FOUR ESSAYS IN HEALTH ECONOMICS

FOUR ESSAYS IN HEALTH ECONOMICS

By NADINE CHAMI, B.A. (Hons.), M.A.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University

McMaster University DOCTOR OF PHILOSOPHY (2019) Hamilton, Ontario
(Economics)

TITLE: Four Essays in Health Economics

AUTHOR: Nadine Chami, B.A. (Hons.), M.A. (McMaster University)

SUPERVISOR: Dr. Arthur Sweetman

NUMBER OF PAGES: xii, 143

Abstract

This thesis addresses health-policy relevant questions regarding quantity and quality of service delivery in primary healthcare using health administrative data from the province of Ontario. It is comprised of four chapters that explore the following questions: (1) What is the impact of switching from an enhanced fee-for-service (EFFS) payment model to a blended capitation payment model on the specialist referral rates of primary care physicians? (2) What are the rates of inappropriate laboratory testing in the province of Ontario? (3) What are the costs and determinants (physician and practice characteristics) of these inappropriate tests? (4) What is the impact of primary care payment structure on the quantity (number and cost) and the quality (appropriateness) of clinical laboratory testing?

Fee-for-service (FFS) payment systems give physicians an incentive to treat patients on the margin of being referred, whereas in capitation systems physicians do not have a financial incentive to treat such marginal patients. Chapter 1 empirically examines how these two payment systems affect referral rates. The results show an increase in specialist visits upon a switch from an EFFS model to a blended capitation model when the physician is listed as the referring physician in the data, but no change in total specialist visits for these physicians' patients. This change is not observed immediately upon switching payment models. Physicians paid by blended capitation who practice in an interdisciplinary health team have fewer specialist visits per rostered patient compared to EFFS physicians, despite an increase in their patients' specialist visits after joining the interdisciplinary team.

Using a definition of inappropriateness that quantifies ordering clinical laboratory tests too often or too soon following a previous test, Chapter 2 examines the rates of inappropriate laboratory testing for nine selected analytes in Ontario. The chapter finds that the percentage of inappropriate tests ranges from 6% to 20%. Moreover, between 60% and 85% of the time, the physician ordering an inappropriate test is the same physician who ordered the previous test. The findings also show that specialists are more likely than primary care physicians to order repeat tests too soon.

Chapter 3 examines the costs and determinants associated with the rates of inappropriate laboratory utilization. The associated costs of inappropriate/redundant laboratory testing for the selected analytes ranges between 6 – 20% of the total cost of each test.[1] Statistical analyses of the association of physician and practice characteristics with inappropriate testing are done using a logit model. Conditional upon the variables within the model, male physicians, physicians trained outside of Canada, older physicians, and a younger patient population are all shown to be associated with less inappropriate testing. Primary care physicians in group practices and in payment models with pay-for-performance (P4P) incentives are less likely to order inappropriate tests and specialist physicians are twice as likely to order inappropriately compared to FFS primary care physicians. Differences in physician, practice and patient characteristics, however, explain only a small amount of the variation in inappropriate utilization.

Chapter 4 examines how physicians' laboratory test ordering patterns change following a switch from an FFS payment model enhanced with P4P to a blended capitation payment

[1]The cost of each test performed is provided in the Ontario Health Insurance Plan (OHIP) database.

model, and the differences in ordering patterns between traditional staffing and interdisciplinary teams within the blended capitation model. Using a propensity score weighted fixed-effects specification to address selection, the chapter estimates that a mandatory switch to capitation would lead to an average of 3% fewer laboratory requisitions per patient. Patients' laboratory utilization also becomes more concentrated with the rostering physician. More importantly, using diabetes-related laboratory tests as a case study, physicians order 3% fewer inappropriate/redundant tests after joining the blended model and 9% fewer if they joined an interdisciplinary care team within the blended model.

Acknowledgements

I would like to express my deep gratitude for all the help my supervisor Professor Arthur Sweetman has provided throughout my graduate journey, including academic guidance, career advice, and financial support. Soon after finishing my undergraduate degree at McMaster University in 2014, I began working with Professor Sweetman as a research assistant at the Centre for Health Economics and Policy Analysis (CHEPA), which is where I truly realized my passion for research. Professor Sweetman was the first to introduce me to, and encourage, the idea of pursuing a PhD. Thank you for your patience and constant encouragement, and for investing so much of your time in your students and ensuring their future success.

I would also like to thank my thesis committee members, Professor Michael Veall and Professor Katherine Cuff for their advice, support, and feedback. Also, I would like to thank Dr. Andrew C. Don-Wauchope and Dr. Janet E. Simons for their collaboration and providing their clinical expertise in two of my thesis chapters. My thanks also go to Gioia Buckley, a research associate at CHEPA, who dedicated a lot of her time in helping me understand the data used in this thesis.

I am also grateful to the Department of Economics staff for all the support they provide to the program and students. I would like to extend my thanks and appreciation to all the professors in the Department of Economics who have played a role in my journey at McMaster University for the past 7½ years, starting as an undergraduate student in 2011. Moreover, I'd like to thank my peers at McMaster and all my friends who have made this

journey a lot easier.

Finally, I would like to thank my parents and siblings for their continuous love and support, and for always believing in me. I hope I make you proud.

Declaration of Academic Achievement

Chapters 1 and 4 are co-authored with Professor Arthur Sweetman. Chapters 2 and 3 are co-authored with Professor Arthur Sweetman, Dr. Janet E. Simons, and Dr. Andrew C. Don-Wauchope. The material in this dissertation consists of my own research. I conducted all the empirical analysis and was responsible for writing the manuscripts.

Table of Contents

List of Tables

List of Figures

Introduction

The Primary Care Reform in Ontario began in 1999 and launched a series of new primary care payment models in order to increase access and improve the quality and delivery of primary care services across the province (Marchildon and Hutchison, 2016). Among these new models, the enhanced fee-for-service (EFFS) primary care model, known as the Family Health Group (FHG) was first introduced in Ontario in mid-2003, followed by the blended capitation model, known as the Family Health Organization (FHO), in late 2006. Fee-for-service (FFS) models are often criticized for having incentives to over-provide the quantity of services but under-provide the quality of care. Physicians in capitated models are criticized for the under-provision of services, low tolerance of risk and ‘cream-skimming’, which involves the avoidance or refusal to enroll very ill patients, and/or those with complex health conditions (Barros, 2003). Although the EFFS model and blended capitation model are, respectively, predominantly FFS and capitation payment models, they include patient rostering, or enrolment, incentives, bonuses, and pay-for-performance (P4P) incentives targeting particularly preventative care and chronic disease management. The blended capitation model also incorporates FFS payments. McGuire (2011) and Robinson (2001) argue that such blended models of FFS and capitation can correct for issues that arise from pure FFS and pure capitation payment models and therefore may be preferred.

In addition to using new primary care payment models, the province introduced in 2005 an interdisciplinary health care group or practice known as the Family Health Team (FHT). The objective of the FHT was to create a more patient-centered medical home promoting comprehensive and interdisciplinary services such as chronic disease management, counseling, health education and palliative care. Interdisciplinary practices are not payment models, and physicians who join these practices are paid the same as those in a blended capitation model. However, the nature of the practice, such as pre-selection of FHT physicians, greater onsite resources, and opportunity for collaboration between other health professionals in the team, suggests that physicians in interdisciplinary teams might differ from those in non-interdisciplinary practices.

This thesis addresses health-policy relevant questions regarding quantity and quality of service delivery in primary healthcare. It is comprised of four chapters that explore the following questions: (1) What is the impact of switching from an EFFS payment model to a blended capitation payment model on the specialist referral rates of primary care physicians? (2) What are the rates of inappropriate laboratory testing in the province of Ontario? (3) What are the costs and determinants (physician and practice characteristics) of these inappropriate tests? (4) What is the impact of primary care payment structure on the quantity (number and cost) and the quality (appropriateness) of clinical laboratory testing? The thesis contributes to a large existing literature on differences in physician payment structure, especially between FFS and capitation payments. Furthermore, it also introduces a new literature on laboratory test utilization in Ontario and is intended to inform policymakers on the overutilization of laboratory testing in the province.

Theory suggests that FFS payment systems give physicians an incentive to treat patients on the margin of being referred in order to collect a fee for that patient instead of sending them to a specialist; whereas in capitation systems physicians do not have a financial incentive to treat such marginal patients since they will not receive any additional fees for services already included in the capitation basket and there are no penalties for referrals to specialists[2] (Robinson, 2001). Chapter 1 examines the impact of switching from an enhanced fee-for-service (EFFS) payment model to a blended capitation payment model on the specialist referral rates of primary care physicians. It looks at two types of specialist visits: (1) specialist visits associated with a referral from a primary care physician, and (2) all visits to specialists by the primary care physician's patients, regardless of whether or not the referring physician was listed as the referring physician.

One challenge faced in this chapter is that physicians voluntarily choose which payment model to join, potentially creating selection bias. This bias is controlled for by adopting an estimation strategy using a difference-in-differences fixed effect model employing propensity score weights. This technique is sometimes called double robust estimation (Wooldridge, 2010). Non-parametric percentile-t cluster-bootstrap hypothesis testing is performed on the relevant coefficients in all estimations because the generated weights are estimates and not true parameters. The hypothesis testing uses methods of resampling and pivoting (correcting for scale) of the t-statistic (Hall and Wilson, 1991). Inference is based on p-values generated from the bootstrapped t-distribution.

The results show an increase in specialist visits where the blended capitated physician was associated with a referral, but no change in total specialist visits for these physicians' patients. This implies a concentration of patients with their rostering physician, where

[2] Enrolment of patients is required for capitated physicians and they suffer a financial penalty if their patients visit another primary care physician, but there are no financial ramifications for specialist visits.

patients may no longer be seeking referrals from other primary care physicians in the province. This change is not observed immediately upon switching payment models. Blended capitated physicians in an interdisciplinary group have fewer specialist visits per rostered patient compared to EFFS physicians, despite an increase in specialist visits for their rostered patients after joining the interdisciplinary health team.

The rest of the thesis examines a less explored topic in Canadian healthcare, the utilization of medical laboratory testing, specifically inappropriate, or redundant, testing. Laboratory testing is ordered through a primary care gatekeeping system, where laboratory tests cannot be done for patients without the request of a primary care physician. Almost all outpatient laboratory testing in Ontario is done at community medical labs and reimbursed by the provincial government on a per-test basis, capped at an annual maximum (Ndegwa 2011). Studies have shown rapid increase in diagnostic testing in Canada (Naugler, 2014) and initiatives such as "Choosing Wisely" are highlighting the concerns of unnecessary tests, treatments, and procedures that could potentially do more harm than good. An increase in unnecessary testing increases the risk of false positives, resulting in an increase in patient anxiety and unintended patient morbidity, as well as burdening the healthcare system with costs that could be avoidable. "Choosing Wisely" aims to encourage physicians and patients to discuss unnecessary tests and treatment and raise awareness of the issue of overutilization of tests that do not add value to or may negatively impact patient health outcomes. Inappropriate testing can refer to the over- or under-utilization of medical laboratory testing. In this thesis, inappropriate diagnostic testing is defined in terms of the time interval between tests; that is, ordering a test too soon or too often following the previous order of the same test. Recommendations for repeat testing are based on the average effect of the analyte over several days, weeks, or months in the body. Therefore, repeating the test before the recommended interval does not convey accurate or reliable results.

Chapters 2 and 3 identify the rate, costs, and determinants of inappropriate laboratory test utilization of nine analytes in the province of Ontario and explore how physician and practice characteristics are associated with inappropriate diagnostic testing to better understand how primary care drives the quantity and quality of laboratory test utilization. The analytes selected for consideration are thyroid stimulating hormone (TSH), hemoglobin A1c (HbA1c), lipid profile, serum protein electrophoresis (SPEP), immunofixation, quantitative immunoglobulins (QI), Vitamin D, Vitamin B12, and folate. These are selected as they have generally accepted guidelines on the appropriate frequency of measurement and represent either high volume or high-value tests.

Chapter 2 finds that for the nine selected analytes, the percentage of inappropriate tests ranges from 6% to 20%. We also examine the theory that patients are possibly being seen by multiple physicians who do not have the patient's medical history and are therefore ordering the laboratory test without the knowledge of when the previous test was done. However, surprisingly, between 60% and 85% of the time, the ordering physician of an inappropriate test is the same physician who ordered the previous test. Findings also show that specialists are more likely than primary care physicians to order repeat tests too soon.

For the same set of nine analytes, Chapter 3 examines the costs associated with the rates of inappropriate laboratory utilization in Ontario. In line with the magnitude of results found in Chapter 2, the associated cost of inappropriate/redundant laboratory testing for the selected analytes ranges between 6 – 20% of the total cost of each test. In all, the annual cost of inappropriate tests for these nine analytes across the province between 2006 and 2010 is almost 12 million Canadian dollars (2010-dollars), representing 9% of the total cost of these tests.

Chapter 3 also aims to understand selected correlates of inappropriate testing, using measures of physician and practice characteristics. Statistical analyses of the association of physician and practice characteristics with inappropriate testing are done using a logit model. Conditional upon the variables in the model, male physicians, physicians trained outside of Canada, older physicians, and a younger patient population are shown to be associated with less inappropriate testing. Primary care physicians in group practices and in payment models with P4P incentives are less likely to order inappropriate tests and specialist physicians are twice as likely to order inappropriately compared to traditional FFS primary care physicians.

Chapter 4 then examines how physicians' laboratory test ordering patterns change following a switch from an FFS payment model enhanced with P4P to a blended (interdisciplinary and non-interdisciplinary) capitation model. The impacts of primary care payment structure on the quantity and quality of clinical laboratory testing is a topic that, to our knowledge, has not been explored in the health care literature. There is no financial gain or loss to the primary care physician ordering the laboratory test. However, it is assumed that switching to a blended capitation model with a more patient-centered medical home concept and rostering requirements, would lead to a decrease in physician laboratory utilization as patient care becomes more focused on a single practice or with a single primary care physician, and the physician becomes more familiar with their patients' medical history.

Following the estimation strategy from the first chapter, Chapter 4 shows that after a switch from the EFFS to the blended capitation model, physicians are estimated to order around 3% fewer laboratory requisitions per patient. Patients' laboratory utilization also becomes more concentrated with the rostering physician. More importantly, using diabetes-related laboratory tests as a case study, physicians order 3% fewer inappropriate/redundant tests after joining the blended model and 9% fewer if they joined an interdisciplinary care team within the blended model.

References

Barros P. 2003. Cream-skimming, incentives for efficiency and payment system. *Journal of Health Economics* 22(3): 419–443.

Choosing Wisely Canada, (2017). http://www.choosingwiselycanada.org/ (accessed January 10, 2017).

Ferraro, S., and Panteghina, M. 2017. The role of the laboratory in ensuring appropriate test requests. *Clin Biochem*, 50(10–11):555–61.

Hall P, Wilson S.R. 1991. Two guidelines for bootstrap hypothesis testing. *Biometrics* 47(2): 757–762.

Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47, 5–86.

Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, Kerr E. 2015. "Choosing Wisely": A growing international campaign. *BMJ Qual Saf*, 24(2):167–74.

Lippi, G., Brambilla, M., Bonelli, P., Aloe, R., Balestrino, A., and Nardelli, A. 2015. Effectiveness of a computerized alert system based on re-testing intervals for limiting the inappropriateness of laboratory test requests. *Clin Biochem*, 48(16–17):1174–6.

Liu, Z., Abdullah, A., Baskin, L., Lewis, G., Kelter, G., and Naugler, C. 2012. An Intervention to reduce laboratory utilization of referred-out tests. *Lab Med*, 43(5):164–7.

Marchildon, G.P. and Hutchison, B. 2016. Primary care in Ontario, Canada: New proposals after 15 years of reform. *Health Policy*, 120 (7): 732-738.

McGuire T.G. 2011. In The Oxford Handbook of Health Economics, Glied S, Smith P.C (Ed.), Physician agency and payment for primary medical care. Oxford University Press: Oxford: 602-619.

Naugler, C. 2014. A perspective on laboratory utilization management from Canada. *Clin. Chim. Acta*., 427, 142–144. doi:10.1016/j.cca.2013.09.022.

Ndegwa S. 2011. Funding of laboratory testing in Canada. *Canadian Agency for Drugs and Technologies in Health* (*Environmental Scan)*, 19, 1–5.

Robinson J. C. 2001. Theory and practice in the design of physician payment incentives. *The Milbank Quarterly* 79(2): 149–177.

Wooldridge, J.M., 2010. Econometric analysis of cross section and panel data, 2nd ed, MIT Press Books. The MIT Press, Cambridge, MA.

Chapter One

Primary Care Physicians’ Specialist Referral Rates in Ontario: Blended Capitation versus Enhanced Fee-for-Service

1.1 Introduction

This study aims to understand the impact of transitioning from an enhanced FFS (EFFS) payment model to a blended capitation payment model on specialist referral rates and whether a blended model alleviates the incentives associated with “pure” capitation models on the referral behaviour of primary care physicians. Fee-for-service (FFS), capitation with pay-for-performance (P4P), and salary are the most popular physician payment models in most countries (Robinson, 2001; Li et al., 2011). Identifying changes in referral patterns upon switching payment models is important in identifying the incentives each payment model has on the referral behaviour of physicians. Physicians in an FFS payment system have an incentive to treat patients that are considered on the margin of being referred and collect a fee for that service rather than sending those patients to a specialist. On the other hand, capitated physicians have an incentive to send marginal patients to a specialist rather than performing a service that is already included in the capitated fee where the physician does not receive any additional payment for the service rendered. Robinson (2001, 149) argues that FFS encourages “‘ping-pong’ referrals among specialists" while capitation is "a narrow scope of practice that refers out every time-consuming patient".

FFS models are often criticized for having incentives to over-provide the quantity of services but under-provide the quality of care. Empirical evidence shows that FFS physicians provide significantly more services (Hennig-Schmidt et al., 2011; Kralj and Kantarevic, 2013), and more patient visits (Gosden, 2000; Kralj and Kantarevic, 2013; Shimmura, 1988), than those in capitated models. On the other hand, physicians in capitated models are criticized for the under-provision of services, low tolerance of risk and 'cream-skimming', which involves the avoidance or refusal to enroll very ill patients, and/or those with complex health conditions (Barros, 2003). Hence, it is often argued that physicians in a FFS model are more willing to treat more time-consuming patients since they can charge for numerous services. Capitation does not depend on the number of services provided, and therefore incentivizes physicians to focus on preventative care and health education for patients, which suggests an overall increase in quality of care provided to patients if the physician assumes that the patient will remain in his/her roster for a considerable amount of time.

McGuire (2011) provides theoretical and empirical evidence showing that a blended capitation payment system can correct for issues associated with pure FFS and pure capitation payment systems; and suggests that enrolment requirements can create incentives to accept the treatment of patients with complex health conditions. Similarly, Robinson (2001) argues that the blending of elements of FFS and capitation outperform the payment methods in their non-optimal pure forms. Findings of Kantarevic and Kralj

(2016) imply that the best approach is to offer physicians a menu of contracts tailored to physician type, instead of a single 'optimal' contract. Therefore, the effect of switching from FFS to capitation on physician referral rates may be less pronounced, or even avoided, in a blended capitation payment model compared to pure capitation.

This study focuses on a distinction between two types of specialist visits that to our knowledge has been unexamined in the literature: (1) specialist visits associated with a referral from a primary care physician, and (2) all visits to specialists by the primary care physician's patients. The results show an increase in specialist visits where the blended capitated physician was associated with a referral, but no change in the total number of specialist visits for these physicians' patients, whether or not the visit was associated with the rostered physician. This change is not observed immediately upon switching payment models. Blended capitated physicians in an interdisciplinary group have fewer specialist visits per rostered patient compared to EFFS physicians, despite an increase in specialist visits after joining the interdisciplinary health team. These results suggest a concentration of patient services with the rostering physician and blended capitation seems to be successfully reducing the incentive for specialist referrals inherent in pure capitation.

The rest of the paper is organized as follows. A brief institutional background is presented in Section 1.2, and Section 1.3 presents a literature review. Section 1.4 describes the type of referral used in the study. The data are discussed in Section 1.5, and Section 1.6 describes the empirical strategy used. Section 1.7 presents the results and Section 1.8 concludes.

1.2 Institutional Background

The Canadian health care system is publicly funded and offers universal coverage to all its citizens and permanent residents. Each province and territory is responsible for the administration and delivery of health care services, following guidelines set by the federal government. Primary care physicians self-select their method of remuneration given different payment models designed by the provincial government. In order to obtain the benefits while avoiding problems associated with pure FFS and pure capitation, the Canadian province of Ontario modified its pure FFS models to incorporate P4P, and then introduced blended capitation models that incorporate capitation with a FFS component and P4P as part of a major primary healthcare reform established in 1996.

In July 2003, an EFFS model, known as the Family Health Group (FHG), was introduced. The main method of remuneration is FFS and although enrolling, or the rostering of, patients is not required of FHG physicians, it is encouraged through incentives that include a comprehensive-care premium. The blended capitation model, Family Health Organization (FHO), was introduced on November 1st 2006. The Ontario blended capitation models are predominantly capitation and sometimes referred to as “doubly blended” since they include (1) a “capitation basket” of most common services provided to enrolled patients and a FFS component for services outside the capitation basket, and (2) blended FFS and capitation for services inside the capitation basket, where the FFS component is in the form of a shadow-billing premium of 10% (increased to 15% in 2010) initially used as an incentive for physicians to submit billing claims for capitated services (Sweetman and Buckley, 2014). Physicians additionally receive FFS payments for services for non-enrolled patients.[3] The capitation rate is calculated by multiplying the base rate with a specific age-sex ratio.[4] Enrolment of patients is required for capitated physicians and they suffer a financial penalty if their patients visit another primary care physician, but there are no financial ramifications for specialist visits.[5] Both payment models require group practices with a minimum of three physicians, after-hours care obligations and P4P incentives targeting, in particular, preventative care and chronic disease management.

An interdisciplinary health care group known as the Family Health Team (FHT), was established in April 2005, and is described by Hutchison et al. (2011, 266) as "the provincial government's flagship initiative in primary health care renewal", promoting comprehensive and interdisciplinary services such as chronic disease management, counseling, health education and palliative care. FHTs are closely related to the concept of a Patient-Centered Medical Home (PCMH) which have become increasingly popular in the United States (Kralj and Kantarevic, 2013). They consist of a team of family physicians, nurses, and interdisciplinary health professionals that may include dieticians, psychologists, pharmacists, social workers and others. FHTs are not a physician-payment model, but a selected group of capitated and salary-based primary care physicians in

[3] These services are capped at $52,883 at the time of the study.
[4] The base rate is equal to $139.12 and there are 38 age-sex specific ratios that vary between 0.44 and 3.57.
[5] Enrolment requirement of 2,400, 3,200 or 4,000 patients for groups of three, four or five FHO physicians, respectively (Sweetman and Buckley, 2014).

Ontario.[6] This can imply that FHO physicians who join a FHT are significantly different than non-interdisciplinary FHO physicians. Liddy et al. (2014, 2) suggest that interdisciplinary capitated physicians are likely to have lower referral rates than non-interdisciplinary capitated physicians "as providers in this [interdisciplinary capitated] model have greater onsite resources and opportunity for collaboration than non-interdisciplinary practices".

1.3 Literature Review

Referral patterns are clearly affected by the remuneration scheme. However, results from the literature are contrasting (Gosden et al., 2000). Iversen and Luras (2000) find that the referral rates of general practitioners (GPs) are lower for those in a fixed-salary payment system combined with FFS payments than those in a capitation payment system with a reduced FFS component. Krasnik et al. (1990) find that, upon the introduction of a partial FFS payment method to a salary payment method, an increase in the provision of services by GPs results in a decrease in referral rates. On the other hand, Davidson et al. (1992) find that referrals to specialists for children in the Medicaid program in the United States decreased for capitation physicians compared to FFS physicians. On the one hand, Kralj and Kantarevic (2013) find, in contrast with our results, that physicians in the blended capitation FHO model have about 3% fewer referrals per rostered patient than physicians in the EFFS FHG model. On the other hand, Sarma et al. (2018) find that switching from the blended FFS FHG model to the blended capitation FHO model increases referrals to

[6] Unlike any physician that joins a payment model of their choice, physicians who wish to join a FHT are required to submit an application to the Ontario Ministry of Health and Long-Term Care (MOHLTC) and are either accepted or rejected.

specialists by about 5% to 7% per annum. However, the distinction of this study compared to the Kralj and Kantarevic (2013) and Sarma et al. (2018) papers is that our study includes different measures of outcome, controlling for an association with an interdisciplinary health team, distinguishing between individual-level and system-level specialist visits, employing the method of virtual rostering, and identifying when the referral was made with respect to the time of the switch.

In addition to the differences listed above between our study and the one done by Kralj and Kantarevic (2013) and Sarma et al. (2018), the most important one that we believe explains the contrasting results in physician-level referrals comes from different patient populations. Kralj and Kantarevic's study includes all rostered patients of the physicians in the study period, whereas this study includes only continuously rostered patients for the identified physicians. That is, patients that are formally rostered with the same primary care physician each and every year of the sample period. By looking at only continuously rostered patients, the continuity of care is similar for all these patients, and the physician is familiar with the patient's medical record before and after switching from the EFFS model to the blended capitation model. This continuity of care creates a more exclusive GP-patient relationship compared to part-timed rostered patients, which may affect referral decisions. Focussing only on continuously rostered patients eliminates the problems associated with attrition, patients the roster with more than one physician, and enables us to identify the physician who most likely initiated the referral for patients' specialist visits that do not have a listed referring physician on record. On the other hand, there could be selection bias associated with patients that remain continuously in one roster. An advantage in the study done by Sarma et al. (2018), although similar in study design, is that their study includes two extra years of data after the switch (up to 2013 compared to 2011 in our study) estimating a longer-term impact of the switch on the specialist referral rate, and may explain the difference in results upon switching from the EFFS model to blended capitation.

Limitations of some of the studies include small sample sizes, not controlling for selection bias or cream-skimming, and self-reported referral rates. Additionally, our data is panel, following the same set of physicians (either switching to blended capitation or remaining in EFFS) across 5 years, while other studies use cross-sectional data. Differences among studies, such as payment model features, techniques applied, and issues tackled, need to be taken into consideration when comparing results. Li et al. (2011), for example, find that primary care physicians in Ontario do not respond to financial incentives for all services. This suggests either that the large number and/or small value of these incentives have little effect on the behaviour of primary care physicians, since physicians may take several years to adapt to a new complex system. Factors other than the remuneration scheme should also be considered when analyzing specialist referral rates. Studies find that the variation of GP referral rates is attributable to patient and physician practice factors (Franks et al., 2000). Similarly, O'Donnell (2000) finds that factors affecting the variation of GP referral rates include patient, physician and practice characteristics and access to specialty care. In fact, Hackl et al. (2015) find that more patients were referred by their physicians to specialists that were within their personal network, especially to high-quality specialists, than to those outside their network. Iverson et al. (2005) find that physicians with practices located in areas with a larger population have higher referral rates compared to those located in less populated areas.

Interdisciplinary teams are also shown to have an effect on physician referral rates. Liddy et al. (2014) find in their two-year cross-sectional study that physicians in an interdisciplinary blended capitated model in Ontario have a 3.5% lower referral rate than physicians in a non-interdisciplinary capitated model; and physicians in an EFFS model are found to have lower referral rates than both interdisciplinary and non-interdisciplinary capitated models. Although this study does not examine changes in referral patterns upon a change in payment models, it shows that primary care payment models are significantly associated with referral rates. Additionally, Chung et al. (2010) find that primary care physicians in medical practices with nurse practitioners and physician assistants (NP-PA), which has a higher capitation rate than practices without NP-PA, refer less than primary care physicians in practices without NP-PA.

1.4 Definition of a Referral

1.4.1 Gatekeeping System

Since the first point of contact for non-emergency health care is with a primary care physician, unnecessary and costly referrals to specialists are assumed to be reduced in a gatekeeping system (Scott and Jan, 2011). Canada, among other OECD countries, has a gatekeeping system where referrals to specialists must be, or are normally, made by a GP. Fertig et al. (1993) find that the variation in referral rates to hospitals among GPs is not found to be explained by unnecessary referrals, but find it more difficult to address whether a specialist referral is appropriate. O'Donnell (2000, 469) states that "there is a lack of consensus about what constitutes an appropriate referral, and the use of guidelines has had only limited success in altering referral behaviour". However, risk averse or less-experienced physicians may refer 'excessively' to confirm a diagnosis or for a second opinion. In fact, Forrest et al. (2006) find a significant positive association between risk aversion and higher primary care physician referral rates and an increase in the risk of a referral being made by a capitated primary care physician.

Although minimal, there may be referrals from one GP to another. For example, according to The College of Family Physicians in Canada, most referred-to GPs are special-interest or focused-practice family physicians.[7] Referrals to GP specialists are not included in our analysis.[8] Additionally, specialist visits where the same specialist was associated with the referral are not included.[9]

1.4.2 Wait Times to See a Specialist

A caveat in this model is that the time the referral was made by the primary care physician is not known. Instead, we observe the date the patient saw the specialist and the primary care physician associated with the referral. Therefore, it is possible that the physician could have switched models during the patient's wait time to see a specialist. Median wait times in Ontario vary greatly by specialty, ranging between 28 days for

[7] Some family physicians in Canada focus their practice in areas such as psychotherapy, obstetrics and gynaecology, sports medicine, and dermatology (The College of Family Physicians in Canada) http://www.cfpc.ca/CPFM/

[8] Prior to 2009, physicians practicing in a harmonized model in Ontario received an impact on their access bonus when referring to a focused-practice GP. These physicians became eligible for an exemption from the access bonus impact in 2009. For more information please refer to the 2008 Physician Services Agreement at https://www.oha.com/CurrentIssues/keyinitiatives/PhysicianandProfessionalIssues/Documents/2008%20Physician%20Services%20Agreement.pdf.

[9] The specialist would also have a specialty in family medicine and is therefore able to refer patients to himself/herself.

pediatrics and 103 days for neurosurgery (Thind et al., 2012). To control for this lag, visits made in the first 1.5 years after the time of the switch and visits made the following 2.5 years are identified. Therefore, the first 1.5 years of the switch is considered a "transition" year and may not reflect the true behaviour of physicians who switch models.

1.4.3 Individual-Level and System-Level Specialist Visit

Unlike previous studies in the literature, we distinguish between two types of specialist visits. An individual-level specialist visit, which we name "listed referral", includes those where the referring physician was listed in the data as the physician who initiated the referral for the patient. Due to gatekeeping practices in Ontario, a specialist visit should be associated with a referring physician, but at times the referring physician is not listed. System-level specialist visits include all patient visits with a specialist, regardless of whether the referring physician is listed or not.[10] It is important to distinguish between these two types of specialist visits in order to identify whether the physician is listed as the referring physician more or less often upon joining the blended capitation model, or if the total number of specialist visits by the patient within the health care system is changing. An increase in individual-level, or listed referrals but no change in the number of total specialist visits suggests a decrease in referrals elsewhere in the system, primarily from GPs other than the patients' rostering GP. This could indicate a shift of services with the rostering physician due to enrolment requirements and physician incentives of the blended capitation model. As mentioned earlier, we identify all specialist visits as being

[10] In the Ontario Health Insurance Plan claims database, the physician who performed the service is identified as the "attending physician" and the physician who initiated a referral, if applicable, is identified as the "referring physician".

initiated by the rostering physician since all patients are continuously rostered with the same physician throughout the entire sample period.

1.5 Data

1.5.1 Data Sources and Sample of Physicians in the Analysis

The Ontario Health Insurance Plan (OHIP) database includes referrals made by all physicians in Ontario who submit billing claims, which includes essentially all relevant physicians in the province. The unit of analysis is the primary care physician. All physicians in the sample are affiliated with the EFFS model as of April 1st 2006 and either remain in an EFFS model or switch to the blended capitation model by the end of the sample period, March 31st 2011.[11] The analysis focuses on referrals for patients that are formally and virtually rostered with the same primary care physician each and every year of the sample period. This allows us to identify the total number of specialist visits for rostered patients and prevents issues that arise from part-time rostering and rostering with more than one physician such as identifying which physician made the referral when a primary care physician is not associated with a patient's specialist visit.

Our sample of physicians includes 3,101 physicians who formally roster their patients. Approximately 44% of these physicians switched to the blended capitation model by the end of the sample period, of which 11% were also affiliated with an interdisciplinary health team. The referral rate of a physician for each year is determined by the number of referrals divided by the number of continuously rostered patients. All physicians in the

[11] Physicians that switch from a FHG to a FHO after more than 3 days of leaving a FHG are excluded from the sample to include only physicians practicing for the entire sample period. Physicians that switch from a FHG to a FHO remain in a FHO for the rest of the sample period.

sample have a roster size of at least 100 patients in all years of the sample period. The administrative datasets used in the analysis are discussed in Appendix 1.1.

1.5.2 Virtual Rostering

Pure capitation models are believed to lead to cream-skimming behaviour, inducing physicians to avoid the care or enrolment of patients with complex health conditions. Despite patient enrolment requirements and incentives as well as a cap for FFS payments in a blended capitation model, cream-skimming may still arise in these models. Physicians in the blended capitation model may choose to treat patients with more complex conditions, but not roster them, in order to receive FFS income for these services (incentive effect). Rudoler et al. (2015) find that primary care physicians in Ontario who roster more medically and socially 'complex' patients are less likely to join a blended capitation model compared to a standard FFS model and EFFS models which suggests that physician self-selection into payment models should be taken into account when studying the impact of changes in remuneration schemes on physician behaviour. The method of virtual rostering is used to help control for this patient selection and identify any cream-skimming which can affect physician referral behaviour since patients with more complex health conditions are expected to have more specialist visits. If results in the virtual roster differ substantially from that of the official roster, this might suggest the presence of patient selection.

Virtual rostering is employed by adopting an algorithm developed by the Ontario Ministry of Health and Long-Term Care (MOHLTC) where patients are assigned to family physicians who have the greatest dollar value in total billings for that patient in the previous 2 years (Zhang, 2015). This strategy identifies patients that the physician selected to treat, either with complex and expensive care, or patients that sought the physician regularly, but were not enrolled in the physician's "official" roster. The analysis includes referrals of patients that appear in the physicians' virtual roster each and every year of the sample period. The physician referral rate is calculated by dividing the total number of referrals by the number of virtually rostered patients. There are 3,177 physicians who virtually roster their patients continuously over the sample period of which 3,035 both formally and virtually roster their patients.

1.5.3 Summary Statistics

Summary statistics are presented in Table 1.1. We can see many similarities in the averages for FHGs and FHOs, but there are also clear differences. For example, FHO physicians on average have a larger total roster size than FHG physicians which can be explained by patient enrolment requirements for FHOs. Shimmura (1988) argues that a FFS system incentivizes physicians to practice in urban areas while a capitated system has the effect of evenly distributing physicians according to population. In this study, the results indicate that there are fewer blended capitated physicians located in urban areas than EFFS physicians; however, the distribution of blended capitated physicians, similar to the distribution of EFFS physicians, is not evenly spread in areas of Ontario. Furthermore, previous findings have shown that a greater percentage of interdisciplinary blended capitated physicians are located in non-major urban and rural areas compared to non-interdisciplinary blended capitated physicians (Rudoler et al., 2014).

Figure 1.1 shows the trend in referrals for patients continuously rostered with their physician. Before the introduction of the blended capitation FHO model, physicians who eventually switch to a FHO but do not join the interdisciplinary FHT, on average, have fewer referrals per patient than doctors who remain in the EFFS FHG model. This gap could reflect practice style, average health status of patients, and/or other factors. By 2010, physicians who are exclusively in a FHO have higher referral rates than physicians in a FHG. While both differences are insignificant, FHO-FHT physicians have significantly fewer referrals than FHO and FHG physicians. Figure 1.2 shows the trend in specialist visits of continuously rostered patients. Specialist visits are statistically significantly higher for FHG physicians than FHO physicians and FHO-FHT physicians across all sample periods.[12]

1.5.4 Referral Behaviour Prior to the Introduction of the Blended Capitation Model

As a preliminary to the main analysis, Ordinary Least Squares (OLS) is used to estimate the difference in referral rates in 2006, which is prior to the introduction of the FHO model. This is done to identify whether or not physicians had different referral behaviours before potentially switching payment models. Results are presented in Table 1.2. The dependent variable is the log of referrals per rostered patient. "FHOdoc" is a dummy variable that takes on the value of 1 if the physician ever switches to the FHO model during the sample period and is not affiliated with a FHT, and 0 otherwise. "FHTdoc" is a dummy variable that equals 1 if a FHO doctor is also affiliated with a FHT, and equals 0

[12] There is also an upward trend in specialist referrals for all physicians but the aging of this group of patients needs to be taken into account since patient age is fixed as of March 31st 2011 in the analysis.

otherwise. Results show different referral rate behaviours for FHO-FHT physicians, but no statistically significant differences between FHG and FHO physicians before joining the FHO model.

1.6 Empirical Strategy

A specification using difference-in-differences with fixed effects and propensity score weights is used to estimate the impact of switching from an EFFS model to a blended capitation model on a physician's referral rate. The main model used is:

$$log(R_{it}) = \alpha_i + \lambda_t + \beta X_{it} + \delta FHO_{it} + \mu FHT_{it} + u_{it} \quad (1)$$

where R_{it} is referrals per formally or virtually rostered patient for physician *i* in year t; α_i is the set of physician fixed effects; λ_t is the set of year fixed effects; X_{it} are physician practice characteristics that vary over time (patient age distribution, percentage of male patients, and total roster size per year); FHO_{it} is the treatment indicator which measures the percentage of the year in which a physician is affiliated with a FHO; FHT_{it} measures the percentage of the year in which the physician is affiliated with a FHT; and u_{it} is the error term.

In a second model presented in equation (2), the FHO and FHT variables are interacted with dummy variables that identify referrals made in the first 1.5 years of the switch and referrals made after 1.5 years of the switch.[13] This is done to control for the lag between the date the referral was made and the date the patient saw the specialist. In addition, the

[13] For example, if a physician switched to FHO in June 2008, all first year referrals include those made between June 2008 and December 2008, and the later referrals include those made after December 2008.

physician may not immediately change his/her referral behaviour after switching to a FHO.

$$log(R_{it}) = \alpha_i + \lambda_t + \beta X_{it} + \delta_1 FHO\_firstyear_{it} + \delta_2 FHO\_lateryears_{it} + \mu_1 FHT\_firstyear_{it} + \mu_2 FHT\_lateryears_{it} + u_{it} \quad (2)$$

Propensity score matching is used to generate the weights which are then employed in the fixed effects difference-in-differences estimation. This technique is sometimes called double robust estimation, where only one model needs to be correctly specified to obtain a consistent estimator and efficiency is increased if both models are correctly specified. The identifying assumption of the covariates is conditional independence, where a change in the physician's referral rate is a causal impact of a change in the physician's remuneration scheme. An alternative interpretation is as a conditional correlation (formally a covariance) if causality is not assumed. The existence of a causal impact can be supported if physicians are observed to change their referral behaviour once switching to the blended capitation model compared to their behaviour prior to the introduction of the blended capitation model.

As discussed earlier, primary care physicians in Ontario who roster more medically and socially 'complex' patients are found to less likely join a blended capitation model compared to a standard FFS model and EFFS models (Rudoler et al. 2015). The two-step approach helps mitigate the issue of selection bias. The first step is using propensity score matching on the sample of physicians prior to the introduction of the FHO, fiscal year 2006, to match control and treatment groups based on observed characteristics, such as age, sex and location and controlling for the income gain (or loss) of the physicians if they were to switch from a FHG to a FHO. The results of the matching are presented in Table 1.3.

The comparison group refers to physicians who remain in a FHG throughout the entire sample period and the treatment group refers to physicians who eventually switch to a FHO at any point in the sample period. The weighted matched sample for the comparison group presents statistics for the sample of FHG physicians in 2006 after using the generated weights. There is a clear difference between physicians who decide to remain in a FHG and physicians who decide to switch to a FHO based on observable characteristics in the year prior to the introduction of the FHO model. For example, FHG physicians would, on average, suffer an income loss of around $25,000 if they were to switch to a FHO. This incomparability is mitigated by employing the two-step estimation procedure.

The propensity scores are generated using a logit regression, and then weights are obtained from those scores using a local linear regression as a matching estimator and a bi-weight kernel with a bandwidth of 0.2. Sensitivity tests using normal, uniform, and tricube kernels as well as alternative bandwidths lead to no substantive changes in the results. In addition, the results are effectively unchanged using kernel and nearest-neighbour matching estimators instead of local linear regression.[14] Furthermore, propensity score matching is used with trimming (at a 5% trimming level) and common

[14]Methods for bootstrapping the standard errors have been found to be invalid when using the nearest-neighbour estimator, and therefore the local linear regression and kernel estimators are preferred (Abadie and Imbens, 2008). Fan (1993) demonstrates the advantageous properties of local linear regression.

support procedures, but sensitivity tests not applying trimming or common support show no substantive changes.[15]

Non-parametric percentile-t cluster-bootstrap hypothesis testing is performed on the relevant coefficients in all estimations because the generated weights are estimates and not true parameters. The hypothesis testing uses methods of resampling and pivoting (correcting for scale) of the t-statistic, and relies on two guidelines presented by Hall and Wilson (1991).[16] Non-parametric paired bootstrapping over 999 iterations is used on the entire process involving generating the weights from propensity score matching and the difference-in-differences fixed effects estimation using the generated weights. Inference is based on p-values generated from the bootstrapped t-distribution.

1.7 Results

Table 1.4 presents the results of equation (1). The first panel includes referrals for patients who are continuously and officially rostered with the physician. The second panel includes referrals for virtually rostered patients. Estimation results from the OLS model in Table 1.2 suggest that FHO-FHT physicians are changing their referral behavior once joining an interdisciplinary team. Although the referral rate of FHO-FHT physicians does not change, their total number of specialist visits is 1.85% fewer than that of FHG physicians. When comparing results in Table 1.2 and Table 1.4, physicians had fewer

[15] Restricting the sample to treated observations with common support among the comparison excludes three FHO physicians and trimming 5% of the treated sample excludes 68 FHO physicians.
[16] The first guideline under the hypothesis test H_0: $b=b_0$ against H_1: $b \neq b_0$ is resampling $b_n^*-\hat{b}$, not $b_n^*-b_0$, where b_n^* is the bootstrap coefficient for the n^{th} iteration and $\hat{b}$ is the estimated coefficient. The second guideline is basing the bootstrap distribution on $(b_n^*-\hat{b})/ s.e_n^*$, not $(b_n^*-b_0)/s.e$ where $s.e_n^*$ is the bootstrap standard error for the n^{th} iteration and s.e is the standard error of the coefficient estimate (Hall and Wilson, 1991).

listed referrals and specialist visits prior to joining an interdisciplinary team compared to when they switched models. This indicates that physicians increase their listed referrals and specialist visits once they join an interdisciplinary team, which could be a result of a concentration of patients with their rostering physician.

Table 1.5 presents the results of equation (2). Results in the first column show that FHO physicians have 2.18% more listed referrals than FHG physicians when focussing on specialist visits associated with a referral that were made at least 2.5 years after switching to the blended capitation model. However, results from the second column show no change in total specialist visits for these patients. Similar to results in Table 1.4, physicians in an interdisciplinary team have fewer total specialist visits than FHG physicians once they switch payment models. Results are qualitatively similar for the sample of virtually rostered patients, suggesting no evidence of cream-skimming. These results indicate that physicians do not change their referral behaviour immediately upon switching payment models, either due to long wait times to see a specialist or other unexplained factors.

Models with different specifications are shown to have similar results to those of Table 1.5.[17] The increase in listed referrals when physicians switch from a FHG to a FHO is slightly higher for male physicians, younger physicians, physicians with a smaller roster size, and physicians with practices in rural areas. Additional results show that FHO and FHO-FHT physicians have respectively 1.5% and 4.5% fewer specialist visits per

[17] Specifications include male and female physicians, physicians aged 35-55 and greater than 55, physicians with less than 25 years of experience and greater or equal to 25 years of experience, physicians with roster size less than 500, 500-1000 and greater than 1000, and physicians with practices in major urban, non-major urban and rural areas.

officially rostered patient; and 1.7% and 4.5% fewer services per officially rostered patient than FHG physicians. Results of first-time referrals show approximately 3% and 2% more listed referrals per officially rostered patient for FHO and FHO-FHT physicians, respectively, compared to FHG physicians, with no change in total number of specialist visits.[18]

1.8 Conclusion

The focus of this study is the impact on specialist referral rates of primary care physicians of switching from a predominantly FFS payment model to a capitation-based payment. The expectation is that capitation-based physicians refer more than FFS-based physicians because they are not remunerated per service provided and therefore have an incentive to send patients to a specialist. This effect is expected to be smaller in a blended payment system compared to pure FFS or pure capitation. Propensity score weighted difference-in-differences estimation with fixed effects is employed in the analyses of administrative data from Ontario, Canada over the period 2006-2010. The sample looks at physician-level referrals and system-level referrals for patients who are continuously enrolled with the physician across the sample period.

The results show an increase in the number of specialist visits when the blended capitated physician is associated with a referral, but no change in the number of total specialist visits for these physicians' patients. However, this increase is not shown to happen immediately after the physician joins a blended capitation system. Additionally,

[18] First specialist visit where the physician is associated with the referral for each continuously rostered patient per year.

physicians in an interdisciplinary blended capitation model have fewer specialist visits per rostered patient compared to EFFS physicians, despite an increase in specialist visits after joining the interdisciplinary health team. In line with McGuire's (2011) argument, these results could indicate a commitment of blended capitation physicians to their patients who no longer seek other physicians for medical care or a specialist referral. Cream-skimming is also shown to be less of a concern in a blended capitation model since results from the virtual roster of patients are similar to that of the official roster. In line with these results, Kantarevic and Kralj (2016) find no incentive of patient risk-selection for primary care physicians in the blended capitation model in Ontario. The blended capitation model seems to be successfully reducing the incentive for specialist referrals inherent in pure capitation. Little evidence exists that suggests whether the enhanced FFS or blended capitation model optimizes quality of care and results in better patient health outcomes. Future research may entail the study of the impact of a transition between these models and the effects of joining an interdisciplinary health team on physician behavior and the quality of patient care.

1.9 References

Abadie A, Imbens G. 2008. On the failure of the bootstrap for matching estimators. *Econometrica* 76(6): 1537–1557.

Barros P. 2003. Cream-skimming, incentives for efficiency and payment system. *Journal of Health Economics* 22(3): 419–443.

Chung K, Yang D, Lee J.H. 2010. Determinants of primary care physcians' referral pattern: a structural equation model approach. *International Journal of Public Policy* 5: 259–271.

Davidson S, Manheim L, Werner S, Hohlen M, Yudkowsky B, Fleming G. 1992. Prepayment with office-based physicians in publicly funded programs: results from the children's medicaid program. *Pediatrics.* 89(4 Pt 2): 761–7.

Fan J. 1993. Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics.* 21(1): 196-216.

Fertig A, Roland M, King H, Moore T. 1993. Understanding variation in rates of referral among general practitioners: are inappropriate referrals important and would guidelines help to reduce rates? *British Medical Journal (Clinical Research Edition)* 307(6917): 1467–1470.

Franks P, Williams G. C, Zwanziger J, Mooney C, Sorbero M. 2000. Why do physicians vary so widely in their referral rates? *Journal of General Internal Medicine* 15(3): 163–168.

Forrest C.B, Nutting P.A, Von Schrader S, Rohde C, Starfield B. 2006. Primary care physician specialty referral decision making: patient, physician, and health care system determinants. *Medical Decision Making : An International Journal of the Society for Medical Decision Making* 26(1): 76–85.

Gosden T, Forland F, Kristiansen I, Sutton M, Leese B, Giuffrida A, Sergison M, Pedersen L. 2000. Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians. *Cochrane Database of Systematic Reviews* 10: 1-29.

Hackl F, Hummer M, Pruckner G.J. 2015. Old boys' network in general practitioners' referral behavior? *Journal of Health Economics* 43: 56-73.

Hall P, Wilson S.R. 1991. Two guidelines for bootstrap hypothesis testing. *Biometrics* 47(2): 757–762.

Hennig-Schmidt H, Selten R, Wiesen D. 2011. How payment systems affect physicians' provision behaviour – an experimental investigation. *Journal of Health Economics* 30(4): 637–646.

Hutchison B, Levesque J.F, Strumpf E, Coyle N. 2011. Primary health care in Canada: systems in motion. *The Milbank Quarterly* 89(2): 256-288.

Iversen T.O.R, Luras H. 2000. The effect of capitation on GPs' referral decisions. *Health Economics* 210(1): 199–210.

Iverson G, Coleridge S, Fulda K, Licciardone J. 2005. What factors influence a family physician's decison to refer a patient to a specialist? *The International Electronic Journal of Rural and Remote Health Research, Education, Practice and Policy* 5(415): 1-7.

Kantarevic J, Kralj B, 2016. Physician payment contracts in the presence of moral hazard and adverse selection: the theory and its application in Ontario. *Health Economics* 25: 1326–1340.

Kralj B. 2000. Measuring 'rurality' for purposes of health-care planning: an empirical measure for Ontario. *Ontario Medical Review* 1-9.

Kralj B., Kantarevic J. 2013. Quality and quantity in primary care mixed-payment models: evidence from family health organizations in Ontario. *Canadian Journal of Economics* 46(1): 208–238.

Krasnik A, Groenewegen P.P, Pedersen P.A, Scholten P, Mooney G, Gottschau A, Damsgaard M.T. 1990. Changing remuneration systems: effects on activity in general practice. *British Medical Journal* 300: 1698–1701.

Li J, Hurley J, Decicca P, Buckley G. 2011. Physician response to pay-for-performance – evidence from a natural-experiment. *Health Economics* 19: 1300–1317.

Liddy C, Singh J, Kelly R, Dahrouge S, Taljaard M, Younger J. 2014. What is the impact of primary care model type on specialist referral rates? A cross-sectional study. *BMC Family Practice* 15(1): 1–8.

McGuire T.G. 2011. In The Oxford Handbook of Health Economics, Glied S, Smith P.C (Ed.), Physician agency and payment for primary medical care. Oxford University Press: Oxford: 602-619.

O'Donnell C. A. 2000. Variation in GP referral rates: what can we learn from the literature? *Family Practice* 17(6): 462–471.

Robinson J. C. 2001. Theory and practice in the design of physician payment incentives. *The Milbank Quarterly* 79(2): 149–177.

Rudoler D, Laporte A, Barnsley J, Glazier R.H, Deber R.B. 2014. Paying for primary care : a cross-sectional analysis of cost and morbidity distributions across primary care payment models in Ontario, Canada. *Social Science & Medicine* 124: 18–28.

Rudoler D, Deber R.B, Barnsley J, Glazier R.H, Dass A.R, Laporte A. 2015. Paying for primary care: the factors associated with physician self-selection into payment models. *Health Economics* 24: 1229–1242.

Sarma, S., Mehta, N., Devlin, R.A., Kpelitse, K.A., Li, L. 2018. Family physician remuneration schemes and specialist referrals: Quasi-experimental evidence from Ontario, Canada. *Health Economics* 27(10): 1533-1549.

Scott A, Jan S. 2011. In The Oxford Handbook of Health Economics, Glied S, Smith P.C (Ed.), Primary care. Oxford University Press: Oxford: 463-485.

Shimmura K. 1988. Effects of different remuneration methods on general medical practice: a comparison of capitation and fee-for-service payment. *The International Journal of Health Planning and Management* 3(4): 245–258.

Sweetman A, Buckley G. 2014. Ontario's experiment with primary care reform. *University of Calgary, School of Public Policy, Research Paper* 7(11): 1–39.

Thind A, Stewart M, Manuel D, Freeman T, Terry A. 2012. What are wait times to see a specialist? An analysis of 26,942 referrals in southwestern Ontario. *Healthcare Policy* 8(1): 80–92.

The College of Family Physicians in Canada (Available from: http://www.cfpc.ca/CPFM/) [Accessed March 15 2016]

Zhang X. 2015. Validation for virtual rostering algorithm. Working Paper, Department of Economics, McMaster University, Hamilton.

2008 Physician Services Agreement (Available at: https://www.oha.com/CurrentIssues/keyinitiatives/PhysicianandProfessionalIssues/Documents/2008%20Physician%20Services%20Agreement.pdf.) [Accessed December 6 2015]

Table 1.1: Summary Statistics, Listed Referrals, Fiscal Years 2006-2010

	Official Roster			Virtual Roster		
	FHO	FHO & FHT	FHG	FHO	FHO & FHT	FHG
Physicians	1,031	331	1,739	1,044	300	1,828
Patients with at least 1 referral	467	389	393	517	458	478
Listed Referrals	1,351	1,092	1,149	1,554	1,333	1,422
Specialist visits	3,992	3,265	3,436	4,396	3,866	4,034
Referrals per patient	1.39	1.27	1.40	1.74	1.64	1.70
Visits per patient	3.49	3.24	3.75	3.48	3.36	3.73
Services per patient	4.83	4.43	5.46	4.82	4.61	5.44
Physician Age	53	52	55	53	52	55
Male Physicians	65%	60%	65%	64%	62%	66%
Roster Size	1,560	1,320	1,408	1,557	1,414	1,374
Virtual Roster Size	1,529	1,260	1,547	1,534	1,399	1,542
Daily Visits	27	25	31	27	27	31
Daily Services	38	35	44	38	37	44
Annual Working Days	248	236	247	248	244	247
Years of Practice	24	23	25	24	23	24
Male Patient Referrals	38%	37%	39%	37%	38%	40%
Place of Graduation						
Canada	85%	85%	73%	84%	83%	73%
Foreign	15%	15%	27%	16%	17%	27%
Area of Practice						
Major Urban	48%	40%	58%	49%	32%	58%
Non-Major Urban	48%	50%	40%	48%	56%	40%
Rural	4%	10%	2%	3%	12%	2%
Patient Age Distribution						
0-19	7%	7%	7%	5%	6%	5%
20-39	15%	15%	15%	14%	13%	15%
40-59	37%	36%	38%	38%	36%	39%
60-79	32%	33%	31%	33%	35%	32%
80 plus	9%	9%	9%	10%	10%	9%

Notes: This table presents the averages and percentages of the variables across the sample period 2006-2010. "FHO-Not in FHT" refers to physicians that eventually switch to a FHO at any point in the sample period and are not affiliated with a FHT. "FHO-FHT" refers to physicians that eventually switch to a FHO at any point in the sample period and are also affiliated with a FHT. "FHG" refers to physicians that remain in a FHG throughout the entire sample period.

Table 1.2: OLS Model, 2006 (Pre-FHO)

	Official Roster		Virtual Roster	
	Listed Referrals	Specialist Visits	Listed Referrals	Specialist Visits
FHOdoc	.0036	-.0174	.0191	.0195
	(.0172)	(.0120)	(.0153)	(.0119)
FHTdoc	**-.0734*****	**-.0519*****	-.0190	**-.0327****
	(.0238)	(.0163)	(.0210)	(.0167)
N	3,102	3,102	3,177	3,177
R^2	.1745	.3740	.0997	.2961

Notes: This table reports the coefficients from an OLS model. The dependent variable is the log of referrals per rostered patient. The model controls for daily services and daily visits, annual working days, roster size (and roster size squared), percentage of referrals made for patients in different age categories, percentage of referrals made for male patients, physician age, place of graduation from medical school, if the primary care physician had any other specialty, and physician years of experience. Robust standard errors are in the parentheses. *N* represents the number of observations in the estimation. *p<0.10, ** p<0.05, *** p<0.01

Table 1.3: Matching Results, Fiscal Year 2006

	Treatment (FHO+FHT)	Comparison (FHG)	
		Matched + Unmatched Sample	Weighted Matched Sample
Number of Physicians	1361	1741	1359
Daily Visits	29	**31*** **	29
Daily Services	39	**45*** **	40
Annual Working Days	249	250	251
Roster Size	1387	**1296*** **	1406
Virtual Roster Size	1502	**1573*** **	1534
Average Physician Age	53	**55*** **	53
Male physicians	63%	65%	63%
Years of Practice	23	**25*** **	23
Income Gain	$23,286	**-$25,448*** **	$17,895
Geographic Area of Practice			
Major Urban	46%	**58%*** **	47%
Non-major Urban	49%	**40%*** **	47%
Rural	5%	**2%*** **	6%
Place of Graduation			
Canada	85%	**73%*** **	84%
Foreign	15%	27%	16%

* p<0.10, ** p<0.05, *** p<0.01

Table 1.4: Propensity Weighted Difference-in-Differences Fixed Effects, Model One

	Official Roster		Virtual Roster	
	Listed Referrals	Specialist Visits	Listed Referrals	Specialist Visits
FHO	.0087	-.0041	.0043	-.0026
	(.0173)	(.0064)	(.0095)	(.0060)
FHT	-.0033	**-.0185****	**-.0185****	**-.0175*****
	(.0173)	(.0119)	(.0145)	(.0100)
R^2				
Within	.0871	0.2845	0.1447	0.4355
Between	.0118	0.1279	0.0137	0.2407
Overall	.0060	0.0669	0.0063	0.0582
N	3102	3102	3177	3177
T	5	5	5	5
N*T	15510	15510	15885	15885

Notes: This table reports the coefficients from equation (1). The dependent variable is the log of referrals per patient. Pairs cluster bootstrap standard errors are reported in parentheses, but the asterisks are based on the preferred cluster percentile-t bootstrap methodology (* p<0.10, ** p<0.05, *** p<0.01). *N* represents the units of observations in the estimation, *T* represents the number of time periods in the sample, and *N*T* represents the number of total observations. All specifications control for daily services, daily visits, annual days worked, roster size, percentage of referrals made for enrolled patients in each age category, percentage of referrals made for male patients, and year fixed effects.

Table 1.5: Propensity Weighted Difference-in-Differences Fixed Effects, Model Two

	Official Roster		Virtual Roster	
	Listed Referrals	Specialist Visits	Listed Referrals	Specialist Visits
FHO_firstyear	-.0039	**-.0064***	-.0079	-.0060
	(.0111)	(.0062)	(.0095)	(.0061)
FHO_lateryears	**.0218*****	-.0016	**.0181*****	.0001
	(.0133)	(.0076)	(.0109)	(.0071)
FHT_firstyear	-.0072	**-.0195*****	**-.0274*****	-.0102
	(.0169)	(.0120)	(.0145)	(.0113)
FHT_lateryears	.0022	**-.0172****	-.0076	**-.0233*****
	(.0218)	(.0151)	(.0174)	(.0114)
R^2				
Within	.0884	0.2846	0.1495	0.4358
Between	.0116	0.1285	0.0132	0.2420
Overall	.0059	0.0673	0.0060	0.0586
N	3102	3102	3177	3177
T	5	5	5	5
N*T	15510	15510	15885	15885

Notes: This table reports the coefficients from equation (2). The dependent variable is the log of referrals per patient. Pairs cluster bootstrap standard errors are reported in parentheses, but the asterisks are based on the preferred cluster percentile-t bootstrap methodology (* p<0.10, ** p<0.05, *** p<0.01). *N* represents the units of observations in the estimation, *T* represents the number of time periods in the sample, and *N*T* represents the number of total observations. All specifications control for daily services, daily visits, and annual days worked, roster size, percentage of referrals made for enrolled patients in each age category, percentage of referrals made for male patients, and year fixed effects.

Figure 1.1: Referrals per Patient by Primary Care Model, 2006-2010

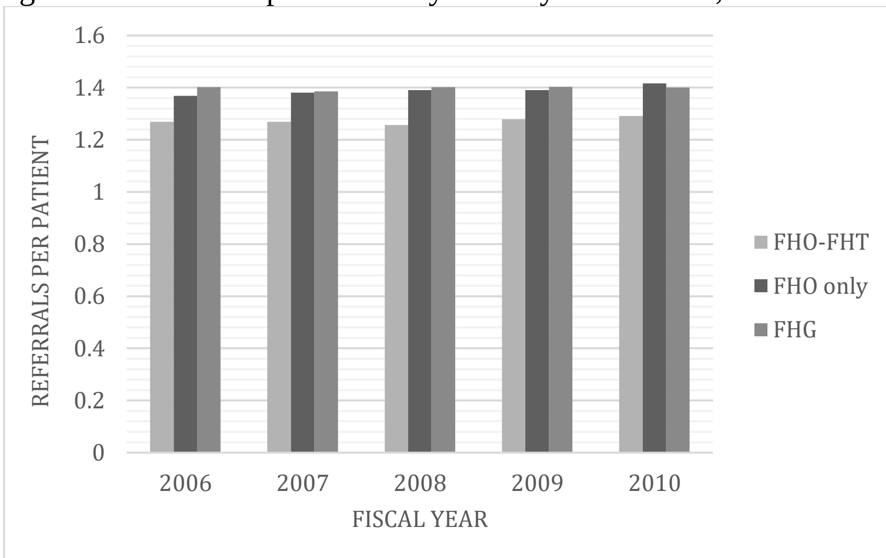
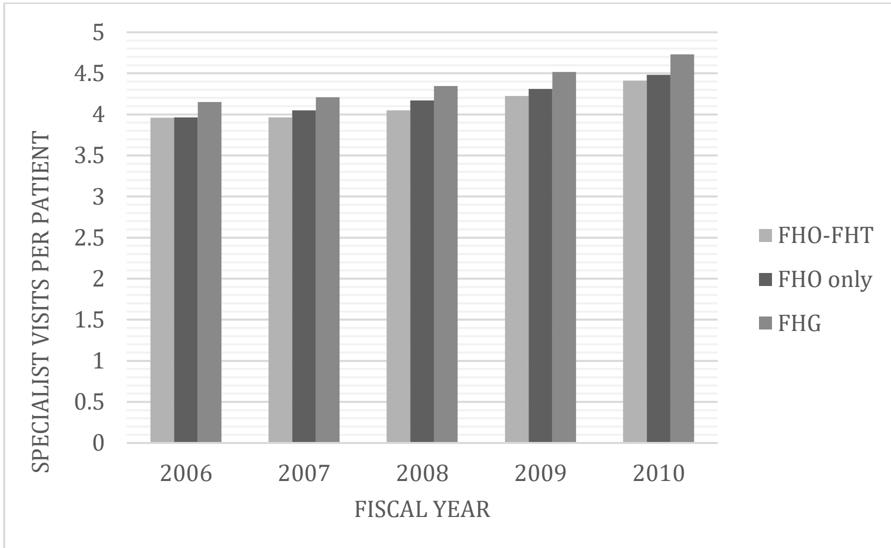


Figure 1.2: All Specialist Visits per Patient by Primary Care Model, 2006-2010



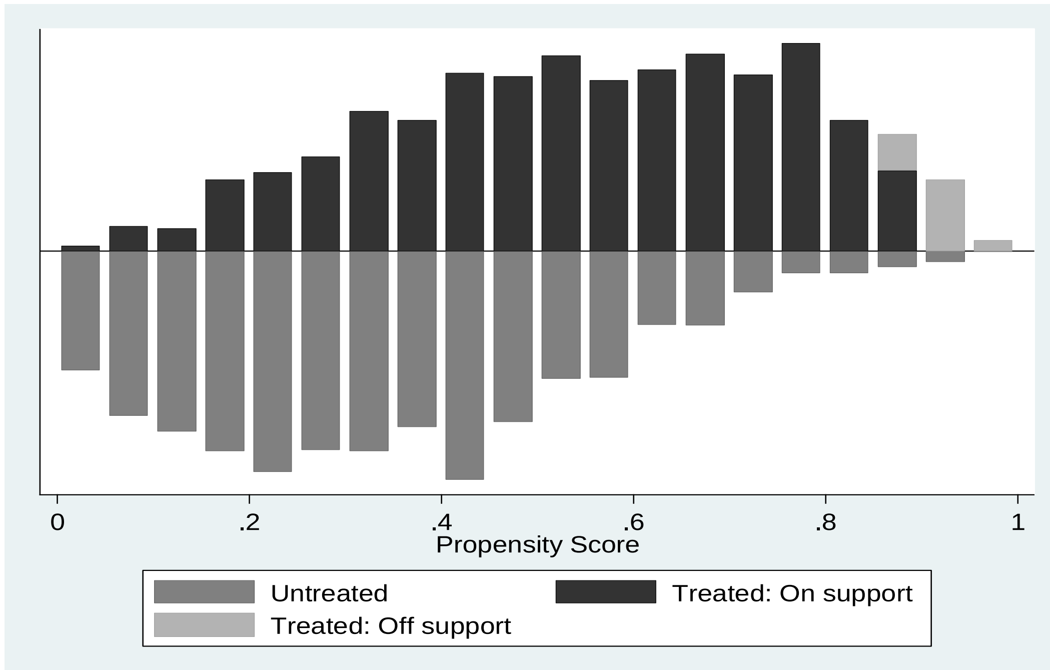
Appendix 1.1 – List of administrative data sets used in the study

- Ontario Health Insurance Plan Claims Database (OHIP): Administrative data on physician claims
- Client Agency Program Enrolment (CAPE): patient-physician enrolment data
- Registered Persons Database (RPDB): Information on patients' age and sex restricting the sample to only individuals who are eligible for Ontario health insurance coverage which excludes transients, visitors to Ontario, temporary residents, new residents within 3 months of establishing residency, members of the Canadian Forces, serving members of the RCMP,[19] and persons in federal prison.
- Corporate Provider Database (CPDB): Information on physician characteristics such as age, sex, and specialty as well as the primary care models in which the physician is affiliated with
- Rurality Index of Ontario (RIO): Information on rurality which assigns each practice an RIO score that ranges between 0 and 100, where 0 is the most urban and 100 is the most rural (Kralj 2000)

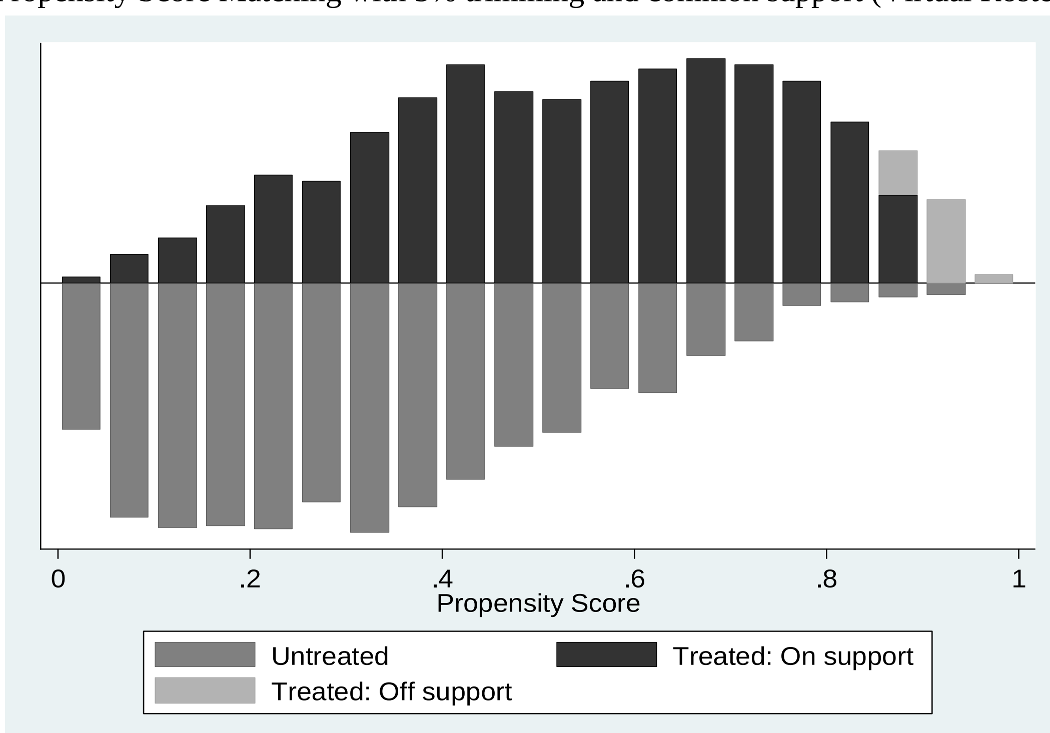
[19] Prior to June 29, 2012, serving members of the RCMP were also excluded from the definition of insured persons under the *Canada Health Act* but the *Jobs, Growth and Long-term Prosperity Act* amended the *Canada Health Act* and repealed that exclusion (Canada Health Act Annual Report 2013-2014). The data used in this study are from 2006-2011, so serving members of the RCMP are not included in the RPDB.

Appendix 1.2 – Propensity Score Matching Figures

Propensity Score Matching with 5% trimming and common support (Official Roster)



Propensity Score Matching with 5% trimming and common support (Virtual Roster)



Chapter Two

Rates of Inappropriate Laboratory Test Utilization in Ontario

2.1 Introduction

Appropriate utilization of resources is required to ensure that health care remains sustainable and to promote the best outcomes for patients (Institute of Medicine, 2010). While diagnostic test utilization has long been an important issue (Bates et al., 1991; Stair, 1998; Wilson, 2002) health care systems increasingly face challenges relating to aging populations, rapid introduction of new technology and changing educational environments for healthcare providers and initiatives such as "Choosing Wisely" are highlighting relevant issues. "Choosing Wisely" is a global campaign aimed at encouraging physicians and patients to discuss unnecessary tests and treatment and raise awareness of the issue of overutilization of tests that do not add value to or may negatively impact patient health outcomes. It began in the United Stares in 2012 and was launched as "Choosing Wisely Canada" in 2014, organized by the Canadian Medical Association, University of Toronto and St. Michael's Hospital (Toronto).[20] Recent studies have suggested that the use of diagnostic testing is increasing even beyond what can be explained by changing patient demographics or availability of new tests (Naugler, 2014).

[20] Please see https://choosingwiselycanada.org/about/ for more details.

Knowledge in healthcare continues to expand at rates too rapid for physicians to keep pace, creating uncertainty when ordering and interpreting laboratory tests. One study found that family physicians faced uncertainty when ordering diagnostic tests in 14.7% of patient encounters, but were unlikely to seek advice from laboratory professionals (Hickner et al., 2014). As well, medical education systems have replaced much of the basic science and physiology teaching with tools and guidelines for the contextual application of knowledge. The detailed knowledge of pathophysiology that enables full understanding of the diagnostic testing process is often lost, and so physicians may not appreciate some of the nuances surrounding certain tests.

The future of medical practice includes access to ever expanding lists of diagnostic tests. Appropriate access to these new tests will be essential to facilitate optimal health care for patients and efficient use of resources. Developing the role of laboratory specialists in guiding selection and the targeting of optimal test utilization may be one way of improving system efficiency. There are several reasons a laboratory test can be considered inappropriate, and historically there has been debate regarding relevant evidence standards (van Walraven and Naylor, 1998). Tests which are ordered despite an extremely low chance that they will show positive results (very low pre-test probability), tests which will not change management or diagnostic decisions regardless of the result, and tests which are repeated too frequently or too soon are some examples of inappropriate laboratory tests. The reasons for inappropriate test ordering may include 'defensive' medical practice, lack of knowledge or training on the part of the ordering physician, or lack of medical record integration resulting in the repetition of tests by different physicians who are unaware of previous orders or results. Different models of primary care practice have been adopted in Ontario and while these models may change individual ordering practice, they may also reduce redundant orders by multiple physicians given the shift to group practices sharing medical records and patient rostering (sometimes called enrollment). In addition, patient related factors such as location, travel and convenience may also impact the time between testing. Physicians have an important gatekeeping role in the Canadian public system where provincial governments fund effectively all medical laboratory tests. No billable laboratory test can be ordered without a request from a physician. Recent work for the Canadian province of Alberta documents primary care physicians' perceptions of their own role in the context of the various parties involved in misutilization and highlights the need for feedback such as that presented here (Thommasen et al, 2016).

In any effort to improve utilization of resources, there is a desire to avoid negative impacts on clinical outcome. Of the many targets for reducing diagnostic testing, targeting tests ordered redundantly or too frequently is least likely to negatively affect patient care. For many commonly ordered tests, there are clearly defined and accepted intervals at which repeat testing is appropriate. By considering the frequency with which tests are re-ordered prematurely or unnecessarily, and identifying patterns which contribute to inappropriate orders, it is possible to identify ways to target the reduction of these inappropriately utilized laboratory tests.

2.2 Methods

We study one clearly defined category of excessive laboratory utilization where inappropriate is defined solely in terms of the time interval between tests. The analytes selected for consideration are thyroid stimulating hormone (TSH), hemoglobin A1c (HbA1c), lipid profile, serum protein electrophoresis (SPEP), immunofixation (IFE), quantitative immunoglobulins (QI), Vitamin D, Vitamin B12, and folate. These are selected as they have generally accepted guidelines on the appropriate frequency of measurement and represent either high volume or high-value tests. The definition of an inappropriate test for each analyte and the rationale is outlined in Table 2.1. Two types of inappropriateness exist: a test that is too soon following a previous test and having two or more tests within one year. Repeat tests done within 8 weeks of one another are considered inappropriate for TSH, and within 3 months for HbA1c, lipid profile, SPEP, IFE, QI, Vitamin D and Vitamin B12. Vitamin D and Vitamin B12 are also considered inappropriate if more than two tests were done in a year, and the guideline for inappropriate folate tests is a repeat test within one year (Garber et al., 2012; Canadian Diabetes Association Clinical Practice Guidelines Expert Committee, 2013; Sacks et al, 2011; Genest et al., 2009, Anderson et al., 2009, Shehata et al., 2010, Brown and Josse, 2003; Smellie et al., 2005). The guidelines used are contemporaneous to the time period of data used in the study. In some cases, more up to date guidelines are now available (Papaioannou et al., 2010, Anderson et al., 2016). The updated guideline for lipid profile does not affect the interval of repeat testing; the updated recommendation, based on general and community-based studies, is that non-fasting lipid testing is acceptable (Anderson et al., 2016). For Vitamin D testing, the updated guideline based on a consensus of an expert panel, suggests that testing should not be repeated within 3-4 months or if an optimal level (≥ 75 nmol/L) is achieved (Papaioannou et al., 2010).

Population data from the near universal and public Ontario Health Insurance Plan (OHIP) dataset for the years 2006-2010 (i.e., April 1, 2006 to March 31, 2011) are employed in this study. In Ontario, the government funds all nine medical tests included in this study. This means that all the relevant provincial testing is captured in the government database, aside from a few exceptions. Individuals outside of the provincial system include members of the military, the national police force and inmates in federal prisons. There is also a small market for private pay health evaluations offered by employers for executives and other employees. Overall, for the vast majority of the population these data encompass the universe of outpatient laboratory testing. This implies that the OHIP dataset holds all testing for all out-of-hospital patients by all physicians. However, our estimates represent a lower bound on total excess testing since the data do not include the test results and therefore guidelines for the minimum estimate of over-utilization were used. Aside from only considering one category of inappropriateness we cannot compare the timing of the tests in our out-of-hospital dataset with respect to in-hospital testing since we do not have access to in-hospital data. The tests are fulfilled by private medical laboratories reimbursed on a per-test basis.

The data were obtained by a grant entitled 'Harnessing Evidence and Values for Health System Excellence' from the Government of Ontario through a Ministry of Health and Long-Term Care's Health System Research fund. No personal identifiers of patients or physicians are included in the data. All analyses were done using the statistical software STATA (version 13; STATA corp.). Research ethics was through the Hamilton Integrated Research Ethics Board (#11-086-C).

2.3 Results

We report, in Table 2.2, the number of laboratory tests done in the province for each of the analytes listed above, tests per capita where the denominator is the population of Ontario obtained from Statistics Canada (2016, CANSIM Table *051-0001*), the number and percentage of inappropriately ordered laboratory tests, the number of patients who had the test, the percentage of patients who had at least one inappropriately ordered test, and the percentage of patients with more than one inappropriately ordered test. The date used in the analysis is the "service date", which is the date when the test is completed by the laboratory. In total, between 2006 and 2010 these 9 tests were ordered 75,072,601 times by 23,743 physicians for Ontario's population of about 13.5 million people.

For the nine analytes studied, 6-20% of tests done in Ontario from 2006-2010 were inappropriate based on the defined minimum testing interval (Table 2.2). This represents over 1 million inappropriate tests being done annually. Data on a year-to-year basis are presented in the appendix. Figure 2.1 illustrates that a large proportion of inappropriate tests were conducted sooner than 2 weeks prior to the minimum interval definition of 8 weeks or 3 months. That is, most of the inappropriate orders cannot be attributed to patients having bloodwork done slightly sooner than necessary, perhaps in preparation for a physician visit close to the relevant testing guideline threshold. For the three analytes using the alternative definition of inappropriate, when two or more tests per year contravenes the guidelines – Vitamin D, Vitamin B12, and folate – a considerable number of tests were repeated excessively in a year (Figure 2.2), reaching almost 40% of Vitamin D tests done three times within one year.

Between 60% and 85% of the time, the ordering physician of an inappropriately timed test is the same physician who ordered the previous test (Figure 2.3). Table 2.3 shows that a large percentage (>70%) of physicians who order TSH, HbA1c, lipid profiles, Vitamin B12 and folate do so inappropriately at least once across fiscal years 2006 to 2010, and whether the inappropriate tests are most ordered by primary care physicians (PCPs), specialist physicians, or PCPs who also hold speciality certification varies by analyte. For example, around 6% of all TSH tests ordered by PCPs are inappropriate, and almost 91% of PCPs (out of 22,701) ordered at least one inappropriate TSH test given that they ordered at least one TSH test. Table 2.3 shows that specialists order far fewer laboratory tests compared to PCPs; however, the tests that they order are on average more likely to be inappropriately ordered than those of PCPs.

2.4 Discussion

This study demonstrates that a large proportion of physicians (Table 2.3) make inappropriate orders, with approximately 80% of physicians who ordered a test making at least one incorrect order for HbA1C and lipid profile. This large proportion of physicians choosing to order tests in a way that is not consistent with best practice guidelines highlights that some form of support is required to facilitate appropriate ordering.

Similarly, a study done in Alberta finds that 16% of six types of laboratory tests are repeated earlier than necessary (Morgen and Naugler, 2015). Another study from the VA health system in the United States found that 30% of patients being followed for diabetes had repeat HbA1cs done within 90 days of a previous test (Laxmisan et al., 2011). While there are reasons why some of what we are deeming to be inappropriate in this study may be valid medically, the high frequency of inappropriate ordering observed suggests that there are opportunities for various targeted interventions aimed at changing practice behaviour.

We found that between 6-20%, depending on analyte, of tests ordered were inappropriate based on repeat criteria. This is consistent with other studies on the topic. A 2013 systematic review and meta-analysis found that overutilization in repeat testing was 7.8% (95% CI 2.5%-12.5%) for all analytes (Zhi et al., 2013). For this study, we chose a few analytes which were either high volume or high value tests. As demonstrated in the review by Zhi et al. (2013), repeat testing constitutes only a small portion of the overutilization of laboratory tests. They found that the overutilization rate of initial testing was much higher at 44% than that for repeat testing. This suggests that there are significantly more inappropriate tests than identified in this study.

Prior to our analysis, we presumed that duplicity across multiple providers would be associated with a much larger share of inappropriate testing than was actually observed. There are a number of reasons for this, including that physicians may not be aware of what other physicians caring for the same patient have ordered, and issues of privacy if patients do not want one physician to know that they are seeing a second physician for the same or similar issue. However, this is not the case, as on average for the tests we consider, the re-ordering was done by the same physician over 70% of the time (Figure 2.3). Only TSH was repeatedly ordered by the same physician less than 70% of the time, and at 60% still suggests that physicians need help in applying minimum retesting intervals. This is consistent with previous findings which show that provider continuity resulted in more frequent testing than when patients were seen by multiple physicians (van Walraven et al., 2006). Furthermore, primary care group practices in Ontario allow for sharing of patient records and in practices with mandatory patient rostering, physicians are able to see when their patients visit another primary care physician. The role of the primary care model in inappropriate laboratory utilization is discussed in the next chapter.

The laboratory tests considered in this study represent tests used for diagnosis, monitoring and screening, although the data used in this study do not specifically identify the purpose of the test. Furthermore, the data we use do not allow us to probe for results and many of the tests can be used for more than one reason. For diagnosis and screening, a single test result is most often accepted while in monitoring, repeat tests are recommended at certain time intervals depending on biology and/or treatment interventions. Thus the proportion of "appropriate testing" will not reflect only those people with monitoring requirements and thus the actual proportion of patients who get inappropriate testing is most likely quite a lot higher than the 6-20% range identified by this study.

A reduction in the number of tests ordered inappropriately in this manner can be achieved by a variety of interventions that do not impose large costs on the healthcare system. One study found that an intervention requiring a clinical justification for high-value tests ordered by health care professionals resulted in a 50% reduction in referred-out tests that were deemed unnecessary (Liu, 2012). Another recommendation would be to reconsider the frequency of visits to monitor change in therapy, which may result in less frequent ordering and a lower likelihood of re-ordering in the time frames defined for inappropriate testing. A three-month follow up might trigger patients to get laboratory tests redone after only 8-10 weeks of therapy, whereas a four-month follow up would allow physicians to see the full effect of therapy as laboratory tests would be done 12-14 weeks after changes made at the previous visit. This would both reduce unnecessary utilization and improve the ability of a physician to see the full effect of therapy. Policy implications suggested by Naugler and Wyonch (2019) that could be employed to reduce inappropriate laboratory testing include (1) physician education with mandatory audit and feedback of laboratory usage, (2) adjusting incentives in primary care to align with improving inappropriate testing, (3) targeted administrative intervention such as modifying requisition orders to adhere to clinical guidelines more closely, and (4) develop provincial formularies.

Other interventions could include the computer aided auto-substitution of results from the most recent relevant test for orders which violate pre-determined minimal testing intervals. Of course, this could be overridden by clinicians with clinical justification. One study of an automated computer ‘pop-up’ for each test ordered within a specific time interval for inpatients found that 77% of flagged tests were cancelled by the ordering physician once flagged (Lippi et al., 2015). One successful intervention involved in-person visits and review of ordering data with community physicians (Bunting and van Walraven, 2004). With the advent of new electronic medical records (EMRs) and other technologies, these methods could be adapted to be less resource intensive but still potentially quite effective (Love et al., 2015). Systems which automatically generate 'report cards' for individual physicians which highlight how often they are requesting tests which are flagged as potentially inappropriate versus other physicians with similar practices may also be effective.

Perhaps the most effective intervention would be one at a policy level, in which reimbursements for tests deemed inappropriate are limited, or require additional input from the physician to be covered. This would need to be integrated into a physician's EMR system to be effective without increasing the documentation burden on physicians or the laboratories that see large volumes of patients.

The proportion of inappropriate testing is different between different tests (6-20%) (Table 2.2) and many of these only occur once per patient with a smaller proportion occurring with higher frequency. One of the reasons for this may be that physicians are seeking to confirm an unexpected result with a repeat test. Although we are unable to test for this, in community practice the expected rates of this type of repeat testing are less than 2% for a test repeated within a week for HbA1c and HDL-cholesterol (van Walraven and Raymond, 2003).

Limitations of this study include only having access to the billing data, which do not allow for assessments of the clinical (or medical) need for early repeat testing. Since the definition of inappropriateness is defined based on the average effect of the analyte, this implies that some tests may be done earlier or later; however, it is difficult to assess which tests these refer to without knowing the clinical result of the test. The data set over-represents testing by primary care physicians since many specialists practice within the hospital setting and hospital testing is not included in this data set. The strengths include the number of physicians and tests, and the almost universal coverage of the data base for community-based testing in Ontario.

2.5 Conclusion

Relatively large proportions of the tests ordered for the nine analytes studied are inappropriate according to accepted best practice guidelines. These tests are mostly ordered by the same physician, and both primary care and speciality physicians exhibit this inappropriate testing practice. Most inappropriate repeats occur only once in the defined time frame, but multiple repeat tests are also quite common. This study demonstrates that there are areas for improvement in the ordering of medical laboratory tests that will have no negative, and likely have some positive, impacts for patients. Given the difficulty faced by many physicians in keeping up with changing guidelines, and with standardizing practice between practitioners, it is recommended that systemic, technology-based approaches for preventing unnecessary repeat testing are investigated by the funding agencies.

2.6 References

Anderson, K.C., Alsina, M., Atanackovic, D., Biermann, J.S., Chandler, J.C., Costello, C., et al. 2015. Multiple Myeloma, Version 2.2016: Clinical Practice Guidelines in Oncology., J. Natl. Compr. Canc. Netw. 13, 1398–435.

Anderson, T.J., Grégoire, J., Pearson, G.J., Barry, A.R., Couture, P., Dawes, M., Francis, G.A., et al. 2016. 2016 Canadian Cardiovascular Society Guidelines for the Management of Dyslipidemia for the Prevention of Cardiovascular Disease in the Adult, Can. J. Cardiol. 32, 1263–1282.

Bates, D.W., Boyle, D.L., Rittenberg, E., Kuperman, G.J., Luf, N.M., Menkin, V., Winkelman, J.W., and Tanasijevic, M.J. 1991. What proportion of common diagnostic tests appear redundant?, Am. J. Med. 104, 361–368.

Brown, J.P., and Josse, R.G. 2003. 2002 clinical practice guidelines for the diagnosis and management of osteoporosis in Canada, 167, S1–S34.

Bunting, P.S., and Van Walraven, C. 2004. Effect of a Controlled Feedback Intervention on Laboratory Test Ordering by Community Physicians, Clin. Chem. 50, 321–326.

Canadian Diabetes Association Clinical Practice Guidelines Expert Committtee. 2013. Canadian Diabetes Association 2013 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada, Can. J. Diabetes. 37, S1–S212.

Choosing Wisely Canada, Choosing Wisely Canada. 2017. http://www.choosingwiselycanada.org/ (accessed January 10, 2017).

Garber, J.R., Cobin, R.H., Gharib, H., Hennessey, J.V, Klein, I., Mechanick, J.I, Pessah-Pollack, R., Singer, P.A., and Woeber, K.A. 2012. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association., Endocr. Pract. 18, 988–1028.

Genest, J., McPherson, R., Frohlich, J., Anderson, T., Campbell, N., Carpentier, A., Couture, P., et al. 2009. 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of cardiovascular disease in the adult - 2009 recommendations., Can. J. Cardiol. 25, 567–79.

Hickner, J., Thompson, P.J., Wilkinson, T., Epner, P., Sheehan, M., Pollock, A.M., Lee, J., Duke, C.C., Jackson, B.R., and Taylor, J.R. 2014. Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results., J. Am. Board Fam. Med. 27, 268–74.

Institute of Medicine (US). 2010. The Healthcare Imperative: Lowering Costs and

Improving Outcomes, National Academies Press (US), Washington, DC.

Laxmisan, A., Vaughan-Sarrazin, M., and Cram, P. 2011. Repeated hemoglobin A1C ordering in the VA health system, Am. J. Med. 124, 342–349.

Lippi, G., Brambilla, M., Bonelli, P., Aloe, R., Balestrino, A., Nardelli, A., Ceda, G.P., and Fabi, M. 2015. Effectiveness of a computerized alert system based on re-testing intervals for limiting the inappropriateness of laboratory test requests, Clin. Biochem. 48, 1174–1176.

Liu, Z., Abdullah, A., Baskin, L., Lewis, G., Kelter, G., and Naugler, C. 2012. An Intervention to Reduce Laboratory Utilization of Referred-Out Tests, Lab. Med. 43, 164–167.

Love, A.A, McKinney, Z.J, Sandoval, Y., Smith, S.W, Kohler, R., Murakami, M.M., and Apple, F.S. 2015. Electronic Medical Record-Based Performance Improvement Project to Document and Reduce Excessive Cardiac Troponin Testing, Clin. Chem. 61, 498–504.

Morgen, E.K., and Naugler, C. 2015. Inappropriate repeats of six common tests in a Canadian city: A population cohort study within a Laboratory Informatics framework, Am. J. Clin. Pathol. 144, 704–712.

Naugler, C. 2014. A perspective on laboratory utilization management from Canada, Clin. Chim. Acta. 427, 142–144.

Naugler, C., Wyonch, R. 2019. What the doctor ordered: Improving the use and value of laboratory testing. C.D. Howe Institute. 533, 1–22.

Papaioannou, A., Morin, S., Cheung, A.M., Atkinson, S., Brown, J.P., Feldman, S., Hanley, D.A., Hodsman, A., Jamal, S.A., Kaiser, S.M., Kvern, B., Siminoski, K., and Leslie, W.D. 2010. 2010 clinical practice guidelines for the diagnosis and management of osteoporosis in Canada: Summary, Can. Med. Assoc. J. 182, 1864–1873.

Sacks, D.B., Arnold, M., Bakris, G.L., Bruns, D.E., Horvath, A.R., Kirkman, M.S., Lernmark, A., Metzger, B.E., and Nathan, D.M. 2011. Position statement executive summary: Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus, Diabetes Care. 34, 1419–1423.

Shehata, N., Palda, V., Bowen, T., Haddad, E., Issekutz, T.B., Mazer, B., Schellenberg, R., Warrington, R., Easton, D., Anderson, D., and Hume, H. 2010. The Use of Immunoglobulin Therapy for Patients With Primary Immune Deficiency: An Evidence-Based Practice Guideline, Transfus. Med. Rev. 24, S28–S50.

Smellie, W.S.A., Wilson, D., McNulty, C.A.M., Galloway, M.J., Spickett, G.A., Finnigan, D.I., Bareford, D.A., Greig, M.A., and Richards, J. 2005. Best practice in primary care pathology: review 1., J. Clin. Pathol. 58, 1016–24.

Stair, T.O. 1998. Reduction of redundant laboratory orders by access to computerized patient records, J. Emerg. Med. 16, 895–897.

Thommasen, A., Clement, F., Kinniburgh, D.W., Lau, C.K., Guo, M., Viczko, J., Guggisberg, K., Thomas, R.E., Turin, T.C., Wesenberg, J.C., Abdullah, A., Hnydyk, W.S., and Naugler, C. 2016. Canadian family physician knowledge and attitudes toward laboratory utilization management., Clin. Biochem. 49, 4–7.

van Walraven, C., Cernat, G., and Austin, P.C. 2006. Effect of provider continuity on test repetition, Clin. Chem. 52, 2219–2228.

van Walraven, C., and Naylor, D. 1998. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits, JAMA. 280, 550–558.

van Walraven, C., and Raymond, M. 2003. Population-based Study of Repeat Laboratory Testing, Clin. Chem. 49, 1997–2005.

Wilson, M.L. 2002. Appropriate use of clinical microbiology tests, Clin. Lab. Med. 22, 491–503.

Zhi, M., Ding, E.L, Theisen-Toupal, J., Whelan, J., and Arnaout, R. 2013. The landscape of inappropriate laboratory testing: A 15-year meta-analysis, PLoS One. 8(11).

Table 2.1
Definition of inappropriate testing by analyte.

Analyte	**Definition of inappropriate test**	**Rationale/Reference**
TSH	Repeat test within 8 weeks	AACE Hypothyroidism Guidelines
HbA1c	Repeat test within 3 months	Canadian Diabetes Association Guidelines, AACC Position Statement
Lipid Profile	Repeat test within 3 months	CCS Dyslipidemia Guidelines
SPEP	Repeat test within 3 months	NCCN Clinical Practice Guidelines in Oncology (Multiple Myeloma)
IFE	Repeat test within 3 months	NCCN Clinical Practice Guidelines in Oncology (Multiple Myeloma)
QI	Repeat test within 3 months	Primary Immunodeficiency Guidelines
Vitamin D	Repeat test within 3 months OR >2 tests in one year	Osteoporosis Canada Guidelines
Vitamin B12	Repeat test within 3 months OR >2 tests in one year	Best Practice in Primary Care Pathology: Review 1
Folate	Repeat test within 1 year	Best Practice in Primary Care Pathology: Review 1

Table 2.2

Description of inappropriate tests by analyte, 2006 – 2010.

	Number of tests fulfilled	Patients with at least 1 fulfilled test	Tests per capita	Number of inapp. tests	% of inapp. tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
TSH	4,699,160	3,538,268	0.3466	369,251	7.70	6.87	1.69
HbA1c	2,655,414	1,645,902	0.1957	462,270	17.75	16.02	5.87
Lipid Profile	4,641,416	3,469,559	0.3427	473,781	10.23	9.22	2.56
SPEP	127,164	111,635	0.0047	7,967	6.27	4.86	0.95
IFE	63,411	53,690	0.0047	6,165	9.73	8.78	1.52
QI	67,677	58,094	0.0050	5,885	8.71	5.90	1.42
Vit D	371,687	326,690	0.0272	25,502	6.88	5.80	1.31
Vit B12	2,130,219	1,789,908	0.1567	174,620	8.22	7.08	1.61
Folate	321,954	290,739	0.0239	63,873	19.89	17.14	3.39

[a] denominator is the number of tests fulfilled

[b] denominator is the number of patients with tests fulfilled

Inapp.: inappropriate

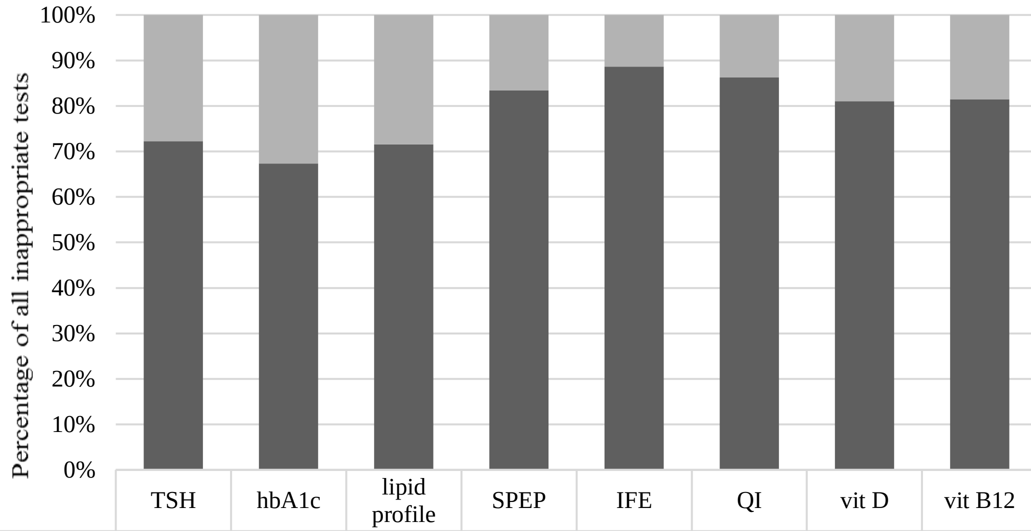
Table 2.3

Inappropriate repeat test by physician type.

	# of MDs ordering lab test	% of each MD type that ordered at least one inapp. test			% of tests ordered by each MD type that is inapp.		
		PCP	Specialist	PCP with specialty	PCP	Specialist	PCP with specialty
TSH	22,701	90.9	70.4	71.0	6.2	18.8	14.3
HbA1c	19,751	92.3	63.0	65.2	16.1	27.5	22.5
Lipid Profile	21,592	90.7	62.0	66.1	8.8	23.8	18.7
SPEP	12,495	35.5	38.4	37.3	4.2	8.7	8.6
IFE	9,662	41.3	42.2	43.2	9.3	9.5	10.6
QI	11,599	30.2	35.7	36.9	7.0	8.8	10.6
Vit D	13,961	58.8	36.5	40.5	6.0	12.8	10.1
Vit B12	19,782	89.0	57.7	61.2	7.5	19.0	12.7
Folate	16,399	78.4	53.4	59.9	19.1	22.9	24.4

Inapp.: inappropriate

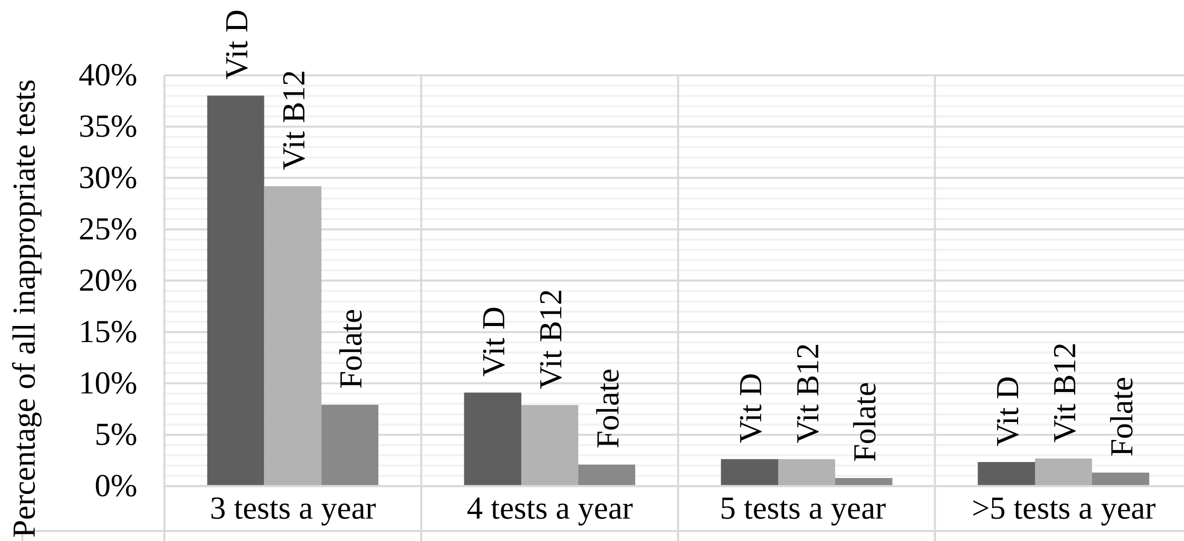
Figure 2.1. Percentage of inappropriate tests fulfilled near guideline's minimum interval between tests.



	TSH	hbA1c	lipid profile	SPEP	IFE	QI	vit D	vit B12
Tests fulfilled 1-2 weeks before guideline	0.2773741	0.3268338	0.2845536	0.1654366	0.1132883	0.1374775	0.189342	0.1850884
Tests fulfilled 3 or more weeks before guideline	0.7226259	0.6731662	0.7154464	0.8345634	0.8867117	0.8625225	0.810658	0.8149116

The percentage of 2006-2010 repeat testing in the last 2 weeks of the recommended interval compared to those undertaken earlier. Earlier than the last two weeks is defined as: 0-6 weeks for TSH, and 0-2½ months for HbA1c, lipid profile, SPEP, IFE, QI, Vitamin D, and Vitamin B12. Approximately 67% of all inappropriate HbA1c tests are processed more than 2 weeks prior to the threshold.

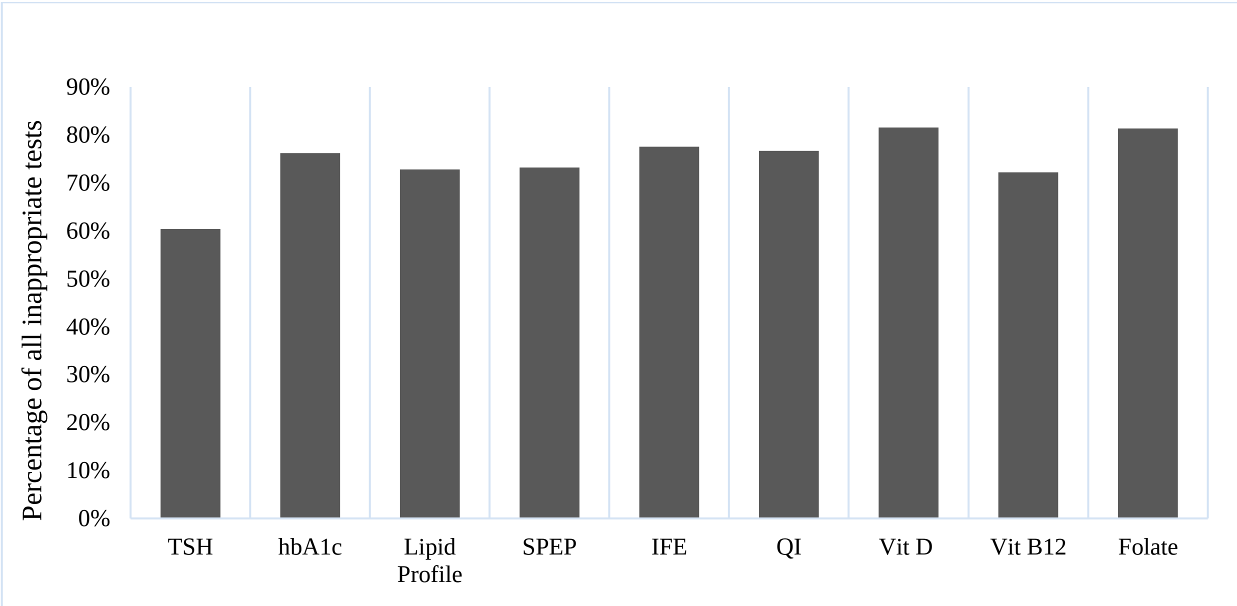
Figure 2.2. Percentage of inappropriate tests fulfilled more than twice per year.



	3 tests a year	4 tests a year	5 tests a year	>5 tests a year
Vit D	38.03%	9.11%	2.63%	2.33%
Vit B12	29.21%	7.88%	2.62%	2.68%
Folate	7.92%	2.11%	0.80%	1.31%

The average percentage of inappropriate tests using only the criteria of being conducted more than twice within a year across fiscal years 2006-2010 for Vitamin D, Vitamin B12 and Folate. This figure does not encompass all inappropriate tests done for these three analytes, only those repeated more than twice within the same year. Omitted inappropriate tests for Vitamin D and B12 include repeat tests within 3 months and done less than 3 times within the same year. Omitted inappropriate tests for Folate include tests done twice within the same year.

Figure 2.3. Percentage of inappropriate testing done by same ordering physician.



The average percentage of all inappropriate tests across fiscal years 2006-2010 where the ordering physician also ordered the previous test.

Appendix 2.1

Table A.2.1.1: TSH Lab Tests

Fiscal Year	Number of labs fulfilled	Patients with at least 1 fulfilled test	Labs per capita	Number of inapp. lab tests	% of inapp. lab tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
2006	3,885,709	2,985,294	0.2930	301,243	7.01	6.58	1.64
2007	4,212,731	3,218,251	0.3142	323,826	7.69	6.63	1.63
2008	4,994,661	3,729,219	0.3686	398,082	7.97	6.99	1.74
2009	5,289,816	3,924,633	0.3867	429,351	8.12	7.23	1.79
2010	5,112,885	3,833,941	0.3707	393,752	7.70	6.94	1.67

Table A.2.1.2: HbA1c Lab Tests

Fiscal Year	Number of labs fulfilled	Patients with at least 1 fulfilled test	Labs per capita	Number of inapp. lab tests	% of inapp. lab tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
2006	1,922,246	1,122,422	0.1449	396,355	20.62	16.83	6.12
2007	2,209,310	1,330,807	0.1648	419,974	19.01	17.69	6.75
2008	2,819,771	1,779,837	0.2081	473,542	16.79	15.27	5.62
2009	3,126,133	1,971,474	0.2286	520,430	16.65	15.47	5.60
2010	3,199,610	2,024,970	0.2320	501,051	15.66	14.84	5.27

Table A.2.1.3: Lipid Profile Lab Tests

Fiscal Year	Number of labs fulfilled	Patients with at least 1 fulfilled test	Labs per capita	Number of inapp. lab tests	% of inapp. lab tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
2006	4,342,008	3,248,274	0.3274	467,687	10.77	9.60	2.71
2007	4,453,268	3,324,212	0.3321	473,914	10.64	9.55	2.69
2008	4,688,482	3,498,209	0.3460	476,703	10.17	9.13	2.56
2009	4,866,237	3,613,384	0.3558	499,608	10.27	9.33	2.60
2010	4,857,083	3,663,715	0.3522	450,995	9.29	8.50	2.25

Table A.2.1.4: Serum Protein Electrophoresis Lab Tests

Fiscal Year	Number of labs fulfilled	Patients with at least 1 fulfilled test	Labs per capita	Number of inapp. lab tests	% of inapp. lab tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
2006	120,572	106,163	0.0045	7,570	6.28	4.68	0.86
2007	123,485	108,482	0.0046	7,618	6.17	4.86	0.94
2008	129,014	112,803	0.0048	8,227	6.38	4.92	1.01
2009	132,701	116,594	0.0049	8,318	6.27	4.93	0.97
2010	130,047	114,131	0.0047	8,101	6.23	4.90	0.99

Table A.2.1.5: Immunofixation Lab Tests

Fiscal Year	Number of labs fulfilled	Patients with at least 1 fulfilled test	Labs per capita	Number of inapp. lab tests	% of inapp. lab tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
2006	57,613	48,611	0.0043	5,763	10.00	9.01	1.49
2007	62,383	52,672	0.0047	6,034	9.67	8.79	1.47
2008	65,324	54,872	0.0048	6,605	10.11	9.06	1.62
2009	67,651	57,601	0.0049	6,434	9.51	8.58	1.55
2010	64,085	54,693	0.0046	5,988	9.34	8.44	1.45

Table A.2.1. 6: Quantitative Immunoglobulin Lab Tests

Fiscal Year	Number of labs fulfilled	Patients with at least 1 fulfilled test	Labs per capita	Number of inapp. lab tests	% of inapp. lab tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
2006	59,902	50,851	0.0045	5,423	9.05	6.18	1.45
2007	63,230	53,837	0.0047	5,641	8.92	6.42	1.50
2008	67,310	58,001	0.0050	5,752	8.55	5.66	1.39
2009	73,618	63,568	0.0054	6,323	8.59	5.70	1.43
2010	74,325	64,211	0.0054	6,284	8.45	5.52	1.32

Table A.2.1.7: Vitamin D Lab Tests

Fiscal Year	Number of labs fulfilled	Patients with at least 1 fulfilled test	Labs per capita	Number of inapp. lab tests	% of inapp. lab tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
2006	69,812	61,423	0.0053	5,076	7.27	5.98	1.45
2007	148,928	130,625	0.0111	9,539	6.41	5.41	1.21
2008	420,487	363,788	0.0310	28,492	6.78	5.68	1.38
2009	737,834	641,666	0.0539	49,941	6.77	5.83	1.29
2010	481,372	435,946	0.0349	34,462	7.16	6.10	1.20

Table A.2.1.8: Vitamin B12 Lab Tests

Fiscal Year	Number of labs fulfilled	Patients with at least 1 fulfilled test	Labs per capita	Number of inapp. lab tests	% of inapp. lab tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
2006	1,187,548	1,020,883	0.0895	106,389	8.96	7.52	1.75
2007	1,509,437	1,297,586	0.1126	113,965	7.55	6.42	1.43
2008	2,436,666	2,030,115	0.1798	201,774	8.28	7.14	1.71
2009	2,865,632	2,376,945	0.2095	246,643	8.61	7.53	1.73
2010	2,651,811	2,224,009	0.1923	204,331	7.71	6.78	1.44

Table A.2.1.9: Folate Lab Tests

Fiscal Year	Number of labs fulfilled	Patients with at least 1 fulfilled test	Labs per capita	Number of inapp. lab tests	% of inapp. lab tests[a]	% of patients with at least 1 inapp. test[b]	% of patients with > 1 inapp. test[b]
2006	381,334	344,345	0.0288	74,526	19.54	16.96	3.35
2007	382,854	347,057	0.0286	74,608	19.49	16.99	3.25
2008	330,090	297,365	0.0244	65,176	19.74	17.08	3.41
2009	289,924	261,178	0.0212	60,129	20.74	17.79	3.53
2010	225,570	203,750	0.0164	44,926	19.92	16.89	3.43

[a] denominator is the number of tests fulfilled
[b] denominator is the number of patients with tests fulfilled
Inapp.: inappropriate

Chapter Three

Costs and determinants of inappropriate laboratory test utilization

3.1 Introduction

Appropriate clinical laboratory testing has long been recognized as an important driver of high- quality medical care (Bates el al., 1991; Stair, 1998; Wilson, 2002), whereas inappropriate testing is a drain on healthcare resources and sometimes associated with less than optimal care delivery (Chami et al., 2017; Morgen and Naugler, 2015; Zhi et al., 2013; Peterson and Rodin, 1987). The issue is starting to be highlighted and addressed through initiatives such as the "Choosing Wisely" campaign and advances in information technology that facilitate appropriate medical testing (Levison et al., 2015; Ferraro and Panteghina, 2017; Lippi et al., 2015; Liu et al., 2012). Initiatives to promote appropriate laboratory testing may support changes in ordering practice. "Choosing Wisely" aims to encourage physicians and patients to discuss unnecessary tests and treatment and raise awareness of the issue of overutilization of tests that do not add value to or may negatively impact patient health outcomes. This study aims to understand selected correlates of inappropriate testing, using measures of physician and practice characteristics, and the costs of inappropriate laboratory utilization in the Canadian province of Ontario. It focusses on the subset of inappropriate tests that were repeated earlier than recommended by the relevant clinical guideline (Chami et al., 2017).

3.2 Methods

3.2.1 Design

We focus on nine analytes, representing high-value or high-volume tests, with clearly defined and generally accepted guidelines based on the time interval between tests: thyroid stimulating hormone (TSH), hemoglobin A1c (HbA1c), lipid profile, serum protein electrophoresis (SPEP), immunofixation (IFE), quantitative immunoglobulins (QI), Vitamin D, Vitamin B12, and folate. The guideline and reference for each analyte can be found in Table 3.1. The guidelines are selected based on the time period of the data, although some guidelines have been updated (Garber et al., 2012; Canadian Diabetes Association Clinical Practice Guidelines Expert Committee, 2013; Sacks et al, 2011; Genest et al., 2009, Anderson et al., 2009, Shehata et al., 2010, Brown and Josse, 2003; Smellie et al., 2005; Papaioannou et al., 2010, Anderson et al., 2016).

3.2.2 Data

Data used in this study were obtained from the Ontario Health Insurance Plan (OHIP) dataset, a nearly universal public provincial health insurance plan funded by the Government of Ontario. The data contain the number and cost of virtually all out-of-hospital laboratory testing from April 1, 2006 until March 31, 2011 (fiscal years 2006–2010) ordered by all physicians and fulfilled by community medical labs for the population of Ontario. Exceptions include members of the military, the national police force and inmates in federal prisons, along with the small minority of individuals that opt to pay privately for health evaluations offered by their employers. The date of the test in the data, the “service date”, reflects the date that the test was processed by the community medical lab.

Per capita test costs are calculated by dividing their cost by the Ontario population obtained from Statistics Canada (2016, CANSIM Table *051-0001*). All cost are inflation-adjusted to 2010-dollar values based on Statistics Canada consumer price index (V41690973 series). Data on a year-to-year basis (not adjusted for inflation) are presented in Appendix 3.2.

3.2.3 Variables

The data also contain physician and practice characteristics including practice location, physician age, sex, years of practice, medical school graduation and specialty, the physician payment model at the time the laboratory test was done, average daily visits and services, and the year and month the laboratory test was done to potentially identify seasonality and trends in inappropriate testing. A list of all variables, and their definitions, is provided in Appendix 3.1. No personal identifiers of patients or physicians are included in the data.

Characteristics of physicians include the payment model with which the physician is associated. Physicians in Ontario are paid by a universal public provincial health insurance plan and primary care physicians voluntarily select their payment model. Primary Care Reform has introduced a sequence of new payment models that combine FFS with capitation and/or pay-for-performance (P4P) incentives and other bonuses (Sweetman and Buckley, 2014). Models with enhanced FFS enhanced with P4P are known as Family Health Groups (FHGs) and Community Care Models (CCMs). Blended capitation with FFS and P4P are known as Family Health Organizations (FHOs) and Family Health Networks (FHNs). Physicians practicing in specialized models in Northern Ontario known as the Rural and Northern Physician Group Agreement (RNPGA) also have salary components. The FHG, FHN, and FHO models are group practices of three or more primary care physicians who share patient medical records. Other primary care payment models include smaller salary or capitation models and few specialty-specific payment models that target specific population groups or geographical areas. Specialists in Ontario are primarily paid FFS.

3.2.4 Statistical Analysis

Initially, we tally the number and cost of inappropriate laboratory tests done for the selected analytes. The entire dataset of the nine analytes used in this study includes over 74 million observations (individual labs). Multivariate estimation to determine the physician and practice characteristics associated with inappropriate laboratory testing is undertaken using logit estimation, where the dependent variable is set to one if the test is inappropriate and zero otherwise. Standard errors are clustered on the 22,242 physicians in the data, meaning the standard errors are not assumed to be independent across tests ordered by the same physician and the degrees of freedom are related to the number of physicians not tests.

Results from logistic regression models, conditional on (adjusting for) the full set of variables available, and 13 unconditional models including, in turn, each individual set of regressors in isolation, are presented to permit comparisons. That is, a single logistic regression with the full set of variables is estimated on the sample of tests for all nine analytes and is presented in the conditional model in Table 3.4. The unconditional model Table 3.4 presents (in a single column to conserve space) results from 13 logistic regressions; one for each set of independent variables as indicated by the horizontal lines. The latter identify the unconditional association on the probability of ordering an inappropriate test where we test the association of a set of explanatory variables at a time with the outcome. For both columns, for sets of indicator (sometimes called dummy) variables the omitted variable is in parenthesis. Results for all analytes using logit estimation on a 20% random sample are presented in Appendix 3.1. A table of a list of explanatory variables used in the analyses is also included in Appendix 3.1. For comparison, Appendix 3.1 also presents results for the 100% sample of observations using a linear probability model (LPM) on all analytes (conditional and unconditional on variables). The LPM provides results that are interpreted as changes in probabilities rather than odds ratios. LPM models were also estimated on each separate analyte with no qualitative changes observed. All analyses are done using the statistical software STATA (version 14; STATA corp.). Research ethics was through the Hamilton Integrated Research Ethics Board (#11-086-C).

3.3 Results

Means, with indicator variables represented by percentages, of the variables are presented in Table 3.2. For example, almost 11% of all tests in the sample are ordered inappropriately, and the majority of tests in the sample are ordered by primary care physicians, male physicians, those that graduated from a Canadian medical school, and those that have practices in major urban areas.

Table 3.3 reports the average annual total cost of laboratory tests for each analyte, the cost per capita, and the cost and share of the cost of inappropriately ordered laboratory tests across the sample period 2006 – 2010. The cost of inappropriate tests for the nine selected analytes ranges between 6 – 20% of the total cost of the test. In all, the average annual cost of inappropriate tests for these nine analytes across the province between 2006 and 2010 is almost 12 million Canadian 2010 dollars, representing 9% of the total cost of these tests. This, of course, ignores costs (financial and otherwise), associated with lower quality care.

Results of the logit estimation on all nine analytes combined, in Table 3.4, indicate that, on average and conditional on the other regressors, specialists are twice as likely to order inappropriate tests as FFS primary care physicians. This is in line with descriptive results from the previous chapter that show that specialists order far fewer lab tests compared to PCPs; however, the tests that they order are on average more inappropriately ordered than those of primary care physicians. Specialists might be ordering more inappropriately to confirm false positives or due to lack of sharing of patient records by the primary care physician. Similarly, primary care physicians in a specialty-specific payment model are more likely to order inappropriately compared to FFS primary care physicians. Within primary care, physicians in group practices receiving FFS payments with P4P incentives, or capitation payments with an FFS component or other primary care payment models are less likely to order inappropriate tests compared to physicians in a solo-practice, traditional FFS model with no P4P incentives. The rationale for choosing physicians in an FFS payment model with no P4P incentives as the reference group is to compare them to physicians in a blended capitated model (FHG, FHO, and FHN), salary payment model (RNPGA) and FFS payment model with P4P incentives (CCM). Male physicians and physicians who graduated from a non-Canadian medical school are less likely to order inappropriate tests than female and Canadian-trained physicians, respectively.

A small negative association of years of practice and inappropriate testing exists, where more years of practice represents slightly lower odds of ordering inappropriate tests. There are also increasing odds of inappropriate testing with increasing patient age, where younger patients are less likely to be ordered an inappropriate test by a physician compared to older patients. Older patients may have more chronic conditions that require more laboratory testing compared to younger patients; perhaps ordering one test prompts the physician to order multiple tests, where some of these tests are inappropriate. There is also a significant downward trend in years from 2006 to 2010. With respect to seasonality, there are slightly higher odds of inappropriate testing in the months of April, June, July, and December compared to the beginning of the year, which represent summer and holiday months.

The low pseudo R^2 (0.0359) indicates that the physician and practice characteristics included in the estimation do not explain a large share of the variation in inappropriate testing. The pseudo R-squared used in this estimation is McFadden’s pseudo R-squared,

$R^2_{McFadden} = 1 - \frac{log(Lc)}{log(Lnull)}$ where L_c equals the maximized likelihood value from the current fitted model and L_{null} equals the maximized likelihood value from the null model.

The unconditional associations of the independent variables with the probability of ordering an inappropriate test, also displayed in Table 3.4, show no qualitative changes in the results for most variables. However, for years of practice the results in the unconditional model show that physicians with more years of practice are more likely to order inappropriately than physicians with fewer years of practice, whereas the effect is switched when controlling for other variables in the conditional model. Additionally, the magnitude of primary care physician payment models compared to FFS is larger in the unconditional model. No qualitative changes are observed by restricting the sample to male or female physicians compared to the sample of all physicians and all patients as seen in Appendix 3.1. In addition, no qualitative changes are observed for tests done for male or female patients done separately.

3.4 Discussion

The cost of inappropriate testing of the nine selected analytes with clear and generally accepted testing guidelines is large across the fiscal years 2006–2010, ranging between 6–20% of the total cost. Physician and practice characteristics are seen to have statistically significant associations with the probability of ordering an inappropriate test. The most prominent results show that specialists are about twice as likely to order inappropriate tests as FFS primary care physicians. Primary care physicians in alternative payment models are less likely to order inappropriate tests compared to traditional FFS primary care physicians. Other physician and practice characteristics such as male physicians, physicians trained outside of Canada, older physicians, and a younger patient population are shown to be associated with less inappropriate testing.

Other studies have found an association between physician and practice characteristics and the degree (quantity) of laboratory utilization, defined as the number of laboratory tests ordered– with no indication of whether these tests were appropriately ordered (Verstappen et al., 2004; Freeborn et al., 1972; Kristiansen and Hjortdahl, 1992; Taylor, 2005). However, these were in line with our results on inappropriate testing where studies found that group practices are associated with lower levels of tests ordered, female physicians requested tests more often, tests were more often ordered as patient's age increases, most tests were found to be ordered for patients aged 65-84, and younger physicians were high utilizers of labs whereas older physicians were low utilizers (Verstappen et al., 2004; Freeborn et al., 1972; Kristiansen and Hjortdahl, 1992; Taylor, 2005). Years of practice could also be correlated with age of the physician, where older physicians have more experience and therefore may need fewer tests to confirm a diagnosis compared to younger, less experienced physicians.

Patients receiving a laboratory test require a referral from a physician. Physicians in Ontario are not paid a fee for ordering a laboratory test; however, the community medical laboratory that fulfilled the test is reimbursed on a per-test basis by the provincial government with a cap. Although it can be argued that the physician still receives a payment for patient visits when ordering a laboratory test, and there might be an incentive to schedule a follow-up visit with the patient, this incentive is diminished in practices where laboratory results are often communicated by a nurse to the patient through telephone, especially when the physicians in the practice are paid by capitation. Furthermore, different payment models in Ontario may have different incentives and bonuses for physicians that might affect the quantity and quality of inappropriate testing. For example, physicians practicing in the blended capitation model are more responsive to the Diabetes Management Incentive (DMI), a financial incentive promoting the continuous treatment and management of diabetic patients including adherence to clinical guidelines, compared to enhanced FFS physicians (Kantarevic and Kralj, 2013). Furthermore, Chapter 4 discusses the impacts of switching from an FFS primary care model enhanced with P4P to a blended capitation model with rostering requirements on the quantity and quality of laboratory test utilization.

The data are limited to out-patient laboratory tests only, so in-hospital labs are not included in the study. In addition, a minimum bound of inappropriate testing was used since the results of the clinical tests are not provided. The study also ignores costs (financial and otherwise) associated with lower quality care. No clinical data is included in the data set nor are the actual results of the tests available. Furthermore, although the study found that physicians in FFS primary care payment models are more likely to order inappropriate tests compared to other primary care payment models, this over-utilization reflects only 1% of the potential cost associated with over-utilization as only 1% of the physicians in the sample are paid FFS.

3.5 Conclusion

Inappropriate repeat laboratory testing may be translated into health care cost savings for the provincial government. Without links to clinical outcomes it is difficult to predict the clinical impact of changing test ordering practices. This study also suggests that physician characteristics, including age, gender, time in practice as well as location of training, specialty and the nature of their payment model are associated with different patterns of repeat testing among physicians. However, the physician and practice variables included in the model did not explain the variability. Thus, further data analysis with different variables is required to better understand the driving factors of inappropriate repeat ordering by physicians. The next chapter investigates further the role of primary care payment models as a driver of the quantity and quality of laboratory utilization.

3.6 References

Anderson, K.C., Alsina, M., Atanackovic, D., Biermann, J.S., Chandler, J.C., Costello, C., et al. 2015. Multiple Myeloma, Version 2.2016: Clinical Practice Guidelines in Oncology., J. Natl. Compr. Canc. Netw. 13, 1398–435.

Anderson, T.J., Grégoire, J., Pearson, G.J., Barry, A.R., Couture, P., Dawes, M., Francis, G.A., et al. 2016. 2016 Canadian Cardiovascular Society Guidelines for the Management of Dyslipidemia for the Prevention of Cardiovascular Disease in the Adult, Can. J. Cardiol. 32, 1263–1282.

Bates, D.W., Boyle, D.L., Rittenberg, E., Kuperman, G.J., Luf, N.M., Menkin, V., Winkelman, J.W., and Tanasijevic, M.J. 1991. What proportion of common diagnostic tests appear redundant?, Am. J. Med. 104, 361–368.

Brown, J.P., and Josse, R.G. 2003. 2002 clinical practice guidelines for the diagnosis and management of osteoporosis in Canada, 167, S1–S34.

Canadian Diabetes Association Clinical Practice Guidelines Expert Committtee. 2013. Canadian Diabetes Association 2013 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada, Can. J. Diabetes. 37, S1–S212.

Chami N., Simons J.E., Sweetman A., Don-Wauchope A.C. 2017. Rates of inappropriate laboratory test utilization in Ontario. Clin Biochem. 50(15):822-827.

Ferraro, S., and Panteghina, M. 2017. The role of the laboratory in ensuring appropriate test requests. *Clin Biochem*, 50(10–11):555–61.

Freeborn D.K., Baer, D., Greenlick, M.R., and Bailey, J.W. 1972. Determinants of medical care utilization: Physicians' use of laboratory services. Am J Public Health. 62(6):846–53.

Garber, J.R., Cobin, R.H., Gharib, H., Hennessey, J.V, Klein, I., Mechanick, J.I, Pessah-Pollack, R., Singer, P.A., and Woeber, K.A. 2012. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association., Endocr. Pract. 18, 988–1028.

Genest, J., McPherson, R., Frohlich, J., Anderson, T., Campbell, N., Carpentier, A., Couture, P., et al. 2009. 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of cardiovascular disease in the adult - 2009 recommendations., Can. J. Cardiol. 25, 567–79.

Kantarevic, J., and Kralj, B. 2013. Link between pay for performance incentives and physician payment mechanisms: evidence from the diabetes management incentive in

Ontario. Health Econ. 22(12):1417–39.

Kristiansen, I.S., and Hjortdahl, P. 1992. The general practitioner and laboratory utilization: Why does it vary? Fam Pract. 9(1):22–7.

Levinson W., Kallewaard. M., Bhatia, R.S., Wolfson, D., Shortt, S., and Kerr, E.A. 2015. “Choosing Wisely”: A growing international campaign. BMJ Qual Saf. 24(2):167–74.

Lippi, G., Brambilla, M., Bonelli, P., Aloe, R., Balestrino, A., Nardelli, A., Ceda, G.P., and Fabi, M. 2015. Effectiveness of a computerized alert system based on re-testing intervals for limiting the inappropriateness of laboratory test requests, Clin. Biochem. 48, 1174–1176.

Liu, Z., Abdullah, A., Baskin, L., Lewis, G., Kelter, G., and Naugler, C. 2012. An Intervention to Reduce Laboratory Utilization of Referred-Out Tests, Lab. Med. 43, 164–167.

Morgen, E.K., and Naugler, C. 2015. Inappropriate repeats of six common tests in a Canadian city: A population cohort study within a Laboratory Informatics framework, Am. J. Clin. Pathol. 144, 704–712.

Papaioannou, A., Morin, S., Cheung, A.M., Atkinson, S., Brown, J.P., Feldman, S., Hanley, D.A., Hodsman, A., Jamal, S.A., Kaiser, S.M., Kvern, B., Siminoski, K., and Leslie, W.D. 2010. 2010 clinical practice guidelines for the diagnosis and management of osteoporosis in Canada: Summary, Can. Med. Assoc. J. 182, 1864–1873.

Peterson S.E., and Rodin, A.E., 1987. Prudent laboratory usage, cost containment, and high quality medical care: are they compatible? Hum Pathol. 18: 105-8.

Sacks, D.B., Arnold, M., Bakris, G.L., Bruns, D.E., Horvath, A.R., Kirkman, M.S., Lernmark, A., Metzger, B.E., and Nathan, D.M. 2011. Position statement executive summary: Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus, Diabetes Care. 34, 1419–1423.

Shehata, N., Palda, V., Bowen, T., Haddad, E., Issekutz, T.B., Mazer, B., Schellenberg, R., Warrington, R., Easton, D., Anderson, D., and Hume, H. 2010. The Use of Immunoglobulin Therapy for Patients With Primary Immune Deficiency: An Evidence-Based Practice Guideline, Transfus. Med. Rev. 24, S28–S50.

Smellie, W.S.A., Wilson, D., McNulty, C.A.M., Galloway, M.J., Spickett, G.A., Finnigan, D.I., Bareford, D.A., Greig, M.A., and Richards, J. 2005. Best practice in primary care pathology: review 1., J. Clin. Pathol. 58, 1016–24.

Stair, T.O. 1998. Reduction of redundant laboratory orders by access to computerized patient records, J. Emerg. Med. 16, 895–897.

Sweetman, A., and Buckley, G. 2014. Ontario's Experiment with Primary Care Reform. Univ Calgary, Sch Public Policy, Res Pap. 7(11):1–39.

Taylor, J.B. 2005. Relationships among patient age, diagnosis, hospital type, and clinical laboratory utilization. Clin Lab Sci. 18(1):8–15.

Verstappen, W.H.J.M., ter Riet, G., Dubois, W.I., Winkens, R., Grol, R.P.T.M., and van der Weijden, T. 2004. Variation in test ordering behaviour of GPs: Professional or context-related factors? Fam Pract. 21(4):387–95.

Wilson, M.L. 2002. Appropriate use of clinical microbiology tests, Clin. Lab. Med. 22, 491–503.

Zhi, M., Ding, E.L, Theisen-Toupal, J., Whelan, J., and Arnaout, R. 2013. The landscape of inappropriate laboratory testing: A 15-year meta-analysis, PLoS One. 8(11).

Table 3.1
Definition of inappropriate testing by analyte.

Analyte	**Definition of inappropriate test**	**Rationale/Reference**
TSH	Repeat test within 8 weeks	AACE Hypothyroidism Guidelines
HbA1c	Repeat test within 3 months	Canadian Diabetes Association Guidelines, AACC Position Statement
Lipid Profile	Repeat test within 3 months	CCS Dyslipidemia Guidelines
SPEP	Repeat test within 3 months	NCCN Clinical Practice Guidelines in Oncology (Multiple Myeloma)
IFE	Repeat test within 3 months	NCCN Clinical Practice Guidelines in Oncology (Multiple Myeloma)
QI	Repeat test within 3 months	Primary Immunodeficiency Guidelines
Vitamin D	Repeat test within 3 months OR >2 tests in one year	Osteoporosis Canada Guidelines
Vitamin B12	Repeat test within 3 months OR >2 tests in one year	Best Practice in Primary Care Pathology: Review 1
Folate	Repeat test within 1 year	Best Practice in Primary Care Pathology: Review 1

Table 3.2 Means of variables, 2006-2010	
Variable	**Mean (range)**
Part 1: Dependent Variable	
Inappropriate tests	10.5%
Part 2: Independent Variables	
Practice location	
Rural area	3.0%
Non-major urban area	42.2%
Major urban area	54.8%
Physician specialty	
Primary care	86.0%
Specialist	4.3%
GP specialist	1.1%
Primary care with another specialty	8.6%
Primary care payment model (if primary care physician specialty=1)	
FFS	1.0%
FHG	54.1%
FHO	19.3%
FHN	5.1%
RNPGA	0.4%
CCM	5.2%
Other primary care model	14.9%
Male physicians	68.6%
Physician age	54.7
Place of medical school graduation	
Canada	68.0%
USA	5.9%
UK	3.5%
other	26.8%
Years of practice	23.6
Percentage of tests for male patients	41.3%
Patient age (years)	55.0
Annual number of tests per patient	2.0

Daily visits	31.3
Daily services	44.9
Percentage of annual tests by month	
January	8.7%
February	7.9%
March	9.2%
April	8.6%
May	8.7%
June	8.5%
July	7.9%
August	7.7%
September	8.4%
October	8.8%
November	8.9%
December	6.7%

Means of indicator (dummy) variables are expressed as percentages rather than proportions.

Table 3.3
Description of average annual cost of inappropriate tests by analyte, 2006 – 2010 (2010 inflation adjusted dollars)

	Cost of tests fulfilled[a] **($/year)**	**Cost per capita**	**Cost of inapp. tests**[a] **($/year)**	**Cost of inapp. tests**[b] (%)
TSH	47,277,000	3.49	3,714,000	8%
HbA1c	5,145,000	0.38	906,000	18%
Lipid Profile	12,335,000	0.91	1,260,000	10%
SPEP	2,421,000	0.18	152,000	6%
IFE	4,554,000	0.34	446,000	10%
QI	1,418,000	0.10	120,000	8%
Vit D	19,507,000	1.43	1,338,000	7%
Vit B12	31,472,000	2.32	2,581,000	8%
Folate	7,644,000	0.56	1,516,000	20%
Total	131,772,000	1.95	12,035,000	9%

[a] Cost rounded to the nearest thousand

[b] (Cost of inappropriate tests*100%)/(total cost of all tests for each analyte)

Inapp.: inappropriate

Table 3.4: Logit Model for all Analytes
Dependent Variable: Inappropriate Test

	Conditional Model			Unconditional Model		
	odds ratio	robust s.e.	95% confidence interval	odds ratio	robust s.e.	95% confidence Interval
Practice location (major urban)						
Rural	1.02	.032	0.96, 1.09	.98	.033	0.92, 1.05
Nonmajor urban	**1.06*** **	.019	1.02, 1.10	**1.07*** **	.024	1.02, 1.11
Physician specialty (primary care)						
Specialist	**2.01*** **	.097	1.83, 2.21	**2.61*** **	.071	2.47, 2.75
Primary Care (PC) & specialty	**1.19*** **	.074	1.05, 1.34	**2.07*** **	.064	1.95, 2.20
GP specialist	1.04	.049	0.97, 1.17	1.04	.060	0.93, 1.16
Primary care physician payment model (FFS)						
FHG	**0.70*** **	.030	0.64, 0.76	**.42*** **	.011	0.39, 0.44
FHO	**0.67*** **	.027	0.62, 0.72	**.41*** **	.010	0.39, 0.43
FHN	**0.67*** **	.028	0.62, 0.73	**.44*** **	.012	0.42, 0.53
RNPGA	**0.77*** **	.049	0.69, 0.89	**.47*** **	.027	0.42, 0.53
CCM	**0.66*** **	.034	0.60, 0.73	**.38*** **	.016	0.35, 0.42
other PC model	**0.83*** **	.038	0.76, 0.91	**.51*** **	.021	0.47, 0.55
PC-specialty model	**1.37*** **	.092	1.20, 1.56	1.02	.038	0.95, 1.10
Male MD	**0.84** **	.065	0.72, 0.97	**.86** **	.065	0.74, 0.99
MD age*male	**1.00** **	.001	1.00, 1.01	**1.01*** **	.001	1.00, 1.01
Medical school location (Canada)						
USA	**0.81* **	.090	0.66, 1.02	1.07	.091	0.91, 1.27
UK	**0.92* **	.043	0.84, 1.01	1.04	.056	0.94, 1.16
Other school	**0.79*** **	.019	0.76, 0.83	**.79*** **	.020	0.75, 0.83
Years of practice (<=5 years)						
5<years of practice<=10	**0.91*** **	.024	0.86, 0.96	**1.11*** **	.041	1.03, 1.19
10<years of practice<=20	**0.88*** **	.023	0.83, 0.93	**1.10*** **	.039	1.03, 1.18
20<years of practice<=30	**0.91*** **	.025	0.86, 0.96	**1.15*** **	.035	1.09, 1.22
30<years of practice<=40	0.95	.036	0.88, 1.02	**1.26*** **	.045	1.18, 1.35
Years of practice>40	1.05	.062	0.94, 1.18	**1.65*** **	.087	0.09, 0.10

Patient male	**1.02*** **	.006	1.01, 1.03	**1.05*** **	.009	1.03, 1.07
Patient age (40<=patient age<60)						
Patient age<20	**0.65*** **	.027	0.59, 0.70	**.73*** **	.030	0.67, 0.79
20<=patient age<40	**0.83*** **	.008	0.81, 0.84	**.83*** **	.009	0.82, 0.85
60<= patient age<80	**1.54*** **	.011	1.51, 1.56	**1.56*** **	.015	1.53, 1.59
patient age>=80	**2.04*** **	.032	1.98, 2.11	**2.04*** **	.035	1.98, 2.11
Daily visits	**0.99*** **	.003	0.99, 1.00	**.99*** **	.002	0.99, 1.00
(daily visits)^2	**1.00*** **	.000	1.00, 1.00	**1.00*** **	.000	1.00, 1.00
Daily services	**1.01*** **	.001	1.00, 1.01	.99	.001	0.99, 1.00
(Daily services)^2	**1.00*** **	.000	1.00, 1.00	**1.00*** **	.000	1.00, 1.00
Year fixed effects (2006)						
2007	**0.95*** **	.004	0.95 0.96	**.94*** **	.004	0.94, 0.96
2008	**0.92*** **	.006	0.90, 0.92	**.90*** **	.006	0.89, 0.91
2009	**0.92*** **	.007	0.90, 0.93	**.90*** **	.007	0.89, 0.92
2010	**0.85*** **	.008	0.83, 0.86	**.85*** **	.007	0.83, 0.86
Month fixed effects (January)						
February	**0.94*** **	.003	0.93, 0.95	**.93*** **	.005	0.93, 0.94
March	**0.96*** **	.003	0.95, 0.97	**.94*** **	.004	0.93, 0.95
April	**1.01*** **	.004	1.00, 1.02	**1.02*** **	.005	1.02, 1.04
May	1.00	.003	0.99, 1.01	**1.02*** **	.005	1.02, 1.03
June	**1.03*** **	.004	1.02, 1.04	**1.04*** **	.005	1.03, 1.05
July	**1.05*** **	.004	1.03, 1.05	**1.05*** **	.005	1.04, 1.06
August	1.00	.003	0.99, 1.01	1.00	.005	0.99, 1.01
September	**0.94*** **	.003	0.93, 0.95	**.96*** **	.005	0.95, 0.96
October	**0.94*** **	.003	0.94, 0.95	**.95*** **	.005	0.94, 0.96
November	**0.97*** **	.003	0.95, 0.97	**.98*** **	.005	0.97, 1.00
December	**1.13*** **	.004	1.12, 1.14	**1.11*** **	.006	1.02, 1.13
Constant	**0.13*** **	.007	0.12, 0.14			
Pseudo R-squared		.0359				
Observations		74,459,516			74,459,516	
Clusters (number of physicians)		22,242			22,242	

Heteroskedasticity consistent standard errors are employed. (* p<0.10, ** p<0.05, *** p<0.01).

Appendix 3.1

Table A.3.1.1 List of independent variables

Variable	**Description**
Major urban	physician practices in major urban area; reference group
rural	physician practices in rural area
Nonmajor urban	physician practices in non-major urban area
Specialist	Physician has a specialty other than primary care
PC	physician’s specialty is primary care
PC with specialty	physician has an additional specialty along with primary care
GP specialist	Physician is a general practitioner (GP) specialist
FFS	physician is in fee-for-service practice; reference group
FHG	physician is in a Family Health Group
FHO	physician is in a Family Health Organization
FHN	physician is in a Family Health Network
RNPGA	physician is in a Rural and Northern Physician Group Agreement
CCM	physician is in a Comprehensive Care Model
Other PC model	physician is in another primary care model, other than the ones mentioned above
PC-spec model	physician is in a specialized primary care payment model
Male MD	Equals 1 if physician is male and 0 if female
md_age*male	interaction variable between MD age and sex
MD school-Canada	graduated from a medical school in Canada; reference group
MD school-USA	graduated from a medical school in the United States
MD school-UK	graduated from a medical school in the United Kingdom
MD school-other	graduated from a medical school in any other country
yrs_prac<=5	years of practice is less than or equal to 5; reference group
5<yrs_prac<=10	years of practice is greater than 5 but less than or equal to 10
10<yrs_prac<=20	years of practice is greater than 10 but less than or equal to 20
20<yrs_prac<=30	years of practice is greater than 20 but less than or equal to 30
30<yrs_prac<=40	years of practice is greater than 30 but less than or equal to 40
Years>40	years of practice is greater than 40
Patient male	equals 1 if the lab was ordered for a male patient and equals 0 if the lab was ordered for a female patient
Patient age<20	lab was ordered for a patient younger than 20 years
20<=patient age<40	lab was ordered for a patient aged between 19 and 40
40<=patient age<60	lab was ordered for a patient aged between 39 and 60; reference group
60<=patient age<80	lab was ordered for a patient aged between 59 and 80
patient age>=80	lab was ordered for a patient equal to or older than 80 years
Daily visits	number of daily visits in the year
Daily services	number of daily services in the year
year dummy variables	5 dummy variables for years 2006-2010; reference group is 2006

month dummy variables	12 dummy variables for the month the lab test is ordered; reference group is January

Table A.3.1.2: All lab tests (20% sample)

Logistic regression	Number of obs = 14,887,924
	Wald chi2(46) = 12283.24
	Prob > chi2 = 0.0000
Log pseudolikelihood = -4829254	Pseudo R2 = 0.0359

(Std. Err. adjusted for 20,321 clusters in ref_doc_no)

inapp_test	Odds Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
rural	1.02593	.0324508	0.81	0.418	.9642588	1.091545
nonmajor_urban	1.056648	.0193872	3.00	0.003	1.019325	1.095337
specialist	2.009355	.0967187	14.50	0.000	1.828457	2.20815
PC_MD_spec	1.185174	.0738468	2.73	0.006	1.048925	1.339119
GP_spec	1.064288	.050083	1.32	0.185	.970518	1.167118
FHG	.697493	.0299691	-8.38	0.000	.64116	.7587755
FHO	.6660615	.0270471	-10.01	0.000	.6151048	.7212396
FHN	.6685733	.0278748	-9.66	0.000	.6161123	.7255013
RNPGA	.7814687	.0504124	-3.82	0.000	.6886535	.8867933
CCM	.6607177	.0340385	-8.04	0.000	.597261	.7309164
other_PC	.829878	.0381843	-4.05	0.000	.7583136	.9081962
other_PC_spec	1.365694	.0920296	4.62	0.000	1.196723	1.558522
md_male	.8343703	.0647461	-2.33	0.020	.7166493	.9714289
md_age_male	1.003261	.0014676	2.23	0.026	1.000389	1.006142
md_school_USA	.8222961	.0908622	-1.77	0.077	.6621738	1.021138
md_school_UK	.9234855	.0433577	-1.70	0.090	.8422986	1.012498
md_school_other	.7942146	.0190034	-9.63	0.000	.7578286	.8323477
yrs_prac_10	.9085485	.0244643	-3.56	0.000	.8618427	.9577855
yrs_prac_20	.8787455	.0235043	-4.83	0.000	.8338646	.926042
yrs_prac_30	.9083724	.0256218	-3.41	0.001	.8595175	.9600043
yrs_prac_40	.9518569	.0361957	-1.30	0.194	.8834938	1.02551
yrs_prac_40plus	1.051373	.0623936	0.84	0.399	.9359277	1.181058
pat_male	1.01795	.005694	3.18	0.001	1.00685	1.029171
pat_age_0_19	.6459865	.0276282	-10.22	0.000	.5940438	.7024711
pat_age_20_39	.8288456	.0080214	-19.40	0.000	.8132721	.8447174
pat_age_60_79	1.536828	.0113377	58.25	0.000	1.514767	1.559211
pat_age_80plus	2.042086	.032253	45.20	0.000	1.97984	2.10629
daily_vis	.9931056	.0027177	-2.53	0.011	.9877932	.9984465
daily_vis2	1.000078	.0000138	5.64	0.000	1.000051	1.000105
daily_srv	1.005747	.001377	4.19	0.000	1.003052	1.00845
daily_srv2	.9999794	3.73e-06	-5.51	0.000	.9999721	.9999867
yr7	.9540716	.0044802	-10.01	0.000	.9453308	.9628933
yr8	.9117373	.0062452	-13.49	0.000	.8995788	.9240601
yr9	.9164157	.0074065	-10.80	0.000	.9020136	.9310479
yr10	.8486644	.0080593	-17.28	0.000	.8330146	.8646082
feb	.9415057	.0045318	-12.52	0.000	.9326652	.95043
mar	.9583777	.0045549	-8.95	0.000	.9494918	.9673468
apr	1.007227	.0051268	1.41	0.157	.9972285	1.017326
may	1.00396	.0051577	0.77	0.442	.9939013	1.01412
june	1.02836	.005161	5.57	0.000	1.018294	1.038525
july	1.044514	.0052774	8.62	0.000	1.034222	1.054909
aug	.9959923	.0050985	-0.78	0.433	.9860494	1.006035
sept	.9400369	.0046021	-12.63	0.000	.93106	.9491003
oct	.944158	.0045966	-11.80	0.000	.9351917	.9532103
nov	.9634985	.0047045	-7.62	0.000	.9543218	.9727634
dec	1.128315	.0056283	24.20	0.000	1.117337	1.1394
_cons	.1289476	.0066195	-39.90	0.000	.1166049	.1425967

Table A.3.1.3: Male physicians (20% random sample)

```
Logistic regression                              Number of obs     = 10,216,833
                                                 Wald chi2(44)     =    8125.19
                                                 Prob > chi2       =     0.0000
Log pseudolikelihood =   -3410258                Pseudo R2         =     0.0348

                           (Std. Err. adjusted for 13,178 clusters in ref_doc_no)
```

inapp_test	Odds Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
rural	.9991051	.0400017	-0.02	0.982	.9237006	1.080665
nonmajor_urban	1.044529	.0244463	1.86	0.063	.9976978	1.093559
specialist	1.870051	.1192633	9.82	0.000	1.650319	2.119041
PC_MD_spec	1.227103	.0891958	2.82	0.005	1.064164	1.414989
GP_spec	.9444994	.0543261	-0.99	0.321	.8438047	1.05721
FHG	.6995833	.0385215	-6.49	0.000	.628014	.7793087
FHO	.6413442	.0335508	-8.49	0.000	.5788448	.7105919
FHN	.6521489	.0350179	-7.96	0.000	.5870032	.7245245
RNPGA	.7297125	.0628972	-3.66	0.000	.6162869	.8640137
CCM	.6508538	.0419421	-6.66	0.000	.5736284	.7384758
other_PC	.8312678	.0485033	-3.17	0.002	.7414375	.9319817
other_PC_spec	1.272817	.0976318	3.14	0.002	1.095152	1.479305
md_school_USA	.7515861	.0893646	-2.40	0.016	.5953465	.9488285
md_school_UK	.9398029	.0460638	-1.27	0.205	.8537205	1.034565
md_school_other	.8266674	.023566	-6.68	0.000	.7817455	.8741707
yrs_prac_10	.9358723	.0366776	-1.69	0.091	.8666771	1.010592
yrs_prac_20	.9127376	.0333064	-2.50	0.012	.8497379	.9804081
yrs_prac_30	.9874032	.0341322	-0.37	0.714	.9227213	1.056619
yrs_prac_40	1.035165	.0380792	0.94	0.347	.9631582	1.112555
yrs_prac_40plus	1.191289	.0691951	3.01	0.003	1.063104	1.33493
pat_male	1.000945	.0067622	0.14	0.889	.9877791	1.014287
pat_age_0_19	.6428863	.0342298	-8.30	0.000	.5791791	.713601
pat_age_20_39	.7712198	.0102489	-19.55	0.000	.7513917	.7915711
pat_age_60_79	1.516495	.013616	46.38	0.000	1.490042	1.543418
pat_age_80plus	2.015527	.0399203	35.39	0.000	1.938784	2.095308
daily_vis	.9907766	.0033394	-2.75	0.006	.9842531	.9973434
daily_vis2	1.000085	.0000177	4.82	0.000	1.000051	1.00012
daily_srv	1.006098	.0016282	3.76	0.000	1.002912	1.009294
daily_srv2	.9999794	4.44e-06	-4.64	0.000	.9999706	.9999881
yr7	.9488179	.0052843	-9.43	0.000	.9385172	.9592316
yr8	.9138324	.0076444	-10.77	0.000	.8989717	.9289387
yr9	.9190596	.008965	-8.65	0.000	.9016553	.9367998
yr10	.8533273	.010119	-13.38	0.000	.8337232	.8733924
feb	.9495963	.0053837	-9.12	0.000	.9391029	.9602071
mar	.952424	.0054366	-8.54	0.000	.9418278	.9631395
apr	1.00633	.0063282	1.00	0.316	.9940029	1.01881
may	.9977831	.0061158	-0.36	0.717	.985868	1.009842
june	1.027724	.0062225	4.52	0.000	1.015601	1.039993
july	1.051277	.0064244	8.18	0.000	1.038761	1.063945
aug	.9974682	.0062358	-0.41	0.685	.9853207	1.009765
sept	.9393024	.0054859	-10.72	0.000	.9286115	.9501163
oct	.9380377	.0052509	-11.43	0.000	.9278024	.9483859
nov	.9727126	.0057801	-4.66	0.000	.9614495	.9841075
dec	1.131162	.0068944	20.22	0.000	1.117729	1.144756
_cons	.1310599	.0092582	-28.77	0.000	.1141144	.1505218

Table A.3.1.4: Female physicians (20% random sample)

```
Logistic regression                               Number of obs     =  4,679,352
                                                  Wald chi2(44)     =    5776.13
                                                  Prob > chi2       =     0.0000
Log pseudolikelihood =   -1416748                 Pseudo R2         =     0.0376

                         (Std. Err. adjusted for 7,131 clusters in ref_doc_no)
```

inapp_test	Odds Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
rural	1.00584	.045032	0.13	0.897	.9213404	1.098089
nonmajor_urban	1.066096	.0253942	2.69	0.007	1.017468	1.117048
specialist	2.343122	.1564225	12.75	0.000	2.05575	2.670666
PC_MD_spec	1.014569	.0730939	0.20	0.841	.8809626	1.168438
GP_spec	1.246363	.1172306	2.34	0.019	1.03653	1.498673
FHG	.7194593	.0460947	-5.14	0.000	.6345576	.8157206
FHO	.7541346	.0464528	-4.58	0.000	.66837	.8509044
FHN	.7383896	.0464899	-4.82	0.000	.6526689	.8353688
RNPGA	.9378489	.0843238	-0.71	0.475	.7863207	1.118577
CCM	.7122553	.0597746	-4.04	0.000	.6042273	.8395975
other_PC	.8478763	.0604556	-2.31	0.021	.7372926	.9750462
other_PC_spec	1.846962	.2304814	4.92	0.000	1.446229	2.358734
md_school_USA	1.00907	.0990679	0.09	0.927	.8324393	1.223179
md_school_UK	.8987128	.0991202	-0.97	0.333	.7240038	1.115581
md_school_other	.7516342	.0244167	-8.79	0.000	.70527	.8010463
yrs_prac_10	.9052769	.0308703	-2.92	0.004	.8467498	.9678493
yrs_prac_20	.8737999	.0298656	-3.95	0.000	.8171819	.9343406
yrs_prac_30	.8490187	.0286067	-4.86	0.000	.7947618	.9069796
yrs_prac_40	.9766577	.0445746	-0.52	0.605	.8930866	1.068049
yrs_prac_40plus	.9857371	.1207358	-0.12	0.907	.7753603	1.253195
pat_male	1.053251	.0094063	5.81	0.000	1.034975	1.071849
pat_age_0_19	.6474901	.0443808	-6.34	0.000	.5660951	.7405884
pat_age_20_39	.9275469	.0125586	-5.55	0.000	.9032562	.9524909
pat_age_60_79	1.576798	.0180198	39.85	0.000	1.541873	1.612515
pat_age_80plus	2.068018	.0379792	39.56	0.000	1.994904	2.143812
daily_vis	1.006417	.0059506	1.08	0.279	.9948214	1.018148
daily_vis2	.9999637	.0000605	-0.60	0.548	.9998452	1.000082
daily_srv	.9981086	.0040096	-0.47	0.637	.9902807	1.005998
daily_srv2	1.000038	.0000335	1.12	0.261	.999972	1.000103
yr7	.949263	.0089823	-5.50	0.000	.9318202	.9670324
yr8	.9004908	.0105278	-8.97	0.000	.8800913	.9213631
yr9	.8961098	.012501	-7.86	0.000	.8719402	.9209494
yr10	.8221034	.0122668	-13.13	0.000	.798409	.8465009
feb	.9433995	.0082392	-6.67	0.000	.9273884	.9596871
mar	.9612072	.0080049	-4.75	0.000	.9456453	.9770252
apr	1.036489	.0093675	3.97	0.000	1.01829	1.055012
may	1.028212	.0091888	3.11	0.002	1.010359	1.04638
june	1.043182	.0094189	4.68	0.000	1.024883	1.061807
july	1.056018	.0093279	6.17	0.000	1.037893	1.07446
aug	1.017963	.0094187	1.92	0.054	.9996694	1.036592
sept	.9462799	.0083582	-6.25	0.000	.9300392	.9628043
oct	.9470663	.0084489	-6.10	0.000	.9306507	.9637715
nov	.9761104	.0083956	-2.81	0.005	.9597933	.9927049
dec	1.129107	.0106481	12.88	0.000	1.108428	1.15017
_cons	.1121179	.0081496	-30.10	0.000	.0972305	.1292847

Table A.3.1.5: LPM Model – All Tests
Dependent Variable: Inappropriate Test

	Conditional Model			Unconditional Model		
	coefficient	robust s.e	95% confidence interval	coefficient	robust s.e	95% confidence interval
Practice location (major urban)						
Rural	.001	.0028	-.005, .006	-.001	.0030	-.007, .005
Nonmajor urban	**.005*****	.0017	.002, .008	**.006*****	.0021	.002, .010
Physician specialty (primary care)						
Specialist	**.088*****	.0060	.077, .100	**.117*****	.0042	.109, .125
PC with specialty	**.016*****	.0061	.004, .028	**.081*****	.0042	.073, .090
GP specialist	.002	.0042	-.006, .010	.004	.0050	-.006, .013
Primary care physician payment model (FFS)						
FHG	**-.033*****	.0046	-.042, -.024	**-.102*****	.0037	-.109, -.095
FHO	**-.037*****	.0044	-.046, -.029	**-.104*****	.0036	-.111, -.097
FHN	**-.038*****	.0044	-.047, -.029	**-.097*****	.0037	-.105, -.090
RNPGA	**-.026*****	.0061	-.038, -.014	**-.094*****	.0059	-.106, -.082
CCM	**-.038*****	.0051	-.048, -.028	**-.108*****	.0043	-.117, -.099
other PC model	**-.018*****	.0049	-.028, -.009	**-.084*****	.0048	-.094, -.075
PC-specialty model	**.048*****	.0079	.032, .063	.004	.0057	-.008, .015
Male MD	**-.016*****	.0071	-0.030, -.002	**-.016****	.0074	-.031, -.002
MD age*male	**.0003****	.0001	.000, .001	**.001*****	.0001	.0003, .0008
Medical school location (Canada)						
USA	**-.021***	.0122	-.044, .001	.006	.0088	-.011, .023
UK	**-.008***	.0046	-.017, .001	.004	.0055	-.007, .015
Other school	**-.020*****	.0021	-.024, -.016	**-.022*****	.0021	-.026, -.018
Years of practice (<=5 years)						
5<years of practice<=10	**-.008*****	.0022	-.012, -.004	**.009*****	.0031	.002, .015
10<years of practice<=20	**-.012*****	.0022	-.017, -.008	**.008*****	.0030	.003, .014
20<years of practice<=30	**-.010*****	.0024	-.015, -.005	**.012*****	.0025	.008, .017
30<years of practice<=40	**-.006***	.0033	-.012, .001	**.021*****	.0032	.015, .028
Years of practice>40	.006	.0060	-.006, .018	**.051*****	.0059	.039, .062
Patient male	**.002*****	.0005	.001, .003	**.005*****	.0008	.003, .006
Patient age (40<=patient age<60)						
Patient age<20	**-.033*****	.0027	-.038, -.028	**-.022*****	.0026	-.027, -.016
20<=patient age<40	**-.014*****	.0007	-.015, -.012	**-.013*****	.0007	-.015, -.012
60<= patient age<80	**.040*****	.0007	.039, .042	**.043*****	.0009	.041, .044
patient age>=80	**.075*****	.0019	.071, .079	**.076*****	.0020	.072, .080
Daily visits	**-.001*****	.0003	-.002, -.001	**-1.22*****	.2235	-1.66, -.787
(daily visits)^2	**.000*****	.0000	.000, .000	**.012*****	.0023	.008, .017
Daily services	**.001*****	.0001	.000, .001	**-.315*****	.1416	-.593, -.038

(Daily services)^2	**.000*** **	.0000	.000, .000	**.004*** **	.0010	.002, .005
Year fixed effects (2006)						
2007	**-.005*** **	.0004	-.005, -.004	**-.005*** **	.0004	-.006, -.004
2008	**-.008*** **	.0006	-.010, -.007	**-.009*** **	.0006	-.011, -.008
2009	**-.008*** **	.0007	-.010, -.007	**-.009*** **	.0007	-.010, -.008
2010	**-.015*** **	.0008	-.017, -.014	**-.015*** **	.0008	-.017, -.014
Month fixed effects (January)						
February	**-.005*** **	.0002	-.006, -.005	**-.007*** **	.0003	-.007, -.006
March	**-.004*** **	.0003	-.004, -.003	**-.005*** **	.0003	-.006, -.005
April	**.001*** **	.0003	.000, .002	**.002*** **	.0003	.002, .003
May	.0004	.0003	-.000, .001	**.002*** **	.0003	.002, .003
June	**.002*** **	.0003	.002, .003	**.003*** **	.0003	.003, .004
July	**.005*** **	.0003	.004, .005	**.004*** **	.0003	.004, .005
August	.0001	.0003	-.001, .001	**-.001* **	.0003	-.001, .000
September	**-.005*** **	.0003	-.006, -.005	**-.004*** **	.0003	-.005, -.003
October	**-.006*** **	.0003	-.006, -.005	**-.005*** **	.0003	-.005, -.004
November	**-.003*** **	.0003	-.003, -.002	**-.002*** **	.0003	-.003, -.002
December	**.012*** **	.0003	.011, .012	**.011*** **	.0003	.010, .011
Constant	**.123*** **	.0052	.112, .133			
Pseudo R-squared		.0270				
Observations		74,459,516			74,459,516	
Clusters (number of physicians per year)		22,242			22,242	

(* p<0.10, ** p<0.05, *** p<0.01).

Table A.3.1.6 LPM by each independent variable

X variable	Practice location	Specialty	GP spec.	Payment model	MD sex	MD age	MD school	Years of practice	Patient sex	Patient age
Rural	-.0010									
Nonmajor urban	**.0060*** **									
Specialist		**.1169*** **								
PC w/ specialty		**.0813*** **								
GP specialist			-.0037							
FHG				**-.1019*** **						
FHO				**-.1038*** **						
FHN				**-.0974*** **						
RNPGA				**-.0941*** **						
CCM				**-.1083*** **						
other PC model				**-.0844*** **						
PC-spec model				.0037						
Male					**.0143*** **					
35<age<=45						**.0083** **				
45<age<=55						**.0066** **				
55<age<=65						**.0090*** **				
65<age<=75						**.0202*** **				
age>75						**.0437*** **				
USA							.0061			
UK							.0041			
other school							**-.0218*** **			
5<yrs prac<=10								**.0086*** **		
10<yrs prac<=20								**.0084*** **		
20<yrs prac<=30								**.0125*** **		
30<yrs prac<=40								**.0212*** **		
Yrs prac>40								**.0505*** **		
Male patients									**.0048*** **	
Patient age<20										**-.0215*** **
20<=patient age<40										**-.0131*** **
60<=patient age<80										**.0426*** **
Patient age>=80										**.0763*** **

Table A.3.1.6 (continued)

LPM by each independent variable

X variable	Daily visits/1000	Daily services/1000	Year	Month
Daily visits	**-1.225*** **			
(daily visits)^2	**.0125*** **			
Daily services		**-.3155** **		
(Daily services)^2		**.0036*** **		
2007			**-.0048*** **	
2008			**-.0094*** **	
2009			**-.0090*** **	
2010			**-.0151*** **	
February				**-.0068*** **
March				**-.0054*** **
April				**.0024*** **
May				**.0022*** **
June				**.0034*** **
July				**.0044*** **
August				**-.0006* **
September				**-.0039*** **
October				**-.0046*** **
November				**-.0021*** **
December				**.0105*** **

(* p<0.10, ** p<0.05, *** p<0.01).

Appendix 3.2: Cost of Inappropriate Test Utilization in Ontario

Table A.3.2.1: TSH Lab Tests

Fiscal Year	Cost of labs fulfilled	Cost per capita	Cost of inapp. lab tests	% Cost of inapp. lab tests[a]
2006	38,159,620	2.880	2,958,336	0.078
2007	41,369,189	3.087	3,179,960	0.077
2008	49,047,686	3.616	3,909,203	0.080
2009	51,946,245	3.795	4,216,255	0.081
2010	50,208,932	3.640	3,866,671	0.077

Table A.3.2.2: HbA1c Lab Tests

Fiscal Year	Cost of labs fulfilled	Cost per capita	Cost of inapp. lab tests	% Cost of inapp. lab tests[a]
2006	4,304,415	0.325	887,543	0.206
2007	4,560,918	0.340	866,998	0.190
2008	5,142,666	0.380	863,640	0.168
2009	5,651,870	0.413	940,908	0.166
2010	5,441,414	0.395	852,112	0.157

Table A.3.2.3: Lipid Profile Lab Tests

Fiscal Year	Cost of labs fulfilled	Cost per capita	Cost of inapp. lab tests	% Cost of inapp. lab tests[a]
2006	11,245,809	0.844	1,211,317	0.108
2007	11,533,964	0.858	1,227,437	0.106
2008	12,143,168	0.893	1,234,661	0.102
2009	12,603,554	0.921	1,293,985	0.103
2010	12,579,845	0.914	1,168,077	0.093

Table A.3.2.4: Serum Protein Electrophoresis Lab Tests

Fiscal Year	Cost of labs fulfilled	Cost per capita	Cost of inapp. lab tests	% Cost of inapp. lab tests[a]
2006	2,222,001	0.168	140,067	0.063
2007	2,294,422	0.171	141,548	0.062
2008	2,396,384	0.177	152,752	0.064
2009	2,465,543	0.180	154,500	0.063
2010	2,415,966	0.175	150,467	0.062

Table A.3.2.5: Immunofixation Lab Tests

Fiscal Year	Cost of labs fulfilled	Cost per capita	Cost of inapp. lab tests	% Cost of inapp. lab tests[a]
2006	3,993,701	0.301	404,935	0.101
2007	4,358,868	0.325	428,200	0.098
2008	4,563,973	0.337	469,084	0.103
2009	4,765,417	0.348	450,845	0.095
2010	4,509,874	0.327	420,631	0.093

Table A.3.2.6: Quantitative Immunoglobulin Lab Tests

Fiscal Year	Cost of labs fulfilled	Cost per capita	Cost of inapp. lab tests	% Cost of inapp. lab tests[a]
2006	1,210,816	0.091	103,418	0.085
2007	1,270,910	0.095	108,368	0.085
2008	1,389,048	0.103	116,765	0.084
2009	1,516,320	0.111	130,756	0.086
2010	1,526,710	0.111	128,234	0.084

Table A.3.2.7: Vitamin D Lab Tests

Fiscal Year	Cost of labs fulfilled	Cost per capita	Cost of inapp. lab tests	% Cost of inapp. lab tests[a]
2006	3,609,280	0.272	262,429	0.073
2007	7,699,578	0.574	493,166	0.064
2008	21,739,178	1.604	1,473,036	0.068
2009	38,146,018	2.789	2,581,950	0.068
2010	24,886,932	1.804	1,781,685	0.072

Table A.3.2.8: Vitamin B12 Lab Tests

Fiscal Year	Cost of labs fulfilled	Cost per capita	Cost of inapp. lab tests	% Cost of inapp. lab tests[a]
2006	17,195,691	1.296	1,540,509	0.090
2007	21,856,622	1.630	1,650,213	0.076
2008	35,282,924	2.604	2,921,688	0.083
2009	41,494,351	3.034	3,571,391	0.086
2010	38,398,199	2.784	2,958,697	0.077

Table A.3.2.9: Folate Lab Tests

Fiscal Year	Cost of labs fulfilled	Cost per capita	Cost of inapp. lab tests	% Cost of inapp. lab tests[a]
2006	8,665,961	0.653	1,696,615	0.196
2007	8,821,309	0.658	1,718,838	0.195
2008	7,632,844	0.563	1,506,418	0.197
2009	6,730,155	0.492	1,394,787	0.207
2010	5,233,618	0.379	1,041,987	0.199

[a] (cost of inappropriate tests)/(total cost of all tests for each analyte)
Inapp.: inappropriate

Chapter Four

Payment Models in Primary Healthcare: A Driver of the Quantity and Quality of Medical Laboratory Utilization

4.1 Introduction

Payment models' influences on various aspects of medical practice is the subject of a large literature (e.g., Barham and Milliken, 2015; Barros, 2003; Clemens and Gottlieb, 2014; Helmchen and Lo Sasso, 2010), however the structure of remuneration's impacts on some important facets of clinical practice are still not understood. Addressing one such issue – the impacts of remuneration's structure on the quantity (number and cost) and especially the quality (appropriateness) of clinical laboratory (lab) testing – we focus on general practitioners (GPs) switching from fee-for-service enhanced with pay-for-performance (EFFS) to a blended capitation model, with some of those in blended capitation subsequently shifting to interdisciplinary healthcare teams.[21] As pointed out by Chalkley and Listl (2018) who study the structure of remuneration's impact on the incidence of X-rays in dentistry, the imperative to balance clinical harms and benefits should not be influenced by payment models. The ordering of laboratory tests provides a well-defined context to assess payment models' impacts on clinical practices that have no

[21] A related economic literature addressing medical screening and imaging does exist (e.g., Abaluck et al. 2016; Aas 2009; Hackl et al. 2015), but we focus on clinical laboratories.

direct financial impact on physicians, but where testing may be redundant/costly or even, using our definition of inappropriate, potentially produce biased evidence regarding patient health.[22]

Given that there is no financial gain or loss to the primary care physician ordering the laboratory test, how can payment models affect a physician's decision when ordering a test? The shift to a more patient-centered blended capitation model may play an important role in the utilization of medical services. Zhang and Sweetman's evidence of more exclusive GP-patient relationships in the physician's practice is related to the concept of a "medical home" (e.g., Agha, Frandsen, and Rebitzer, 2017; McGuire, 2011; Barros, 2003; Ellis and McGuire, 1986). It is expected that subsequent to joining the blended capitation payment model, physician laboratory utilization would decrease as patient care became more focused on a single practice or with a single GP, and GPs became more familiar with their patients' medical history. During the period under study, the province also introduced an interdisciplinary health team practice model open only to physicians paid by capitation (or salary, but physicians in a salary payment model are not included in this study), which allows us to compare an additional alternative. Interdisciplinary care teams face a trade-off: their introduction might improve testing efficiency as a result of specialization and increased contact time, or efficiency may decrease because of coordination complexities.

[22] Our measure of inappropriateness focuses on tests that are ordered too soon. Not ordering useful tests is also problematic, but the data available to us cannot identify this.

Alexander (2012) argues that primary care clinical laboratory utilization is particularly important since GPs initiate a large share of healthcare costs and most of their decisions are influenced by test results. Both over- and underutilization are problematic, and a large literature in medicine examines approaches to improving appropriateness and efficiency (e.g., Naugler, 2014; Zhi, et al., 2013). There is also recent interest in this topic because of the Choosing Wisely campaign, which addresses clinical technology utilization with a focus on reducing overuse (Bhatia et al., 2015; Levinson et al., 2015).[23]

We use Canadian administrative data from the province of Ontario, Canada to assess whether switching from an EFFS payment system with premiums for voluntary patient enrolment (also called rostering) to a blended capitation payment system with mandatory patient rostering, and in some cases also an interdisciplinary health team, affects the volume and/or (in)appropriateness of primary care physicians' utilization of out-of-hospital medical laboratory tests. Importantly, we have data on effectively all practicing physicians in the province so we can examine system-wide responses to the payment model change. The transitions on which we focus are part of a program of primary care reform surveyed by Marchildon and Hutchison (2016), McLeod, Buckley, and Sweetman (2016), and Sweetman and Buckley (2012). In fiscal year 2000/01 about 98% of GPs were paid traditional fee-for-service (FFS), but by 2013/14 this had fallen to around 28%, with the remainder voluntarily opting into one of the reformed payment models that was sequentially introduced.

[23] For the US see http://www.choosingwisely.org/, and for Canada see http://www.choosingwiselycanada.org/.

Most research focuses on the largest transitions: traditional FFS to EFFS, EFFS to blended capitation, and blended capitation to capitation plus an interdisciplinary health team. We address the last two transitions, since those three models roster patients. Relevant research on this program of reform includes Rudoler et al. (2015) and Kantarevic and Kralj (2016) who study GP choice among the menu of primary care payment models available in Ontario and observe important selection on observables. Li, Hurley, DeCicca, and Buckley (2014) focus on the introduction of the EFFS model in 2003 and find that the shift from traditional FFS to FFS with pay-for-performance had negligible to modest impacts on the targeted services in the short run. In contrast, Kantarevic and Kralj (2013) focus on diabetes management and the subsequent transition from EFFS to the blended capitation model introduced in 2006, and observe increased use of the diabetes billing code. Kralj and Kantarevic (2013) further look at changes in the total number of services provided per patient caused by the EFFS to blended capitation shift. They observe a short-run reduction of about 7% together with improvements in incentivized preventative care targets. Focusing on the same transition, but addressing the details of the blended model, which pays rostering capitated physicians a blend for fee codes inside a capitated basket combined with FFS for codes outside the basket, Zhang and Sweetman (2018) find that GPs in the capitation model reduce billings inside the capitated basket while increasing (shifting to) billing outside of it. Simultaneously, services provided by GPs outside of the rostering group diminish, providing some evidence of improved continuity of care. Additional relevant areas of interest are referrals to specialists for, and emergency department use by, rostered patients (e.g., Sarma, Mehta, Devlin, Li, and Kpelitse, 2016; Glazier, Zagorski, and Rayner, 2012). Ontario's blended capitation does not financially penalize either activity and, they appear to be used as mechanisms to shift in-the-basket work outside of the primary care practice. Overall, with the partial exception of Li et al. (2014) studying the short run for the earliest transition, all the studies discussed find modest GP behavioral responses as a function of payment system incentives.

To address physician selection into the new payment model, we employ propensity score weighting in the context of a difference-in-differences fixed-effects model with a useful, and we argue sufficient, set of covariates to satisfy the conditional independence assumption. To address patient selection into primary care practices we focus on those continuously rostered with the same physicians throughout the study period. Notably, studying the same policy change, Kantarevic and Kralj (2014) find no evidence of cream skimming (patient "dumping" from blended capitation), which if present would threaten a causal interpretation. This is a particular set of patients, but it allows a well-defined economic question to be clearly answered. Readers who judge the identifying assumptions to be met may interpret the results as causal,[24] while others may view them as well controlled conditional covariances. Inference is conducted using a non-parametric clustered percentile-t bootstrap. To provide context, we also present results for other sets of patients seen by each GP, although these do not have a causal interpretation since the patient population is shifting endogenously.

[24] Discussions of econometric and interpretational issues are by Athey and Imbens (2017) and Imbens (2015).

For continuously rostered patients, and focussing on quantity, we estimate that switching to blended capitation causes a reduction in testing. The decline comes entirely from physicians other than the rostering GP, which we interpret as improving continuity of care/medical home status. No additional change in testing rates is observed for GPs who subsequently shift to the interdisciplinary capitated model. Turning to the (arguably more important) quality side, a diabetes-related laboratory test is used as a case study since it is both common and clinically significant, and there is a clear guideline categorizing as inappropriate a test that is within three months of a previous test (we classify a test that follows too closely on the heels of an earlier test as inappropriate; other forms of inappropriateness exist, but cannot be measured in our data). Blended capitation, both without and especially with an interdisciplinary team, causes a reduction in the proportion of inappropriately ordered tests, suggesting improved quality of care caused by a change in the payment model.

For patients only rostered for part of the period under study, an increase in all types of testing is observed by newly capitated GPs not in interdisciplinary practice. On the quality side, despite no change in the number of the diabetes tests ordered, the share of inappropriate tests increases. One possible explanation is that newly rostered patients are given more extensive initial examinations in the blended model than the EFFS one. While the findings are interesting, in contrast to continuously rostered patients there is no causal interpretation given the shifting patient population, so interpretation is fraught.

The rest of the paper is organized as follows. Section 4.2 provides an institutional background and a brief descriptive overview of laboratory utilization in Ontario. Section

4.3 presents the data and Section 4.4 the estimation methods. Results for the "quantity" analysis for all laboratory tests are presented in Section 4.5, and those for the "quality" analysis are in Section 4.6. Section 4.7 discusses the results and concludes.

4.2 Clinical Laboratory and Primary Care

4.2.1 Clinical Laboratory Literature Review

In the US many perceive both significant overuse and inappropriate use of many technologies throughout the healthcare system (Garber and Skinner, 2008; aforementioned Choosing Wisely). Feldman et al. (2013) find that presenting laboratory fees to the physician leads to a modest reduction in testing. However, it is not known to what extent this reduction comprises inappropriate tests.

In the Canadian context Naugler (2014) and Rockey et al. (2014) find disproportionately increasing laboratory costs compared to other medical expenditures for Alberta, while McGrail et al. (2011) and Sivananthan et al. (2012) find an increase in laboratory expenditures in British Columbia, especially for the elderly population in the late 1990s and early 2000s. Turning to objective measures of inappropriateness, a recent study done with the Ontario data source we use identifies the percentage of inappropriate testing done for 9 tests as ranging between 6% and 20% (Chami et al., 2017). A similar study in Alberta of 6 common tests identifies 16% of tests as inappropriate (Morgen and Naugler, 2015).

4.2.2 Institutional Background

Medically necessary hospital and physician services, including laboratory testing, are part of Canada's set of provincial public universal Medicare systems funded entirely through taxation (Lavis, 2016). Patients are not permitted to pay fees, nor can they obtain such tests without a physician's order, and physicians are unlikely to be aware of/consider costs. Essentially all off-reserve, non-federally incarcerated, civilian patients are covered by the system and included in our dataset.

Outpatient laboratory testing in the province of Ontario is done almost entirely at private medical labs reimbursed by submitting claims, capped at an annual maximum, to the Ontario Health Insurance Plan (OHIP) on a per-test basis (Ndegwa 2011). Medical laboratory tests are done out-of-office, and primary care physicians' income is not directly affected by laboratory utilization. There may, however, be an indirect effect, such as a fee-for-service physician generating a subsequent patient visit to review test results.

In July 2003, an EFFS model, known as the Family Health Group (FHG), was introduced in Ontario as part of a series of primary care reforms. These GPs are paid predominantly FFS, but this is enhanced with pay-for-performance (P4P) and incentives to voluntarily roster patients. Patient rostering seeks to establish a one-to-one relationship involving a bilateral commitment between the patient and the primary care physician, where the patient agrees to seek primary care from their rostering physician's group practice, and the physician agrees to provide comprehensive care to the rostered patient (with exceptions for travel/emergencies).

Family Health Organizations (FHOs), a blended capitation model, were introduced November 1st 2006 and unlike the EFFS model rostering regular patients is mandatory. The capitated physician receives a financial penalty in the form of a claw-back in their "access bonus" (an annual amount for each rostered patient) for each service in the capitated basket that rostered patients receive from physicians outside the rostering GP's group. This blended model has a "capitated basket" of core services plus outside the basket FFS billing. The capitated basket contains commonly used billing codes that encompass a large share of services. Further, these models are sometimes referred to as "doubly" blended since in addition to fee codes being inside/outside of the capitated basket, services in the capitated basket also attract a 10% (increased to 15% in 2010) premium based on the FFS fee schedule. For patients not rostered with the capitated group (including the new patients and those of other rostering GPs) each capitated GP bills the public insurer the full FFS amount, but "in the basket" services for non-enrolled patients are capped at $52,883 per year. While physicians and patients face few regulatory restrictions in selecting each other, this restriction puts a limit on the size of the non-capitated portion of the practice that a capitated GP can operate. The blended model also has P4P that is very similar to EFFS. Both are group/team payment models requiring a minimum of three physicians who share medical records, and provide after-hours care.

Family Health Teams (FHTs) are interdisciplinary primary healthcare groups initiated in April 2005 (Hutchison, Levesque, Strumpf, and Coyle, 2011) resembling what are sometimes termed "medical homes". They encourage chronic disease management, counseling, health education and palliative care, and consist of a team of primary care physicians, nurses and interdisciplinary health professionals such as pharmacists, dieticians, and psychologists (Sweetman and Buckley, 2012). This blended capitation plus an interdisciplinary team structure is not a physician payment model, but participating GPs are required to be paid by capitation or salary with the model under study being the most common. Hence, we observe transitions from EFFS to blended capitation, and/or interdisciplinary blended capitation and can address how both capitation and joining an interdisciplinary health team affects a physician's out of hospital laboratory utilization. GPs choosing to switch models systematically differ from those who do not, and our analysis attempts to take this selection into account.

4.3 Data

4.3.1 Data Sources

We employ a census of out-of-hospital laboratory tests ordered by all physicians in Ontario obtained from claims submitted by private medical laboratories to the Ontario Health Insurance Plan (OHIP) for fiscal years 2006/2007 to 2010/2011 (April 1, 2006 to March 31, 2011).[25] These data include the ordering date, test ordered, and (anonymized) indicators that allow patients and ordering physicians to be linked to tests, but not the results of the tests. Data on physician, practice and patient characteristics are from the Ontario Ministry of Health and Long-Term Care (MOHLTC) data sources listed in Appendix 4.1. The unit of analysis is the primary care physician (i.e. GPs). In related work Zhang and Sweetman (2018) initially employ both the physician and the patient

[25] Research ethics was through the Hamilton Integrated Research Ethics Board (#11-086-C).

(clustered by GP) as units of analysis in complementary specifications, but find qualitatively similar results so the focus of this study is on the former approach. We explored this dual approach but rejected it since the results were remarkably similar, adding complexity and length without additional substance. All GPs in the sample are EFFS as of April 1st, 2006 and either remain in the EFFS model or switch to blended capitation by the end of the sample period.[26] A laboratory requisition is a form filled out by the GP indicating which tests are required. In this study, we differentiate between laboratory requisitions completed, tests ordered, and the cost of those tests. Each requisition includes all tests ordered by a GP for the same patient on the same day.

4.3.2 Summary Statistics

The trend in the number of physicians in each primary care model can be seen in Figure 4.1. All 3,200 GPs are in the EFFS model in 2006, but by 2010 about 42% (1,360) switched to one of the blended capitation models, with 1,033 blended capitation GPs, and 327 interdisciplinary blended capitation GPs.

Table 4.1 presents panel (five year) averages of the annual average of each dependent variable. All subsequent tables, except 4.3 and 4.7, have similar structures. For each GP who is in the EFFS model in the first year of our panel, the dependent variables are calculated for sets of patients based on their relationship to that GP. The first column represents all patients, both rostered and non-rostered, seen by each GP. The second and third columns are restricted to the sub-set of continuously rostered patient-GP pairs, with

[26] In order to focus on physicians practicing for the entire sample period, those who join the blended capitation model more than three days after leaving the EFFS model are excluded from the sample. Physicians who switch from EFFS to blended capitation remain in blended capitation for the rest of the sample period.

column 2 showing orders by the rostering physician and column 3 orders by all MDs in the province (both primary care and specialists, including the rostering GP) for the patients in the aforementioned GP-patient pair. This reflects one advantage of our data; we see all labs ordered for all patients by all physicians in the province. For each sampled GP, the next to last column averages across patients rostered less than five years (e.g., new patients) in the years when they are rostered. Finally, for each sampled GP, the rightmost column tabulates tests ordered for patients never rostered with her/him during the sample period (though they may have been rostered with a different GP).

Across the columns of Table 4.1, the number of GPs varies since some, for example, have no continuously rostered patients. The number of GPs in Table 4.1 differs from that in Figure 4.1 since Table 4.1 conforms with the trimming associated with propensity score matching, discussed below. Column 1, for example, shows that on average GPs completed 2,117 requisitions per year for all patients seen, or 1.14 per patient, and ordered 14,916 individual tests with a cost of $91,938. The table shows that the averages vary greatly across the five columns. For example, the number of lab requisitions per patient and the number of labs ordered per patient ranges between 0.55 and 1.86 lab requisitions per patient and 3.37 and 12.54 labs ordered per patient for the samples of never rostered patients and continuously rostered patients whose tests were ordered by all physicians in the province, respectively.

Figure 4.2.1 presents trends in laboratory requisitions by our sample of GPs, whereas Figure 4.2.2 is restricted to continuously rostered patients. Unlike Figure 4.1, which categorizes GPs by payment model on a year-by-year basis, Figures 4.2.1 and 4.2.2 (and some subsequent analyses) categorizes GPs according to their terminal primary care model. EFFS physicians complete statistically significantly more laboratory requisitions per continuously rostered patient than blended capitation physicians in all years of the sample, and blended capitation physicians similarly order more than interdisciplinary blended capitation physicians.[27]

To understand, in a purely descriptive manner, pre-existing differences in relevant variables across physicians who will, and will not, eventually switch models, Table 4.2 uses ordinary least squares (OLS) on 2006 cross-sectional data. The dependent variable is the number of laboratory requisitions per patient. "BCap_{2010}" (BCap-Team_{2010}) are indicators that equal 1 if the physician is in the blended capitation (interdisciplinary team) model in 2010, and 0 otherwise. Physicians who will in future join interdisciplinary blended capitation complete statistically (and clinically) significantly fewer laboratory requisitions per patient prior to the new model being introduced. Although not shown, results for the number of labs and the cost of labs are similar. This is consistent with Rudoler et al.'s (2015) observation that GPs who enrol more medically and socially 'complex' patients are less likely to join a blended capitation model. Clearly, physicians' self-selection matters.

4.4 Estimation Strategy

Double robust estimation that combines propensity score weighting with regression adjustment (see, e.g., Wooldridge, 2010, section 21.3.4; Imbens, 2015) is employed. First,

[27] The increasing trend may be explained by the aging of the sample.

propensity scores are estimated. From these, weights are generated for the comparison group (i.e., those who remain in the EFFS model) by matching. The goal is weighted mean characteristics of the comparison group that are statistically indistinguishable from those of the treated group. These weights are then employed in a second stage that estimates a difference-in-differences model with physician fixed-effects. This offers protection against misspecification in either stage since only one of the two models needs to be correctly specified to obtain consistent estimates. If both stages are correctly specified, then the estimates are asymptotically efficient. Moreover, the approach addresses selection bias. The identifying assumption for a causal interpretation is that the treatment is independent of the error term conditional on the *X*s (including the fixed-effects), which is variously called the conditional independence assumption, selection on observables or unconfoundedness. As discussed below, we believe this may be plausible in this context.

Propensity score matching uses the cross-section of data prior to the introduction of blended capitation. Observed measures include: physician age and sex, place of medical school graduation (Canada/USA/UK/other), practice location based on indicators for health region and a rurality index, average patient age, percent male patients, roster size, years of practice, and a specialty additional to primary care. Importantly for identification, controls also include the annual number and total cost of labs ordered, the number of patients seen both daily and annually, and billing days per year. Finally, and also crucial for identification, a variable measuring the "income gain" from switching is included in the propensity matching estimation; using data from the year prior to the new model being introduced, this variable calculates how much the GP's income would increase (or decrease) if pre-switch total annual services were remunerated under blended capitation instead of EFFS. That is, it looks at the billing codes actually submitted in the first year of our panel, the year before the blended capitation model was introduced, and it calculates the difference between the observed cost of total billings under EFFS, and that which would have been paid (counterfactually) if those same billing codes were processed using the blended capitation payment scheme. This is a measure of the cost of switching payment models holding practice patterns, and the practice population, constant. Quadratic terms in the continuous variables are also included. This approach is similar to that used by Kantarevic and Kralj (2011) but uses more extensive practice characteristics. The propensity scores are estimated using logit/probit regressions.[28]

Weights are obtained from the propensity scores using a local linear regression matching estimator with a bi-weight kernel and a bandwidth of 0.2.[29] The comparison group is re-weighted to render its characteristics, on average, similar to those of the treatment group. Sensitivity tests using alternative bandwidths, and normal, uniform, and tricube kernels for the local linear regression, as well as kernel and nearest-neighbour matching estimators, result in no qualitative changes in the estimates. Because of a lack of common support, and high weights for some on-support observations, 5% of the treated sample is trimmed; however, sensitivity tests applying no trimming or only employing treated observations with common support show no substantive changes.

[28] We find trivial differences between logits and probits. Experiments with a nonparametric replacement for the logit were undertaken since Frölich, Huber, and Wiesenfarth (2015) show this can sometimes improve performance, but they made little difference and the logit imposes substantially less computational cost.

[29] Fan (1993) demonstrates the advantageous properties of local linear regression.

The panel fixed-effect regression is:

$$log(L_{it}) = \alpha_i + \lambda_t + \beta X_{it} + \delta BCap_{it} + \mu BCap\text{-}Team_{it} + u_{it} \quad (1)$$

where L_{it} is the log of the number of laboratory requisitions, labs ordered, or total cost of labs ordered per patient by physician *i* in year *t*. Physician fixed-effects are denoted α_i, while λ_t is a set of year fixed-effects; X_{it} are time varying physician and practice characteristics. $BCap_{it}$ is the treatment indicator measuring the proportion of year *t* in which physician *i* is affiliated with the blended capitation model, and $BCap\text{-}Team_{it}$ similarly measures affiliation with the interdisciplinary blended capitation model; these indicators are mutually exclusive and each equals 0 or 1, except in transition years when they sum to one. The error term, u_{it}, is clustered on the physician.

Inference follows from bootstrapping the entire process involving propensity score matching, generating weights, and the weighted difference-in-differences fixed-effects model (using 999 replications). Non-parametric percentile-t cluster bootstrap hypothesis testing with re-centred t-statistics is employed.[30] Inference is based on the p-values from the bootstrapped t-distribution; no standard errors are generated by this method.

Ideally, a falsification test examining the common trend (sometimes called bias stability) assumption would be undertaken for difference-in-differences analysis. However, this is not feasible for two interrelated reasons: the period preceding the start of our data was not stable since the EFFS model grew rapidly after its introduction in 2003, and rostering was

[30] If H_0: b=0, $\hat{b}$ is the estimate of b, b_n^* is the bootstrap coefficient for the n^{th} resample, and $se(b_n^*)$ is the standard error of the resampled coefficient estimated using the cluster robust formula, then we simulate the distribution of $t=(\hat{b}-0)/se(\hat{b})$ using $t^*=(b_n^*-\hat{b})/se(b_n^*)$.

a relatively new innovation for these GPs. Following the continuously rostered group "backwards" is not feasible. This problem was also faced by the previously cited studies of this policy change.

4.5 Results

Propensity score logistic regression results are presented in the appendix. Distributions of propensity scores are in Figures 4.3.1 and 4.3.2. Each histogram reflects the entire relevant sample, including the 5% trimmed. About 4.5% of the trimmed are off-support (less than 0.1% of the entire sample). Trimming beyond those off-support is undertaken since some of the comparison group observations would otherwise have very high weights. Following Imbens (2015), we view this as more robust although our findings do not reflect those GPs most likely to switch to the blended model.

Summary statistics, in 2006, for the treatment (those who switch to blended capitation), and unweighted and weighted comparison (those who stay in the EFFS model), groups for physicians who ordered laboratory tests for all patients are presented in Table 4.3. Statistically and economically significant differences, especially for the income gain, between groups are observed prior to weighting. On average, physicians who remain in an EFFS model have more daily visits and services, and see more unique patients although fewer are rostered. EFFS stayers are slightly older, practice in more urban settings, and, given their practice style, would suffer an average income loss of $33,595 if they were to switch to a blended capitation model as opposed to an average gain of $17,783 for switchers. Matching can be seen to successfully balance the pre-switch characteristics of the groups, and the inclusion of pre-switch dependent variables and the income gain variable address normally unobserved heterogeneity that makes causal inference more credible.

Panel A of Table 4.4 reports the treatment variables' coefficients from the fixed-effects model where the dependent variable is the (ln) number of laboratory requisitions per patient. Coefficients for total tests ordered are in panel B, and total cost in panel C. The panels display broadly consistent results. Relative to GPs who stay in EFFS, on average, those who join the blended capitation model have a 3-4% reduction in testing for all patients seen (rostered or not). In contrast, no change occurs in laboratory intensity for those who join the interdisciplinary health team, although as seen in Table 4.2 this latter group orders fewer tests initially.

Continuously rostered patients, in the second and third columns, show no reduction in laboratory intensity by the rostering physician but there is a 3-4% drop in testing by all GPs and specialists in the province. Again, no such reduction is observed for the interdisciplinary blended capitation switchers but, as noted, Table 4.2 shows they ordered fewer tests initially. For blended capitation, this is consistent with one of the desired goals of capitation: to create a medical home where patients concentrate service receipt. The reduction in testing by all physicians in the province represents the trade-off resulting from the reduction in visits to non-rostering GPs observed by Zhang and Sweetman (2018) and the increase in referrals to specialists observed by Sarma et al. (2018). However, labs per visit must go up since Zhang and Sweetman find visits per year declines.

Column 4 shows blended capitation physicians having an almost 4% increase in laboratory requisitions for their patients rostered only part of the study period compared to EFFS physicians, which would be consistent with patient selection out of rostering, and other

explanations, as discussed in Section 7. In contrast, on average patients who see either blended capitation or interdisciplinary blended capitation physicians, but are not rostered with them, experience an approximately 8-10% reduction in per patient laboratory utilization ordered by these physicians.

Overall, payment models appear to influence laboratory utilization. Blended capitation concentrates testing with the rostering physician with total testing reduced – a shift towards the idea of a medical home.

4.6 Appropriate Laboratory Utilization: The Case of HbA1c tests

Inappropriate testing is an indicator of poor-quality care. Recommendations for repeat testing are based on the average effect of the analyte over several days, weeks, or months in the body. Therefore, repeating the test before the recommended interval does not convey accurate or reliable results, leading to unnecessary patient visits to the laboratory and increased costs that could have been avoided from the over-utilization of laboratory test utilization. Moreover, the increase in the number of tests done for a patient increases the number of false-positive results, perhaps inducing “Ulysses syndrome” (Rang, 1972). This syndrome refers to the consequences initiated by an unnecessary or inappropriate investigation or misinterpretation of results in healthy patients. These patients are consequently led down a path of further unnecessary health interventions, referrals, and treatments which can cause increased patient anxiety and unintended patient morbidity (Essex, 2005).

Changes in one aspect of the appropriateness of HbA1c (i.e., hemoglobin A1c, sometimes called glycated hemoglobin or glycohemoglobin) testing as a function of primary care reform is our focus. HbA1c was chosen because it is associated with diabetes, which is a common ambulatory care sensitive chronic condition where proper disease management has important implications for health and healthcare utilization. Also, HbA1c has generally accepted guidelines regarding the minimum time interval between tests. A subsequent test should not be repeated too soon following an earlier one since each test represents an average effect over about three months and changes in HbA1c do not reflect short-term changes in the amount of glucose in the blood (Laxmisan, Vaughan-Sarrazin, and Cram, 2011). Following Morgen and Naugler (2015), we employ a minimum bound of three months between tests to define the second as inappropriate. This operationalization has the advantage of being well-defined and it can be implemented in our data. Other types of inappropriateness, such as failing to test when doing so would be beneficial or redundantly testing beyond the three month minimum are not captured by our measure, which is therefore not inclusive of all measures of inappropriateness.[31] Our measure also captures some redundant testing, which is potentially relevant since Chami

[31] Patients not meeting their glycemic goals are normally tested more regularly than those meeting them. Diabetic patients meeting their glycemic goals may be tested every six months and non-diabetic patients should be tested every one to three years depending on their risk profile (Canadian Diabetes Association Clinical Practice Guidelines Expert Committtee, 2013).

et al. (2017) show that roughly 75% of inappropriate tests, using this measure, are ordered by the same physician who ordered the preceding one.

The sample of primary care physicians is similar to the earlier analysis, but physicians who do not order HbA1c tests are dropped (as are observations, the number of which varies from model to model, where there is no variation in the dependent variable raising an identification problem given the fixed-effect specification). Summary statistics for various dependent variables are presented in Table 4.5. Perhaps the most surprising element of Table 4.5 is that fully 14-16% of tests are ordered inappropriately with continuously rostered patients not very different from other patient samples in the table. Comparing summary statistics in Table 4.5 with those of Table 4.1 show that the annual averages of the dependent variables for all laboratory tests and the sample of HbA1c tests are quite different, where the highest average labs ordered per patient for HbA1c tests is for the sample of all patients rather than continuously rostered patients. This may suggest that looking at all laboratory tests in the province may mask some heterogeneity between different types of tests.

Results from OLS models using cross-sectional data from 2006 (before the capitated model) are reported in Table 4.6. In this part of the analysis we omit the number of requisitions completed and the total cost of tests since we are looking at only one test. For each patient with at least one test, Table 4.6, panel A, presents results where the dependent variable is the total number of HbA1c tests ordered per patient and panel B shows the share that is inappropriate. GPs who eventually shift payment models evidently have different practice patterns and/or patient populations. Especially, patients of physicians who switch to the capitated model have fewer tests ordered by non-rostering physicians, and those who switch to the interdisciplinary blended capitation model are more likely to order inappropriately for their non-continuously rostered patients while their continuously rostered patients are more likely to have inappropriate tests ordered by other physicians in the province.

Table 4.7 presents matching results focussing on variables not in Table 4.3. Again, the reweighting successfully matches the differences between the treated and control groups.

Panel results in panel A of Table 4.8 show no changes in the propensity to order HbA1c tests for rostered patients (despite the overall decrease seen in Table 4.4), which likely reflects the increased attention to diabetes associated with this switch documented by Kantarevic and Kralj (2013).

However, remarkably, for continuously rostered patients the share of tests that is inappropriate declines following the blended capitation, and especially the interdisciplinary blended capitation, transitions. Of course, as seen in Table 4.6, interdisciplinary blended capitation physicians had some higher rates of inappropriate ordering prior to the switch. Conversely, for blended capitation switchers, the move drives an increase in inappropriate testing for patients who are only rostered part of the study period. Although we can only speculate, plausibly newly rostered patients in the blended capitation model are more likely to be given a standardized initial exam without taking into account recent test results. Nevertheless, overall, the move to the new payment models appears to reduce the number of inappropriately ordered HbA1c tests for continuously rostered patients.

4.7 Discussion and Conclusion

How do primary care physician payment models affect the quantity and appropriateness of medical laboratory utilization, which are markers of continuity and quality of patient care? Focusing on continuously rostered patients, using a propensity score weighted difference-in-differences fixed-effects specification to address selection issues there is a short-run reduction of about 3% in laboratory requisitions, tests and the cost of tests ordered by GPs after switching to a capitated model. Patients' laboratory utilization also becomes concentrated with the rostering physician, since their continuously rostered patients have fewer total laboratory tests done by GPs in the province, with no change in the number of laboratory tests done by the rostering physician. The decline in total laboratory utilization in the province and concentration of services reflects a shift towards a more patient-centered medical home practice. The reduction in testing for all physicians in the province can therefore be interpreted as a positive effect on primary care practice, demonstrating an improvement in continuity of care for patients through coordinated and integrated care.

More importantly, using diabetes-related laboratory tests as a case study, after joining the blended model physicians order 3% fewer inappropriate/redundant tests, and the addition of an interdisciplinary care team makes the reduction about 9%. How should these findings be interpreted? We believe that the combination of fixed-effects for physicians, studying stable patient-physician rostered pairs established prior to the introduction of the blended capitation model, and propensity score reweighting with the set of variables employed (included demographics, pre-switch measures reflecting practice style/composition, and the income gain from switching) support the plausibility that the unobserved heterogeneity is not correlated with treatment.

Any omitted variable bias would need to arise from measures that vary over time and are coincident with the switch from enhanced fee-for-service to blended capitation. One such threat can be interpreted as a form of cream-skimming. While the blended model is designed to alleviate incentives for cream-skimming relative to a "pure" capitation payment system, those incentives are not eliminated. GPs in blended capitation have an incentive to de-enroll (i.e., remove from their roster) high service intensity patients, and this may reduce the ordering of clinical laboratory tests. Despite rostering being mandatory, blended capitation GPs could perhaps continue to serve such patients inside their own practice as non-rostered FFS patients. However, the cap on FFS payments for in-the-basket services to non-enrolled patients places an upper limit on the exercise of this option. Alternatively, such patients could be shifted out of capitated, and into non-capitated, practices. Studying these issues is beyond the scope of this paper, however Kantarevic and Kralj (2014) pursue the issue for the two payment models under study using the same data sources for a subset of our time period and they find no evidence of cream skimming – that is, the shifting of "complex and vulnerable" patients out of blended capitation practices. If their results apply, and they may, then our findings may be interpreted causally. On the other hand, Table 4 shows that lab ordering increases for non-continuously rostered patients. This is consistent with blended capitation GPs shifting high intensity patients to the FFS side of their practice, although other interpretations are possible. In the end we cannot definitively conclude that our results should be interpreted causally for continuously enrolled patients. Although such an interpretation is conceivable, if selection resulting from cream skimming (or other factors) occurred then that understanding would be undermined.

In contrast, endogeneity renders a causal interpretation difficult for patients not rostered for the entire length of the panel or never rostered with the GP providing care. As mentioned, for those rostered in some years there is an increase in laboratory utilization, and also an increase in the share of HbA1c tests that are inappropriate We can only speculate as to the cause, but beyond the patient shifting mentioned above, one possibility is that newly rostered patients are given extensive initial checkups. In contrast, there is a marked decline in laboratory utilization for patients never rostered and no change in the share of tests that are inappropriate. These results likely combine causal and selection effects, and are best interpreted as “descriptive”, that is, conditional covariances from a well-balanced sample.

Clinical relevance, or economic significance, is also important. The magnitudes of these estimates are modest in size, but it would be unrealistic to expect much larger impacts. Studying the same transition, Kralj and Kantarevic (2013) find that physicians switching to blended capitation provide 6-7% fewer services, and are 7-11% more likely to attain preventive care quality targets for diabetes. These findings are the same order of magnitude. Further, the overall share of inappropriate tests is around 15%, so a drop of 3-9% of all tests represents an appreciable fraction of all inappropriately ordered HbA1c tests. Future research exploring impacts of economic incentives on other laboratory tests, perhaps combined with educational initiatives such as those by Choosing Wisely, would be worthwhile.

4.8 References

Aas, E. 2009. Pecuniary Compensation Increases Participation in Screening for Colorectal Cancer. *Health Economics*, *18*, 337–354.

Abaluck, J., Agha, L., Kabrhel, C., Raja, A., and Venkatesh, A. 2016. The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care. *American Economic Review*, *106*(12), 3730–3764.

Agha, L., Frandsen, B. R., and Rebitzer, J. B. 2017. Causes and Consequences of Fragmented Care Delivery: Theory, Evidence, and Public Policy. *NBER Working Paper 23078*.

Alexander, C. B. 2012. Reducing Healthcare Costs through Appropriate Test Utilization. *Critical Values*, *5*, 6–8.

Athey, S., and Imbens, G. 2017. The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspecitves*, *31*(2), 3–32.

Barham, V., and Milliken, O. 2015. Payment mechanisms and the composition of physician practices: Balancing cost-containment, access, and quality of care. *Health Economics*, *24*, 895–906.

Barros, P. P. 2003. Cream-skimming, incentives for efficiency and payment system. *Journal of Health Economics*, *22*(3), 419–443.

Bhatia, R. S., Levinson, W., Shortt, S., Pendrith, C., Fric-Shamji, E., Kallewaard, M., and Kerr, E. A. 2015. Measuring the effect of Choosing Wisely: an integrated framework to assess campaign impact on low-value care. *BMJ Quality & Safety*, *24*(8), 523–31.

Canadian Diabetes Association Clinical Practice Guidelines Expert Committtee. 2013. Canadian Diabetes Association 2013 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. *Canadian Journal of Diabetes*, *37*(suppl 1), S1–S212.

Chalkley, M., and Listl, S. 2018. First do no harm – The impact of financial incentives on dental X-rays. *Journal of Health Economics*, *58*, 1–9.

Chami, N., Simons, J. E., Sweetman, A., and Don-Wauchope, A. C. 2017. Rates of inappropriate laboratory test utilization in Ontario. *Clinical Biochemistry*, *50*(15), 822–827.

Clemens, J., and Gottlieb, J. D. 2014. Do physicians' financial incentives affect treatment patterns and patient health? *American Economic Review*, *101*(4), 1320–1349.

Ellis, R. P., and McGuire, T. G. 1986. Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*, *5*, 129–151.

Fan, J. 1993. Local Linear Regression Smoothers and Their Minimax Efficiencies. *The Annals of Statistics*, *21*(1), 196–216.

Feldman, L. S., Shihab, H. M., Thiemann, D., Yeh, H. C., Ardolino, M., Mandell, S., and Brotman, D. J. 2013. Impact of providing fee data on laboratory test ordering: a controlled clinical trial. *JAMA Intern Med*, *173*(10), 903–908.

Frölich, M., Huber, M., and Wiesenfarth, M. 2015. The finite sample performance of semi- and nonparametric estimators for treatment effects and policy evaluation. *Institute for the Study of Labor (IZA) Working Paper 8756*.

Garber, A. M., and Skinner, J. 2008. Is American health care uniquely inefficient? *The Journal of Economic Perspectives*, *22*(4), 27–50.

Glazier, R. H., Zagorski, B. M., and Rayner, J. 2012. Comparison of primary care models in Ontario by demographics, case mix and emergency department use, 2008/09 to 2009/10 (ICES Investigative Report). *Institute for Clinical Evaluative Sciences (ICES)*. Toronto, ON: Institute for Clinical Evaluative Sciences Investigative Report.

Hackl, F., Halla, M., Hummer, M., and Pruckner, G. J. 2015. The Effectiveness of Health Screening. *Health Economics*, *24*(8), 913–935.

Helmchen, L. A., and Lo Sasso, A. T. 2010. How sensitive is physician performance to alternative compensation schedules? Evidence from a large network of primary care clinics. *Health Economics*, *19*(11), 1300–1317.

Hutchison, B., Levesque, J.-F. F., Strumpf, E., and Coyle, N. 2011. Primary health care in Canada: Systems in motion. *Milbank Quarterly*, *89*(2), 256–288.

Imbens, G. W. 2015. Matching methods in practice: Three examples. *The Journal of Human Resources*, *50*(2), 373–419.

Kantarevic, J., Kralj, B., and Weinkauf, D. 2011. Enhanced fee-for-service model and physician productivity: Evidence from Family Health Groups in Ontario. *Journal of Health Economics*, 30(1), 99–111

Kantarevic, J., and Kralj, B. 2013. Link Between Pay For Performance Incentives and Physician Payment Mechanisms: Evidence from the Diabetes Management Incentive in Ontario. *Health Economics*, *22*(12), 1417–1439.

Kantarevic, J., and Kralj, B. 2014. Risk selection and cost shifting in a prospective physician payment system: Evidence from Ontario. *Health Policy*, 115(2–3), 249–257

Kantarevic, J., and Kralj, B. 2016. Physician Payment Contracts with Moral Hazard and Adverse Selection: Theory and Application to Ontario. *Health Economics*, 25(10), 1326–134

Kralj, B., and Kantarevic, J. 2013. Quality and quantity in primary care mixed-payment models: Evidence from family health organizations in Ontario. *Canadian Journal of Economics*, *46*(1), 208–238.

Lavis, J. (Ed.). 2016. *Ontario's health system: key insights for engaged citizens, professionals and policymakers*. Hamilton, ON: McMaster Health Forum.

Laxmisan, A., Vaughan-Sarrazin, M., and Cram, P. 2011. Repeated hemoglobin A1C ordering in the VA health system. *American Journal of Medicine*, *124*(4), 342–349.

Levinson, W., Kallewaard, M., Bhatia, R. S., Wolfson, D., Shortt, S., and Kerr, E. A. 2015. "Choosing Wisely": A growing international campaign. *BMJ Quality & Safety*, *24*(2), 167–74.

Li, J., Hurley, J., Decicca, P., and Buckley, G. 2014. Physician response to pay-for-performance: Evidence from a natural experiment. *Health Economics*, *23*(11), 962–978.

Marchildon, G. P., and Hutchison, B. 2016. Primary care in Ontario, Canada: New proposals after 15 years of reform. *Health Policy*, *120*(7), 732–738.

McGrail, K. M., Evans, R. G., Barer, M. S. L., Kerluke, K. J., and McKendry, R. 2011. Diagnosing senescence: Contributions to physician expenditure increases in British Columbia, 1996/97 to 2005/06. *Healthcare Policy*, *7*(1), 41–54.

McGuire, T. G. 2011. Physician Agency and Payment for Primary Medical Care. In G. Sherry and P. Smith (Eds.), *The Oxford Handbook of Health Economics* (First). Oxford: Oxford University Press, USA.

McLeod, L., Buckley, G., and Sweetman, A. 2016. Ontario primary care models: a descriptive study. *CMAJ Open*, *4*(4), E679–E688.

Morgen, E. K., and Naugler, C. 2015. Inappropriate repeats of six common tests in a Canadian city: A population cohort study within a Laboratory Informatics framework. *American Journal of Clinical Pathology*, *144*(5), 704–712.

Naugler, C. 2014. A perspective on laboratory utilization management from Canada. *Clinica Chimica Acta*, *427*, 142–144.

Ndegwa S. 2011. Funding of laboratory testing in Canada. *Canadian Agency for Drugs and Technologies in Health* (*Environmental Scan)*, 19, 1–5.

Rockey, M.-J., Naugler, C., and Sidhu, D. 2014. Laboratory Utilization Trends: Past and Future. *Canadian Journal of Pathology*, *92*(2), 65–71.

Rudoler, D., Deber, R., Barnsley, J., Glazier, R. H., Dass, A. R., and Laporte, A. 2015. Paying for Primary Care: The Factors Associated with Physician Self-selection into Payment Models. In *Health Economics* (Vol. 24, pp. 1229–1242).

Sarma, S., Mehta, N., Devlin, R. A., Li, L., and Kpelitse, K. A. 2018. Family Physician Remuneration Schemes and Referrals to Specialists: Evidence from Ontario. *Health Economics*, 27(10):1533-49.

Sivananthan, S. N., Peterson, S., Lavergne, R., Barer, M. L., and McGrail, K. M. 2012. Designation, diligence and drift: understanding laboratory expenditure increases in British Columbia, 1996/97 to 2005/06. *BMC Health Services Research*, *12*(1), 472.

Sweetman, A., and Buckley, G. 2012. Ontario's experiment with primary care reform. In *School of Public Policy, University of Calgary* (Vol. 7, pp. 1–38).

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge, MA: The MIT Press.

Zhang, X., and Sweetman, A. 2018. Blended capitation and incentives: Fee codes inside and outside the capitated basket. *Journal of Health Economics*, *60*, 16–29.

Zhi, M., Ding, E. L., Theisen-Toupal, J., Whelan, J., and Arnaout, R. 2013. The landscape of inappropriate laboratory testing: A 15-year meta-analysis. *PLOS ONE*, *8*(11).

Table 4.1: Annual Average of Dependent Variables per Physician, 2006-2010

Patient Sample	All Patients	Continuously rostered patients		Non-continuously rostered patients	Never rostered patients
Tests ordered by	GP seeing patient	GP seeing patient	All physicians in province	GP seeing patient	GP seeing patient
Lab requisitions	2,117	1,135	1,636	612	337
Lab requisitions per patient	1.14	1.28	1.86	1.12	0.55
Patients (with at least 1 lab ordered)	932	464	546	265	189
Labs ordered	14,916	8,224	11,077	4,458	1,978
Labs ordered per patient	7.94	9.22	12.54	7.72	3.37
Lab cost ($)	91,938	49,229	70,909	27,660	13,194
Lab cost per patient ($)	49.11	55.77	81.02	47.74	22.31
Number of GPs	3,132	2,870	2,870	3,070	3,129

Note: Sample after imposing common support by 5% trimming in propensity score matching.

Table 4.2: Pre-FHO Model of All Tests using Cross-sectional 2006 Data

Patient Sample	All Patients	Continuously rostered patients		Non-continuously rostered patients	Never rostered patients
Tests ordered by	GPs in Sample	GPs in Sample	All physicians in province	GPs in Sample	GPs in Sample
$BCap_{i, 2010}$	.0331	**.0505****	.0252	-.0101	**-.1124*****
	(.16)	(.04)	(.22)	(.61)	(.01)
$BCap\text{-}Team_{i, 2010}$	**-.1381*****	-.0720	**-.0818****	**-.1572*****	**-.3001*****
	(.00)	(.14)	(.02)	(.00)	(.00)
R^2	.3121	.1071	.1814	.2393	.2711
No. of GPs	3,132	2,870	2,870	3,070	3,129

Notes: This table reports the coefficients from an OLS model. P-values from robust standard error are in the parentheses. The dependent variable is the log of lab requisitions per patient for the sample. "$BCap_{i,2010}$" (and $BCap\text{-}Team_{i,2010}$) are indicators that equal 1 if the physician is in the blended capitation (or interdisciplinary/team blended capitation) model in 2010, and 0 otherwise. All specifications control for daily services and visits, annual days, roster size, patient and physician age and sex, physician's place of graduation from medical school, and years of practice. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 4.3: Matching Results – All Labs for All Patients, 2006

	Treatment (FHO_{2010} + FHO-FHT_{2010})	Comparison (FHG_{2010}) Full Sample	Comparison (FHG_{2010}) Weighted Matched Sample
Number of GPs	1,360	1,840	1,360
Trimmed (5%)	68	NA	NA
Sample for Regression Analysis	1,292	NA	1,292
Variable Means for Sample after Trimming			
Dependent Variables			
Lab Requisitions	2,102	**2,166***	2,076
Labs Ordered	13,073	**14,584*****	13,056
Lab Cost	76,695	**87,106*****	76,945
Independent Variables			
Unique patients seen by GP	2,249	**2,534*****	2,254
Patients with >= 1 lab	899	**982*****	905
Daily Visits	29	**32*****	29
Daily Services	40	**45*****	40
Annual Working Days	250	249	251
Roster Size	1,385	**1,245*****	1,383
Average Physician Age	53	**55*****	53
Lab requisitions for male patients	36%	**37%*****	35%
Male physicians	63%	65%	62%
Years of Practice	24	**24****	24
Income Gain	$17,783	**-$33,595*****	$17,541
Patient Age Categories‡			
0-19	5%	**6%*****	5%
20-39	22%	**25%*****	22%
40-59	35%	**35%*****	35%
60-79	27%	**24%*****	27%
80 plus	11%	**10%*****	11%
Geographic Area of Practice			
Major Urban	48%	**59%*****	50%
Non-major Urban	48%	**40%*****	45%
Rural	4%	**1%*****	5%
Place of Graduation			
Canada	84%	**73%*****	82%
Foreign	16%	27%	18%

Notes: This table shows the matching results of the variable means for the analysis of all labs for all patients for the year 2006, prior to the introduction of the blended capitation model. The "Treatment" column includes physicians that switch to the blended capitation (FHO) or interdisciplinary blended capitation (FHO-FHT) by the end of the sample period (2010) and the "Comparison" column includes physicians that remain in the EFFS model (FHG) the entire sample period. * p<0.10, ** p<0.05, *** p<0.01. ‡*percentage of lab requisitions made for patients in each age category*

Table 4.4: Panel Regressions of Clinical Laboratory Utilization by Payment Model

Patient Sample	All Patients	Continuously rostered patients		Non-continuously rostered patients	Never rostered patients
Tests ordered by	GPs in Sample	GPs in Sample	All physicians in province	GPs in Sample	GPs in Sample
Panel A: Dependent variable: ln(lab requisitions per patient)					
$BCap_{it}$	**-.0324*****	-.0064	**-.0387*****	**.0356*****	**-.0925*****
	(.00)	(.54)	(.00)	(.006)	(.00)
$BCap\text{-}Team_{it}$	-.0091	.0296	-.0013	.0326	**-.0893***
	(.58)	(.10)	(.91)	(.27)	(.06)
R^2 (Overall)	.1167	.0401	.0003	.0865	.1203
Panel B: Dependent variable: ln(labs ordered per patient)					
$BCap_{it}$	**-.0378*****	-.0141	**-.0360*****	**.0313****	**-.1012*****
	(.00)	(.24)	(.00)	(.02)	(.00)
$BCap\text{-}Team_{it}$	-.0098	.0299	-.0026	.0369	**-.0987****
	(.65)	(.17)	(.83)	(.22)	(.05)
R^2 (Overall)	.1391	.0130	.0532	.0467	.1187
Panel C: Dependent variable: ln(lab cost per patient)					
$BCap_{it}$	**-.0399*****	-.0147	**-.0391*****	**.0301****	**-.0986*****
	(.00)	(.19)	(.00)	(.03)	(.00)
$BCap\text{-}Team_{it}$	-.0280	.0080	-.0164	.0223	**-.1126****
	(.17)	(.72)	(.17)	(.47)	(.02)
R^2 (Overall)	.1009	.0012	.0206	.0155	.1014
No. of GPs	3,132	2,870	2,870	3,071	3,129
Time Periods	5	5	5	5	5

Notes: P-values reported in parentheses are from the cluster percentile-t bootstrap (* $p<0.10$, ** $p<0.05$, *** $p<0.01$). $BCap_{it}$ ($BCap\text{-}Team_{it}$) is the treatment indicator measuring the proportion of year t in which physician i is affiliated with the blended capitation (interdisciplinary blended capitation) model. All specifications control for daily services, daily visits, and annual days worked, roster size, percentage of requisitions made for enrolled patients in each age category, percentage of lab requisitions made for male patients, and physician and year fixed effects.

Table 4.5: Annual Averages of Dependent Variables per Physician – HbA1c tests, 2006-2010

Patient Sample	All Patients	Continuously rostered patients		Non-continuously rostered patients	Never rostered patients
Tests ordered by	GPs in Sample	GPs in Sample	All physicians in province	GPs in Sample	GPs in Sample
HbA1c labs ordered	349	212	216	103	41
Labs ordered per patient	.36	.28	.29	.30	.22
Share inappropriate	.16	.15	.15	.16	.14
Patients with at least 1 HbA1c lab test	234	209	209	83	59
No. of GPs	3,132	2,870	2,870	2,935	2,584

Note: Sample after imposing common support by 5% trimming in propensity score matching

Table 4.6: Pre-FHO Model of HbA1c Tests using Cross-sectional 2006 Data

Patient Sample	All Patients	Continuously rostered patients		Non-continuously rostered patients	Never rostered patients
Tests ordered by	GPs in Sample	GPs in Sample	All physicians in province	GPs in Sample	GPs in Sample
Panel A: Dependent variable: ln(Number of HbA1c tests ordered per patient)					
$BCap_{it}$	-.0250	-.0555	**-.1326*** **	**-.0816** **	-.0372
	(.45)	(.15)	(.00)	(.03)	(.46)
$BCap\text{-}Team_{it}$	.0419	.0611	-.0679	.0574	-.0136
	(.38)	(.23)	(.17)	(.29)	(.85)
R^2	.1850	.1348	.1530	.1765	.1488
N	3,055	2,799	2,819	2,935	2,584
Panel B: Dependent variable: ln(Share of HbA1c tests ordered that are inappropriate)					
$BCap_{it}$	.0126	.0180	.0115	**.0687***	.0227
	(.68)	(.614)	(.74)	(.07)	(.73)
$BCap\text{-}Team_{it}$	**.0834***	.0623	**.0844***	**.1308** **	**.1981** **
	(.06)	(.22)	(.09)	(.02)	(.05)
R^2	.0667	.0457	.0439	.0635	.1430
No. of GPs	2,954	2,563	2,568	2,380	1,077

Notes: This table reports the coefficients from an OLS model. P-values from robust standard error are in the parentheses. "$BCap_{i,2010}$" ($BCap\text{-}Team_{i,2010}$) are indicators that equal 1 if the physician is in the blended capitation (or interdisciplinary blended capitation) model in 2010, and 0 otherwise. All specifications control for daily services and visits, annual days, roster size, patient and physician age and sex, physician's place of graduation from medical school, and years of practice. $^*p<0.10, ^{**}p<0.05, ^{***}p<0.01$.

Table 4.7: Matching Results – Average of all HbA1c labs for all patients, 2006

	Treatment ($BCap_{2010}$ + $BCap\text{-}Team_{2010}$)	Comparison (FHG_{2010})	
		Full Sample	Weighted Matched Sample
Number of GPs	1,257	1,798	1,257
HbA1c labs ordered	238	**267*** **	237
Inappropriate HbA1c labs ordered	46	**53*** **	48
HbA1c labs per patient	0.25	**0.27** **	0.25
Patients >=1 HbA1c test ordered	145	**170*** **	146
Lab requisitions for male patients	48%	48%	48%

Notes: This table shows the matching results of the variable means for the analysis of HbA1c labs for all patients for the year 2006, prior to the introduction of the blended capitation model. The "Treatment" column includes physicians that switch to the blended capitation (BCap) or interdisciplinary blended capitation (BCap-Team) by the end of the sample period (2010) and the "Comparison" column includes physicians that remain in the enhanced fee-for-service (EFFS) model the entire sample period. Matching variables and results are similar to those in Table 4.3, except the ones presented in this table which relate to HbA1c testing. p<0.10, ** p<0.05, *** p<0.01

Table 4.8: Panel Regressions of HbA1c Test Utilization by Payment Model

Patient Sample	All Patients	Continuously rostered patients		Non-continuously rostered patients	Never rostered patients
Tests ordered by	GPs in Sample	GPs in Sample	All physicians in province	GPs in Sample	GPs in Sample
Panel A: Dependent variable: ln(HbA1c tests ordered per patient)					
$BCap_{it}$	-.0110	.0198	.0118	.0166	**-.1351*****
	(.45)	(.27)	(.45)	(.44)	(.00)
$BCap\text{-}Team_{it}$	.0014	.0253	.0293	-.0497	**-.1500*****
	(.95)	(.35)	(.24)	(.12)	(.00)
R^2 (Overall)	.0216	.0701	.0577	.0381	.0417
No. of MDs	3,055	2,799	2,819	2,935	2,584
Panel B: Dependent variable: ln(Share of HbA1c tests ordered that are inappropriate)					
$BCap_{it}$	-.0023	**-.0380***	**-.0339***	**.0459***	-.0361
	(.89)	(.05)	(.05)	(.06)	(.43)
$BCap\text{-}Team_{it}$	**-.0572****	**-.0883*****	**-.0948*****	-.0331	.0123
	(.03)	(.01)	(.00)	(.43)	(.87)
R^2 (Overall)	.0386	.0218	.0271	.0724	.0386
No. of GPs	2,954	2,563	2,568	2,380	1,087
Time Periods	5	5	5	5	5

Notes: P-values are reported in parentheses, based on the cluster percentile-t bootstrap methodology (* p<0.10, ** p<0.05, *** p<0.01). $BCap_{it}$ ($BCap\text{-}Team_{it}$) is the treatment indicator measuring the proportion of year *t* in which physician *i* is affiliated with the blended capitation (interdisciplinary blended capitation) model. All specifications control for daily services, daily visits, and annual days worked, roster size, percentage of labs ordered for enrolled patients in each age category, percentage of labs ordered for male patients, and physician and year fixed effects.

Figure 4.1: Physician Transitions across Primary Care Models

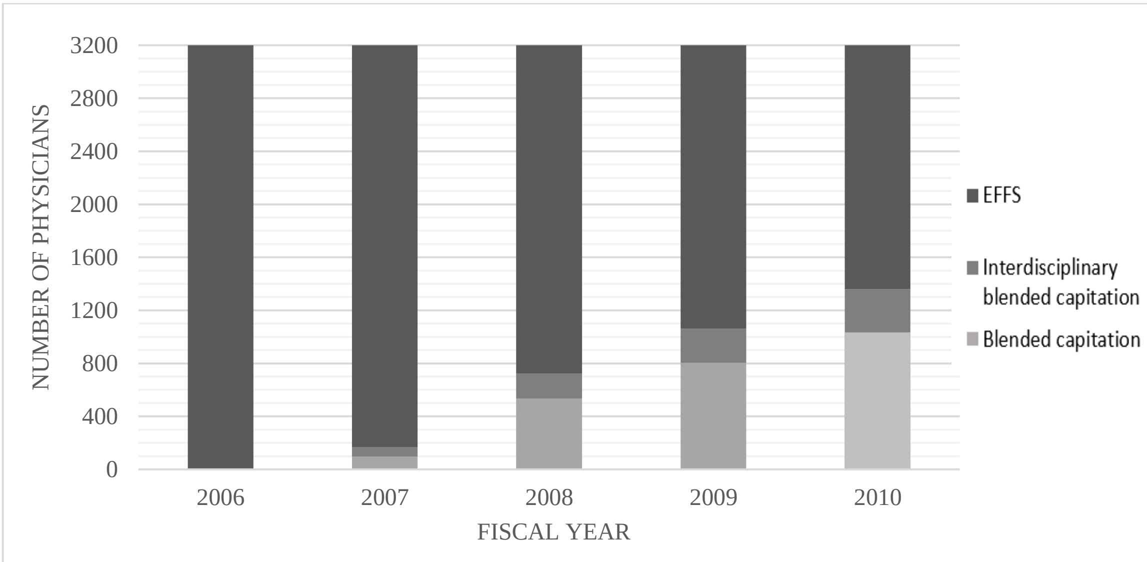


Figure 4.2.1 Lab requisitions per patient (All Patients)

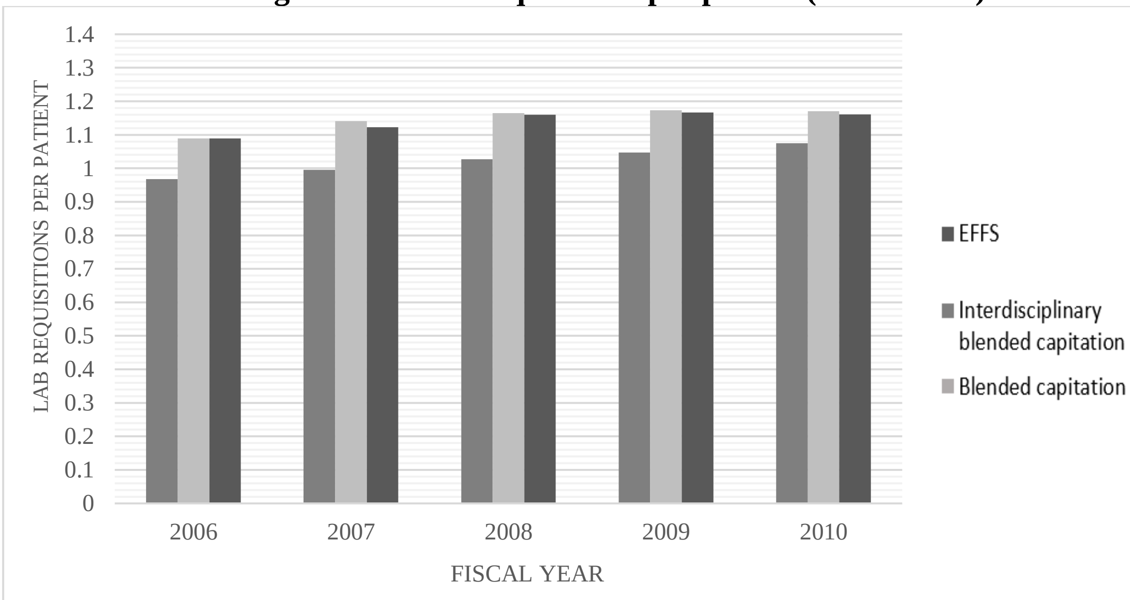


Figure 4.2.2 Lab requisitions per continuously rostered patient

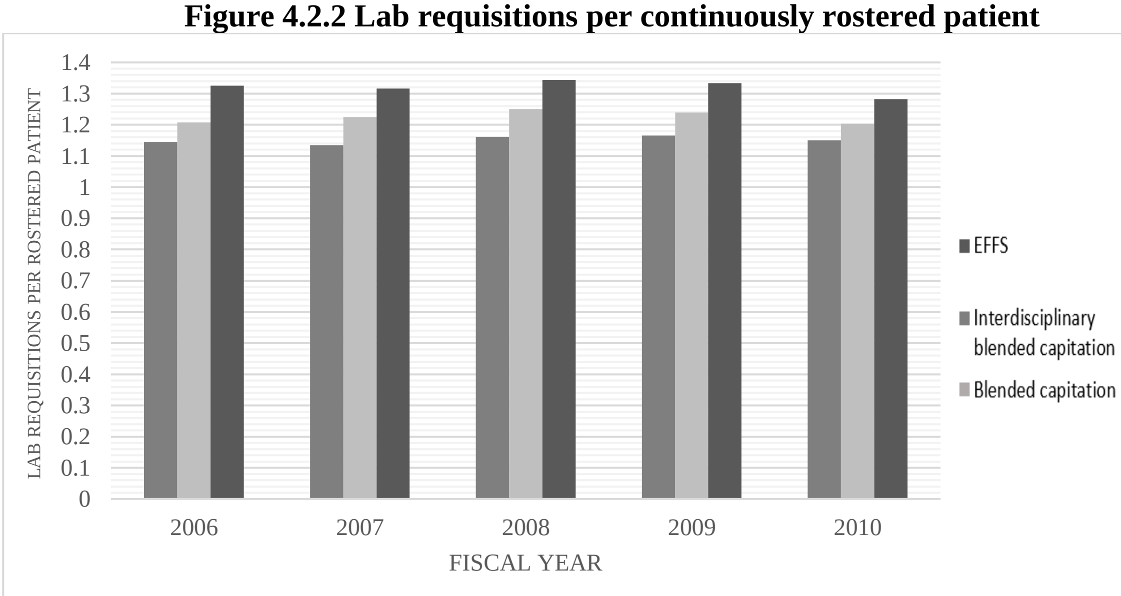


Figure 4.3.1 Propensity Scores with 5% Trimming – All Patients

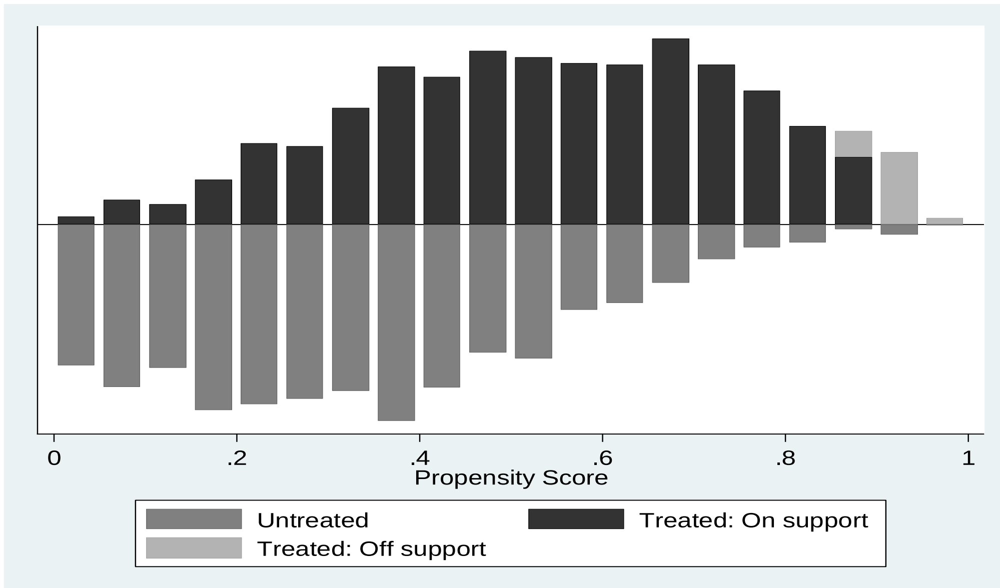
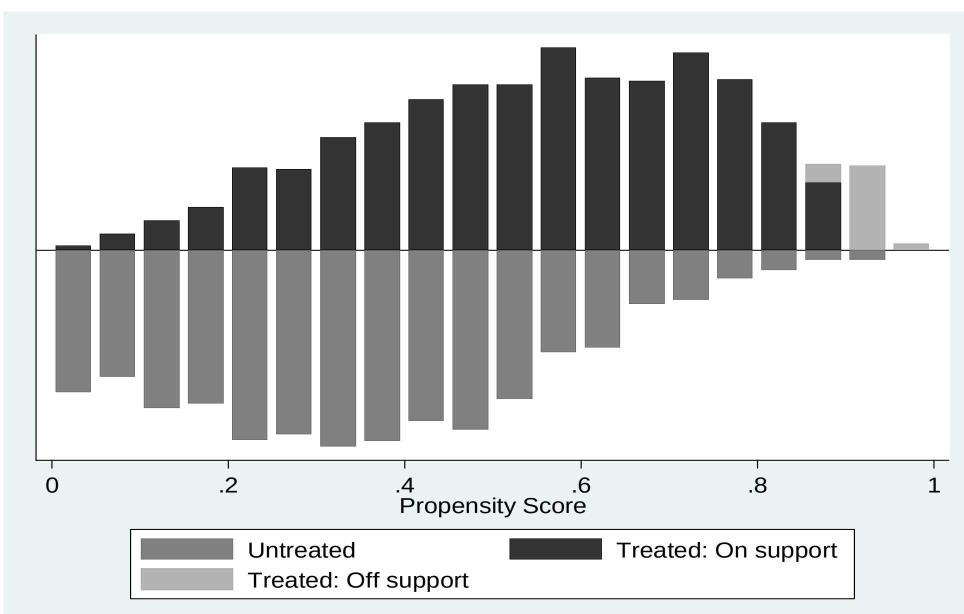


Figure 4.3.2 Propensity Scores with 5% Trimming – Continuously Rostered Patients



Appendix 4.1

List of administrative data sets used in the study

- Ontario Health Insurance Plan Claims Database (OHIP): Administrative data on physician claims
- Client Agency Program Enrolment (CAPE): patient-physician enrolment data
- Registered Persons Database (RPDB): Information on patients' age and sex restricting the sample to only individuals who are eligible for Ontario health insurance coverage which excludes transients, visitors to Ontario, temporary residents, new residents within 3 months of establishing residency, members of the Canadian Forces, serving members of the federal Royal Canadian Mounted Police (RCMP),[32] and persons in federal prison.
- Corporate Provider Database (CPDB): Information on physician characteristics such as age, sex, and specialty as well as the primary care models with which the physician is affiliated.
- Rurality Index of Ontario (RIO): Information on rurality which assigns each practice an RIO score that ranges between 0 and 100, where 0 is the most urban and 100 is the most rural (Kralj 2000).

[32] Prior to June 29, 2012, serving members of the RCMP were also excluded from the definition of insured persons under the *Canada Health Act* but the *Jobs, Growth and Long-term Prosperity Act* amended the *Canada Health Act* and repealed that exclusion (Canada Health Act Annual Report 2013-2014). The data used in this study are from 2006-2011, so serving members of the RCMP are not included in the RPDB.

Conclusion

This thesis addresses health-policy relevant questions regarding quantity and quality of service delivery in primary healthcare. It specifically looks at the effect of switching from a predominantly FFS payment model to a blended capitation payment model on two different outcomes – specialist referral rates and laboratory test utilization, with a focus on inappropriate laboratory testing. The thesis also documents rates and associated costs of inappropriate laboratory testing and explores their determinants using measure of physician and practice characteristics.

Chapter 1 aims to examine the incentives inherent in both FFS and capitation primary care payment models, where theory suggests FFS encourages "ping-pong" referrals among specialists while capitation is "a narrow scope of practice that refers out every time-consuming patient" (Robinson, 2001, 149). However, McGuire (2011) and Robinson (2001) argue that blended models of FFS and capitation can correct for issues that arise from pure FFS and pure capitation payment models, suggesting that blended systems may be preferred. In contrast to what theory may suggest for pure capitation, but in line with McGuire (2011) and Robinson's (2001) arguments regarding blended payment models, results show that the blended capitation model seems to be successfully reducing the incentive for specialist referrals inherent in pure capitation. Results show an increase in specialist visits where the blended capitated physician was listed as the referring physician in the data, but no change in total specialist visits for these physicians' patients. This seems to indicate a concentration of patients with their rostering physician.

Chapter 2 studies the rates of inappropriate laboratory tests in the province of Ontario, specifically looking at redundant testing defined by the interval of time between when the previous test was ordered. Chapter 3 examines the costs of these tests as well as their determinants using physician and practice characteristics. A sizeable proportion and cost (6-20%) of testing for these analytes are inappropriate according to practice guidelines. The overall share and cost of inappropriate test utilization for all nine analytes are respectively 15% and 9% across a time period of five years (fiscal years 2006 – 2010).

Examining the overall magnitude of the impact of switching from the enhanced FFS model to the blended capitation model on the rate and cost of inappropriate laboratory utilization of hemoglobin A1c (HbA1c) tests in Chapter 4, a drop of 3-9% of all tests and costs represent an appreciable fraction of all inappropriately ordered HbA1c tests and their cost, given that the overall share of inappropriate tests is around 15% and the overall cost is 9%. These reductions in laboratory requisitions per patient and appropriate/redundant tests after joining the blended model and concentration of laboratory utilization with the rostering physician suggest a shift towards a more patient-centered medical home practice and advances in the continuity of care for patients through coordinated and integrated care. Also, the blended capitation model seems to be associated with advances in the quality of laboratory utilization through reductions in inappropriate testing. However, caution should be taken when interpreting these results since Chapter 3 suggests that differences in physician and practice characteristics explain only a small amount of the variation in inappropriate utilization. Thus, further data analysis with different variables is required to better understand the driving factors of inappropriate repeat ordering by physicians.

The issue of inappropriate testing is starting to be highlighted and addressed through initiatives such as the "Choosing Wisely" campaign. Such initiatives promote appropriate testing and support change in ordering practices and advances in information technology that facilitate appropriate medical testing. Future research exploring impacts of economic incentives on other laboratory tests and combined with educational initiatives such as those by "Choosing Wisely", would be worthwhile