	(0.034)	
Funding	0.035	0.034	0.031	
	(0.049)	(0.048)	(0.048)	
D2007		-0.126***	-0.129***	-0.126***
		(0.021)	(0.021)	(0.021)
D2008		-0.036	-0.041*	-0.035
		(0.021)	(0.021)	(0.021)
D2009		0.024	0.015	0.024
		(0.021)	(0.021)	(0.021)
D2010		0.120***	0.110***	0.128***
		(0.021)	(0.023)	(0.021)
D2011		0.157***	0.145***	0.163***
		(0.023)	(0.024)	(0.023)
D2012		0.319***	0.306***	0.303***
		(0.024)	(0.026)	(0.025)
D2013		0.261***	0.246***	0.237***
		(0.026)	(0.028)	(0.028)
T			0.035***	
			(0.003)	
T*IND			0.009	
			(0.006)	
DT2011				0.082*
				(0.041)
DT2012				0.183***
				(0.040)
DT2013				0.213***
				(0.046)
N	38 292	38 292	38 292	38 292
ll	-331 884	-331 838	-331 854	-331 825

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Parallel paths : implies equal average outcome change in the two groups, in the absence of treatment. Common trends : implies that in the the absence of treatment, average outcome changes are equal in the two groups for each pre-treatment period. Flexible dynamics : not only linear or polynomial trend form but we can test for possible restrictions on the dynamics. All covariates are included in the model but not shown in this table. T is linear trend standardized at 1 for 2006 year.

Annexe L

Multi-state and multi-episode model

TABLE L.1 – Transition intensities for treated

Initial state	State 1	State 2	State 3	State 4	State 5	State 6
State 1	-0.0898	0.0467	0.0089	0.0259	0.0081	0.0000
State 2	0.0080	-0.0101	0.0000	0.0000	0.0000	0.0021
State 3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
State 4	0.0067	0.0000	0.0000	-0.0084	0.0000	0.0017
State 5	0.0043	0.0000	0.0000	0.0000	-0.0133	0.0089
State 6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

TABLE L.2 – Transition intensities for non treated

Initial state	State 1	State 2	State 3	State 4	State 5	State 6
State 1	-0.0688	0.0368	0.0118	0.0148	0.0052	0.0000
State 2	0.0023	-0.0032	0.0000	0.0000	0.0000	0.0008
State 3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
State 4	0.0022	0.0000	0.0000	-0.0031	0.0000	0.0008
State 5	0.0010	0.0000	0.0000	0.0000	-0.0039	0.0028
State 6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

TABLE L.3 – Mean states duration (days) for treated

	estimates	SE	L	U
State 1	11.1325	0.1820	10.7815	11.4949
State 2	98.0727	5.5522	87.7727	109.5814
State 4	118.0116	9.7669	100.3407	138.7944
State 5	74.9473	7.8139	61.0958	91.9392

TABLE L.4 – Mean states duration (days) for non treated

	estimates	SE	L	U
State 1	14.5335	0.0614	14.4136	14.6543
State 2	308.1555	2.8012	302.7139	313.6950
State 4	319.0078	4.7117	309.9054	328.3776
State 5	256.2186	5.2030	246.2211	266.6220

State 1 : in-hospital ; State 2 : home ; State 3 : in-hospital death ; State 4 : home and CLSC services ; State 5 : other destinations ; State 6 : out-of-hospital death.

Annexe M

Probabilities transition evolution

Transition probability evolution for treatment(Bold line) and control(Dash line) groups in t (days) : hospital discharge

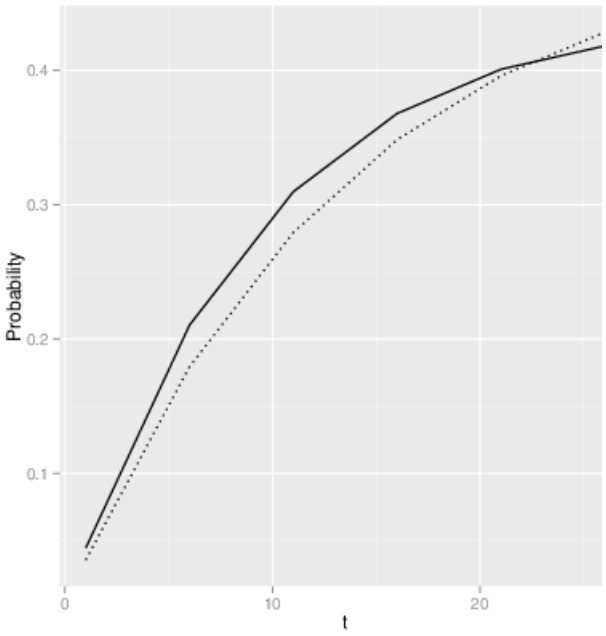


FIGURE M.1 – Discharge at home

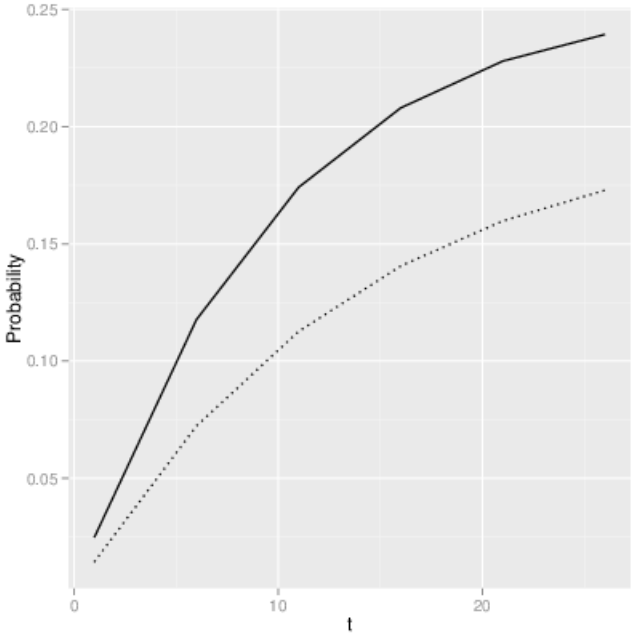


FIGURE M.2 – Discharge at community care

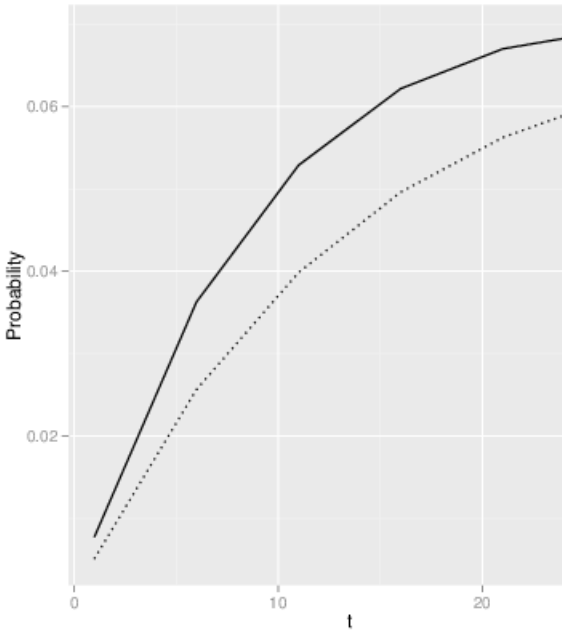


FIGURE M.3 – Discharge at other destinations

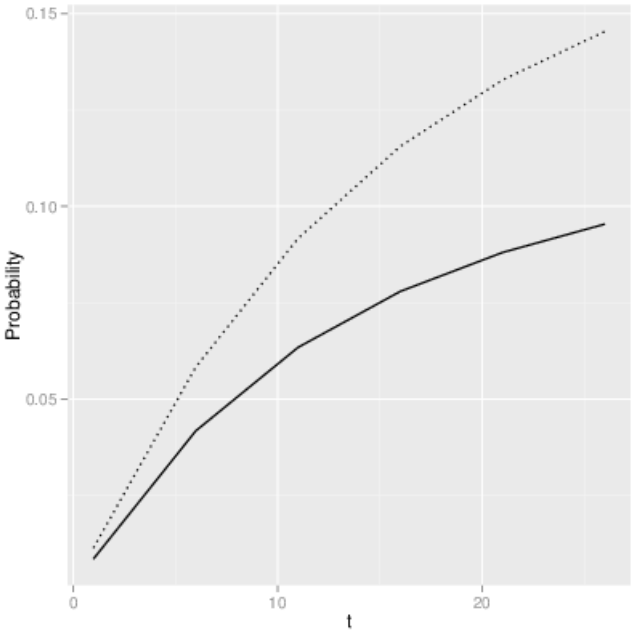


FIGURE M.4 – Hospital death

Transition probability evolution for treatment(Bold line) and control(Dash line) groups in t (days) : out-of-hospital death

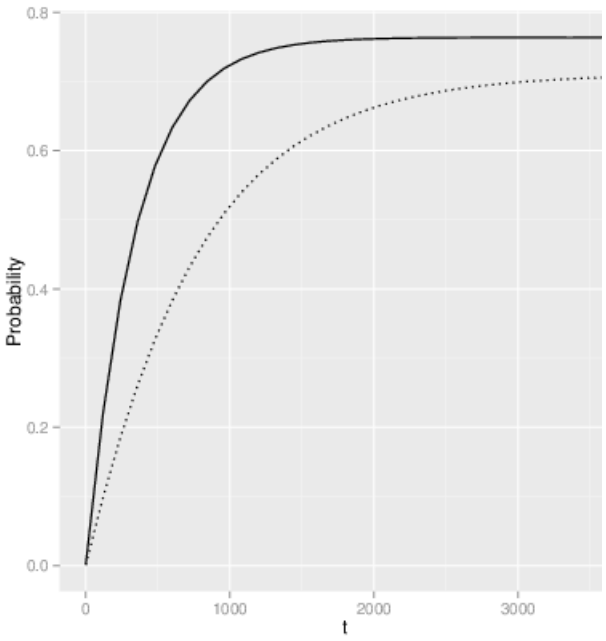


FIGURE M.5 – Death at home

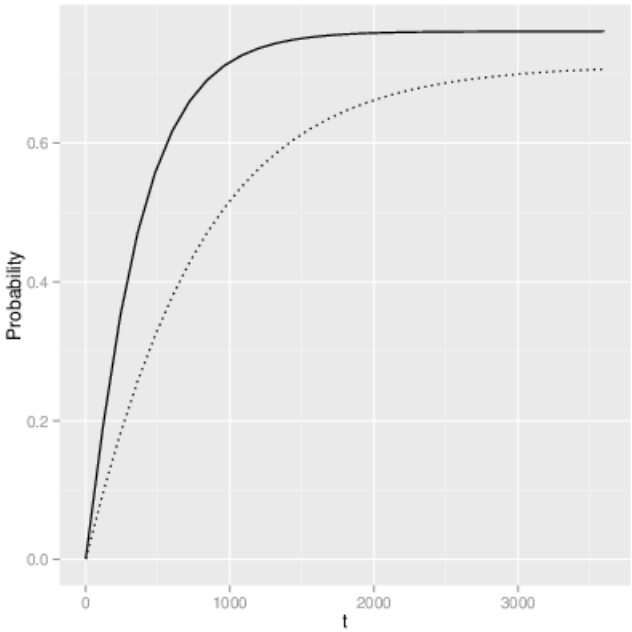


FIGURE M.6 – Death at community care

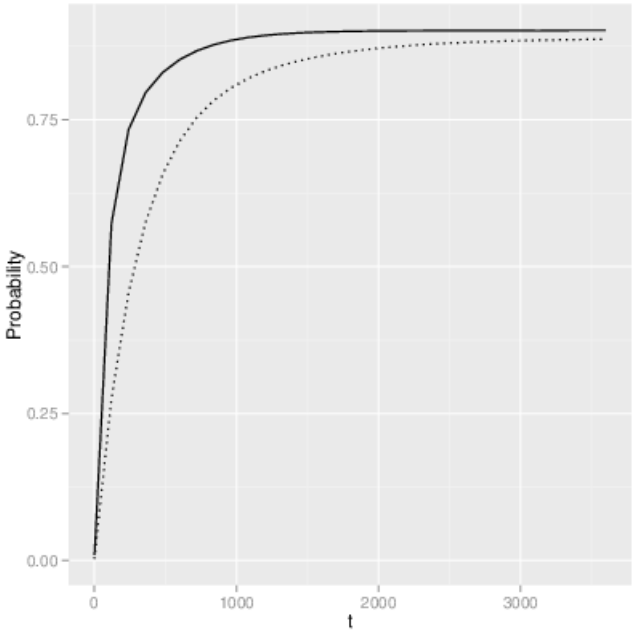


FIGURE M.7 – Death at other destinations

Transition probability evolution for treatment(Bold line) and control(Dash line) groups in t (days) : re-hospitalization

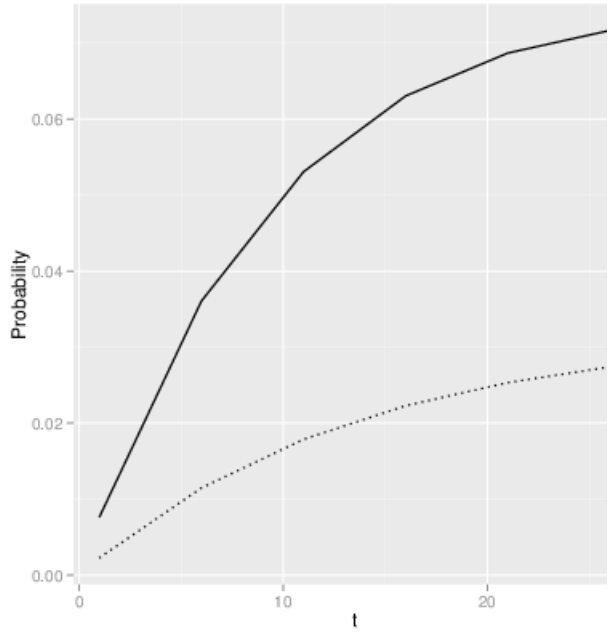


FIGURE M.8 – Admission from home

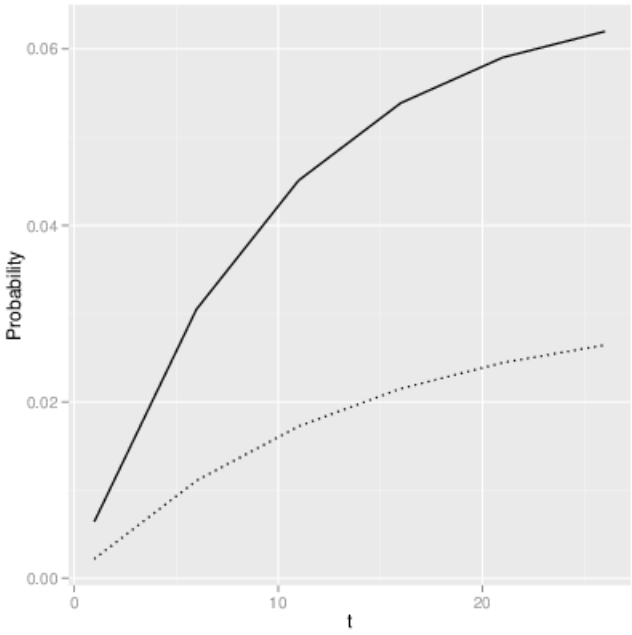


FIGURE M.9 – Admission from community care

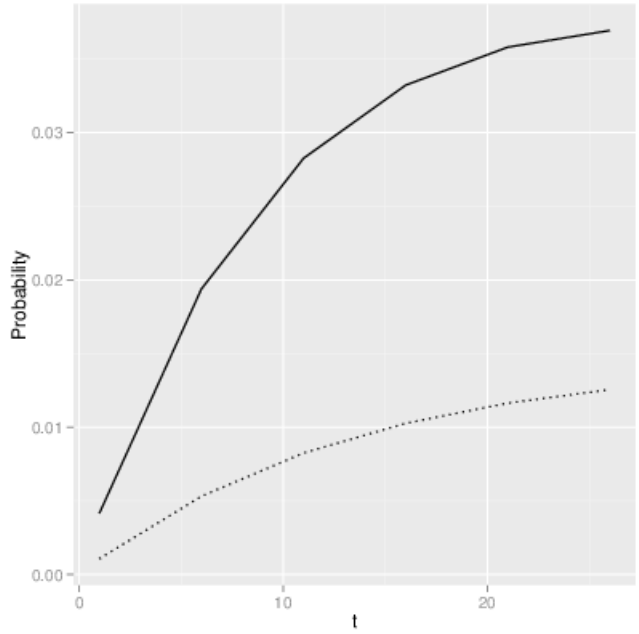


FIGURE M.10 – Admission from other destinations

Annexe N

Prevalence evolution

FIGURE N.1 – States prevalence for control group (blue line for observed values and red line for expected values)

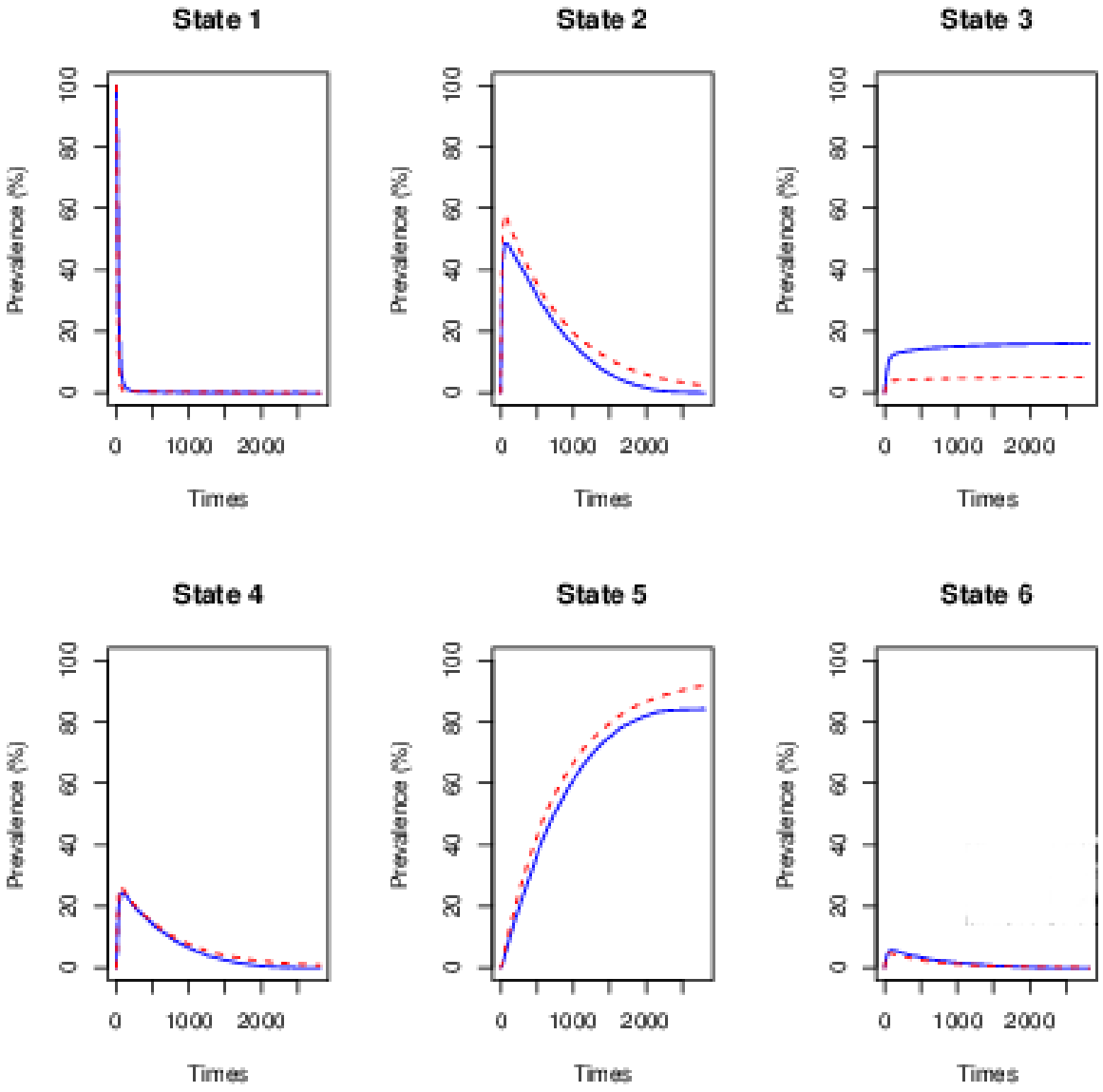
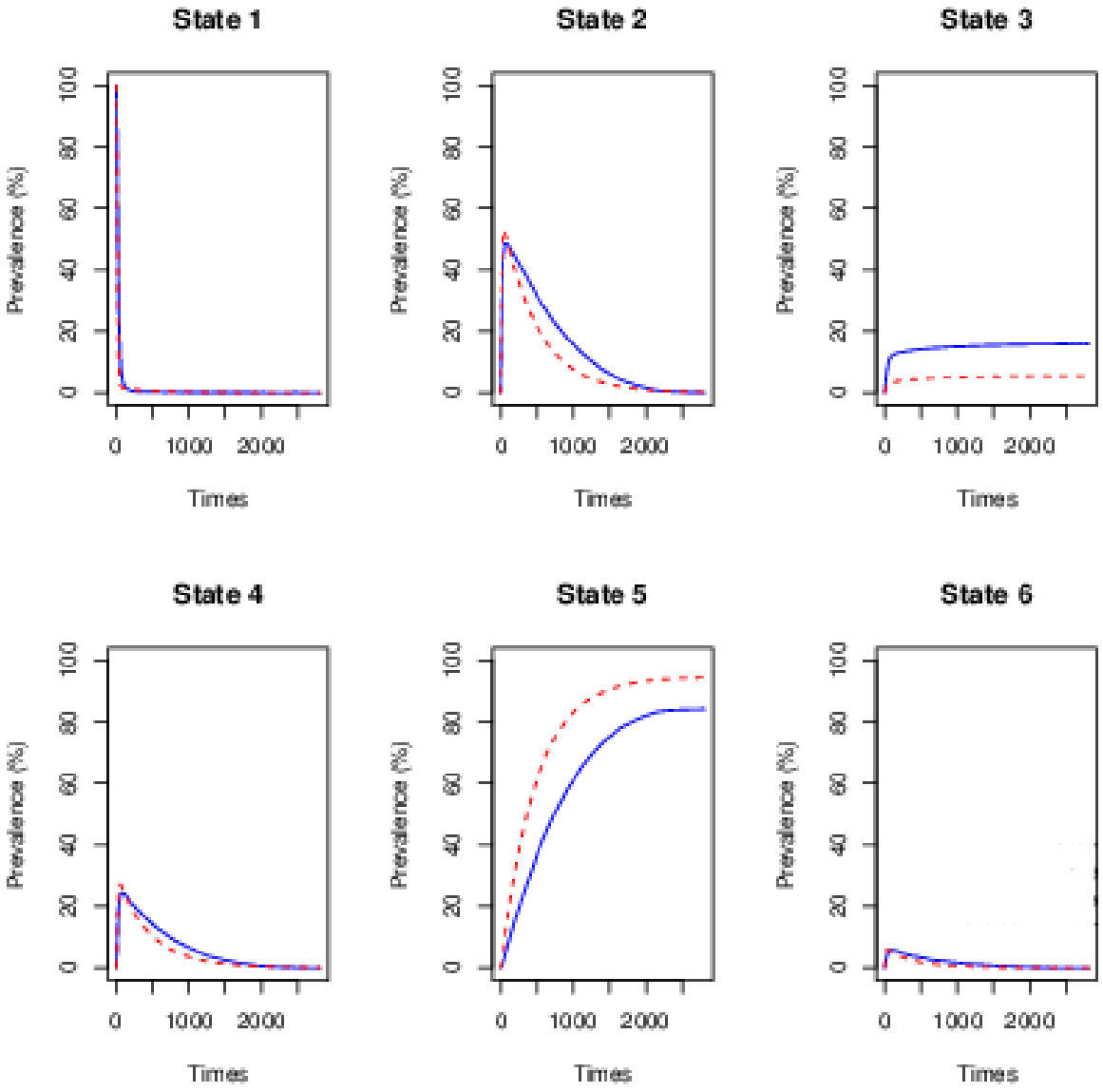


FIGURE N.2 – States prevalence for treatment group (blue line for observed values and red line for expected values)



Annexe O

Reform effect on populational health

TABLE O.1 – Difference in difference regression (Hospitalization rate) - Geographical territory analysis

Variable	Coefficient	Std. Err.
Intercept	-18.890**	2.844
Age	-0.023**	0.006
Gender	13.118**	1.971
Material deprivation	0.000	0.001
Social deprivation	0.004**	0.001
Aboriginal rate	-0.072	0.194
Ind	-0.170*	0.070
Post	-0.080*	0.034
Colonoscopies rate	0.062**	0.001
Treatment	-0.067	0.078
Funding	-0.083**	0.030
N	1293	
R^2	0.793	
$F_{(10,1282)}$	489.916	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Material and social deprivation variables are average rates based on deprivation levels (usually 5 levels for each variable). Aboriginal rate is the rate of aboriginals in each territory. Colonoscopies rate is the rate of colonoscopies exams on a per capita basis and Ind is a dummy for treatment group.

TABLE O.2 – Probit model to test for less invasive treatment - Patient analysis

	(1) Presence of surgery	(2) Approach category
Age	-0.003**	-0.001
	(0.001)	(0.001)
Gender	-0.022	0.044**
	(0.022)	(0.016)
Clinical severity	0.048***	-0.242***
	(0.014)	(0.011)
Post	-0.327***	0.193***
	(0.028)	(0.025)
Ind	0.032	0.213***
	(0.033)	(0.018)
Post2	-1.082***	0.280***
	(0.042)	(0.029)
Treatment	-0.009	0.712***
	(0.053)	(0.041)
Funding	-0.025	-0.170**
	(0.078)	(0.053)
N	22 834	28 497
ll	-8 883	-15 951

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In model (1), outcome variable is a dummy for the presence of surgery in the episode of care for eahc patient : 0 for treatment with surgery and 1 for treatment without surgery. In model (2), outcome variable is the used approach in the episode of care : 0 for laparotomy and 1 for laparoscopy.

Annexe P

Less invasive treatment savings

We considered saving average value proposed by Alkhamesi and al. [1] for right and left colectomies. Costs saving are for ward and care. Additional costs are for operating room. Also, in chapter 2, we demontrated that the clinical protocol increased the use of laparoscopic approach by 71% while additional funding incentives decreases this transition method by -17%. Total reform effect is positive on the use of laparoscopic approach (54%). If we consider 4 021 treated patients in treatment group over the three post treatment years and 2 342 treated patients for only two years, we calculate 2 855 cases of subsitution between the two approches due to treatment variable and -398 cases of substitution due to funding variable. In next table, we present all of these values and the final saving value considered in our case.

Laparoscopy approach savings

Costs			
	Right colectomy	**Left colectomy**	**Mean**
Operating room cost	-781.99$	-1 202.33$	-992.16$
Unit care saving	346.76$	78.84$	212.8$
Ward saving	2 076.75$	2 651.85$	2 364.3$
Total	1 641.52$	1 528.36$	1 584.94$
Substitution cases for treatment			2 855
Savings for treatment			4 525 004$
Substitution cases for funding			-398
Savings for funding			- 630 806$
Total savings			3 894 198$
Total savings after deflation			3 842 810$

Source : Authors' computation using Quebec Health Ministry data about PQDCCR costs.

Annexe Q

Savings in hospitalization costs treatments

Surgery is principal treatment for colorectal cancer. When it is not possible, others treatment can be employed, essentiellay radiotherapy and chiomiotherapy. To estimate saving generated by a decrease in hospitalization rate, we calculate difference between average hospitalization cost for coloreclal surgery and alternative outpatient treatments costs. The average outpatient treatment cost is the average of chimiotherapy cost and radiotherapy cost. We used Quebec financial informations for 2011. The table below shows details of this estimation

Other treatment vs hosptitalization savings

Costs			
	Chimiotherapy	**Radiotherapy**	**Mean**
Unit costs	375.68$	64.80$	
Number of visits/treatments	8	25	
Total outpatient cost	3005.48$	1620.09$	2312.78$
Hospitalization cost			9412.93$
Saving per patient			7100.15$

Source : Authors' computation using Quebec Health Ministry financial data.

Chimiotheray unit cost is based on directs costs including pharmacy costs. These costs are available in functionnal center Hemato-oncology 7060 in accountant Quebec chart. Visits number is calculated in average for all visits related to this functional center. Radiotherapy cost is based also on direct costs. Treatment number is fixed as suggested by Canadian Cancer Society.

To estimate cost for hospitalization with colorectal surgery, we used data on weights for DRG 221

(major small and large bowel procedures). We calculated an average weight for treated patients with this DRG. We then multiplied this average weight by unit cost.

As suugested by Maroun and al. [57], colorectal cancer costs very between 20 000$ and 39 000$ with 60% average proportion due to hospitalization costs.

Annexe R

Sensitivity analysis

The estimated program costs and benefits was deflated based on price consumer index specific to health expenditures. Based on 2011 yer, these discount rates are 2.3% in 2012 and 2.8% in 2013. In this section we assess benefit-cost ratios for two different discount rates : 0.3% and 4.3% in 2012 and 0.8% and 4.8% in 2013.

Costs for different discount rates

	Costs initiaux initiaux	**Additional funding 2012**	**Additional funding 2013**	**Total cost per patient**
	3 872 220 $	608 314 $	508 666 $	
Treated patients	4 021	2 342	2 342	
Rate 1 (2.3%)	-	2.30%	2.80%	
Cost discounted	3 872 220 $	594 637 $	494 812 $	
Cost per patient	963.00 $	253.90 $	211.28 $	1 428 $
Rate 2 (4.3%)	-	4.30%	4.80%	
Cost discounted	3 872 220 $	583 235 $	485 369 $	
Cost per patient	963.00 $	249.03 $	207.25 $	1 419 $
Rate 3 (0.3%)	-	0.30%	0.80%	
Cost discounted	3 872 220 $	606 494 $	504 629 $	
Cost per patient	963.00 $	258.96 $	215.47 $	1 437 $

Benefits for different discount rates

	LOS benefits	Hospital savings	Less invasive approach (T)	Less invasive approach (F)	Total benefit per patient
	3 801 333 $	1 327 728 $	4 525 004 $	(630 806) $	
Treated patients	4021	2342	4021	2342	
Rate 1 (2.3%)	2.3% & 2.8%	2.3% & 2.8%	2.3% & 2.8%	2.3% & 2.8%	
Discounted benefit	3 745 255 $	1 297 127 $	4 458 109 $	(615 300) $	
Benefit per patient	945.37 $	553.85 $	1 108.71 $	(262.72) $	2 345 $
Rate 2 (4.3%)	4.3% & 4.8%	4.3% & 4.8%	4.3% & 4.8%	4.3% & 4.8%	
Discounted benefit	3 703 214 $	1 272 314 $	4 408 067 $	(603 530) $	
Benefit per patient	920.97 $	543.26 $	1 096.26 $	(257.70) $	2 303 $
Rate 3 (0.3%)	0.3% & 0.8%	0.3% & 0.8%	0.3% & 0.8%	0.3% & 0.8%	
Discounted benefit	3 788 967 $	1 322 926 $	4 510 142 $	(627 538) $	
Benefit per patient	942.29 $	564.87 $	1 121.65 $	(267.95) $	2 361 $

We now use different hospitalization costs rates. In the first analysis, we used the cost of 278.05$ for nursing costs. We now replace it by total direct cost of 320$ (including drugs, laboratories, radiology, etc.) and total cost of 406$ including direct and indirect costs. We use also the first scenario discount rate (2.3% & 2.8%).

Benefits for different daily cost for hospitalization

	LOS benefits	**Hospital savings**	**Less invasive approach (T)**	**Less invasive approach (F)**	**Total benefit per patient**
	3 801 333 $	1 327 728 $	4 525 004 $	(630 806) $	
Treated patients	4021	2342	4021	2342	
Daily cost 1	278 $	-	-	-	
Discounted benefit	3 745 255 $	1 297 127 $	4 458 109 $	(615 300) $	
Benefit per patient	945.37 $	553.85 $	1 108.71 $	(262.72) $	2 345 $
Daily cost 2	320 $	-	-	-	
Discounted benefit	4 310 309 $	1 272 314 $	4 458 109 $	(615 300) $	
Benefit per patient	1 071.95 $	543.26 $	1 108.71 $	(262.72) $	2 461 $
Daily cost 3	406 $	-	-	-	
Discounted benefit	5 468 705 $	1 322 926 $	4 458 109 $	(615 300) $	
Benefit per patient	1 360.04 $	564.87 $	1 108.71 $	(262.72) $	2 771 $

Bibliographie

[1] Martin J. Alkhamesi N.A. and Schlachta C.M. Cost-efficiency of laparoscopic versus open colon surgery in a tertiary care center. *Surgical Endoscopy Journal* ., 25(11) :3597–3604, 2011.

[2] Honoré B. and Lleras-Muney A. Bounds in competing risks models and the war on cancer. *Econometrica*, 74(6) :1675–1698, 2006.

[3] Meyer B. Unemployment insurance and unemployment spells. *Econometrica*, 58(4) :757–782, 1990.

[4] Dupuis J.M. Berchi C. and Launoy G. The reasons of general practitioners for promoting colorectal cancer mass screening in france. *Europeen Journal of Health Economics*, 7(2) :91–98, 2006.

[5] Morgenstern H. Berlin T.R., Yu F. and Hurwitz E. Use of a markov transition model to analyse longitudinal low-back pain data. *Statistical Methods in Medical Research*, 12 :321–31, 2003.

[6] Mutevedzi P. Newell M. Bor J., Moscoe E. and Bärnighausenb T. Regression discontinuity designs in epidemiology : Causal inference without randomized trials. *Epidemiology*, 25(5) :729–737, 2014.

[7] Mervin C. and The University of Queensland Australia. Jackson S..How can we improve waiting time for elective surgery in Australian public hospitals ?. School of Economics, Discussion Paper No. 387. 2009.

[8] Propper C. Contingent valuation of time spent on nhs waiting lists. *The Economic Journal*, 100 :193–199, 1990.

[9] Prosser L. Patel I. Erickson S. Bagozzi R. Chang J., Feed G. and Balkrishnan R. Associations between physician financial incentives and the prescribing of anti-asthmatic medications in children in us outpatient settings. *Journal of Child Health Care*, 17(2) :125–137, 2013.

[10] Lindsay C.M. and Feigenbaum B. Rationing by waiting lists. *American Economic Review*, 74(3) :404–417, 1984.

[11] Hawker G. Freund D. Coyte P., Wright J. and Ho E. Waiting times for knee-replacement surgery in the united states and ontario. *The New England Journal Of Medecine*, 331 :1068–1071, 1994.

[12] Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika*, 69 :239–241, 1982.

[13] Devlin N. Dimakou S., Parkin D. and Appleby J. Identifying the impact of government targets on waiting times in the nhs. *Health Care Management Science*, 12(1) :1–10, 2009.

[14] Fullwood C. Doran T. and Gravelle H. Pay-for-performance programs in family practices in the united kingdom. *New England Journal of Medecine*, 355 :375–384, 2006.

[15] Gelber R.D Weinstein M.C Neumann P.J Earle C.C, Tsai J.S and Weeks J.C. Effectiveness of chemotherapy for advanced lung cancer in the elderly : instrumental variable and propensity analysis. *Journal of Clinical Oncology*, 19(4) :1064–70, 2001.

[16] Gelber R.D Weinstein M.C Neumann P.J Earle C.C, Tsai J.S and Weeks J.C. Effectiveness of chemotherapy for advanced lung cancer in the elderly : instrumental variable and propensity analysis. *Journal of Clinical Oncology*, 19(4) :1064–70, 2001.

[17] Scheppach M. Eijkenaar F., Emmert M. and Schoffiski O. The effect of pay for performance in healthcare : A systematic review of systematic reviews. *Health Policy*, 110(2-3) :115–130, 2013.

[18] Sussman J.B. Deshpande A.R. Sussman D.A. Conte M.L. Dwamena B.A. Rogers M.A. Schoenfeld P.S. Inadomi J.M. Saini S.D. Elmunzer B.J., Singal A.G. and Waljee A.K. Comparing the effectiveness of competing tests for reducing colorectal cancer mortality : a network meta-analysis. *Gastrointest Endoscopy.*, 81(3) :700–709, 2015.

[19] Lee E.T. and Wang J.W. *Rationing by waiting lists*. 3 edition, 2003.

[20] Trombold J. Farmer R.W. and McCafferty M. The impact of colorectal cancer screening in a veteran hospital population. *American Surgeon*, 79(3) :296–300, 2013.

[21] Sutton M. Chalkley M. Sussex J. Farrar S., Yi D. and Scott A. *British Medical Journal*, 2009.

[22] Sutton M. Chalkley M. Sussex J. Farrar S., Yi D. and Scott A. Has payment by results affected the way that english hospitals provide care ? difference-in-differences analysis. *British Medical Journal*, 339(b3047), 2009.

[23] Putter H.and Geskus R.B. Fiocco M. Tutorial in biostatistics : competing risks and multi-state models. *Statistics in Medicine*, 26 :2389–2430, 2007.

[24] Berthelot J.M. White K.J. Coombs B.A. Flanagan W.M., Le Petit C. and Jones-McLean E. Potential impact of population-based colorectal cancer screening in canada. *Chronic Disease Cancer.*, 24(4) :81–88, 2003.

[25] Lacroix G. Fortin B. and Drolet S. Welfare benefits and the duration of welfare spells : evidence from a natural experiment in canada. *Journal of Public Economics*, 88(2004) :1495–1520, 2004.

[26] Cullis J. G. and Jones P. R. Rationing by waiting lists : An implication. *American Economic Review*, 76(1) :250–256, 1986.

[27] Mandelblatt J.S. Mitchell J.M. Weeks J.C. Wang Q. Hadley J., Polsky D. and Hwang Y.T. An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population. *Health Economics*, 12(3) :171–186, 2003.

[28] A. K. Han and Hausman J. A. Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, 5(1) :1–28, 1990.

[29] Allignol A. Murthy A. Wolkewitz M. Beyersmann J. Safran E. Schrenzel J. Pittet D. Harbarth S., De Angelis G. Multistate modelling to estimate the excess length of stay associated with meticillin-resistant staphylococcus aureus colonisation and infection in surgical patients. *Journal of Hospital Infection*, 78(2) :86–91, 2011.

[30] Robinson M.H. Moss S.M. Amar S.S. Balfour T.W. James P.D. Hardcastle J.D., Chamberlain J.O. and Mangham C.M. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet*, 348(9040) :1472–1477, 1996.

[31] Navarro-Lozano S. Heckman J.J, Basu A. and Urzua S. Use of instrumental variables in the presence of heterogeneity and self-selection : an application to treatments of breast cancer patients. *Health Economics*, 16(11) :1133–57, 2007.

[32] Navarro-Lozano S. Heckman J.J, Basu A. and Urzua S. Use of instrumental variables in the presence of heterogeneity and self-selection : an application to treatments of breast cancer patients. *Health Economics*, 16(11) :1133–57, 2007.

[33] Myglegaard P. Hougaard P. and Borch-Johnsen K. Heterogeneity models of disease susceptibility, with an application to diabetic nephropathy. *Biometrics*, 50 :1178–88, 1994.

[34] Siciliani L. Iacone F., Martin S. and Smith P. Modelling the dynamics of a public health care system : evidence from time-series data. *Applied Economics*, 44(23) :2955–2968, 2012.

[35] Lin C.Y. Larkin G.N. Rogge J.D. Imperiale T.F., Wagner D.R. and Ransohoff D.F. Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. *New England Journal of Medicine.*, 343(3) :169–174, 2000.

[36] Sempowski I.P. Effectiveness of financial incentives in exchange for rural and under-serviced area return-of-service commitments : systematic review of the literature. *Canadian Journal of Rural Medicine*, 9(20) :82–88, 2004.

[37] Abbring J. and Van den Berg G. The unobserved heterogeneity distribution in duration analysis. *Biometrika*, 94(1) :87–99, 2007.

[38] Heckman J. and Singer B. Econometric duration analysis. *Journal of Econometrics*, 24 :62–132, 1984.

[39] Hurst J. and Siciliani L. Tackling excessive waiting times for elective surgery : a comparison of policies in twelve oecd countries. *OECD Health Working Paper*, (paper number 6), 2003.

[40] Schreyögg J. and Grabka M.M. Copayments for ambulatory care in germany : a natural experiment using a difference-in-difference approach. *The European Journal of Health Economics*, 11(3) :331–341, 2010.

[41] Hanley J.A. Analysis of mortality data from cancer screening studies : looking in the right window. *Epidemiology.*, 16(6) :786–790, 2005.

[42] Hanson K. Whitty J.M. Jackson P., Timothy P.J. and Ansah E.K. Who benefits from free healthcare ? evidence from a randomized experiment in ghana. *Journal of Development Economics*, 107 :305–319, 2014.

[43] Andersen P. Keiding N. and Klein J. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16 :215–224, 1997.

[44] Solomon D.H. Wright J. Kim S., Losina E. and Katz J.N. Effectiveness of clinical pathways for total knee and total hip arthroplasty : literature review. *Journal of Arthroplasty*, 18(1) :69–74, 2003.

[45] Kingma.M. Economic incentive in community nursing : attraction, rejection or indifference ? *Human Resources for Health*, 1(1) :2, 2003.

[46] Mullen K.J. Can you get what you pay for ? pay-for-performance and the quality of healthcare providers. *Journal of Economics*, 41(1) :64, 2010.

[47] Koo S. Quiroz R. Cooper J.M. Paterno M.D. Soukonnikov B. Kucher, N. and S.Z. Goldhaber. Electronic alerts to prevent venous thromboembolism among hospitalized patients. *The New England Journal of Medicine*, 352(10) :965–977, 2005.

[48] Dague L. The effect of medicaid premiums on enrollment : A regression discontinuity approach. *Journal of Health Economics*, 37 :1–12, 2014.

[49] Huang G. Scheiman J. Chernew M. Ladabaum U., Chopra C. and Fendrick A. Aspirin as an adjunct to screening for prevention of sporadic colorectal cancer. a cost-effectiveness analysis. *Ann Intern Med*, 135(9) :769–781, 2001.

[50] Knudsen A.B. Lansdorp-Vogelaar I. and Brenner H. Cost-effectiveness of colorectal cancer screening - an overview. *Best Practice Research Clinical Gastroenterology.*, 24(4) :439–449, 2010.

[51] Bond J.H. Ahnen D.J. Garewal H. Lieberman D.A., Weiss D.G. and Chejfec G. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. veterans affairs cooperative study group 380. *New England Journal of Medicine.*, 343(3) :162–168, 2000.

[52] Remus D. Lindenauer P.K. and Roman S. Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medecine*, 356(5) :486–496, 2007.

[53] Hansagi H. Lofvendahl S., Eckerlund I. and Hanning M. Waiting for orthopaedic surgery : factors associated with waiting times and patients' opinion. *Journal of the International Society for Quality in Health Care*, 17(2) :133–140, 2005.

[54] Baker M. and Melino A. Duration dependence and nonparametric heterogeneity : a monte carlo study. *Journal of Econometrics*, 96(2) :357–393, 2000.

[55] Sarfaty M. and Feng S. Choice of screening modality in a colorectal cancer education and screening program for the uninsured. *Journal of Cancer Education.*, 21(1) :43–49, 2006.

[56] Church T.R. Snover D.C. Bradley G.M. Schuman L.M. Mandel J.S., Bond J.H. and Ederer F. Reducing mortality from colorectal cancer by screening for fecal occult blood. *The Lancet*, 328 :1365–1371, 1993.

[57] Berthelot J.M. Le Petit C. Dahrouge S. Flanagan W.M. Walker H. Maroun J., Ng E. and Evans W.K. Lifetime costs of colon and rectal cancer management in canada. *Chronic Disease Cancer.*, 24(4) :91–101, 2003.

[58] Jacobs R. Martin S., Rice N. and Smith P. The market for elective surgery : Joint estimation of supply and demand. *Journal of Health Economics*, 26(2) :263–285, 2007.

[59] Rosenthal M.B. Early experience with pay-for-performance : From concept to practice. *Journal of the American Medical Association*, 294(14) :1788–1793, 2005.

[60] Rosenthal M.B. What is the empirical basis for paying for quality in health care ? *Medical Care Research and Review*, 63(2) :135, 2006.

[61] Cavallini A. Micieli G. and Quaglini S. Guideline compliance improves stroke outcome. *Stroke*, 33(5) :1341–7, 2002.

[62] Fraeman K. Teerlink J. Salon J. Sterz R. Durtschi A. Mullahy J., Lissovoy G. and Padley R. Hospital costs for treatment of acute heart failure : economic analysis of the revive ii study. *European Journal of Health Economics*, 11(2) :185–93, 2010.

[63] Esmail N. The private cost of public queues. Published on the Fraser Institute site, 2007.

[64] Eibicha P. Understanding the effect of retirement on health : Mechanisms and heterogeneity. *Journal of Health Economics*, 43 :1–12, 2015.

[65] Hurt C. Steele R. Purushotham A. Love S. Williams M. Thomson R. Patkar, V. and J. Fox. Evidence based guidelines and decision support services : a discussion and evaluation in triple assessment of suspected breast cancer. *British Journal of Cancer*, 95 :1490–1496, 2006.

[66] Hoerger T. Pignone M., Saha S. and Mandelblatt J. Cost-effectiveness analyses of colorectal cancer screening : a systematic review for the u.s. *Annals of internal medicine.*, 137(2) :96–104, 2002.

[67] Andersen P.K. and Keiding N. Multi-state ; models for event history analysis. *Statistical Methods Medical Research*, 11 :91–115, 2002.

[68] Welton N.J. Price M.J. and Ades A.E. Parameterization of treatment effects for meta-analysis in multi-state markov models. *Statistics in Medicine*, 30(2) :140–151, 2011.

[69] Whtnall C. Propper C., Sutton M. and Windmeijer F. Incentives and targets in hospital care : Evidence from a natural experiment. *Journal of Public Economics*, 94(3-4) :318–335, 2010.

[70] Mora R. and Reggio I. Treatment effect identification using alternative parallel assumptions. *Economics Working Papers*, 2012.

[71] Rocha R. and Soares R.R. Evaluating the impact of community-based health interventions : evidence from brazil's family health program. *Health Economics*, 19 :126–158, 2010.

[72] Rocha R. and Soares R.R. Evaluating the impact of community-based health interventions : evidence from brazil's family health program. 2010.

[73] Kay R.A. Markov model for analyzing cancer markers and disease states in survival studies. *Biometrics*, 42 :855–865, 1986.

[74] Town R.J. and Gowrisankaran G. Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics*, 18 :747–767, 1999.

[75] Town R.J. and Gowrisankaran G. Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics*, 18 :747–767, 1999.

[76] Onslow E. Clancy M. Powell J. Roberts A. Hughes K. Coulson D. Roberts H.C., Pickering R.M. and Bray J. The effectiveness of implementing a care pathway for femoral neck fracture in older people : a prospective controlled before and after study. *Age Ageing*, 33, 2004.

[77] Pearson J. Ross K., Carter H. and Guess H. Comparative effciency of prostate-specific antigen screening strategies for prostate cancer detection. *J Amer Med Assoc*, 284(11) :1399–1405, 2000.

[78] Globerman S. A policy analysis of hospital waiting lists. *Journal of Policy Analysis and Management*, 10 :247–262, 1991.

[79] Martin S. and Smith P.C. Rationing by waiting lists :an empirical investigation. *Journal of Public Economics*, 71 :141–164, 1999.

[80] Saquib J. Saquib N. and Ioannidis J. Does screening for disease save lives in asymptomatic adults ? systematic review of meta-analyses and randomized trials. *International Journal of Epidemiology*, 44(1) :264–277, 2015.

[81] Grabka M. Schreyogg J. Copayments for ambulatory care in germany : a natural experiment using a difference-in-difference approach. *The European Journal of Health Economics*, 11(3) :331–341, 2010.

[82] Ward E.M. Siegel R.L. and Jemal A. Trends in colorectal cancer incidence rates in the united states by tumor location and stage. *Cancer Epidemiology Biomarkers Prevention.*, 21 :411–416, 2012.

[83] Grossbart S.R. What's the return ? assessing the effect of $pay-for-performance$ initiatives on the quality of care delivery. *Medical Care Reseach Review*, 63(1 suppl) :29 – 48, 2006.

[84] Stratmann T. and Leive A. Do national cancer screening guidelines reduce mortality ? *Journal of Population Economics*, 28(4) :1075–1095, 2015.

[85] Stratmann T. and Leive A. Do national cancer screening guidelines reduce mortality ? *Journal of Population Economics* ., 28(4) :1075–1095, 2015.

[86] Rainer U. Trends in active life expectancy in germany between 1984 and 2003—a cohort analysis with different health indicators. *Journal of Public Health*, 14, 2004.

[87] Yabroff K. R. Warren J. L. and Meekins A. Evaluation of trends in the cost of initial cancer treatment. *Journal of the National Cancer Institute.*, 100(12) :888–987, 2008.

[88] C.M. Young, P.E. et Womeldorph. Colonoscopy for colorectal cancer screening. *Journal of Cancer*, 4(3) :217–226, 2013.