Big Data Insights into MRI Demand and Wait Time

by

Suting Yang

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Industrial Engineering
University of Toronto

Abstract

Big Data Insights into MRI Demand and Wait Time

Suting Yang
Master of Applied Science
Graduate Department of Industrial Engineering
University of Toronto
2020

The demand for Magnetic Resonance Imaging (MRI) in Ontario has been increasing annually, resulting in longer wait times for patients. This big data study links multiple administrative data sets to characterize and explore drivers of growth in MRI imaging in Ontario between 2008 and 2017. Our retrospective population-based big data study shows an increasing trend in the use of MRI scans, outpacing capacity. Demand increased the greatest amongst family physicians, and there were also wide variations in MRI referral rates among this group. Subgroup analysis demonstrates that family physicians practicing sports medicine ordered more extremities and spine MRIs. Multiple physician characteristics, including years of practice and physician demographics, also impacted the use of MRIs. Overall, 8% family physicians were consistent higher testers, who contributed to nearly 25% of extremities, spine, and brain MRI tests. These findings may help better target interventions to reduce variations in care and overuse.

Acknowledgements

First and foremost, I would like to thank my co-supervisors, Professor Michael Carter and Professor Dionne Aleman, for offering me the opportunity to work on this big data project, and for patiently guiding and encouraging me throughout the research. This work would not have been possible without their endless support and expertise.

I would also like to express my deepest gratitude to my research collaborators at Ontario Health (Cancer Care Ontario) and Trillium Health Partners for their support and feedback throughout this research. I would like to especially thank Dr. Saba Vahid who led this project and took her time out to hear, guide, and keep me on the correct path. I am also extremely grateful to Dr. Ben Fine, Dr. Ali Vahit Esensoy, and Mr. Brian Ho for their valuable clinical insights and feedbacks on my research work. I also like to give a special thank to Dr. Laura Rosella from Dalla Lana School of Public Health, for her extremely useful guidance in risk-adjustment methods and Poisson-based models.

I am thankful to other faculty members and summer research students from Mechanical and Industrial Engineering at University of Toronto, who helped me a lot in my research. I would like to thank Professor Mariano Consens who taught me Data Analytics course and exposed me to the use of Spark. With his course, I understood how to use Spark to handle big data. I would also express my appreciation to Mr. Oscar del Rio who helped me set up the Spark environment in MorLAB cluster, enabling me to process over 600 GB data using parallel processing. I could not have made such a good progress on my work without their help. I am extremely thankful to the three summer research students - Tamara, Parastou, and Cynthia - who helped me with my thesis work. Thanks Tamara and Parastou for their preliminary analyses that helped me get started on my research; thanks Cynthia for the help she offered in the literature review and data analysis that were really useful to my work.

I would also like to express my sincere thanks to my labmates and friends who provided me with endless love and helped me endure the pressure over the past two years. Especially, thanks my labmates - Carolina, Benji, Hamed, Kevin, and Jessen - for their accompanies in the lab and their willingness to listen to my concerns. Thanks my dearest friends - Ming Shi, Yiming Chen, Xiarui Xie, Jiayao Chen, Wei Zeng, Yue Tan, Yutong Ren, Shaun Wang, and Weiran Dai - for comforting me and giving me encouragement when I was depressed. Also, thanks Hachio for reviewing and editing my writings in a much more professional way; thanks Hootan for all the technical support he offered with his expertise.

I am forever indebted to my beloved family members for their unconditional love and support. I want to especially thank my grandpa, Mr. Lianyuan Hua, who was an admirable scientist in Aerospace Engineering and who has stimulated my interest in Engineering since I was young. I want to thank my grandma, Mrs. Jiping Ma, who took great care of me during childhood. I also like to thank my dearest parents, Mr. Jun Yang and Mrs. Min Hua, for raising me up, offering me support and encouragement, and for letting me pursue what I truly love. Thank you for always believing in me and making me who I am today.

Last but not least, I want to give a special thank to my beloved partner, James, for your accompany over five years, and for the endless love and support that you have offered to me, especially when I was on the verge of darkness and desperation during the COVID-19 lockdown period. Thank you James for everything that you have done for me, and I will always be there with you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Magnetic resonance imaging (MRI), an alternative to the traditional ionizing diagnostic techniques, is widely used in the diagnosis of several diseases, such as different forms of cancer, and neurological and cardiac disorders [61]. MRI utilizes a powerful magnetic field and radio waves to generate detailed pictures inside the body, facilitating the diagnosis of disease or the monitoring of its treatment [90]. An MRI scanner is a large tube that contains powerful magnets. During the MRI exam, the patient lies down inside the tube (Figure 1.1a), stays still, and communicates with the MRI technician via the intercom [105]. After the scan, the radiologist examines the images (Figure 1.1b), and prepares a report for the referring physician to discuss with the patient [105]. MRI has several advantages compared to other radiological modalities (e.g., computerized tomography (CT)), including lack of ionizing radiation and excellent ability to identify problems in the joints, soft tissues, ligaments, and tendons [89, 109]. Because of its advantages over traditional radiological modalities, MRI scans are increasingly being ordered, sometimes leading to major bottlenecks in the timely diagnosis of the potential disease [83]. In Ontario, the demand for MRI scans has been increasing annually, outpacing the supply and resulting in longer wait times for patients. Our project is a big data study that links multiple administrative data sets to characterize and explore drivers of growth in MRI imaging in Ontario, between 2008 and 2017. In our study, we first analyze the overall MRI utilization trend, and by patient-specific characteristics. We then examine MRI referral patterns by physician specialties. Lastly, we focus on family physicians in year 2017, and conduct utilization analysis on individual physicians within the group.

(a) Patient undergoing a brain MRI scan [100]

(b) MRI image of the brain [10]

Figure 1.1: Representative pictures of an MRI scan

1.1 Increased MRI usage and wait times

Numerous reports have indicated a significant increase in the use of MRI imaging tests both nationally and internationally [1, 61]. Between 2003 and 2007, MRI usage in the United States (US) Medicare population increased at a rate of 10.6% annually [1]. In Canada, completed MRI tests reached 1.4 million in 2010, doubling the number of MRIs performed in 2003 [21]. Our data shows that in Ontario, between 2008 and 2017, there was an 80% increase in the number of MRI scans. Annual growth in demand (12%) outpaced growth in capacity (7%) between 2012 and 2017.

There are a variety of reasons for the high usage of MRI imaging tests [115]. One reason is "defensive medicine", meaning that physicians order diagnostic measures in order to safeguard against possible accusations of malpractice, rather than to benefit the patients [43]. Another reason is "supply-induced demand", the increase of demand associated with the recent increases in MRI scanners [115]. For example, according to an interview with family physicians, several physicians remarked that because of the increase in scanners, they were ordering scans for minor head injuries more frequently and in a broader spectrum of patients than in the past [115]. A third reason is the remarkable variations in practicing groups; physicians have different decision-making behaviours, and some might overuse MRI scans [114, 115]. Moreover, the sheer number of received requisitions prevents radiologists from discussing the appropriateness of requested MRI tests with referring physicians [115]. As such, it is unrealistic for radiologists to act as gatekeepers for every requested MRI referral.

One of the consequences of the increased MRI demand is longer wait lists [74]. "Wait list" is viewed as a proxy for access to medical care, and the timely accessibility is one of the essential tenets of Canada Health Act [53]. In 2004, the Ontario Ministry of Health and Long Term Care announced Ontario's Wait Time Strategy, with the goal to improve access to healthcare services by reducing the wait times for MRI/CT scans [50]. In the context of this strategy, the Wait Time Information System (WTIS) was introduced by the government to collect and publish the time between arrival date of a scan requisition and scan completition date by hospital [50]. This wait time is illustrated as the Ontario Wait Time (WT2) in Figure 1.2.

The Ontario Ministry of Health has set four priority levels for radiologists to triage patients with a wait time target for each level: priority 1 (emergency, with the target wait time within 24 hours), priority 2 (urgent, with the target wait time within 2 days), priority 3 (semi-urgent, with the target wait time within 10 days), and priority 4 (non-urgent, with the target wait time within 28 days) [75]. These targets are set at the 90^{th} percentile, meaning 90% of patients are expected to receive their scan within the target wait time [75]. However, as of January 2020, only 35% of patients were scanned within the target times [80]. The average wait time for non-urgent MRI scans (priority 4 scans) in Ontario was 62 days, significantly longer than the current 28-day target [80]. The abnormally long wait time for necessary medical treatment may have serious consequences for patients, including extended periods of physical and psychological pain, loss of productivity, deteriorating quality of the life, and even death [6, 94]. The negative impact of a long wait time is even worse for patients with tumors; a study of non-small cell lung cancer demonstrated that due to the exponential growth rate of lung tumors, lengthy wait times delay timely diagnosis, and this delay can be prevented with improved patient flow and reduced wait times [12].

In an effort to mitigate the long wait times in Ontario, supply-side interventions have been carried out, i.e., more diagnostic imaging devices have been produced in order to meet the increasing demand. However, demand continues to surpass capacity, so that the supply cannot catch up with demand at

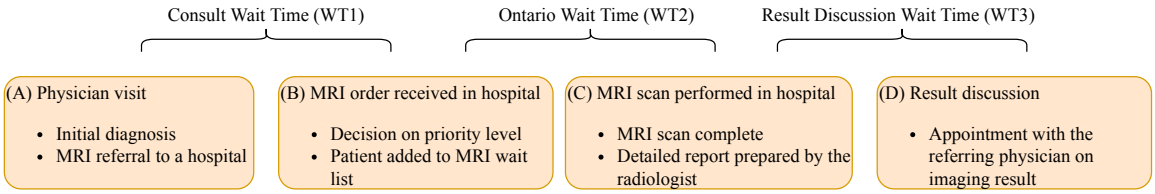


Figure 1.2: Timeline for patients in need of MRI scans

current rate. Simply increasing the capacity will not effectively reduce the waiting list in the MRI orders due to supply-induced demand [57, 67, 94]. Other approaches to reduce the wait time include demand-side interventions, e.g., organizations that help clinicians and patients engage in conversations about unnecessary tests including Choosing Wisely [55, 56] and Right Care [14, 93]. In addition, in 2012, the government of Ontario removed the insurance coverage of diagnostic imaging for uncomplicated lower back pain from the Schedule of Benefits [76]. However, this policy change only resulted in a short-lived reduction in the ordering of single-segment MRI of the spine for both family physicians and specialists [38]. In order to identify additional demand-side interventions, we need to fully understand the MRI utilization patterns in the province. As a result, all the findings in our retrospective study help better target demand-side interventions to reduce variations in care and overuse.

1.2 Related studies

There have been different data-driven approaches to understand the use of MRI and other diagnostic imaging tests both domestically and globally. These studies considered hospital- and patient-specific characteristics. One 2002 study by the Institute for Clinical Evaluative Sciences (ICES) used administrative data in Ontario to conduct some preliminary analyses of MRI scans in Ontario from 1992 to 2001 [53]. They found temporal trends in MRI utilization by body part, as well as (1) considerable age- and gender-adjusted regional variations across the province, (2) high frequency of repeat MRI scans, and (3) variations in MRI referral rates by specialties. This study suggested that it is possible to use multiple administrative data sets to evaluate the correlation of patients, physicians, and system factors with the likelihood that a patient will receive an MRI [53]. Another study by Scheinfeld et al. examined the increasing trend of MRI at a high-volume urban pediatric emergency department [95], finding that MRI use had increased most notably in females, on weekdays, and after hours [95]. A third study was conducted by the research team from National Yang-Ming University and Taipei City Hospital, who examined hospital and diagnosis characteristics with repeat MRI usage within 90 days [18]. Their findings suggested that the repeat use of MRI scans was related to both hospital characteristics and disease types. Specifically, the medical centers had the highest repeat scans, followed by regional hospitals and community centers [18]. Also, repeat CT or MRI was commonly performed for patients with brain or spinal injuries [18]. This knowledge should aid in the review of healthcare policies so that guidelines for repeat scans may be tailored to different hospitals and diseases [18].

Other than the aforementioned hospital- and patient-specific studies, there are other related studies which investigated physicians referral patterns for the diagnostic imaging tests. The most relevant study was conducted by Hall et al. from Queen's University in Ontario, who found that high variation existed in the ordering of imaging tests among Ontario family doctors, and the patients of higher testers were

diagnosed with more thyroid and prostate cancers [42]. The second study, from National Yang-Ming University, suggested that repeat scans are related to physician's specialty, age, sex and type of practice [17]. Specifically, male physicians, physicians aged 41-50 years and internal medicine physicians were associated with more repeat scan referrals [17]. They therefore proposed to set different monitoring standards for repeat scans according to different physician characteristics [17]. A third study was a population-based interrupted time series analysis study conducted by Fine et al., who examined the impact of restricting diagnostic imaging reimbursement for uncomplicated lower back pain in Ontario [38]. They compared physician referral patterns using a time frame of 3 years before and after the policy change, and suggested that the restriction had a stronger impact on family physicians than on specialists, and was more sustained in the use of lumbar spine radiography and CT than spine MRI [38].

Our study is distinct from the existing work in that we examine a variety of recent administrative data sources, including patient, imaging test, and provider data elements. Because of the size and quality of the data sources available, we are able to perform a comprehensive big data analysis on both patient- and physician-specific characteristics associated with the MRI use. The most recent utilization trends and referral patterns in our findings could provide an evidence base for future policy interventions.

1.3 Physician profiling

To understand the variations in MRI utilizations within a physician group (i.e., family physicians), we employ a physician profiling method that compares physicians against their MRI referral rates. Physician profiling is a method that analyzes practice patterns on a specific healthcare service [111]. The practice pattern of a single physician or a group is expressed as a rate: a measure of the use of healthcare resources during a defined period for the population served [52]. The resulting profile of each physician can be compared with a norm that is either based on practice (profiles of other physicians) or based on standards (practice guidelines) [111]. Findings in physician profiling analysis help reduce the variation in performance among physicians and lead to improvements in quality of healthcare [34].

1.3.1 Risk adjustment methods

Current literature suggests that health outcome measures are meaningful only when adequately adjusted for confounding factors, and hence must be risk-adjusted before making comparisons among physicians [26]. A confounding variable is a risk factor associated with the outcome measure [84], and it will cause bias in physician comparison analysis if not removed. Examples of confounding factors include physician practice types and distributions of age/sex of the patient population. There are two major ways to deal with confounding variables at the data analysis stage: stratification and regression [84].

Stratification divides data into strata and layers based on the confounding variable [45]. Data are stratified and analyzed in each stratum [45]. For example, Hall et al. used stratification in their physician test utilization analysis [42]. In their study, they calculated the observed-to-expected (O/E) ratio for each physician, using the entire population as the standard population. For each physician, the observed count was the number of tests he/she referred in the study period, and the expected count was the number of tests that the physician would have referred based on their case mix, if his/her referral behaviour was identical to that observed in the entire population. To obtain the expected count for each physician, they first calculated the global referral rate in each of 14 age/sex strata (e.g., male 40-44) based on the entire study population, then estimated each physician's expected count for each stratum

based on his/her size of patient population, and finally summed up the expected count for each stratum as the expected count for that physician during the study period. They categorized physicians into lower, typical, and higher testers based on the O/E ratio. Stratification is effective when dealing with dichotomous confounding variables [45], but has several limitations . First, stratification is more difficult for continuous variables, such as a specific age or a tumor size [84]. Second, stratification is not able to deal with multiple confounding factors simultaneously [45]. If multiple confounding factors are taken into consideration, each stratum will be very small or disappear (i.e., no patients in the stratum) [84].

Regression analysis helps to resolve the aforementioned two limitations existing in stratification. This method uses all the study data and examines many variables simultaneously, including either continuous or dichotomous variables [72]. Multivariate regression analysis is regarded as the most powerful tool to deal with confounding factors [84], and is the most commonly used method to deal with confounding factors in the medical literature due to its flexibility [84]. There are many types of regression models, and the choice of any particular model depends on the characteristics of the outcome variable, and how the outcome variable is mathematically related to the explanatory variables [33]. The two common regression types for risk adjustment are Poisson regression and logistic regression [33]. In logistic regression, the outcome variable is a binary variable that could be either 0 or 1 [85]. Poisson models are commonly used to model small counts or person-year rates [7]. One example of using regression analysis for risk adjustment comes from a study that compared in-hospital mortality following coronary artery bypass grafting (CABG) among 28 hospitals [26]. In the study, the researchers used the entire patient population to develop a logistic regression model which evaluated the probability of death for a specific patient given their age, gender, comorbidities and their previous heart operation history as risk factors [26]. Then, for the hospital i, with n_i patients, the expected number of post-CABG deaths in the hospital was calculated as [26]

$$E_i = \sum_{j=1}^{n_i} P_{ij} \tag{1.3.1}$$

where P_{ij} is the probability of death for patient j in hospital i given the values of all risk factors, according to the prediction of the logistic regression [26]. Likewise, the observed hospital-specific number of deaths was calculated as [26]

$$O_i = \sum_{j=1}^{n_i} Y_{ij} \tag{1.3.2}$$

where Y_{ij} is a binary variable indicating whether patient j in hospital i was dead following CABG [26]. The researchers then used the O/E ratio for hospital comparison [26], the same measurement used in the study by Hall et al.. One limitation of the regression analysis is that it can only assess a limited number of factors when there is a small number of observations [72]. To deal with this limitation, only variables that are likely risk factors for the outcome of interest should be included [72].

1.3.2 Generalized linear models (GLMs)

Both logistic and Poisson regressions belong to the generalized linear models (GLMs) family. In statistics, GLM is a flexible generalization of ordinary linear regression, which allows response variables to have distributions other than the normal distribution [2]. GLMs were developed by John Nelder and Robert Wedderburn as a way to unify multiple statistical models, including linear regression, logistic regression,

Table 1.1: Common distributions with typical use and link functions

Distribution	Typical uses	Link function	Link name
Normal	Linear-response data	$\mathbf{X}^\top\boldsymbol{\beta} = \mu$	Identity
Bernoulli	Count of single yes/no occurence	$\mathbf{X}^\top\boldsymbol{\beta} = \ln(\frac{\mu}{1-\mu})$	Logit
Binomial	Count of # of "yes" occurences out of N yes/no occurences	$\mathbf{X}^\top\boldsymbol{\beta} = \ln(\frac{\mu}{1-\mu})$	Logit
Poisson	Count of occurences in a fixed amount of time	$\mathbf{X}^\top\boldsymbol{\beta} = \ln(\mu)$	Log

and Poisson regression [73]. The general form of GLM is defined as

$$g(E(Y)) = g(\mu) = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p = \mathbf{X}^\top\boldsymbol{\beta} \tag{1.3.3}$$

where $E(Y)$ is the expected value of the response variable Y, $\mathbf{X}^\top\boldsymbol{\beta}$ is a linear combination of unknown parameters $\boldsymbol{\beta}$, and g is the link function. The link function provides the relationship between the linear predictor and the expected value of response variable, and the choice of the link function depends on the distribution of the dependent variable in the analysis. Table 1.1 lists the commonly used exponential-family distributions and the types of data for which they are typically used, along with the link functions. For logistic regression, the response variable follows a Bernoulli distribution with the Logit link function to the GLM; for Poisson distribution, the response variable follows Poisson distribution with the Log link function to the GLM.

A set of parameters, $\boldsymbol{\beta}$, are often estimated using maximum likelihood estimation (MLE) [65]. In statistics, MLE is a method of estimating the parameters by maximizing a likelihood function [91], a measure of the goodness of fit for a statistical model given a sample of data available. The vector in the parameter space that maximizes the likelihood function is called the maximum likelihood estimate [91], and is chosen as the final parameters fitted to the model. If the likelihood function is differentiable, the maximum likelihood estimate can be determined from the first order conditions, and can be expressed as a closed-form solution. However, for logistic and Poisson regressions, there is no explicit expression for the maximum likelihood estimators, and numerical methods are necessary to find the maximum of the likelihood function. The two common iterative numerical solution methods are the Newton-Raphson algorithm [117] and Fisher's scoring algorithm [60]. Newton-Raphson method is a root-finding algorithm which produces successively better approximations of the roots (zeros) [117]; Fisher's scoring is a hill-climbing algorithm for getting better results through iterations [60].

1.4 Big data analytics frameworks

Big data analytics is the process of examining large and varied data sets, or big data, to uncover information and drive insights [104]. The term "big data" has been applied to data sets whose size are so large that they become difficult to work with using traditional database management systems [35]. Due to its large size, commonly used software tools are unable to store, manage, and process these data within a tolerable time [51]. Scaling is the ability of a system to handle and process large amount of data [102]. In order to resolve the limitations existing in the current software tools, there are two scaling types for big data computing. The first scaling type is called "vertical scaling" (scaling up),

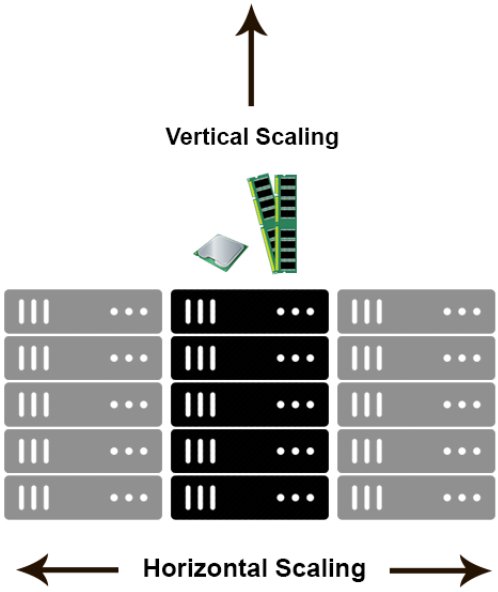


Figure 1.3: Difference between vertical and horizaontal scaling [48]

in which the data resides on a single machine and the scaling is done by adding more processors (e.g., multi-cores) to that single machine [98]. An example of vertical scaling is adding graphics processing units (GPUs) to the local machine to increase the processing power. However, vertical scaling requires a huge amount of financial investment; to manage future workloads, one always needs to add additional resources or upgrade hardware [102]. Another scaling type is "horizontal scaling" (scaling out), which involves distributing the workload across many servers in clusters [62]. It allows for the distributed processing of large data sets across clusters of computers using simple programming models. Figure 1.3 illustrates the difference between the two scaling types.

The big data analytics frameworks that allow horizontal scaling have evolved over time. Apache Hadoop a horizontal scaling framework that uses a MapReduce programming model to process huge data sets across multiple machines [31] (Figure 1.4). A MapReduce program consists of a map procedure, which performs filtering and sorting tasks in each worker of the cluster, and a reduce procedure, which performs a summary operation over all workers. Apache Spark is another horizontal scaling framework which has a programming model similar to MapReduce, but extends it with a data-sharing abstraction called "Resilient Distributed Datasets", or RDDs [118]. RDD enables Spark to perform in-memory computations, which guarantees faster speed in data processing; Spark may be up to 100 times faster than Hadoop's MapReduce [8].

For our study, all the data sets are stored in the Medical Operations Research Laboratory (morLAB) cluster and they in total take up over 660 GB of storage. We deploy a Spark Standalone cluster with 15 workers that enables distributed processing of data sets. The detailed steps of Spark cluster deployment, parallel data processing mechanism, and data file conversion are in Appendix A.

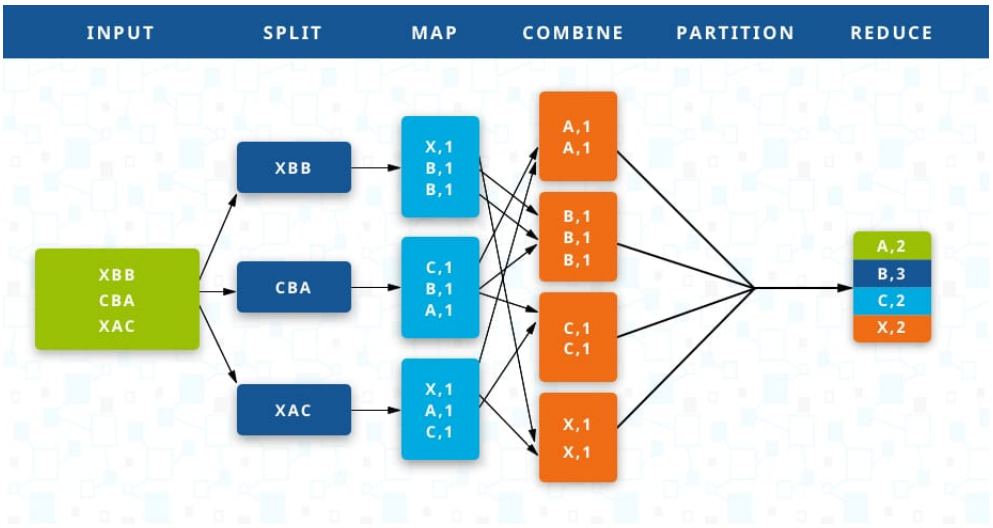


Figure 1.4: An example of Hadoop's MapReduce procedure to sort and count letters [103]

Chapter 2

Data

This study is a retrospective population-based study from 2008/01/01 to 2017/12/31 using records from a variety of administrative data sources in Ontario, Canada. We utilize the data holdings obtained from Ontario Health (Cancer Care Ontario). These data holdings contain data on a range of healthcare activities for the patients in Ontario during the study period, as described below. The data sets are all administrative data sets that have already been collected.

2.1 Data sources

We use the following administrative data sources in our study. The data elements used along with their descriptions are shown in Table 2.1.

1. **Wait Time Information System (WTIS)** contains data on all completed scans in Ontario between 2008 and 2017, and all requested scans between 2012 and 2017.

2. **Ontario Health Insurance Plan (OHIP)** contains data on all physician visits and includes billing codes as well as diagnosis information between 2008 to 2017.

3. **Registered Persons Database (RPDB)** contains demographic information of all Ontario residence that have received a health card for the province's single-payer health coverage with the most current residence information (as of 2018).

4. **Corporate Providers Database (CPDB)** contains sociodemographic and specialization information of registered and active Ontario physicians.

5. **Postal Code Conversion File (PCCF+)** contains geographic information such as income quantile, Local Health Integration Networks (LHINs, representing regional health planning and funding units in Ontario), Census Subdivision (CSD), and immigration tercile. It was pre-linked to RPDB based on three-digit forward sortation area (FSA).

6. **Discharge Abstract Database (DAD)** contains hospitalization information history of each patient between 2008 and 2017.

7. **National Ambulatory Care Reporting System (NACRS)** contains information of Emergency Department (ED) visits of each patient between 2008 and 2017.

Since the data holdings contain confidential patient information, several confidentiality protection procedures were employed. First, all data released to the research team are stripped of Protected Health Information (PHI) by de-identifying Health Insurance Numbers (HINs) as well as truncating postal codes and date of birth data. Second, the data is stored on morLAB cluster, a private cluster in a local physical room with restricted cluster access and restricted directory access. Third, all results and findings in our study are presented at an aggregate level without revealing the information of the individual patient.

2.2 Data linkage

We link different data sources to perform analysis. Figure 2.1 shows the flowchart of the data linkage. Firstly, we link OHIP to RPDB using patient's de-identified HIN, which helps us obtain each patient's age, gender, and other sociodemographic characteristics (e.g., income quntile, immigration tercile) of the residing area. Secondly, we extract all the claims for MRI scans from OHIP data sets between 2008 and 2017, according to the OHIP billing codes. MRI OHIP billing codes are available for head, neck, thorax, abdomen, breast, pelvis, extremities, and spine. For each body part of the MRI scan, a base code for multi-sequence and its corresponding repeat codes are both available. Since the repeat sequence codes as well as the additional MRI-related procedures (cardiac gating, gadolinium, and 3D imaging) are accompanied by a base multi-sequence code [53], these codes are excluded to avoid double-counting referred scans. Also excluded are the claims for biopsies (0.1%) because it is important to focus only on diagnostic studies that required physician's judgment [54]. We additionally exclude the rejected and duplicated MRI claims (5%), according to the OHIP explain code. We link MRI claims in OHIP with CPDB by physician number. As such, we are able to obtain each referring physician's specialty, gender, years of practice, and geographic information. Lastly, we link the MRI claims in OHIP with WTIS data according to the MRI performance date and patient HIN. This linkage helps us retrieve service details and the wait time for each patient record. For the patients who had multiple scans within a day, we match WTIS and OHIP records according to the billing codes (in OHIP) and scanned body parts (in WTIS). Table 2.2 specifies the base codes in OHIP and the corresponding body parts match in WTIS.

In order to find the patient population associated with each referring physician on a yearly basis, for each year, we merge linked data C in Figure 2.1 with linked data G based on referring physician number (in C) and physician number (in G). As such, for each referring physician, we get all claimed patient-physician interactions with complete patient- and physician-specific information.

2.3 Feature extraction

We retrieve a variety of patient- and physician-specific characteristics based on the merged data available. Physicians who do not have a matching record in CPDB (<5%), who did not refer any MRIs during the year, and who had less than 200 interacted patients during the year are excluded from our study. Patients who do not have a matching record in RPDB (<5%) and who did not interact with any referring physicians during the year are excluded from the study.

Table 2.1: Data elements and descriptions

Data Source	Variable Name	Description
WTIS	HIN	De-identified patient health card number
	Wait2	Wait time from the order received date to actual perform date (in minutes), after substracting DART days
	ServiceDetail2	Body part of MRI scan
	PriorityLevel	The priority level for the consultation used to identify similar patients in need of care
	AccessTarget	The maximum recommended wait time in days for the priority level
	OrderReceivedDate	The date the MRI facility receives the request to book a procedure for a patient
	ActualServiceDate	The date the service was performed
	TotalDART	Total Dates Affecting Readiness to Treat (DART) days
OHIP	SERVICE_DATE	Date on which OHIP service was performed
	HIN	De-identified patient health card number
	EXPLAIN_CODE	Indicator of accepted/rejected claims
	HSP_PHYSICIAN_NUMBER	Physician number
	REFERRING_PHYSICIAN_NUMBER	Referring physician number
	BUS_EFFECTIVVE_DATE	Business effective date - Date the business information comes into effect
	FEE_CODE	OHIP fee code
RPDB	HIN	De-identified patient health card number
	DateOfBirth	Birth date
	Sex	Person sex
	Dauid	Dissemination area unique identifier
	CSZEMIZ	Community size and metropolitan influence zones
	CSDuid	Census subdivision unique identifier
	ImmTer	Immigration (foreign-born) tercile
	QAIPPE	Neighbourhood income quintile
CPDB	HSP_PhysicianNumber	Physician number
	Sex	Physician gender
	HSP_SPECIALTY_KEY	Physician specialty key
	HRename	LHIN name
DAD	HIN	De-identified patient health card number
	DateofAdmit	Date of admission to hospital
	DiagnosticCode_1	Main diagnostic code
NACRS	HIN	De-identified patient health card number
	DateOfRegistration	Date of visit
	DiagnosticCode_1	Main diagnostic code
	DiagnosticCode_2-9	Other diagnosted codes

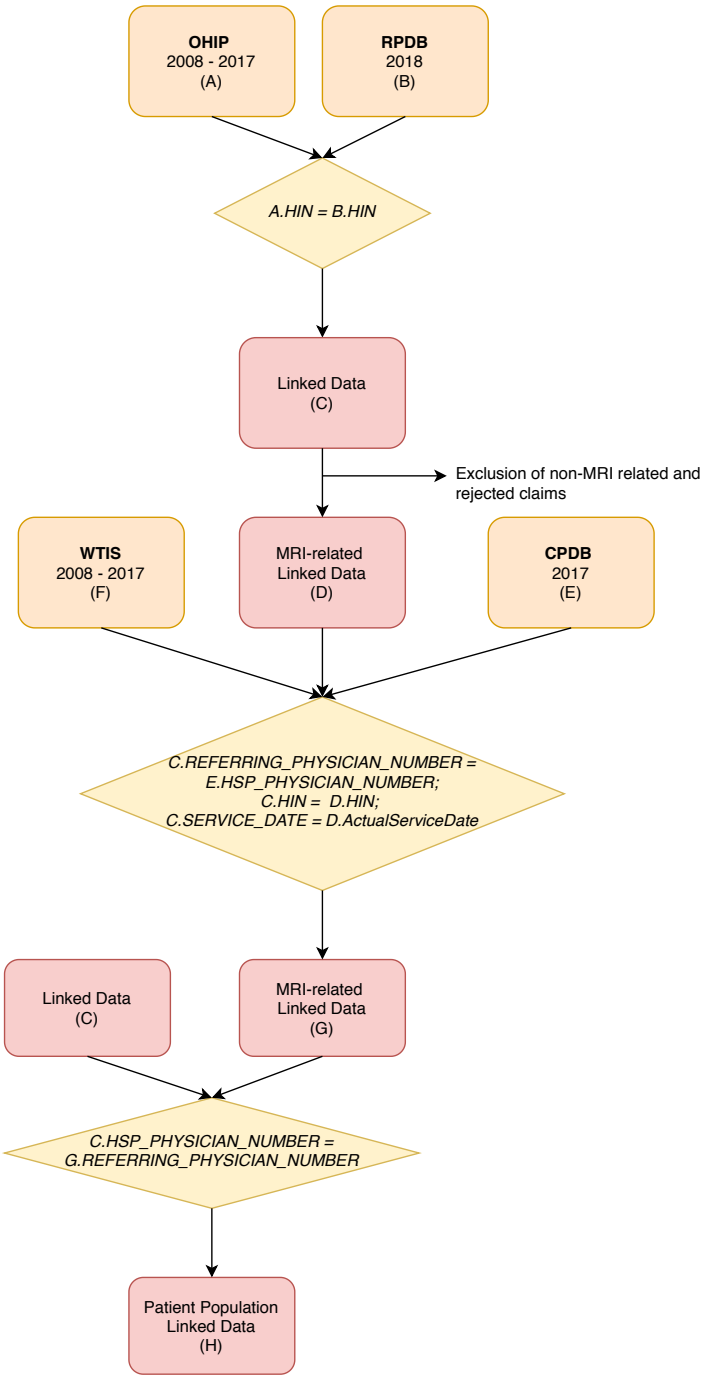


Figure 2.1: Flow chart of data linkage

Table 2.2: Base codes in OHIP and corresponding body parts in WTIS

OHIP Code	Definition	Body part in WTIS
X421	Head	Head (brain)/ head and neck
X431	Neck	Head and neck
X441	Thorax	Thorax / cardiac
X451	Abdomen	Abdomen
X446	Breast	Breast
X461	Pelvis	Pelvis
X471	Extremities or joint(s) – one extremity and/or one joint	Extremities / peripheral vascular
X488	Extremities or joint(s) – two or more extremities, and/or two or more joints same extremity	Extremities / peripheral vascular
X490	Limited spine	Spine
X493	Intermediate spine	Spine
X496	Complex spin	Spine

For each patient, the following features are obtained:

1. Age and sex are retrieved based on the date of birth and sex columns in RPDB. For each record, patient age is calculated from the year of birth to the year of service.

2. Rurality is based on PCCF+ and reported by Metropolitan Influence Zones (MIZ). Table 2.3 shows the community size classification, with MIZ.

3. Neighbourhood income quintile (QAIPPE) is based on PCCF+ at Dissemination Area (DA) level, according to the household-size-adjusted household income per person within the area. Table 2.4 shows the QAIPPE and corresponding explanations.

4. Immigrant (foreign-born) tercile (IMMTER) is based on PCCF+, dividing the immigrant (and non-permanent residence) population into three approximately equal parts. Table 2.5 shows the IMMTER and corresponding explanations.

5. Comorbidity is estimated using Charlson Comorbidity Index (CCI) [16] based on hospital discharge and emergency visits data (DAD and NACRS) with look back of 2 years. Patients with higher scores are associated with greater comorbidities. Appendix B gives an introduction of the CCI, and the pseudo-codes for calculating CCI.

6. For each patient who had an MRI scan between 2008 and 2017, priority level, scanned body part, and the wait time are retrieved based on WTIS. Since 95% of the wait time entries are empty in WTIS, we manually calculate the wait time by calculating the time interval (in days) between order receive date and actual performance date, and then subtracting the total dates affecting readiness to treat (DART) days.

Table 2.3: Community size classification, with MIZ [13]

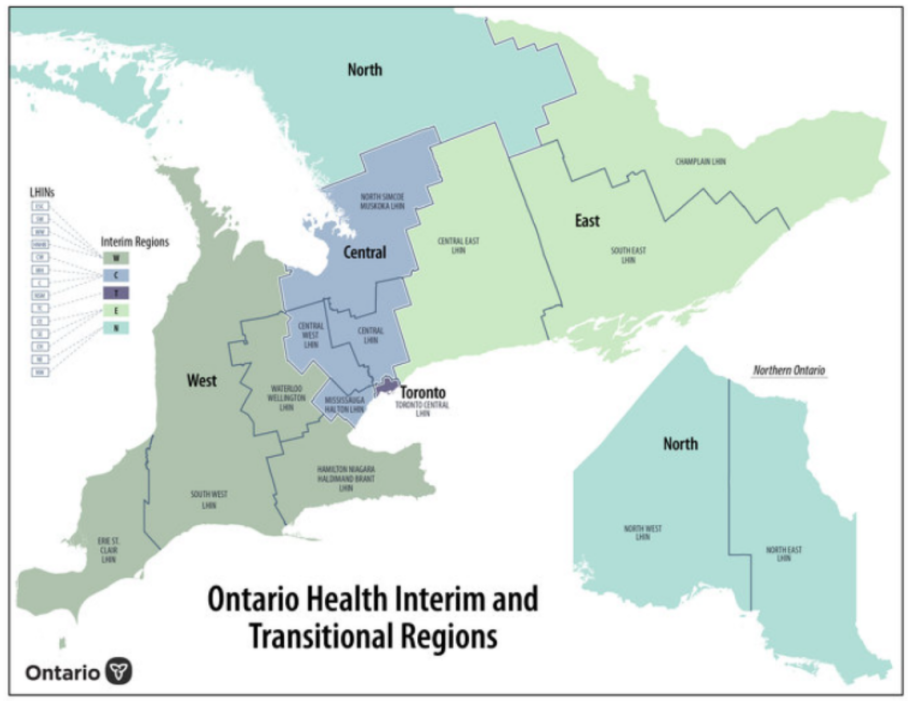
Community size classification	Description
1	1,500,000 +
2	500,000 - 1,499,999
3	100,000 - 499,999
4	10,000 - 99,999 (any census metropolitan areas <100,000)
5	Rural; Strong MIZ
6	Rural; Moderate MIZ
7	Rural; Weak / No MIZ
8	unknown MIZ
9	Missing

Table 2.4: Neighbourhood income per person equivalent [13]

Neighbourhood income	Description
1	Lowest quintile
2	Medium-low quintile
3	Middle quintile
4	Medium-high quintile
5	Highest quintile
9	Missing

Table 2.5: Immigrant (foreign-born) tercile [13]

Immigrant tercile	Description
1	Lowest tercile of foreign-born population
2	Middle tercile of foreign-born population
3	Highest tercile of foreign-born population
9	Missing



North: North West and North East

West: Erie St. Clair, South West, Waterloo Wellington, Hamilton Niagara Haldimand Brant

Central: Central West, Central East, Mississauga Halton, North Simcoe Muskoka

East: Central East, South East and Champlain

Toronto: Toronto Central

Figure 2.2: Assignment of 14 LHINs under the five Interim Regions [71]

For each physician, the following features are obtained:

1. Specialty, sex, years of practice, and working regions are based on the information in CPDB. Year of practice is calculated from the business effective year to the year of service. Each physician's working region is one of the five "Ontario Health Interim and Transitional Regions" assigned by Ontario Health based on LHINs, the health authorities responsible for regional administrative of public health services. Figure 2.2 shows how the existing 14 LHINs are organized under the five Interim Regions.

2. Yearly paneled patients are defined as the unique patients who have seen the physician during a year, and are based on the physician-patient interactions in OHIP. Given the fact that we do not have any data holding that explicitly identifies the physician's rostered patients, we identify the patients who interacted with the physician in OHIP during the year as a proxy of that physician's rostered patients.

2.4 Limitations and assumptions

Several limitations exist in our data linkage process. First, since OHIP only includes completed MRI scans, when linking it with WTIS, the resulting merged data set is a subset of all the demanded scans. Second, for the completed scans, we linked OHIP and WTIS data sets based on the service date (in WTIS) and the service date (in OHIP). However, there are still some instances (<5%) with mismatching OHIP service dates and WTIS actual service dates. Among these mismatching instances, most of them have a difference of ±2 days, and some of them do not have corresponding instances in OHIP at all. These instances are not included in our merged data set. Third, when the combination scan (i.e., more than one body part were scanned in onr MRI procedure) was performed, only one instance with arbitrary picked body part was recorded in WTIS. However, the combination scan instances only account for <1% of the total MRI scans. Fourth, the physicians who do not have a matching record in CPDB (<5%) and the patients who do not have a matching record in RPDB (<5%) are excluded in our analysis. Our assumption is that the MRI utilization patterns and physician referral patterns in our merged data set is a legitimate representation of the overall demand patterns, and the absence of these missing instances does not affect the analysis results.

There are additional limitations in the data extraction process. First, there is no explicit data set that provides each physician's rostered patients. Specifically, previous studies used Client Agency Program Enrolment (CAPE) to identify rostered patients to each primary care physician [42, 59]. However, this data set is unavailable in our study, and instead, we identify all patients who interacted with a physician within a year in OHIP (defined as paneled patients), and assume that the paneled patients are a proxy of that physician's rostered patients. Second, RPDB data contains only one record per HIN with the most current residence information (as of 2018). As a result, the information may not be accurate at the time of the patient's healthcare service. Our assumption is that the patient's residence information does not change over time between 2008 and 2017. Third, the CCI for each DAD record should be determined based on the all 25 diagnostic codes. However, only the main diagnostic code is available in our data set, resulting in an underestimation of the CCI per record. Our assumption is that this underestimation do not affect our key findings when comparing family physicians' uses of MRI with the comorbidities of their patient population.

Chapter 3

Historical trend analysis in Ontario

With our linked data set, we first perform a thorough historical trend analysis of MRI utilizations in Ontario, between 2008 and 2017. Our historical trend analysis is divided into three parts. First, we analyze the overall MRI utilization trend and wait times over the study period. Second, we look deeper into the MRI utilization trends by patients' body parts, age, sex, and residing regions. Third, we investigate physicians' referral patterns by specialties over the study period. We use this analysis to target specific subgroups and time ranges for further investigation.

3.1 Overall MRI wait times and utilizations

For MRI wait times, as mentioned earlier, the patient who will have an MRI is triaged into one of the four different priority levels, and Ontario has different wait time targets for different priority levels. We derive the mean (with standard deviation (stdev)), median, 25%, 75%, and 90% percentiles of wait times, and the percent of scans within the target time of different priority levels over the decade. For overall MRI utilization, we derive the total number of MRIs performed and its cumulative percent of change over the decade. Between 2012 and 2017, we also compare the cumulative increase of MRI requisitions and MRI demand to see to what extent the demand outpaced the supply over the most recent five years in our study.

3.2 MRI utilization by patient characteristics

To investigate MRI utilization by patient characteristics, we first derive the number of MRIs performed by the patient's body part (e.g., brain, head and neck, spine, extremities) along with the cumulative percent of change over the decade. This analysis helps us determine the most frequent types of MRI scans, and the ones with the highest rate of increase. Next, we focus on MRI utilizations by age and sex. Age is aggreagated into 10-year age groups. We compare age- and sex-specific utilization rates per 10,000 population in year 2008 and 2017, using the Ontario population from Statistics Canada as the rate denominator.

We investigate MRI scans by patients' residing regions using small area variation analysis (SAVA) methods. SAVA is a method widely used in health care services research to describe how rates vary across geographic regions [29]. To perform the SAVA analysis, we first aggregate patients' dissemination

areas (DAs) to 49 census divisions (CDs) in Ontario, using the 2016 dissemination area boundary file from Statistics Canada. In Canada, DAs are the smallest standard geographic areas for which all census data are disseminated; CDs are larger units of area representing intermediate geographic areas between the province/territory level and the municipality. We then calculate age-standardized utilization rates per 10,000 population in 2017. The Ontario population in RPDB is used as both the rate denominator and the standard in the adjusted rate calculation. Some of the SAVA statistics are derived to describe the extent of variations in utilization rates across regions. These statistics include extremal quotient (EQ) [30], coefficient of variation (CV) [30], the Chi-squared statistic [112], and systematic component of variation (SCV) [112]. EQ is the ratio between the largest and smallest rates [30], which describes the magnitude of difference between these two extremes. CV is the ratio of the standard deviation to the mean, and shows the extent of variability in relation to the mean utilization rate. The Chi-squared statistic is a testing method with the null hypothesis that there was no significant variation across regions [112]. SCV is another measurement of variation across regions [112]. SCV > 3 is considered significant variation across regions, SCV up to 10 is considered high variation, and SCV > 10 is considered very high variation [4]. Detailed elaboration on the SAVA analysis and the formulas for SAVA statistics is shown in Appendix C.

3.3 Physician referral patterns by specialties

To examine the physicians' referral patterns by specialties, we first count the total number of referred MRIs by specialties to see which physician specialties referred the most MRIs. Second, we obtain each referring physician's referral rate per 1,000 patients. The referral rate per physician is calculated by dividing the total number of referred MRIs by the size of paneled patients. We compare the average referral rates among specialties to see which physician specialties had the most referral rates. Last, to assess inter-physician group variance, we compare the dispersion of referral rates among physician specialties using the quartile coefficient of dispersion [9], a descriptive statistic which measures dispersion and is used to make comparisons between groups of data sets. This statistic is computed using the first (Q1) and third (Q3) quartiles for each group, and is expressed as $(\mathrm{Q3}-\mathrm{Q1})/(\mathrm{Q3}+\mathrm{Q1})$ [9]. The referral patterns help us identify the specialties with high MRI referrals and inter-physician variance, which merit further investigation.

3.4 Results

A high-level review of the data shows that the demand for MRIs consistenty increased year-over-year during the 2008-2017 timeframe in the data set, and there are significant variations in MRI referral rates by physician specialties. Detailed analysis is provided for (1) overall MRI utilization and wait times, (2) patient characteristics associated with MRI referrals, and (3) physician specialties associated with MRI referrals.

3.4.1 Overall MRI wait times and utilzations

Based on the urgency of MRI scans, Ontario has set different MRI wait time targets for different priority levels: 1 day for priority 1 (emergency) patients, 2 days for priority 2 (urgent) patients, 10 days for

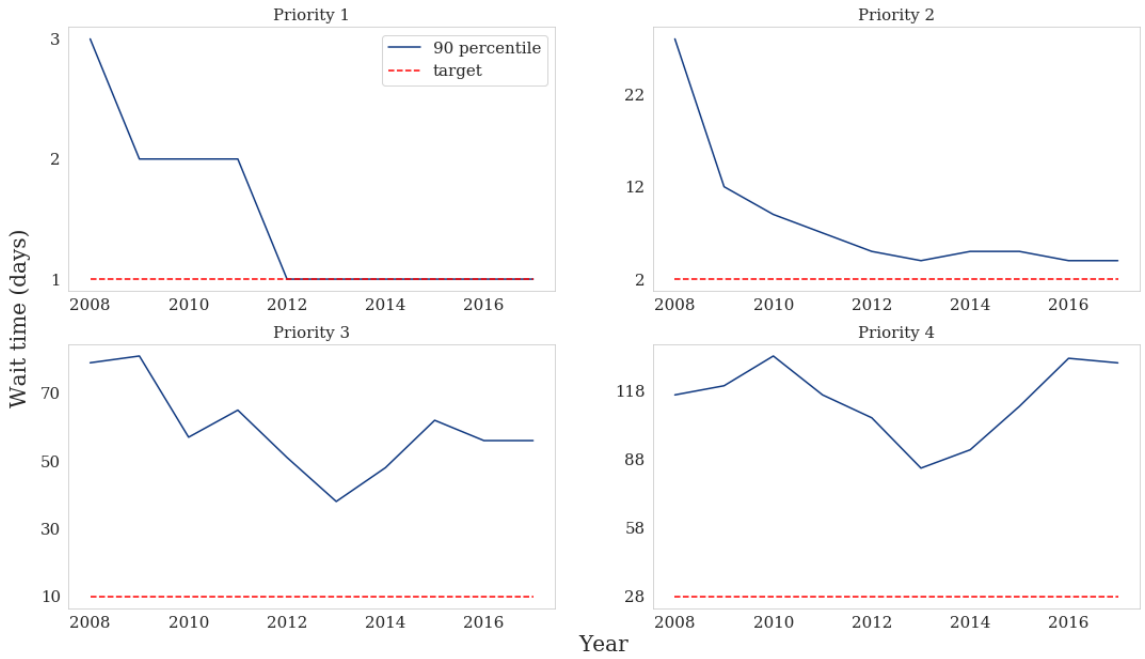


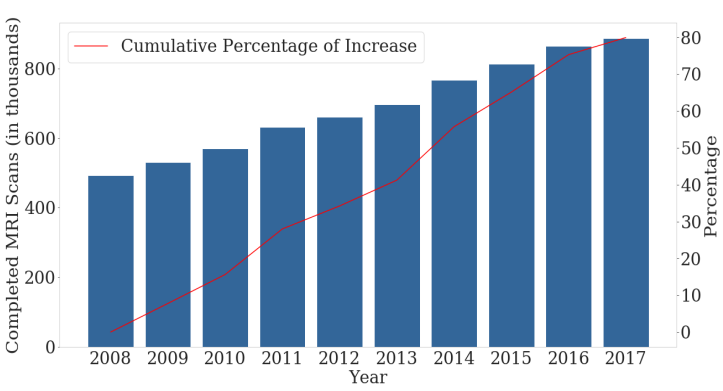
Figure 3.1: 90th percentile of MRI wait times for priority 1-4, 2008-2017

priority 3 (semi-urgent) patients, and 28 days for priority 4 (non-urgent) patients. Figure 3.1 illustrates the 90th percentiles of the wait time along with the targets for each priority level, from 2008 to 2017. Additional detailed statistics of the wait times at each priority level each year is shown in Appendix D. As shown from Figure 3.1, for urgent MRI scans (priority 1 and 2), the wait times had improved over the decade. In 2017, more than 80% of the scans had wait times within the target (94% for priority 1 and 81% for priority 2, 2017). However, for non-urgent MRI scans (priority 3 and 4), the wait time first had an overall improvement from 2008 to 2013, but it started to increase in 2014. In 2017, only 54% and 30% priority 3 and 4 patients were scanned within the target times, respectively, and 90% of the priority 4 scans were completed within 130 days, significantly longer than the current 28-day target.

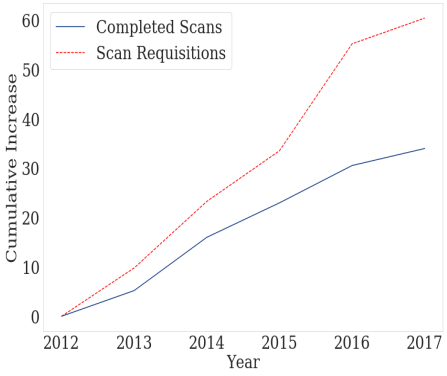
Figure 3.2a shows the total number of MRI scans along with the cumulative percent change over the years. There was a steady increase of MRI scans over the years, with a higher increasing rate from 2013 to 2017. Overall, the MRI scans increased by nearly 80% from 500,000 in 2008 to 880,000 in 2017. Figure 3.2b shows the cumulative increase of MRI requisitions versus completed scans between 2012 and 2017. Demand increase for MRI scans outpaced the capacity growth; since 2012, the number of scan requisitions received per year had grown 12% annually on average, while the number of completed scans per year had only grown 7% per year.

3.4.2 MRI utilizations by patient characteristics

Figure 3.3 presents the number of scans by body parts between 2008 and 2017. The most frequent MRI scans were for head (brain), extremities, and spine (30%, 25%, and 25% of all scans, respectively, in 2017). There was a short-lived decrease for spine MRI scans between 2012 and 2013, but the MRI scans for spine started to increase afterwards. MRI scans for pelvis and abdomen had the greatest cumulative increase during the years (185% and 160% percent change from 2008 to 2017, respectively), but they



(a) Completed scans, 2008-2017



(b) Cumulative increase in demand and supply, 2012-2017

Figure 3.2: Completed and requested scans

merely accounted for a small proportions of all MRI scans (4% and 8%, respectively).

Age- and sex-specific MRI utilization rates per 10,000 population in year 2008 and 2017 are shown in Figure 3.4. In general, MRI utilizations across all age groups increased over the years for both males and females in Ontario. Females had higher utilization rates compared to males across all age groups, except for a slightly lower MRI utilization rate at ages of 70+. The heat maps of age- and sex-specific utilization rates per 10,000 population by body parts in 2017 are shown in Appendix E . Overall, females had higher MRI utilization rates than males for all body parts except for extremities. At ages 20-49, males had higher males had higher utilization in extremities than females.

Figure 3.5 presents the heat map of age-standardized utilization rates per 10,000 population by the 49 census divisions in Ontario, 2017. Although central Ontario (Toronto, Peel, York) had most frequent MRI scans, their utilization rates per 10,000 population were lower than those in northern Ontario (Cochrane, Thunder Bay, Algoma). According to SAVA statistics, the EQ is 2.8, the CV is 0.15, the SVC is greater than 10, and the p-value for the Chi-squared statistic is less than 0.001 (χ^2 = 15,818, df = 48). The statistics indicate a significant variation in the use of MRI across regions in Ontario.

3.4.3 Physician referral patterns by specialties

The distributions of referred MRI scans by specialties between 2008 and 2017 are shown in Figure 3.6. MRI scans were mainly referred by family physicians, neurologists, and orthopaedic surgeons (50%, 11%, and 6% in 2017, respectively). These specialties accounted for 67% total MRI referrals in 2017.

Figure 3.7 shows the comparison of family physicians and specialists on the cumulative increase rate for physician numbers and referred MRI scans, between 2008 and 2017. As illustrated in Figure 3.7a, the number of specialties had a faster increase than that of family physicians; however, Figure 3.7b indicates a greater rate of increase in MRI referrals for family physicians over the decade, especially between 2014 and 2016. Between 2014 and 2017, half of the MRI scans were referred by family physicians.

The mean (with stdev), 25%, 50%, and 75% percentiles of MRI referral rates by physician specialties in 2017 are shown in Table 3.1. We combine the specialties with $< 0.1\%$ contributions to MRI referrals as "Others". These specialties include Cardio-thoracic surgery, dermatology, nurse practitioners, pathology, clinical immunology, nuclear medicine, and community medicine. Neurosurgery, neurology, and therapeutic radiology were the top 3 specialties with highest mean referral rates (346.39, 179.54,

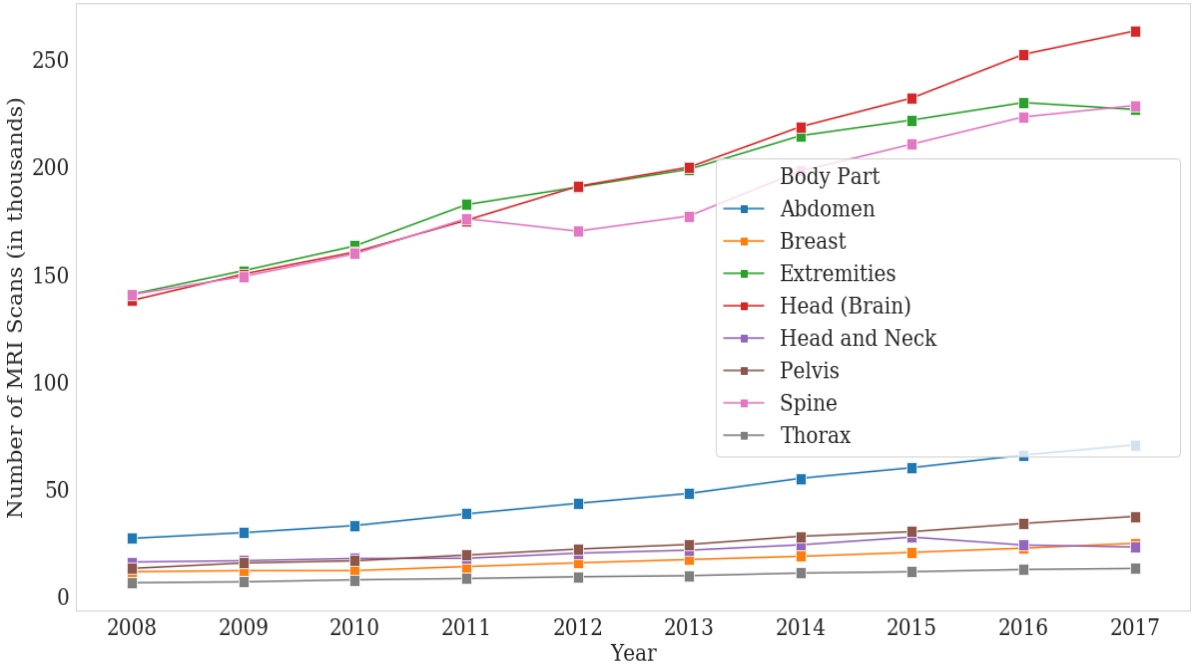


Figure 3.3: Number of MRI scans by body part, 2008-2017

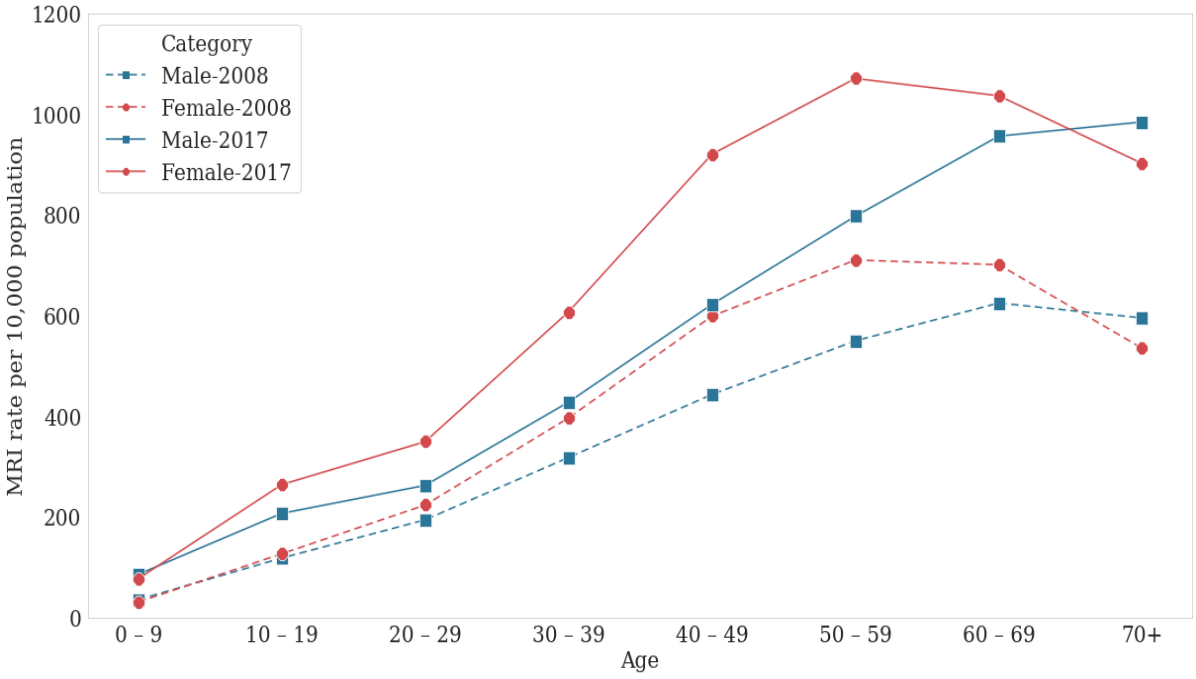


Figure 3.4: MRI utilization rate by patient gender, 2008 and 2017

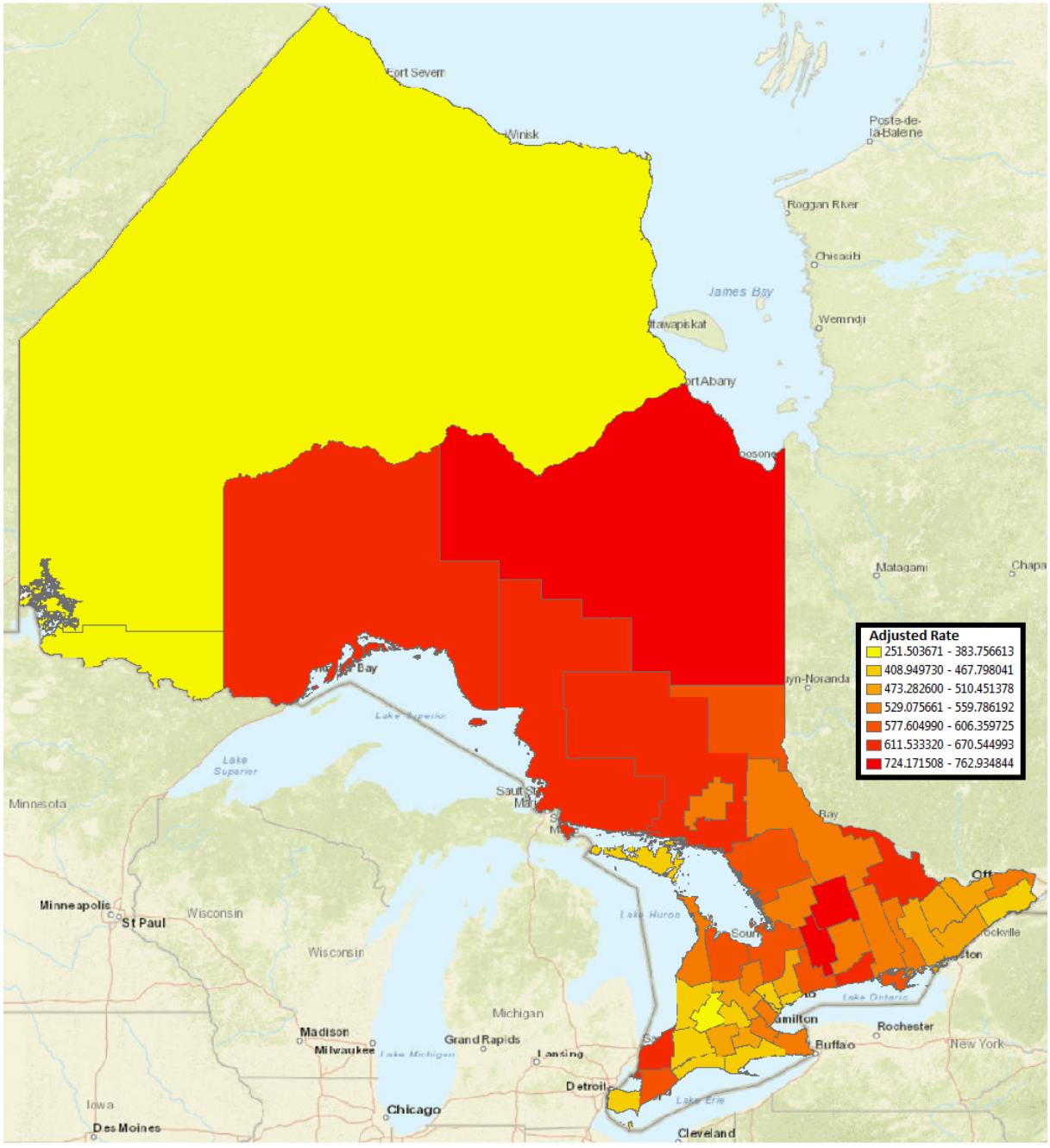


Figure 3.5: Age-standardized MRI rates per 10,000 population by census division, 2017

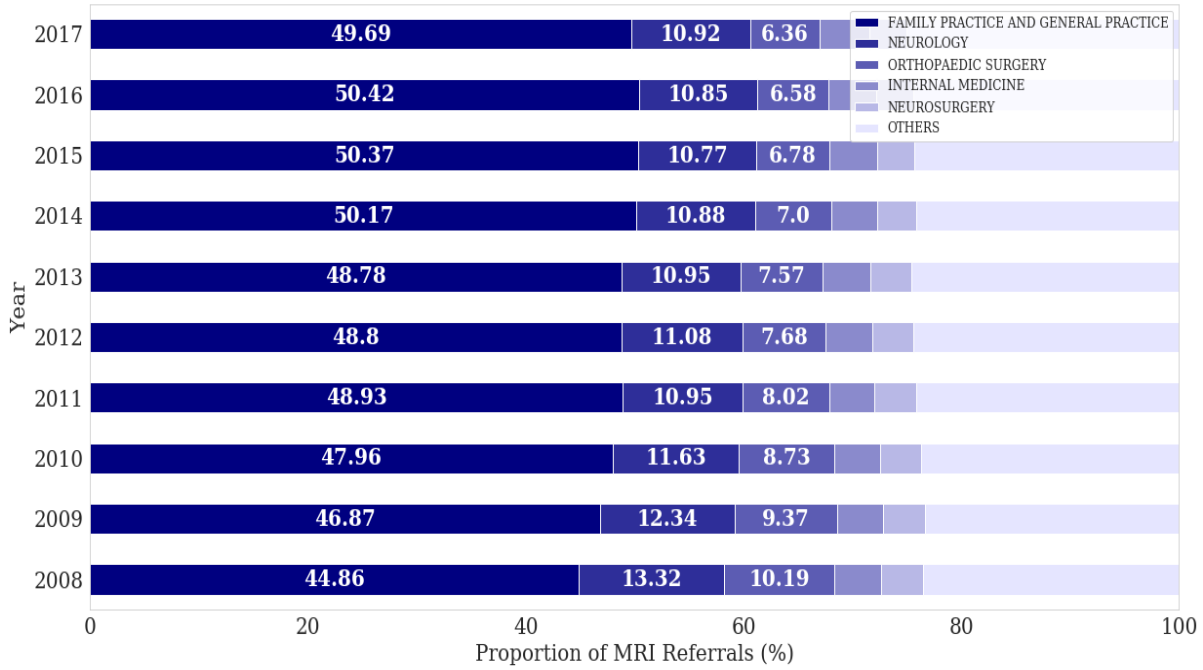
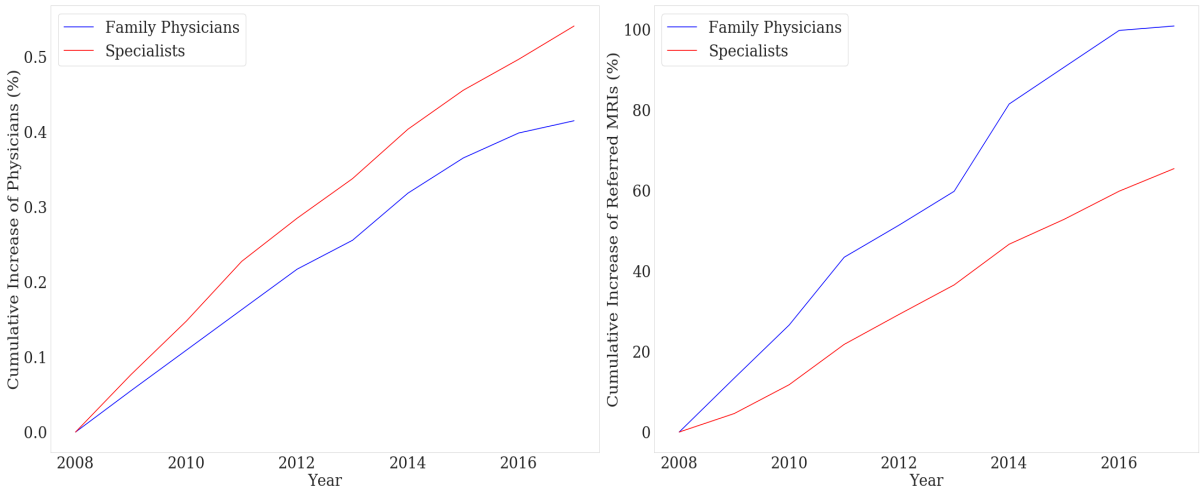


Figure 3.6: Proportion of MRI referrals by specialties, 2008-2017



(a) Cumulative increase of physicians

(b) Cumulative increase of referred MRIs

Figure 3.7: Comparison between family physicians and specialists, 2008-2017

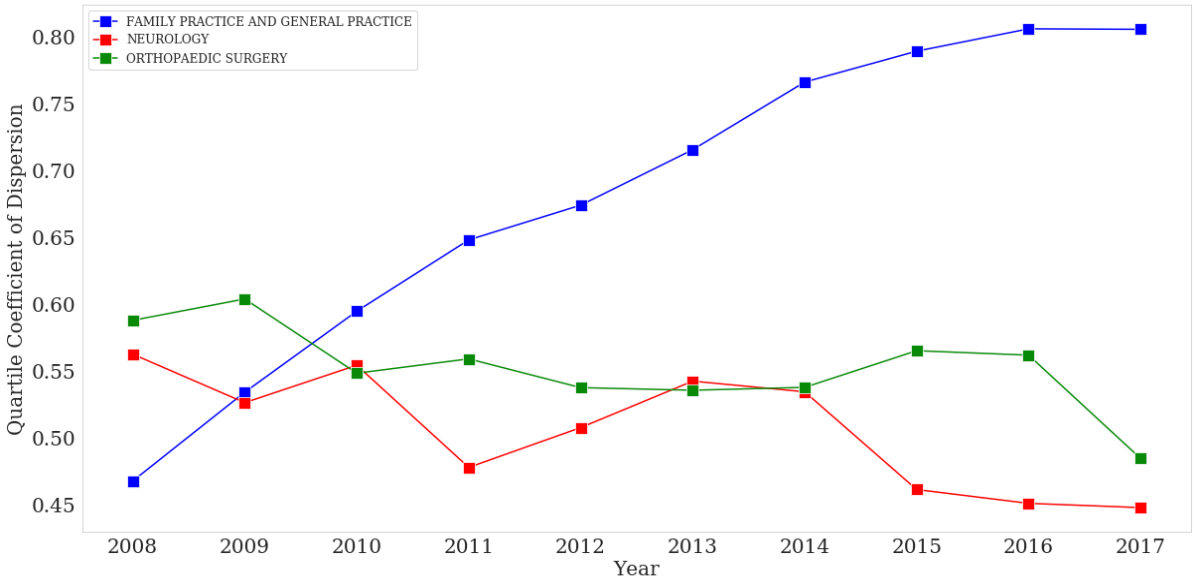


Figure 3.8: Quartile coefficient of dispersion for referral rates, 2008-2017

and 123.46 per 1,000 patients in 2017, respectively). However, they together accounted for a smaller proportion of referred MRIs (16% in 2017) compared to MRI scans referred by family physicians. In addition, although the mean referral rate for family physicians was moderate compared with that for other specialties, it increased by 65% from 13.88 per 1,000 patients in 2008 to 22.9 per 1,000 patients in 2017.

Figure 3.8 shows the inter-physician referral rates variance in terms of quartile coefficient of dispersion, among the three types of physicians (family physicians, neurologists, and orthopaedic surgeons) who contributed most to referred MRI volumes. As illustrated in Figure 3.8, the referral rates variance among family physicians had a constant growth from 2008 to 2017, and by the year 2017, family physicians had over 50% higher variance compared with neurologists and orthopaedic surgeons.

3.5 Discussion

In Ontario, between 2008 and 2017, there was an 80% increase in the number of MRI scans. Despite ongoing efforts and multiple government interventions taken place in Ontario, demand increase for MRI scans still outpaced the capacity of growth. Since 2012, the number of scan requisitions per year had grown 13% annually on average, while the number of completed scans per year had only grown 7%. Given the current growth rate, it is likely that the demand will continue to outpace the capacity. As for the wait times, based on the significant disparity of wait times among different priority levels, it is critical to prioritize MRI requests effectively so that the most-in-need patients will be scanned in time. According to a survey study of public MRI facilities in Canada, less than half of the facilities had explicit, documented criteria to guide the triage process, and the prioritization was usually implicitly assessed by radiologists using a handwritten requisition submitted by the referring physician [36]. Therefore, it is critical to develop a decision support system that generates standardized and reproducible triage decisions. In addition, the wait times for non-urgent patients significantly lagged behind the provincial

Table 3.1: MRI referral rates by specialties per 1,000 patients, 2017

Specialty	Mean (stdev)	25%	50%	75%
Neurosurgery	346.39 (272.86)	164.83	273.78	443.84
Neurology	179.54 (177.68)	75.11	138.67	210.12
Therapeutic Radiology	123.46 (183.78)	28.77	62.85	135.35
Medical Oncology	75.82 (74.06)	35.62	58.68	88.09
Physical Medicine	63.23 (88.89)	18.84	40.57	76.60
Orthopaedic Surgery	60.22 (53.08)	24.88	45.54	79.09
Rheumatology	56.54 (48.87)	24.30	44.44	73.84
Thoracic Surgery	53.27 (35)	25.31	48.20	83.19
Otolaryngology	31.55 (30.3)	13.92	24.03	39.66
Haematology	29.34 (35.3)	6.53	16.98	44.61
General Surgery	28.54 (46.16)	6.89	14.39	28.15
Gastroenterology	26.69 (34.31)	9.95	18.26	32.20
Paediatrics	24.39 (76.48)	2.42	6.06	15.60
Geriatrics	23.28 (24.75)	6.19	15.72	31.88
Family Practice and General Practice	**22.37 (26.07)**	**6.15**	**15.83**	**30.12**
Internal Medicine	22.36 (38.79)	3.13	10.30	27.89
Endocrinology	16.6 (30.73)	4.87	9.78	16.28
Infectious Disease	14.92 (16.4)	5.17	11.04	17.96
Urology	12.61 (13.53)	3.82	8.79	16.35
Anaesthesia	10.42 (26.63)	1.33	3.50	9.84
Others	10.32 (23.15)	0.70	2.10	10.55
Psychiatry	10.25 (17.78)	2.40	4.34	9.53
Nephrology	9.85 (16.41)	2.62	5.95	11.07
Emergency Medicine	8.94 (17.13)	2.44	4.62	8.38
Diagnostic Radiology	8 (30.94)	0.12	0.33	1.72
Plastic Surgery	7.04 (6.74)	2.40	4.88	10.10
Obstetrics and Gynaecology	6.81 (9.85)	1.80	3.66	6.91
Ophthalmology	6.63 (12.86)	1.06	2.54	6.18
Vascular Surgery	6.45 (6.86)	2.04	3.85	7.97
Respiratory Disease	6.18 (15.14)	1.01	2.62	5.47
Cardiology	5.54 (10.07)	1.07	2.42	5.57

standard over the years, with less than half of the patients scanned within the Ontario target. When doing a deeper utilization analysis at the next stage, we need to focus on the utilizations of these non-urgent patients.

The MRI utilizations by the patient's body part, age, sex, and residing region between 2008 and 2017 were consistent with the patterns in Ontario reported by ICES between 1992 and 2001 [53]. The most common MRI scans were for extremities, spine, and brain. It is worth noting that spine MRI scans had a slight decrease in 2012, and then increased from year 2013. This change in spine MRI scans validates the claim that the restriction of spine imaging tests in year 2012 was associated with a short-lived reduction in ordering of spine MRIs [38]. In terms of the patient's age and sex, females had higher MRI utilizations across all age groups below 70 years old. MRI utilization was highest for females with ages between 40 and 69, and this group of patient also contributed most to the growth of MRI use.

When examining the MRI utilization by patients' residing regions, we find that while Greater Toroto Area (City of Toronto, Peel, York, Durham, Halton) had greatest number of MRI scans, they had lower utilization rates per 10,000 population compared to the rates in northern part of the province (Cochrane, Thunder Bay, Subdury). One of the possible factors leading this geographic difference is the disparity in access to MRI services. It was previously reported by ICES that regions with the most MRI scanners per million population were among the ones with the highest utilization, while the areas with lowest number of scanners per million population were among the ones with the lowest utilization [53]. However, the correlation between the regional MRI utilizations and available scanners cannot be validated in our study, since we do not have the information on the number of available scanners in each hospital. Other factors that may cause the regional variance include the patient's severity of illness, patient's health-seeking behaviours and physician's decision-making behaviours [114].

When examining the physician referring patterns by specialties, we find that the specialties with the highest MRI referral rates were neurosurgeons, neurologists, and therapeutic radiologists. However, they together accounted for less than 20% referrals in year 2017. On the other hand, family physicians referred approximately half of MRI scans each year, contributed most to the growth of yearly MRI referrals, and had large inter-physician variance. The finding of high variance in MRI referrals among family physicians is consistent with previous studies. Specifically, a study on diagnostic tests among primary care providers suggested that there were wide variations in the use of diagnostic imaging tests by Ontario family physicians [42]; another study suggested that primary care physicians see patients with wide range of clinical conditions and certainly face challenges in following all related imaging guidelines [116]. The other two studies provided an evidence of the overuse of MRI tests among family physicians, especially in the use of lumbar spine MRIs [37, 82]. Our findings along with the current literature suggest that we need to further investigate family physician's test utilization patterns.

3.6 Conclusion

Our population-based historical trend analysis shows a consistently increasing use of MRI in Ontario between 2008 and 2017. The wait times for non-urgent patients significantly lagged behind the provincial standard over the years, with less than half of the patients scanned within the Ontario target. The most frequent scans were for extremities, spine, and brain, and middle-aged women with ages 40-69 had the highest MRI utilizations. Although having a moderate average referral rate per 1,000 patients compared with other specialties, family physicians had the highest MRI referrals and high within-group variance. Based on our findings, it is important to conduct a deeper analysis on family physicians, and on the non-urgent (priority 3 and 4) patients. As a result, for the next steps, we conduct a physician utilization analysis on individual family physicians over a year, for non-urgent and most referred MRI scans (extremities, spine, and brain). We aim to understand what physician characteristics (e.g., years of practice, region) can cause the variance among this group.

Chapter 4

Family physician test utilization

In Chapter 3, we determine that the most frequent MRI scans were for extremities, spine, and brain, and non-urgent patients with priority levels 3 and 4 had average wait times significantly longer than the Ontario wait time targets. In addition, family physicians had the highest MRI referrals and high within-group variance. We therefore perform a family physician test utilization analysis on non-urgent scans for uncomplicated extremities, uncomplicated spine, and brain MRI tests, during the time period between 2016/07/01 and 2017/07/01. The OHIP billing codes for the selected tests are X471, X490, X493, and X421 in Table 2.2. This investigation includes 11,644 family physicians along with over 10 million patients. The objective is to determine whether the variations in the use of MRI tests were associated with the variations in physician characteristics, and in referred patients' subsequent follow-up visits within six months after their MRI scans. We profile family physicians on their use of MRI scans, and categorize them into low, typical, and high testers according to risk-adjusted referral rates. We then compare tester levels against multiple physician characteristics and patients' subsequent visits.

4.1 Poisson-based risk adjustment models

Before making comparisons among family physicians, we need to first build up risk adjustment models, which help produce the expected number of referral rates per physician based on multiple relevant risk factors. Current literature suggests that risk adjustment models can be built at either an individual or an aggregate level [20]. Most risk adjustment frameworks calculate the risk adjusted score for each individual in a population, and then sum up these predictions to an aggregate level [20]. For example, in a study comparing coronary artery bypass grafting (CABG) among 28 hospitals, the researchers used the logistic regression to predict the probability for each patient's death in a hospital, and summed up the predictions as the expected number of post-CABG deaths in that hospital [26]. However, this level of disaggregation is often not easily available [20], in which case aggregate-level risk adjustment models are built. For example, a study that compared bloodstream infection rates among hospitals stated that due to the unavailability of attribute data on individual patients, they built hospital-level risk adjustment models that directly calculate the expected infection counts for each hospital [107]. Another study conducted by Buajitti et al. also used population-level prevalence estimates rather than individual-level features as risk factors in their risk adjustment models, in order to compare regional variations of premature mortality in Ontario [11]. Even if only aggregate-level information is available, aggregate

models are still useful in adjusting risk and making a fair comparison among providers [20]. In our study, we only have physician-level information on MRI referrals (i.e., number of referrals per 1,000 patients) without explicit patient-level information (i.e., whether a patient was referred an MRI by the physician during a visit). We therefore apply aggregate-level risk adjustment models which calculate the expected referral rate per physician, based on his/her patient population and practice types. The traditional approach to model rates is to use the computed rate for each aggregate unit as a dependent variable in an ordinary least-squares (OLS) regression [81]. However, when the rates are computed from small number of events, this OLS approach should be replaced by the Poisson-based regression, a statistical approach to analyze aggregate rates that solves the problem arising from small populations and low rates [81]. In our study, since the referral rates to be modelled are small with highly skewed distribution, we use Poisson-based regressions as our risk adjustment model type. The candidate Poisson-based models in our study include the basic Poisson regression model and negative binomial (NB) regression model.

4.1.1 The basic Poisson regression model

In statistics, the Poisson regression models belong to GLMs with the logarithm as the link function, and are used to model counts or small rates. Poisson regression assumes the dependent variable, Y, has a Poisson distribution, and the logarithm of its expected value can be expressed by a linear combination of unknown parameters. The basic Poisson regression for counts is expressed as

$$\ln(\lambda_i) = \mathbf{X}_i^\top \boldsymbol{\beta} \tag{4.1.1}$$

$$P(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \tag{4.1.2}$$

Equation 4.1.1 is a regression equation that relates the natural logarithm of mean (expected) number of events for case i, $\ln(\lambda_i)$, to the sum of explanatory variables, $\mathbf{X}_i^\top$, multiplied by regression coefficients, $\boldsymbol{\beta}$. Equation 4.1.2 indicates the probability that y_i, the observed number of events for case i, follows the Poisson distribution with the mean (expected) number of events, λ_i, from Equation 4.1.1. The role of the natural logarithm in Equation 4.1.1 is comparable to the logarithmic transformation of the dependent variable, a transformation method commonly used in analysis of aggregate rates [81].

Since in our study, our interest is in the MRI referral rates rather than the counts of MRI referrals of each physician, we alter the basic Poisson regression so that it models the referral rates rather than counts. In this case, if λ_i is the expected number of MRI referrals for a given physician, then λ_i/n_i would be his/her corresponding referral rate per 1,000 patients, where n_i is the patient population size (in thousands) for that physician. We can thus derive a variation of Equation 4.1.1 to model referral rates, which is expressed as

$$\ln\left(\frac{\lambda_i}{n_i}\right) = \mathbf{X}_i^\top \boldsymbol{\beta} \tag{4.1.3}$$

$$\ln(\lambda_i) = \ln(n_i) + \mathbf{X}_i^\top \boldsymbol{\beta} \tag{4.1.4}$$

$\ln(n_i)$ is the natural logarithm of the population at risk, which is also called an offset to the regression model in Equation 4.1.1. By giving the offset a fixed coefficient of one, Poisson regression becomes a

model for rates, rather than for counts of MRI referrals.

4.1.2 Overdispersion and NB regression model

The basic Poisson model is appropriate only if the probability model of Equation 4.1.2 matches the data, which requires that the residual variance be equal to the fitted values, λ_i. To meet this requirement, one assumption of the Poisson model is that λ_i is the true rate for each case, and the explanatory variables account for all of the meaningful variations among the aggregate units. If not, the differences between the fitted and true rates will inflate the residual variance. However, it is very unlikely that this assumption will be valid, as the Poisson regression will not explain all of the variations in MRI referral rates. Residual variance will also be greater than λ_i if the assumption of independence among MRI referrals is violated. In our case, we cannot guarantee that the MRI referrals are independent, as one patient can have multiple MRI scans referred by the same physician. The dependence would increase the variability in referral rates beyond λ_i. Due to these two reasons, "overdispersion" in which residual variance exceeds λ_i is unavoidable, and applying the basic Poisson regression to such data can produce inaccurate expected rate for each physician.

There are different ways to solve the problem of overdispersion. NB regression is the best known and most widely available Poisson-based regression model that allows for overdispersion [81]. NB regression combines the Poisson distribution in Equation 4.1.2 with a gamma distribution of the unexplained variations in the expected event counts, λ_i. This combination produces the NB distribution, and replaces the initial Poisson distribution. The formula for the NB distribution is expressed as

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \phi)}{y_i!\Gamma(\phi)} \frac{\phi^{\phi}\lambda_i^{y_i}}{(\phi + \lambda_i)^{\phi - y_i}} \tag{4.1.5}$$

where Γ is the gamma function (a continuous version of the factorial function), and ϕ is the reciprocal of the residual variance [40]. In NB regression, the GLM regression form remains the same as the basic Poisson model, with Equation 4.1.1 for counts of MRI referrals and Equation 4.1.4 for MRI referral rates. Therefore, although the response probabilities associated with the fitted value differ from the basic Poisson regression model, the interpretation of regression coefficients remains the same.

4.1.3 Interpretation of model coefficients

In Poisson-based model, the coefficient estimates can be interpreted in terms of the rate ratio, or relative risk. For example, consider a comparison of two univariate Poisson models - one for a given explanatory variable x, and one after increasing x by 1. The two models can be expressed as

$$\ln(\lambda_1) = \beta_0 + \beta_1 x \tag{4.1.6}$$

$$\ln(\lambda_2) = \beta_0 + \beta_1 (x + 1) \tag{4.1.7}$$

where λ_1 in Equation 4.1.1 is the expected count/rate for model 1, and λ_2 in Equation 4.1.7 is the expected count/rate for model 2. The difference between the logarithm of λ_1 and λ_2 is expressed as

$$\begin{aligned} \ln(\lambda_2) - \ln(\lambda_1) &= \beta_1 \\ \ln\left(\frac{\lambda_2}{\lambda_1}\right) &= \beta_1 \\ \frac{\lambda_2}{\lambda_1} &= e^{\beta_1} \end{aligned} \tag{4.1.8}$$

The derived result suggests that by exponentiating the coefficient, we can obtain the multiplicative factor by which the mean response changes. The quantity of change, e^{β_1} in Equation 4.1.8, is referred as a rate ratio or relative risk, and represents a percent change in the response for one unit change in the explanatory variable x.

4.1.4 Methods of standardization

Standardization is a method used to compare rates of a given disease/outcome by removing the influence of factors that may confound the comparison [108]. There are two major standardization methods; one is used when the available "standard" is the structure of a reference population (direct method) and the other one is used when the "standard" is a set of expected rates (indirect method) [108]. Direct standardization can be used only when the stratification is chosen as the risk adjustment method. The SAVA analysis in Appendix C is an example of direct standardization, by which the age-adjusted MRI utilization rate in each CD is interpreted as the rate that the CD would have if they had the same age distribution as the overall population. Indirect standardization can be used when applying either the stratification or regression as the risk adjustment method. The studies performed by Hall et al. [42] and DeLong et al. [26] introduced in Section 1.3.1 both used indirect standardization, by which they calculated the expected rates using stratification [42] and regression [26], and compared each provider's observed value against what was expected. In our study, since we use the regression as our risk adjustment method to produce the expected referral rate for each physician, we apply indirect standardization that compares each physician's observed and expected referral rates in order to identify low- and high-ordering physicians.

4.2 Multi-class logistic regression

In statistics, the logistic regression is a regression model that uses a logistic function to model binary dependent outcome. Mathematically, a binary logistic regression has a dependent variable with two possible values labelled as "0" and "1". The logarithm of odds (log-odds) for the value labelled "1", is a linear combination of one or more independent variables; the independent variables can be either categorical or continuous variables. The logistic regression for binary outcomes is expressed as

$$P(Y_i = 1|\mathbf{X}_i) = \pi \tag{4.2.1}$$

$$\ln\left(\frac{\pi}{1-\pi}\right) = \mathbf{X}_i^\top \boldsymbol{\beta} = \eta \tag{4.2.2}$$

$$\pi(\eta) = \frac{e^{\eta}}{1 + e^{\eta}} \tag{4.2.3}$$

Equation 4.2.1 defines π as the probability of the outcome being labelled "1", given all the explanatory variables. Equation 4.2.2 is a regression equation that relates the log-odds for the value labelled "1", $\ln(\frac{\pi}{1-\pi})$, to a linear predictor, η, defined as the sum of explanatory variables, $\mathbf{X}_i^{\top}$ multiplied by regression coefficients, $\boldsymbol{\beta}$. Equation 4.2.3 defines the logistic function, which finds the probability, π, by inverting equation 4.2.2. The logistic function is an S-shaped sigmoid function with horizontal asymptotes at 0 and 1, allowing the logistic function to vary between 0 and 1.

In a binary logistic regression model, the dependent variable has two levels of either "0" or "1". If the outcome has more than two classes, the data can be modeled as multinomial logistic regression, or ordinal logistic regression if the multiple categories are ordered (e.g., the outcome of low, medium, or high). Multinomial logistic regression is an extension of binary logistic regression that allows for more than two categories of the dependent variable [99]. It creates multiple binary logistic regression analyses (i.e., $M-1$ regression analyses for modelling the outcome with M categories), one for each pair of outcomes. In multinomial logistic regression, one category of the outcome is selected as a baseline comparison group that is later compared against each of the other categorical groups. One assumption of the multinomial logistic regression is that the categories are nominal, meaning that there is no order to the categories of the outcome variable. The downside is that the ordering information is lost when using this approach. If the outcome is ordered, an alternative approach to model the multi-class outcome would be the proportional odds ordinal logistic regression (also called ordered logistic regression) [64]. The ordered logistic regression is an ordinal regression that models for ordinal dependent variables [64]. It is also an extension of the logistic regression model, allowing for more than two ordered response categories. However, the model only applies to data that meet the proportional odds assumption, the meaning of which can be exemplified as follows. Suppose fitting an ordered logistic regression model for the data with outcome of 1 (lowest), 2, or 3 (highest). The ordered logistic regression creates two binary models with events of $y > j$ v $y \leqslant j$, $j = 1, 2$. The logistic function can then be written as [113]

$$P(Y_i > j) = \frac{e^{\alpha_j + \mathbf{X}_i^{\top}\boldsymbol{\beta}}}{1 + e^{\alpha_j + \mathbf{X}_i^{\top}\boldsymbol{\beta}}}, \quad j = 1, 2 \tag{4.2.4}$$

The proportional odds assumption requires that the $\boldsymbol{\beta}$'s (but not the α's) be the same for all values of j [113]. In other words, the effect of the predictors on the odds of being (1) (2 or 3) v 1 and (2) 3 v (2 or 1) are the same. This assumption is often violated in the real-world data sets.

In logistic regression, the coefficient calculated for each predictor determines the adjusted odds ratio (OR) for the outcome associated with a 1-unit increase, when other predictors are being controlled [106]. The coefficient of each predictor provides a measure of magnitude of its influence on the outcome of interest [106]. Specifically, considering a binary logistic regression model with the form

$$\ln\left(\frac{\pi}{1-\pi}\right) = \text{logodds}_{(Y=1)} = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p \tag{4.2.5}$$

Let w be the odds that $Y = 1$ based on X_1,...,X_p, then

$$w = e^{(\beta_0 + \beta_1 X_1 + ... + \beta_p X_p)} \tag{4.2.6}$$

The interpretation of β_1 is illustrated as follows. By holding $X_2, ..., X_p$ fixed, the adjusted OR that

$Y = 1$ at $X_1 = a$ to $X_1 = b$ is

$$\frac{w_a}{w_b} = e^{\beta_1(a-b)} \tag{4.2.7}$$

From Equation 4.2.7, if X_1 increases by 1 unit and holding all other X's constant, the odds that $Y = 1$ changes by a multiplicative factor of e^{β_1}.

4.3 Goodness-of-fit evaluation

There are two popular fit statistics, Akaike's Information Criteria (AIC) [3] and Schwarz's (Bayesian Information) Criteria (BIC) [96] that combines log-likelihood with a penalty, and are useful for comparing different types of models with the same response and same data. In a statistical model, let k be the number of estimated parameters in the model, and let $\hat{L}$ be the maximum value of the likelihood function for the model. Then the AIC value of the model is expressed as [3]

$$\text{AIC} = 2k - 2\ln(\hat{L}) \tag{4.3.1}$$

Given a set of candidate models for the data, the model with the minimum AIC value is preferred. The rule of thumb in AIC is that (1) one model fits better than another if the difference in AIC > 10, and (2) one model is essentially equivalent to another if the difference in AIC is < 2. Similarly, the BIC is expressed as [96]

$$\text{BIC} = k\ln(n) - 2\ln(\hat{L}) \tag{4.3.2}$$

where n is the number of observed data. Like AIC, the model with the lowest BIC is preferred. The BIC generally penalizes additional parameters more strongly than the AIC, though the goodness-of-fit also depends on the size of n and relative magnitude of n and k.

4.4 Analysis pipeline

Our family physician test utilization analysis is divided into five parts. First, we develop physician-level risk adjustment models that adjust for physician's case-mix, and physician's practice type. We use the risk adjustment models to produce the expected referral rate per physician, and compute indirectly standardized rates as observed divided by expected rates. Second, we use the observed-to-expected (O/E) ratio to categorize physicians into low, typical and high testers. Next, we compare physician tester levels against multiple physician-level characteristics to identify any difference in these characteristics among the physician groups. Moreover, we compare tester levels against patients' subsequent visits to see whether higher testers were associated with fewer patients' subsequent visits. Last, we conduct a counterfactual analysis to see how many MRI referrals would have been decreased if all higher testers had referred at typical levels. Figure 4.1 shows the flowchart of the analysis pipeline.

4.4.1 Risk adjustment models

Before making comparisons among individual family physicians on extremities, spine, and brain MRI referrals, we first build up physician-level risk adjustment regression models that calculate the expected

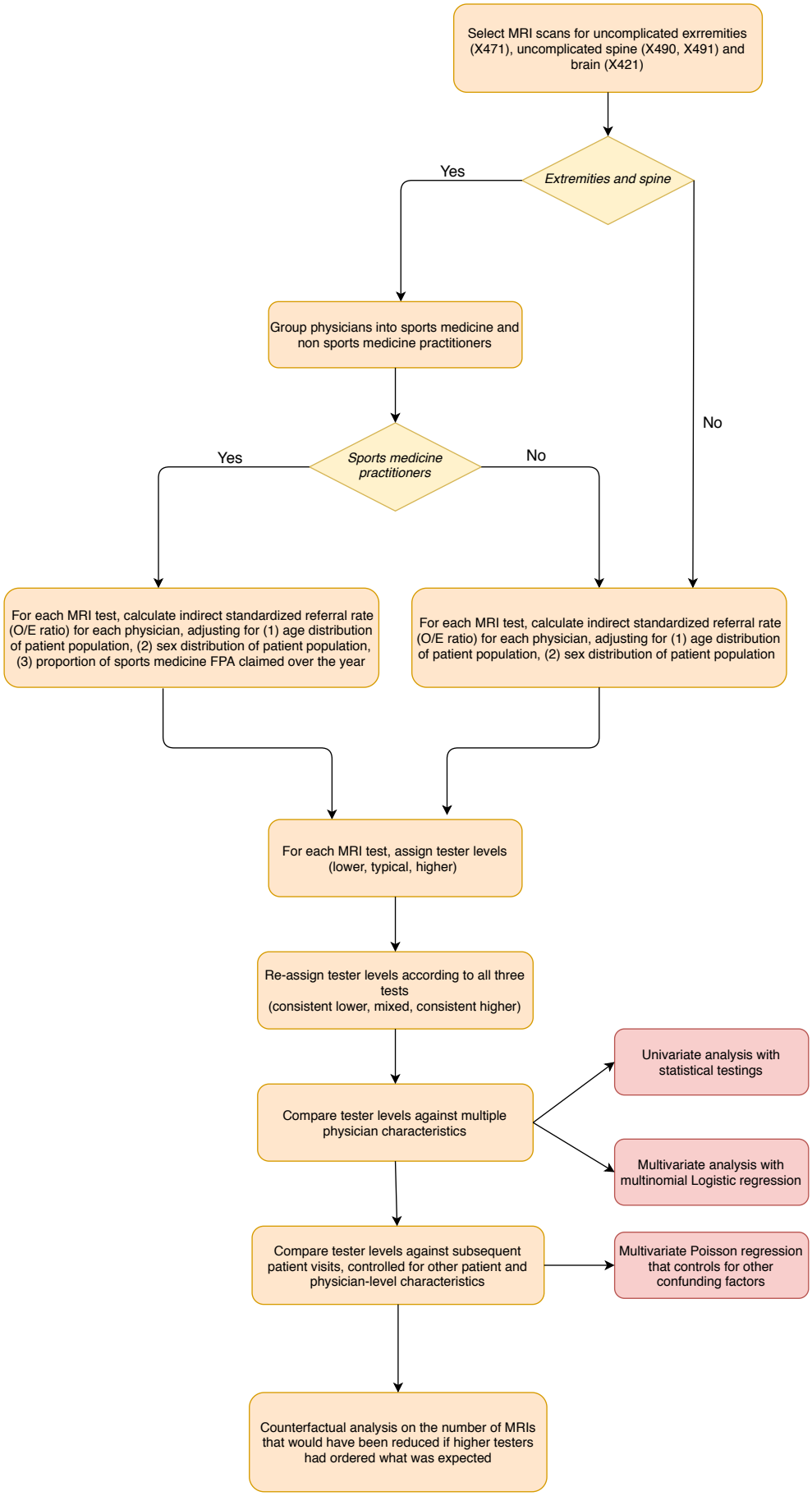


Figure 4.1: Family physician tests analysis pipeline

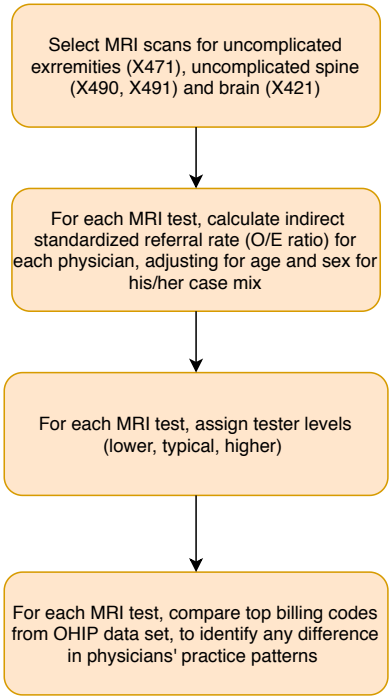


Figure 4.2: Identification of family physicians' practice type differences

referral rate for each physician on each test. At the first stage, we only adjust for patients' case-mix distribution of age and sex, and compare the OHIP billing codes among low, typical, and high testers to identify any difference in physicians' practice types (Figure 4.2, and refer to Appendix G for analysis details). OHIP billing codes are indicators of specific services insured by OHIP and claimed by physicians who provided the service [77]. Our findings show a different practice pattern in high-ordering physicians for extremities and spine. Specifically, high testers for extremities and spine MRIs tended to claim more sports medicine focused practice assessment (FPA) services. Sports medicine FPA (billing code of A917) is a special code claimed by family physicians granted with sports medicine focused practice designation, which requires an additional training and/or experience in sports medicine [78]. As a result, for the model development of extremities and spine MRIs, the physicians are further stratified as sports medicine practitioners and non sports medicine practitioners, based on their claimed billing codes through the year. We identify the physicians as sports medicine practitioners if they claimed at least 30 sports medicine FPA billing codes, and as non sports medicine practitioners otherwise.

We use two types of Poisson-based regression models, basic Poisson regression model and NB regression, to model physicians' referral rates by adjusting for relevant risk factors. We use the AIC value for model comparison, and select the models with smaller AIC as our final risk adjustment models. To model extremities and spine MRI referral rates, we fit the regression models in each stratified group; for brain MRI referral rates, we fit the regression models in the whole family physician population. We use each physician's paneled patients as his/her patient population. In each model, the measure of MRI referrals is the number of complete MRI scans referred by the physician over the study period. The measures of the explanatory variables are based on the physician's patient population and practice types. These explanatory variables include (1) the proportion of patients in each sex group; (2) the proportion of patients in each age group; and (3) the proportion of sports medicine FPA billing codes

over all claimed billing codes through a year, for sports medicine practitioners and for extremities and spine MRI tests. Table 4.1 lists all the risk factors and corresponding levels for each factor. For each risk factor, one of the levels is dropped as the reference to avoid multicollinearity (correlation of explanatory variables). All proportions are normalized into z-scores as final input features to risk adjustment models. For patient population's age and sex distribution, the normalization is based on the distribution across all physicians; for proportion of sports medicine FPA, the normalization is based on the distribution across all sports medicine practitioners. For each physician and each feature, the formula for z-score normalization is expressed as

$$Z_{\text{norm}} = \frac{\text{proportion} - \mu}{\sigma} \tag{4.4.1}$$

where μ is the mean value and σ is the stdev of the proportions. Also included in the model input is the size of each physician's patient population in thousands, which is the population at risk for MRI referrals. For each selected test, we report the percentage of variance explained by the risk adjustment model in terms of R^2. R^2 is a statistical measure that represents the proportion of the variance for a dependent variable (i.e., referral rate in our case) that is explained by independent variables in a regression model. R^2 equals 1 if the predicted values exactly fit into actual values, and equals 0 if the predicted value is just the mean [68]. Real-world risk adjustment models typically fall between 0 and 1 [68].

We choose regression-based risk adjustment over conventional stratification due to the following four reasons. First, it is evident from the literature that compared to stratification, Poisson regression models have several advantages [24]. Not only does the method saves time, but it also provides more reliable estimates than the ones obtained by stratified analysis [24]. Moreover, Poisson regression solves the problem of which external standard distribution to use, since the parameters are estimated internally from the study population [24]. With this method, the choice of one population against another does not affect the size of differences obtained with stratified analysis [24]. Thus, Poisson regression can be considered as a serious alternative to stratified analysis when comparing rates across physicians [24]. Second, for extremities and spine, the risk adjustment models for sports medicine practitioners include more than two relevant risk factors, in which case regression is preferred rather than stratification. Third, the coefficients of Poisson regression models can be interpreted as rate ratios, which help us understand what risk factor(s) significantly impact the outcome measures. Last, while stratification only provides a single summary measure (adjusted rate) for each physician, Poisson regressions are able to generate referral rate estimates with 95% confidence intervals (CIs), which help us more accurately identify the physician outliers whose O/E ratios are significantly above or below the thresholds. Section 4.4.2 elaborates in detail on how the O/E ratios are calculated, and how the physician outliers are identified.

4.4.2 Family physician test utilization

We use the fitted models to create risk-adjusted physician-specific referral rate estimates, adjusting for multiple relevant risk factors. From these models, we obtain the expected referral rates (number of MRI referrals per 1,000 patients) at physician level, and compute indirectly standardized rates as observed divided by expected rates, which are also known as O/E ratios. For each physician and each MRI test, the O/E ratio measures the extent to which his/her actual referral rate deviates from the expected rate. The observed rate is the physician's actual referral rate, calculated as number of complete MRI scans

Table 4.1: Risk factors and levels for risk adjustment models

Risk factor	Level
Patient population age group (%)	0-19
	20-39
	40-59
	60-79
	80+ (reference)
Patient population sex group (%)	Male
	Female (reference)
Proportion of sports medicine FPA practices (for extremities and spine)	

referred divided by the size of his/her paneled patients (in thousands); the expected rate is the referral rate if his/her utilization is identical to that observed in the stratified group (for extremities and spine) or in the entire physician population (for brain). We extract the 95% CI of the expected referral rate in order to derive 95% CI for the O/E ratio, and assign all physicians into lower, typical and higher testers. For each MRI test, we classify those physicians with CI of O/E entirely above 1.5 as higher testers, those with CI of O/E entirely below 0.5 as lower testers, and those with CI of O/E spanning between 0.5 and 1.5 as typical testers. The thresholds are based on the assumption raised by Hall et al. that a 50% increase or decrease in the ordering of routine tests would be of clinical significance [42]. We then use Venn Diagrams to show the overlappings of lower, typical and higher testers among the three MRI tests. We re-assign those family physicians into "consistent lower testers" if they were lower testers for all tests, into "consistent higher testers" if they were higher testers for all tests, and identify the remaining physicians as "mixed testers".

4.4.3 Comparison among tester groups

After assigning tester levels based on all three selected tests, we compare tester groups against multiple physician characteristics to identify possible sources of variations. The analysis unit is physicians, with physician-level characteristics (years of practice, region, rurality, immigration, income, patient population comorbidity) as the key independent variables of interest, and each physician's re-assigned tester level (consistent lower, mixed, consistent higher) as the key dependent variable. We aggregate patient-level rurality, immigration, income, and comorbidity index into physician-level categories based on each physician's patient population; for each physician and each characteristic, we calculate the average index of his/her patient population, and categorize the physician as low/moderate/high based on our observation of the histograms. Appendix F shows the histograms of the four characteristics at physician-level, and table 4.2 provides the thresholds for categorization. Patients with missing/unknown indices are excluded from our calculation. For our analyses, years of practice is treated as a continuous variable; region, rurality, immigration, income, and patient comorbidity are described as categorical variables.

We first perform a univariate analysis to see how each variable of interest individually affected a physician's tester level. In each level of categorical variables, we derive the prevalence of physicians in each tester group to notice any difference in distributions. All physician characteristics are compared

Table 4.2: Thresholds for categorization

Physician characteristic	Level	Range of average index
Rurality	Low	<2.5
	Moderate	2.5-4
	High	>4
Immigration	Low	<1.5
	Moderate	1.5-2.5
	High	>2.5
Income	Low	<2.5
	Moderate	2.5-3.5
	High	>3.5
Comorbidity	Low	<0.15
	Moderate	0.15-0.25
	High	>0.25

amongst tester levels using one-way analysis of variance (ANOVA) for the continuous variable, and Chi-squared Test of Independence for categorical variables. ANOVA is a statistical test to analyze the differences among group means with the null hypothesis that there is no difference among groups. It computes the F statistic as the ratio of between-group variance and within-group variance. If the p-value of F statistic is below the significance level (typically 0.05), the null hypothesis is then rejected and there is evidence supporting the significant differences amongst groups. The Chi-squared Test of Independence is used to determine if there is a significant relationship between two categorical variables. The null hypothesis assumes that there is no association between the two variables. Hypothesis testing for the Chi-squared Test of Independence is similar to ANOVA, where a test statistic (i.e., Chi-squared statistic) is computed and the p-value is compared to a significance level and the degrees of freedom. If the observed Chi-squared test statistic is below the significance level, the null hypothesis is rejected, supporting the association between two variables. After our univariate analysis, we then perform a multivariate logistic regression analysis to explore the adjusted relationship between physician's characteristics and the tester levels. Since our outcome variable is ordered and multi-class, the data can be fitted into both multinomial logistic regression and ordinal logistic regression models. We thus try both model types and select the optimal type based on AIC values. We use adjusted OR with 95% CIs to interpret how each independent variable influences the outcome.

In addition, we want to determine whether variations in the use of MRI tests were associated with variations in patients' subsequent visits. Our hypothesis is that higher testers might be ordering more inappropriate tests, and therefore be potentially associated with fewer patients' follow-up visits per referred MRI scan. To test out our hypothesis, we retrieve all referred scans for extremities, spine, and brain, and the count of patients' subsequent visits within six months after their MRI scans. We again use the Poisson-based regression model to control for other patient and physician-specific features that may influence the subsequent visits. Each data point fitted into the regression model represents an MRI scan history of one of the three MRI tests for a specific patient. In the regression model, the outcome is the number of visits within 6 months after an MRI scan; the exposure variable (variable of interest) is

the tester group of the referring physician, and the confounders are other patient and physician-specific characteristics. Table 4.3 lists all the input features for the regression model. We use the rate ratios with 95% CIs from the model to interpret how tester groups impact the patients' subsequent visits.

4.4.4 Counterfactual analysis

Our counterfactual analysis consists of two "what-if" scenarios. First, we perform the analysis on individual selected tests to understand if those 20% higher testers had ordered what was expected, how many MRI referrals in each test would have been decreased. To perform this analysis, in each test, we use the predicted referral rates (with 95% CI) to calculate the expected number of MRI referrals for those physicians who are identified as higher testers. We then calculate the expected total MRI referrals if those higher testers had ordered what was expected, while the total referrals for typical and lower testers remain unchanged. We report the number and percentage of decrease of MRI referrals compared with the actual volumes. Second, we want to understand if only those 8% consistent higher testers had ordered what was expected, how many total MRI referrals would have been decreased. In this case, we use the predicted rates (with 95% CI) for to calculate the expected number of MRI referrals in each test only for those consistent higher testers. Similarly, we then calculate the expected total MRI referrals and report the number and percentage of decrease.

4.5 Results

Our family physician test utilization analysis shows different referral patterns of family physicians based on their practice types, and there are several physician characteristics that impact the use of MRIs. Detailed analysis is provided for (1) MRI referral rates distribution, (2) model selection and interpretation, (3) physician O/E ratios and tester groups, (4) comparison among tester groups, and (5) counterfactual analysis. The Poisson-based statistical models and multi-class logistic regressions are fit in R version 3.2.3 [88]. We use package `MASS` [110] to call the NB regression and ordered logistic regression functions. We use package `glm2` [63] to call the Poisson regression function. We use package `mlogit` [23] to call the multinomial logistic regression function. All statistical tests are two-sided with 0.05 as significant level.

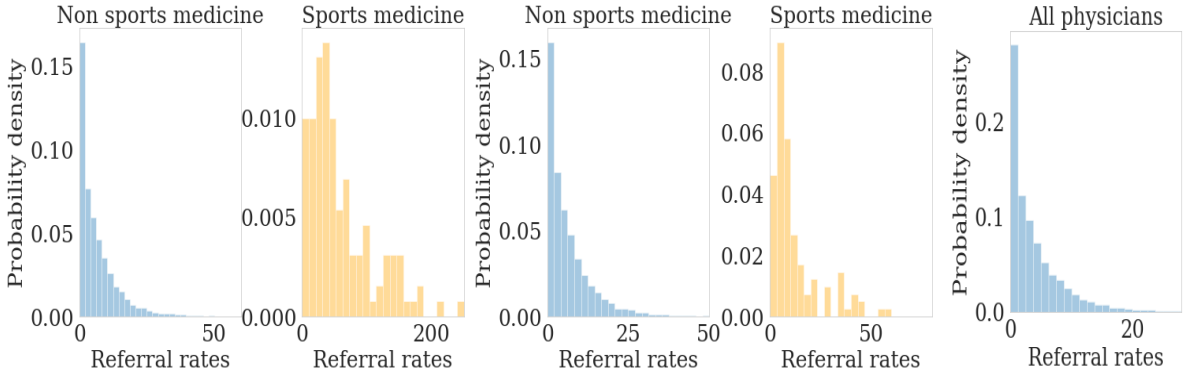
4.5.1 MRI referral rates distribution

Figure 4.3 shows the histograms of the crude referral rates per 1,000 patients by practice types for the three tests. We can see highly skewed distributions across all tests, with the majority of physicians having low rates. The highly skewed distribution suggests that Poisson-based risk adjustment models are preferred against OLS regression for modelling referral rates.

The pointplots of referral rates by practice types for extremities and spine are presented in Figure 4.4. Table 4.4 provides detailed statistics on mean (with stdev) and IQR of referral rates in each group and each test. Sports medicine practitioners were associated with higher MRI referrals for extremities and spine. Family physicians practicing sports medicine referred more extremities (64.1 v 6.6 exams / 1,000 patients) and spine (14.5 v 6.4 exams / 1,000 patients). Such subgroup differences suggest that practice-type-stratified models are necessary for risk adjustment.

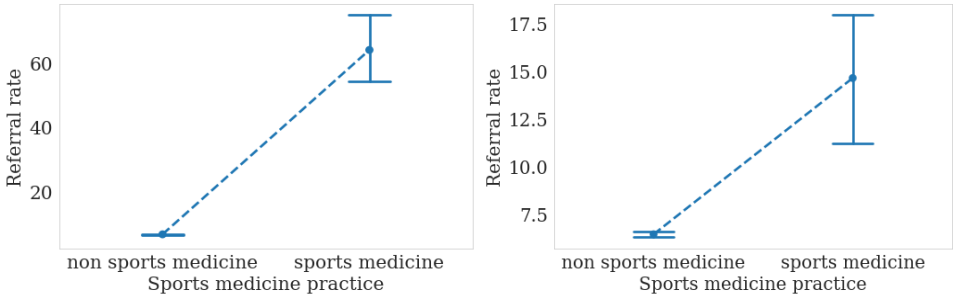
Table 4.3: Explanatory variables for patients' subsequent visits

Explanatory variable	Level
Exposure	
Referring physician tester group	Consistant lower (reference)
	Mixed
	Consistant higher
Counfounder	
Physician characteristics	
Region	Central (reference)
	West
	North
	East
	Toronto
Years of practice	
Sex	Male (reference)
	Female
Patient characteristics	
Age group	0-19
	20-39
	40-59
	60-79
	80+ (reference)
Sex	Male (reference)
	Female
Income index	1 (lowest, reference)
	2
	3
	4
	5 (highest)
Immigration index	1 (lowest, reference)
	2
	3 (highest)
CCI index	0 (lowest, reference)
	1
	2
	>2



(a) Extremities (b) Spine (c) Brain

Figure 4.3: MRI referral rates (per 1,000 patients)



(a) Extremities (b) Spine

Figure 4.4: Mean MRI referral rates (per 1,000 patients) with 95% CI

Table 4.4: Descriptive statistics by body part and physician sub-group

			Referral rate statistics	
Body part	Physician sub-group	No. of physicians	Mean (stdev)	IQR
Extremities	non sports medicine	11,517	6.58 (8.18)	1.16-9.16
	sports medicine	127	64.12 (61.26)	24.05-89.29
Spine	non sports medicine	11,517	6.44 (8.12)	4.98-14.62
	sports medicine	127	14.51 (20.48)	4.55-15.24
Brain	-	11,644	4.21 (5.71)	0.68-5.71

Table 4.5: Comparison of basic Poisson and NB Models

	Akaike information criterion (AIC)					
	Extremities		Spine		Brain	
	Poisson	NB	Poisson	NB	Poisson	NB
Non sports medicine group	155,863	74,056	147,530	74,875	111,885	65,808
Sports medicine group	6,042.6	1,318.6	2,510.2	986.68	-	-

4.5.2 Model selection and interpretation

For our two types of risk adjustment models, basic Poisson regression and NB regression, we use AIC to compare the goodness-of-fit. Table 4.5 presents the AIC of models for each stratified group and each body part. NB models have consistently better performance compared with basic Poisson models for all subgroups. Therefore, we choose NB models as our final risk adjustment models for calculating the expected referral rates. The results from NB models in terms of rate ratios with 95% CI are presented in Appendix H. The impact of each risk factor on referral rates varied across subgroups. In general, physicians who interacted with more female, middle-aged-to-elderly patients (ages 40-79) tended towards more MRI referrals. For extremities, sports medicine practitioners who interacted with more male and patients ages 20-39 tended towards more MRI referrals. Proportions of engagement in sports medicine FPA practice had significant impact on MRI referrals; physicians with higher proportions with sports medicine FPA practices tended to refer more extremities and spine MRIs. Specifically, for extremities MRIs, 1 stdev of increase in proportion of sports medicine practice increased the referral rate by 51% (95% CI 33%-72%); for spine MRIs, 1 stdev of increase in proportion of sports medicine practice increased the referral rate by 53% (95% CI 30%-81%).

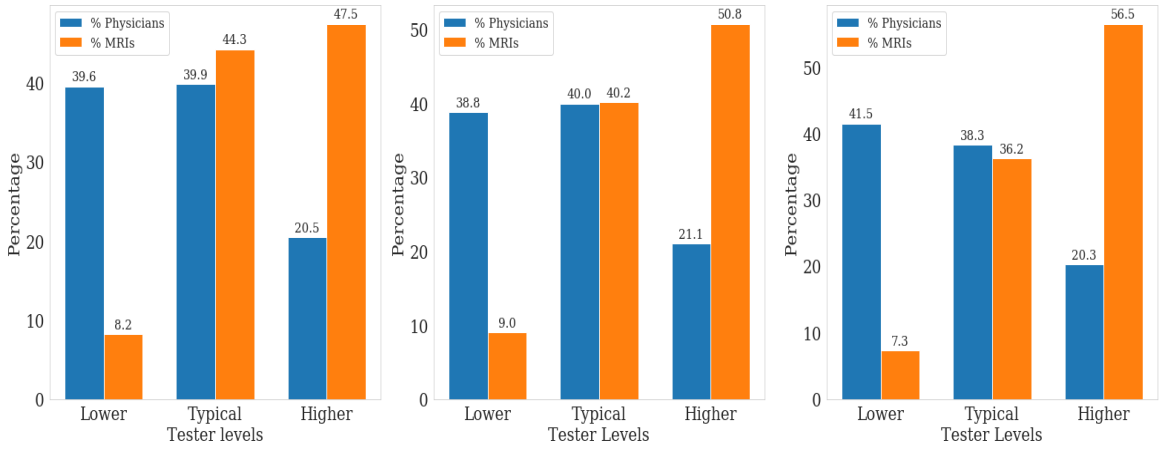
4.5.3 Physician O/E ratios and tester groups

Table 4.6 lists the mean (with stdev), maximum, and IQR values of the O/E ratios, and the percentage of variance explained by risk adjustment models. There was little difference in terms of mean and IQR of O/E ratios among selected tests. The maximum O/E ratio for extremities was the highest; the extreme outlier(s) for extremities referred over 40× as expected. The risk adjustment models explain 36.55%, 5.82% and 6.77% of the referral rates variance for extremities, spine, and brain, indicating that after controlling for physician's case-mix and practice types, there still existed high variance of MRI referrals among family physicians. Figure 4.5 shows the distribution of physician tester groups (lower, typical, and higher) along with their contributions to total MRI scans. 20.5%, 21.1%, and 20.3% physicians were identified as higher testers for extremities, spine, and brain MRIs, who contributed to 47.5%, 50.8%, and 56.5% MRI referrals, respectively. On average, higher testers referred 30% more MRI tests than typical testers, and lower testers referred 80% fewer MRI tests than typical testers.

Figure 4.6 presents the overlapping of lower, typical, and higher testers among all selected tests in Venn diagrams. After re-assigning the tester levels, we have 2,507 (21%) consistent lower testers, 904 (8%) consistent higher testers, and 8,253 (71%) mixed testers. Those 8% consistent higher testers contributed to 23.8% total MRI referrals of the selected tests. Multiple physician-level characteristics are later compared against the tester levels.

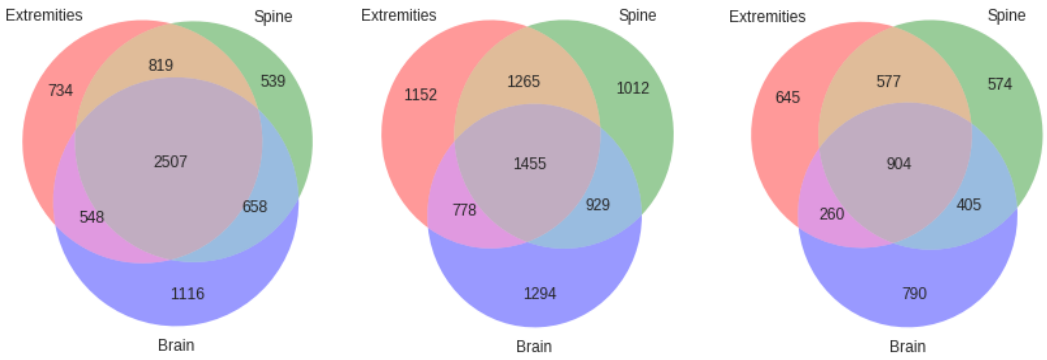
Table 4.6: Descriptive statistics of O/E ratios and % variance explained for selected tests

Body part	Mean (stdev)	Max	IQR	% Variance explained
Extremities	1.02 (1.5)	43.62	0.21-1.39	36.55%
Spine	1.03 (1.34)	37.31	0.24-1.39	5.82%
Brain	1.00 (1.31)	24.82	0.21-1.34	6.77%



(a) Extremities (b) Spine (c) Brain

Figure 4.5: Distributions of lower, typical and higher testers and their contributions to MRI referrals



(a) Lower (b) Typical (c) Higher

Figure 4.6: Venn diagrams of lower, typical and higher testers

4.5.4 Comparison among tester level groups

We compare physician-level characteristics among tester groups. Table 4.7 shows for each level of characteristics, the distributions of consistent lower, consistent higher, and mixed testers. Consistent higher testers had slightly fewer years of practice (mean of 8.93 v 10.28 for consistent lower testers, $p < 0.001$), were more likely to work in northern Ontario (12.10% v 6.97% on average in other parts of Ontario, $p < 0.001$), in high-income areas (7.25% v 4.62% in low-income areas, $p < 0.001$), and in low-immigration areas (9.05% v 3.01% in high-immigration areas, $p < 0.001$). On the other hand, consistent lower testers were more likely to work in urban settings (23.90% v 18.88% in high-rurality areas, $p < 0.001$), in low-income areas (29.41% v 16.54% in high-income areas, $p < 0.001$), and in high-immigration areas (35.44% v 19.53% in low-immigration areas, $p < 0.001$). There was no clear pattern on the relationship between tester levels and comorbidities of their patient population.

We use the model output from multinomial logistic regression to interpret the multivariate analysis result, since it better fits our data set (AIC of 17,238 v 17,576 for ordered logistic regression), potentially indicating that our data set may violate the proportional odds assumption illustrated in Section 4.2. In addition, although our outcome has three levels, we focus more on how the explanatory variables influence the odds of being a consistent higher v consistent lower tester, in which case the multinomial logistic regression gives us the desirable interpretation. The multivariate regression analysis is shown in Table 4.8, showing the same pattern as our univariate analysis. The findings include (1) a 5-year increase in years of practice decreases the odds of being a higher v lower tester by 0.06 (95% CI 0.05-0.07); (2) a change of rurality level from low to high increases the odds of being a higher v lower tester by 0.9 (95% CI 0.39-1.6); (3) a change of income level from low to high increases the odds of being a higher v lower tester by 0.45 (95% CI 0.1-1.1); (4) a change of immigration level from low to high decreases the odds of being a higher v lower tester by 0.79 (95% CI 0.66-0.87); and (5) a change of region from Toronto to North increases the odds of being a higher v lower tester by 2.17 (95% CI 1.01-4). The influences of these factors on the odds of being a mixed v lower tester are similar to that for being a higher v lower tester, but overall, to a smaller magnitude.

Figure 4.7 shows the pointplot of patients' subsequent visits referred by tester groups, where we see that subsequent visits decreased as the tester level goes from consistent lower to consistent higher. Based on the NB regression, when controlling for other patient and physician factors, the subsequent visits of patients referred by consistent higher testers were 10.7% (95% CI 9.26%-12.14%) lower than that for lower testers; the subsequent visits of patients referred by mixed testers were 7.3% (95% CI 5.87%-8.71%) lower than that for lower testers. Rate ratio estimates and CIs from the basic Poisson regression are similar to the NB regression output.

4.5.5 Counterfactual analysis

Table 4.9 shows the counterfactual analysis result for scenario 1, in which case 20% higher testers for each body part had ordered what was expected. The number of MRI referrals by higher testers were 60,884 (47.5%), 22,994 (50.8%), and 41,216 (56.5%) for extremities, spine, and brain, respectively. In each of the selected test, if the 20% higher testers ordered what was expected, the MRI referrals would have been reduced by 37,196 (29%) for extremities, 35,815 (31%) for spine, and 26,365 (36%) for brain. Overall, the total MRI referrals of the three tests combined would have been reduced by 99,376 (31%). Table 4.10 presents the counterfactual analysis result for scenario 2, in which case only 8% consistent

Table 4.7: Physician characteristics by tester groups

	Overall n = 11,644	Lower n = 2,507 (21.49%)	Mixed n = 8,253 (70.76%)	Higher n = 904 (7.75%)	*p*-value
Years of practice					<0.001
Mean (stdev)	9.47 (8.23)	10.28 (8.94)	9.28 (8.11)	8.93 (7.07)	
Median	7	8	7	7	
IQR	3-12	3-17	3-12	4-10	
Region					<0.001
Central	3,519	716 (20.34%)	2,524 (71.72%)	279 (7.92%)	
East	2,880	593 (20.59%)	2,100 (72.91%)	187 (6.49%)	
North	842	142 (16.86%)	598 (71.01%)	102 (12.10%)	
Toronto	1,414	381 (26.94%)	977 (69.08%)	56 (3.96%)	
West	2,948	654 (22.18%)	2,014 (68.31%)	280 (9.50%)	
Immigration level					<0.001
Low	6,547	1,264 (19.31%)	4,684 (71.54%)	599 (9.15%)	
Moderate	4,101	890 (21.70%)	2,936 (71.58%)	275 (6.70%)	
High	996	353 (35.44%)	613 (61.52%)	30 (3.01%)	
Income level					<0.001
Low	1,275	375 (29.41%)	841 (65.95%)	59 (4.62%)	
Moderate	8,066	1,751 (21.71%)	5,637 (69.88%)	678 (8.40%)	
High	2,303	381 (16.54%)	1,755 (76.20%)	167 (7.24%)	
Rurality level					<0.001
Low	6,407	1,518 (23.69%)	4,465 (69.68%)	424 (6.62%)	
Moderate	3,280	642 (19.57%)	2,339 (71.31%)	299 (9.11%)	
High	1,957	347 (17.73%)	1,429 (73.01%)	181 (9.24%)	
Patient comorbidity level					<0.001
Low	3,843	877 (22.82%)	2,735 (71.16%)	231 (6.01%)	
Moderate	3,842	490 (12.75%)	2,944 (76.72%)	408 (10.62%)	
High	3,959	1,140 (28.80%)	2,554 (64.50%)	265 (6.69%)	

Table 4.8: Multinomial logistic regression analysis of the tester levels

	Mixed v Lower		Higher v Lower	
	Adjusted OR (95%CI)	*p*-value	Adjusted OR (95%CI)	*p*-value
Years of practice [1]	0.95 (0.95-0.96)	<0.001	0.94 (0.93-0.95)	0.008
Region				
Toronto (reference)	1	-	1	-
Central	1.34 (1.15-1.56)	<0.001	2.53 (1.83-3.5)	<0.001
East	1.09 (0.91-1.31)	0.364	1.39 (0.96-2.02)	0.085
North	1.22 (1.04-1.42)	<0.001	3.17 (2.01-5)	<0.001
West	0.8 (0.66-0.97)	0.024	1.55 (1.06-2.27)	0.025
Rurality				
Low (reference)	1	-	1	-
Moderate	1.57 (1.33-1.85)	<0.001	1.29 (0.99-1.68)	0.060
High	2.21 (1.81-2.69)	<0.001	1.9 (1.39-2.6)	<0.001
Income				
Low (reference)	1	-	1	-
Moderate	1.06 (0.91-1.23)	0.491	1.42 (1.04-1.94)	0.027
High	1.4 (1.16-1.71)	<0.001	1.45 (1.1-2.1)	0.047
Immigration				
Low (reference)	1	-	1	-
Moderate	0.99 (0.84-1.17)	0.922	0.75 (0.57-0.98)	0.038
High	0.52 (0.42-0.65)	<0.001	0.21 (0.13-0.34)	<0.001
Patient comorbidity				
Low (reference)	1	-	1	-
Moderate	1.72 (1.51-1.95)	<0.001	2.43 (1.98-2.98)	<0.001
High	0.52 (0.46-0.59)	<0.001	0.49 (0.39-0.62)	<0.001

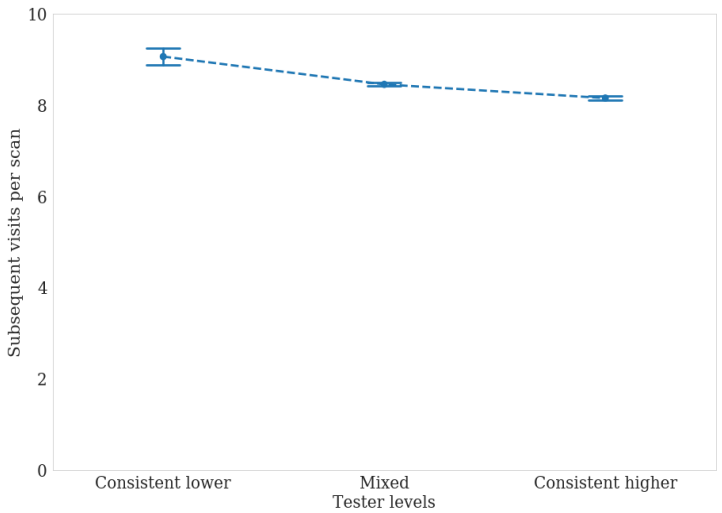


Figure 4.7: Mean subsequent visits per scan within 6 months with 95% CI

Table 4.9: Counterfactual analysis for scenario 1

	Body part		
	Extremities	Spine	Brain
No. total MRI referrals	128,131	115,718	72,946
No. referrals in higher group	60,884	58,809	41,216
No. expected referrals in higher group	23,689	22,994	14,852
(95% CI)	(22,416-25,081)	(22,072-23,987)	(14,278-15,452)
No. expected total MRI referrals	909,35	799,03	46,581
(95% CI)	(89,663-92,328)	(78,982-80,896)	(46,008-47,182)
No. total referrals reduced	37,196	35,815	26,365
(95% CI)	(35,803-38,468)	(34,822-36,736)	(25,764-26,938)
% total referrals reduced	29%	31%	36%
(95% CI)	(28%-30%)	(30%-32%)	(35%-37%)
Overall - No. total referrals reduced (95%CI)			99,376 (96,389-102,142)
Overall - % total referrals reduced (95%CI)			31% (30%-32%)

higher testers had ordered what was expected. The number of MRI referrals by those 8% consistent higher testers were 28,821 (22.5%), 27,627 (23.9%), and 19,016 (26.1%) for extremities, spine, and brain, respectively. If only the consistent higher testers had ordered what was expected and other family physicians had remained unchanged, the overall MRI referrals would still have been reduced by 59,582 (18.8%).

4.6 Discussion

This family physician investigation reveals several insights about patterns of referrals among family physicians. First, the distributions of MRI referrals for extremities, spine, and brain were highly skewed, with the majority of physicians having small referral rates (less than 10 referrals per 1,000 patients). Extremities and spine referral rates were predictable based on the level of engagement in sports medicine FPA, a designation granted to physicians with specialized service in sports medicine. Specifically, sports medicine practitioners referred nearly 10× the number of extremities MRIs per 1,000 patients and over 2× the number of spine MRIs per 1,000 patients. This finding is consistent with the claim that sports medicine practitioners have a clear interest in radiology, and MRI is particularly useful in identifying bone stress, which occurs in a wide variety of sports injuries [66]. In addition, based on the policy and program overview of the FPA group [78], sports medicine practitioners tend to be interacted with more patients with lumbar strain and lumber sprain, increasing their likelihood of referring spine MRIs.

Second, even after controlling for relevant risk factors, there still exited high variations among family physicians; the risk adjustment models could help explain the variations to some extent, but there is still a large proportion of variance that the model could not explain. Individual physicians are exposed to other unpredictable factors related to medico-legal concerns, pressure from patients, and marked variations in their inherent ordering behaviours [115]. After assigning lower, typical, and higher tester levels to

[1]The adjusted OR estimate is for a 5-year increase in the parameter of interest

Table 4.10: Counterfactual analysis for scenario 2

	Body part		
	Extremities	Spine	Brain
No. total MRI referrals	128,131	115,718	72,946
No. referrals in consistent higher group	28,821	27,627	19,016
No. expected referrals in consistent higher group (95% CI)	6,154 (5,802-6,550)	5,793 (5,560-6,045)	3,935 (3,786-4,090)
No. expected total MRI referrals (95% CI)	105,464 (105,112-105,860)	93,884 (93,651-94,136)	57,865 (57,716-58,020)
No. total referrals reduced (95% CI)	22,667 (22,271-23,019)	21,834 (21,582-22,067)	15,081 (14,926-15,230)
% total referrals reduced (95% CI)	17.7% (17.4%-18.0%)	18.9% (18.7%-19.1%)	20.7% (20.5%-20.9%)
Overall - No. total referrals reduced (95%CI)			59,582 (58,779-60,316)
Overall - % total referrals reduced (95%CI)			18.8% (18.6%-19.0%)

family physicians on each selected test, our finding shows that overall, only 20% physicians were higher testers, who contributed to nearly 50% MRI scans. On average, higher testers referred 30% more MRI tests than typical testers, and lower testers referred 80% fewer MRI tests than typical testers. After re-assigning the tester levels according to all three tests, we have 21% consistent lower testers, 8% consistent higher testers, and 71% mixed testers. Those 8% consistent higher testers contributed to nearly 25% total MRI referrals of the selected tests. Several physician-level characteristics were associated with the variations in family physician's MRI test utilizations. First, physicians with fewer years of practice had a higher chance of being a high tester. Previous research has found that less experienced physicians order relatively more diagnostic tests [43, 58, 70], and are associated with more inappropriate imaging referrals [116]. Second, physicians who worked in the northern regions of Ontario had a higher chance of being a high tester. According to an interview study of Ontario family physicians, they complained about the long wait time for specialist consultation, saying that it would be much faster to get an MRI scan than to see a specialist [115]. This situation was exacerbated in northern Ontario, where family physicians sometimes refer an MRI to obviate the need for the specialist referral [114]. Third, physicians who worked in rural areas had a higher chance of being a high tester. This observation is aligned with a study which shows that due to shortage of local specialists and travel challenges for patients within these communities, rural family physicians tend to provide a variety of specialty services [15, 101]. Therefore, rural family physicians may prefer more diagnostic imaging tests that facilitate their diagnoses. Fourth, physicians who worked in high-income areas tended to refer more MRIs. The rationale is that people in higher socioeconmic status (SES) groups may be more likely to have a regular family physician, to be better educated about sophisticated imaging technologies, and to be more assertive healthcare consumers; while poorer health status in lower SES groups may disproportionately affect the utilization of routine radiography procedures [27]. Fifth, physicians working in low-immigration areas tended to refer more MRIs. There is no study that explicitly explains such association, but from a survey result in Statistics

Canada, immigrants in Canada had difficulty in accessing the healthcare services [22]. Another study concludes that new immigrants are much less likely to have a regular doctor [25]. Certain immigrants may prefer having healthcare professionals of their own ethnicity or who speak their language; such preferences may potentially explain some delay in finding a regular doctor [97]. These factors may disproportionately affect immigrants' utilization in MRI tests.

Moreover, after comparing the subsequent patients' visits within 6 months after their MRI scans, we find that the subsequent visits of patients referred by typical and consistent higher testers were 7.3% and 10.7% lower than that for consistent lower testers. This finding validates our assumption that higher testers might be ordering more inappropriate tests, and therefore be potentially associated with fewer patients' follow-up visits per referred MRI scan. However, since there is no explicit indication on whether an MRI referral is appropriate or not in our available data sets, the number of follow-up visits only act as an proxy to the appropriateness of MRI referrals.

Last, our counterfactual analysis indicates that placing restrictions on the higher testers can largely reduce the total MRI referrals. Specifically, when the 20% higher testers in selected test ordered what was expected, the total MRI referrals of the three tests combined would have been reduced by over 30%. Even if we only place restrictions on those 8% consistent higher testers who contributed to approximately 25% of each MRI test, the total MRI referrals would have still been reduced by nearly 20%. However, our counterfactual analysis is only a rough estimate with the assumptions that there are no other risk factors associated with such high MRI referrals.

4.6.1 Strengths and limitations

Our family physician investigation uses a large data set representative of the family physician population and their patient population in Ontario. The results expand on those of previous studies that relied on older data, using a rich data set that allows the comparison of tester levels against multiple physician-related characteristics. However, there still exists certain limitations. First, when doing provider profiling, several previous studies conducted individual patient-level risk adjustment and then aggregated to physician-level estimates [26, 86]. Such patient-level information is unavailable in our study, leading us to conduct physician-level risk adjustment. Therefore, the estimated referral rate for each physician may be less accurate than the estimates from individual-level risk adjustment. This problem is similar to modifiable areal unit problem [39], a problem that affects the statistical results when point-based measures of spatial phenomena are aggregated into districts with summary values (e.g., rates, proportions). Second, when doing physician profiling, we compare each physician's observed rate to a practice-based norm, an expected rate if his/her utilization is identical to that observed in the population. Practice-based norms do not necessarily reflect appropriate care [52]. For example, the mean MRI referral rate for family physicians may be too high or low. Although standards-based norms reflect appropriate care based on sound practice guidelines [52], such a standard guideline is unavailable in our study. Third, there are no data on the appropriateness of MRI referrals made by each physician. We have no information on either patient symptoms and reasons for MRI referrals, or any indication on the results of MRI tests. The information gaps prevent us from conclusively stating whether a higher MRI referral rate leads to better diagnoses and subsequent patient treatment. Last, when doing counterfactual analysis, our assumption is that the relevant risks are well-considered and there are no other risk factors for high MRI referrals. In reality, there might be specific reasons for such high referral rates. In a case study by Lasker et al. about physician profiling on diagnostic imaging tests, the authors provided

several possible reasons for physician outliers who had much higher diagnostic imaging referrals compared to their peers [52]. The potential reasons include (1) they had the wrong specialty designation; (2) they provided imaging tests in their offices whereas most of other physicians in their peer group refer patients who need the tests to a radiologist; and (3) they saw more patients with symptoms that requires diagnostic tests than other physicians in their peer group [52]. Therefore, our counterfactual analysis only provides a rough estimate on potential MRI referrals being reduced when placing restrictions on higher testers.

4.6.2 Recommendations for policy and interventions

Based on our findings and the information gaps, we propose several recommendations for policy and interventions. First, for family physicians, the crude referral rate may not be an appropriate way to measure quality or identify outliers, especially for extremities and spine, where sports medicine practitioners had significantly higher referral rates than other physicians. Therefore, the appropriateness of MRI use may be a more representative measure of quality. Second, based on the high proportion on MRI referrals and high inter-physician variance within family physician group, a restriction might be necessary to reduce the MRI use for this group, however, such restriction's threshold need to be tailored for physicians with different practice types and different case-mix. Third, there still exits wide variations among family physicians even after controlling for the known risk factors associated with the imaging use. Thus, it is recommended to inform physicians who were consistent higher testers in their peer group, and ask them to provide written responses on the rationale of high volume of MRI referrals.

Moreover, the information gaps in our study prevent us from addressing key policy-relevant questions. These information gaps include (1) the unavailability to retrieve patient symptoms and reasons for MRI use, and (2) the unavailability to retrieve clear MRI test results and their impacts on the subsequent treatment of patients. Therefore, we recommend to electronically use a standard form across the province to collect the patient symptoms and reasons for MRI referrals. The collected information would facilitate the development and implementation of clear criteria for appropriate MRI use, as well as the identification of most common symptoms that require MRI scans. We also recommend to collect the information on MRI scan results and subsequent patient treatments, which would help to develop a clear criteria on the appropriateness of MRI scans.

Finally, we provide several recommendations once the above information are collected across the province. First, rather than using handwritten ordering systems, we recommend a web-based computerized physician ordering entry system (CPOE) along with clinical decision support system (CDSS) that guides decision-making and decrease unnecessary MRI referrals. CPOE is a web-based ordering system with standardized requisition form that allows physicians to electronically write full range of orders [19]; CDSS is an analytical tool that converts raw data into useful information to offer safety alerts when potentially inappropriate orders take place, helping clinicians make better decisions [47]. Evidence-based CPOE with embedded point-of-care decision support has been identified as an effective method to ensure appropriate decision at the time of ordering [19, 49, 92]. When specific interventions such as CPOE and CDSS are taken place, we recommend to perform quality improvement studies to assess whether such interventions could reduce practice pattern variations and imporve patient care.

4.7 Conclusion

Our investigation into family physicians' referral rates for extremities, spine, and brain MRIs shows that there were significant variations in family physician's use of MRI, even after controlling for the physician's case mix and practice type. For each selected test, 20% physicians are identified as higher testers, which contributed to 50% MRI referrals. On average, higher testers referred 30% more MRI tests than typical testers, and lower testers referred 80% fewer MRI tests than typical testers. Consistent higher testers had slightly fewer years of practice, worked in rural, low-immigration, and high-income areas. Our counterfactual analysis shows that if the 20% higher testers in each test had ordered what was expected, the total MRI referrals would have been reduced by 31%; even if only the 8% consistent higher testers ordered what was expected, the total referrals would have still been reduced by nearly 20%. One limitation in our study is that there is no explicit indication of the appropriateness for each MRI referral, which prevents us from addressing key policy-relevant questions. We therefore recommend to electronically collect the symptoms and results of MRI scans, and implement a web-based physician ordering entry system embedded with clinical decision support to reduce practice pattern variations and improve patient care.

Chapter 5

Conclusion

MRI is an alternative to traditional ionizing diagnostic techniques, and is widely used in diagnosis of several diseases such as cancer, and neurological and cardiac disorders. In Ontario, Canada, completed MRI tests reached 880,000 in 2017, nearly doubled the number of MRI performed in 2008. One consequence of such an increase in MRI requests is the longer wait time; current average wait times for non-urgent MRI scans are significantly longer than the Ontario target of 28 days. In order to identify appropriate demand-side interventions, we need to fully understand the MRI utilization patterns in the province at both patient- and provider-level. Our research is a retrospective population-based study that focuses on completed MRI scans in Ontario between 2008 and 2017, to investigate how the MRI increased over time, and how different patient and physician groups are contributing to the growth. We contribute to the literature by presenting a comprehensive trend and practice pattern analysis using a variety of recent administrative data sources in Ontario. We link multiple administrative data sources and use merged data sets to perform a series of analysis. Our analysis can be divided into four sections. First, we analyze the overall utilization and wait times over the last decade. Second, we look deeper into the MRI utilization trends by scan body part and patient demographic characteristics. Third, we review physician referral patterns by specialties. Finally, we focus on family physicians and perform a more detailed referral pattern analysis within the group.

Our results show that in Ontario, from 2008 to 2017, there was an 80% increase in the number of completed MRI scans. The most common MRI scans were for extremities, spine, and brain. MRI utilization was the highest for women aged 40 to 69, and this patient group also contributed most to the growth of MRI use. Family physicians had both high MRI referrals and inter-physician variance, and contributed most to the MRI growth, which lead us to perform a deeper investigation into the group for the three referred body parts (extremities, spine, and brain), and for non-urgent patients. We find that physicians who were more engaged in sports medicine FPA referred more extremities and spine MRIs. When controlling for physicians' case-mix and practice type, for each body part, we identify 20% physicians as higher testers, which contributed to 50% MRI referrals. Out of all the physicians, 8% were consistent higher testers, who contributed to nearly 25% MRI referrals for all the three tests. Higher testers had slightly fewer years of practice, and worked in rural, low-immigration, and high-income areas. Our counterfactual analysis shows that if the 20% higher testers in each test had ordered what was expected, the total MRI referrals would have been reduced by 31%; even if only the 8% consistent higher testers ordered what was expected, the total referrals would have still been reduced by nearly

20%.

Several recommendations on policy and interventions are proposed based on our analysis results. First, based on the significant disparity of wait times among different priority levels, it is critical to prioritize MRI requests so that most urgent patients will be served in time. Second, due to the large volume of MRI referrals and high inter-physician variance within family physician group, interventions might be necessary to restrict MRI use for the group. Such restrictions should be tailored for physicians by practice focus (i.e., sports medicine focus or not) and case-mix. Third, the crude referral rate may not be an appropriate way to identify physician outliers, since variations in the use of MRI is clearly multi-factorial. Fourth, we recommend informing physicians who were consistent higher testers in their peer group, and asking them to provide written responses on the rationale of high volume of MRI referrals. Finally, because of the existing information gaps to evaluate the appropriateness of MRI referrals, we recommend to collect the patient symptoms and reasons for MRI referrals, as well as the MRI results and subsequent patient treatment. The collected information would facilitate a web-based decision support system that guides decision-making and decrease unnecessary MRI referrals.

Moving forward, there are several potential future research directions. First, since the current study focus is MRI use rather than wait times, we only perform a high-level analysis on MRI wait times. In the future, we could perform more detailed exploratory analyses and predictive tasks for wait times. The analyses could include, but are not limited to the following: (1) analyze wait times by patient characteristics, such as age, sex, body part, and residing regions; (2) perform a data envelopment analysis (DEA) at the hospital level, using wait time as one of the evaluation metrics; (3) predict long or short wait times according to the patient- and physician-level characteristics, as well as the waiting system characteristics; and (4) investigate how wait time changes in hospitals would impact the likelihood of physicians referring patients to those hospitals. Second, we can examine the repeat use of MRI scans associated with physician, hospital, and diagnosis characteristics. In particular, we can examine whether higher testers were associated with the higher usage of repeat scans. Third, we can perform a deeper analysis on how tester levels influence the patient treatment. Specifically, we can choose a specific test and look into the impact of the test on patients' subsequent referrals and treatments. It was reported that some MRI scans are ordered as "gatekeepers" for other treatments; most Canadian spine surgeons (84%) require imaging studies to accompany all spine-related referrals [46]. Therefore, we can examine whether the patients referred by higher testers of spine MRIs were associated with more spine surgeons. For extremities, similarly, we can examine whether the patients referred by higher testers of extremities MRIs were associated with more orthopedic surgeons. Finally, once the patient symptoms, the reasons for MRI referrals, and the MRI results are electronically collected, we can further build a decision support system that provides in-time evaluation on the appropriateness of MRI referrals and inform physicians if a potential inappropriate referrals are being made.

Appendix A

Handling of Big Data

Due to the size and confidentiality of the data holdings in our study, we store them in the morLAB cluster with restricted cluster access and restricted directory access. The morLAB cluster, named Netu, is a CentOS Linux cluster [69]. It has 15 nodes with four quad-core AMD2354 Opteron processors for a total eight cores per node [69]. Storage is provided by a 36 TB storage component [69]. For our study, the data holdings take up over 660 GB of the storage. We deploy a Spark standalone cluster that enables the distribution of the workloads over 15 nodes. Apache Spark is a unified engine for big data processing with a programming model similar to MapReduce, but extends it with a data-sharing abstract called "Resilient Distributed Datasets", or RDDs [118]. Using this extension, Spark is able to perform different implementations over a common engine, making them easy and efficient to compose. The implementations enabled by Spark includes SQL, streaming, machine learning, and graph processing (Figure A.1) [5, 41, 119]. The sections below provide a detailed illustration of how we use Apache Spark framework to handle big data. Specifically, we introduce how we connect the Spark standalone cluster with Python and Jupyter Notebook, how we use Spark data objects and packages to process data, how Spark uses the "lazy operation" to optimize the processing efficiency, and how we convert the data into a Spark-supported columnar format to speed up queries.

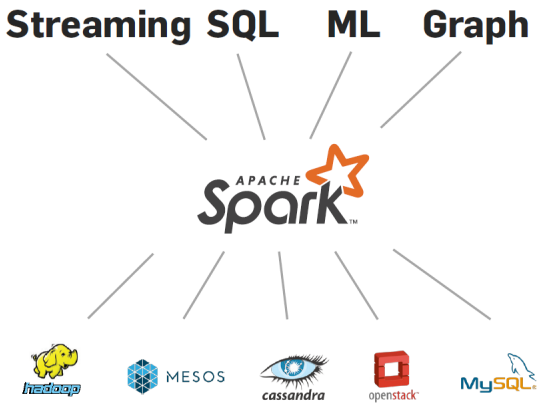


Figure A.1: Apache Spark software stack and implementations over the core engine [120]

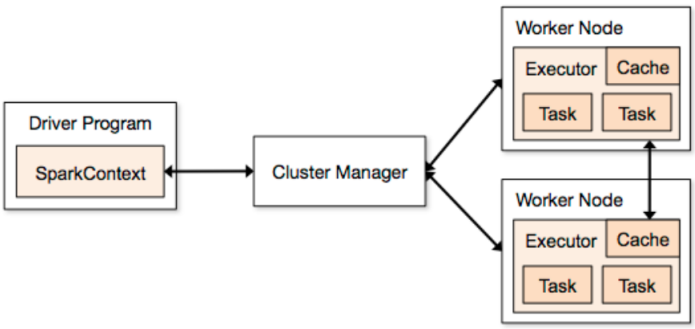


Figure A.2: Workflow of task distribution by Spark Cluster manager [32]

A.1 Setting up Pyspark environment

Spark executors work as independent processors on a cluster, coordinated by the SparkContext object in the main program [32]. Specifically, to perform a task on a cluster, the SparkContext object connects to the cluster manager, which allocates resources across workers (nodes). Once connected, Spark acquires executors on nodes in the cluster. The executors in each node act as processors that store the partitioned data and run the assigned task. Figure A.2 shows the workflow of how the task is distributed by Cluster manager. Spark currently supports three cluster managers: Standalone, Apache Mesos and Hadoop YARN [32]. Since Standalone is a simple cluster manager that is easy to set up, we choose to install this type of cluster manager on our multi-node MorLAB cluster. Each worker in the Spark cluster has 8 cores (1 executor per core) with 30 GB of memory. In total, there are 120 cores (executors) and 450 GB in use.

In Spark, the implementation of the tasks is enabled through an Application Programming Interface (API) in Scala, Java, Python, and R, where users can simply pass local functions to run on the cluster [120]. Python API in Spark, also called PySpark, is a flexible, robust, and easy-to-learn language that is preferred by many of the data scientists, despite of its relatively low processing speed compared to Scala. Jupyter Notebook is a popular application that enables editing, running and sharing Python code into a web view. It is a great tool for data scientists to analyze the data and visualize results. For our study, we enable Python API in Spark and connect it to Jupyter Notebook by adding a function to the bash configuration file of the cluster (Figure A.3). Under the"snotebook" function shown in Figure A.3: part A specifies the Spark path; part B configures PySpark driver by updating Pyspark driver environment variables; part C specifies the Python path; part D specifies the URL of Spark Master as well as the in-use memory for each working when running Spark applications. Once this function is called from the terminal, it will launch the Jupyter Notebook with the Pyspark environment embedded. To run Jupyter Notebook on a remote server, we start an SSH Tunnel and input the following command: *ssh -N -L 8888:localhost:8888 suting@remote*, where *-L* binds the local address: port 1 to a remote address: port 2, and *-N* specifies not to execute a remote command. On the local computer, we navigate to *localhost:8888* and get access to the remote port.

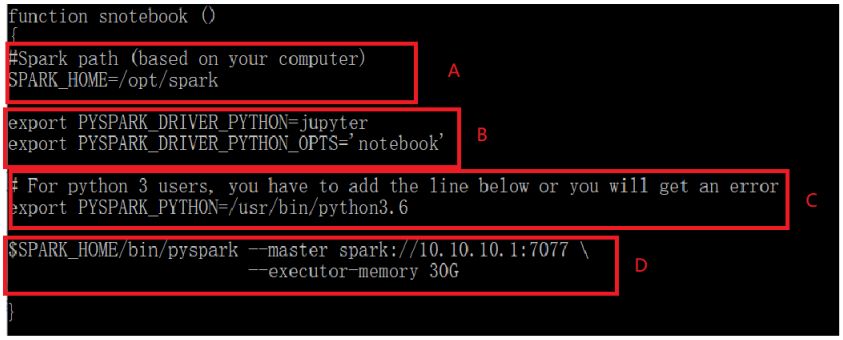


Figure A.3: "snotebook" function in bash configuration file

A.2 Spark data objects and libraries

There are three types of data objects in Spark: Resilient Distributed Datasets (RDDs), dataframes and SQL tables. RDD is a resilient and distributed collection of records that spread over one or many partitions. After reading the data file on disk, Spark driver creates the RDD, divides the RDD into partitions, and distributes the partitions across nodes. Dataframes in Apache Spark prevails over RDD but contains the features of RDD as well. Unlike an RDD, the dataframe is organized into named columns with a clear schema, like a table in relational database. SQL tables are SQL views similar to dataframes, but these tables enable us to execute SQL queries. We use dataframes and SQL tables in our study. Specifically, we load the data files as Spark dataframes, and perform basic Dataframe operations (e.g., join, count, groupby, print schema). We then convert the dataframes into SQL tables and perform more complicated queries (e.g., Online Analytical Processing (OLAP) analysis) to obtain the desired analysis results. In Spark, there are four specialized processing libraries: Streaming, SQL, Machine Learning and Graph. We use two of the four libraries: SQL and Machine Learning libraries in our study. We use SQL library to perform complicated queries over SQL tables, and use Machine Learning library to perform feature engineering and feature normalization.

A.3 Spark operations

There are two operations performed on Spark RDDs: transformation and action. Transformation is a function that produces the new RDD from the existing RDDs. *Filter(), sort(), groupby()* operations are examples of transformation operations. After the transformation when new RDDs are created, we use the action function to work with the actual dataset. *Count(), show(), write()* operations are examples of action operations. In Spark, the operations are performed using a "lazy evaluation" strategy, meaning that Spark will not execute the tasks until we trigger any action (Figure A.4). Using "lazy evaluation", it combines all transformations into a single transformation and executes them together. Spark's "lazy evaluation" strategy has several advantages. First, it saves the computation and increase the processing speed, since only necessary values get compute once the action is triggers. Second, it reduces both time and space complexities, since we do not execute every operation, and the action is triggered only when the data is required. Third, it provides optimization by optimizing the sequence of execution. For example, if a set of transformations include *filter()* and when an action is triggered, Spark performs *filter()* first, and then executes other queries on a much smaller data set.

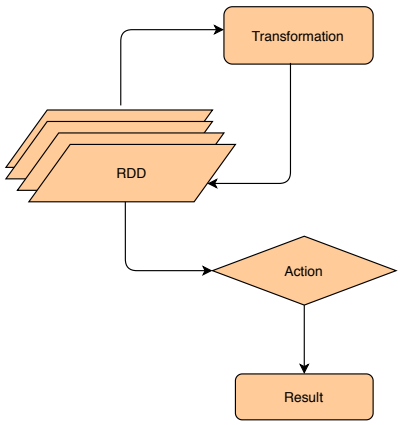


Figure A.4: "Lazy execution" in Spark

Table A.1: Data file sizes before and after parquet conversion

Data holding	csv file size (GB)	parquet file size (GB)
RPDB	2.22	1.3
NACRS	17.7	4
WTIS	19	0.6
OHIP 2008	54	12
OHIP 2009	59	12.8
OHIP 2010	61	13.1
OHIP 2011	60	12.8
OHIP 2012	67	15
OHIP 2013	60	12.8
OHIP 2014	57.9	14
OHIP 2015	60	13.5
OHIP 2016	62	13.6
OHIP 2017	62	13.3

A.4 Parquet format conversion

Apache Parquet is a special storage format supported by Spark. Parquet stores the data in a column-oriented way, where the values of each column are organized so that they are all adjacent, enabling better compression. This format is especially good for queries which only read particular columns from a "wide" table since only needed columns will be read. Compared to a traditional approach (e.g., csv files) where the data is stored in row-oriented way, parquet format is more efficient in terms of both storage and performance. In our study, all the data holdings obtained are originally in csv format. We convert all the csv files into parquet files, which largely reduces the file size and improves the processing speed. Table A.1 shows the the size of each data file (in GB) before and after the parquet format conversion. In total, parquet files reduce the file size by almost 5 times.

Appendix B

Charlson Comorbidity Index (CCI)

The Charlson Comorbidity Index (CCI) is a method of indicating the comorbidities of patients based on the International Classification of Disease (ICD) diagnosis codes which can be found in the administrative data of hospitalization, including Discharge Abstract Database (DAD) and National Ambulatory Care Reporting System (NACRS) [79]. Each comorbidity category is associated with a weight (from 1 to 6) based on the relative risk of the comorbidity, and use in healthcare resources [79]. The original CCI was developed with 19 categories [16], but has been modified to 17 categories [28]. Table B.1 shows each category of the CCI, its corresponding ICD-10 codes in DAD and NACRS diagnostic codes, and the weight associated. The algorithm of calculating the CCI for a patient in a given time period (e.g., 2 years) has two stages. At the first stage, for each hospitalization claim of a patient, a binary indicator (0 or 1) for each CCI category is determined based on all diagnostic codes associated with the claim (Algirthm 1). At the second stage, comorbidity indicators and weighted scores are summarized per patient to get patient-level CCI for the given time period (Algorithm 2).

Table B.1: Categories of Charlson Comorbidity Index

Group	Comorbidity Conditions	ICD-10 Codes [87]	Weight
1	Myocardial infarction	I21.x, I22.x, I25.2	1
2	Congestive heart failure	I09.9,I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0	1
3	Peripheral vascular disease	I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9	1
4	Cerebrovascular disease	G45.x, G46.x, H34.0, I60.x-I69.x	1
5	Dementia	F00.x-F03.x, F05.1, G30.x, G31.1	1
6	Chronic pulmonary disease	I27.8, I27.9, J40.x-J47.x, J60.x-J67.x, J68.4, J70.1, J70.3	1
7	Rheumatic disease	M05.x, M06.x, M31.5, M32.x-M34.x, M35.1, M35.3, M36.0	1
8	Peptic ulcer disease	K25.x-K28.x	1
9	Mild liver disease	B18.x, K70.0-K70.3, K70.9, K71.3-K71.5, K71.7, K73.x, K74.x, K76.0, K76.2-K76.4, K76.8, K76.9, Z94.4	1
10	Diabetes without chronic complication	E10.0, E10.l, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9	1
11	Diabetes with chronic complication	E10.2-E10.5, E10.7, E11.2-E11.5, E11.7, E12.2-E12.5, E12.7, E13.2-E13.5, E13.7, E14.2-E14.5, E14.7	2
12	Hemiplegia or paraplegia	G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0-G83.4, G83.9	2
13	Renal disease	I12.0, I13.1, N03.2-N03.7, N05.2-N05.7, N18.x, N19.x, N25.0, Z49.0-Z49.2, Z94.0, Z99.2	2
14	Cancer	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-C58.x, C60.x-C76.x, C81.x-C85.x, C88.x, C90.x-C97.x	2
15	Moderate or severe liver disease	I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7	3
16	Metastatic solid tumor	C77.x-C80.x	6
17	AIDS/HIV	B20.x-B22.x, B24.x	6

Algorithm 1: Assigning indicators of CCI groups per claim over all his/her claims

input : Each claim in DAD/NACRS in a given time period
output: Indicators of CCI groups for the claim
`/* array for individual CCI group counters */`
set Array CC_GRP(1) - CC_GRP(17) ;
`/* Initialize all CCI group counters to zero */`
for *i = 1 to 17* **do**
  CC_GRP(i) = 0 ;
end
`/* Check each patient record for the diagnosis codes in each CCI group */`
for *i = 1 to 17* **do**
  if *any diagnostic code in ICD-10 codes for CCI group i* **then**
    CC_GRP(i) = 1 ;
  else
    CC_GRP(i) = 0;
  end
end

Algorithm 2: Calculation of CCI per patient in a given time period

input : $n \times 17$ matrix of CCI group indicators over all n claims for a patient in a given time period, with row vectors indicating the CCI group indicators per claim
output: weighted CCI score for the patient
`/* array for summarized CCI group counters */`
set Array total_GRP(1) - total_GRP(17) ;
`/* Initialize total CCI group counters to zero */`
for *i = 1 to 17* **do**
  total_GRP(i) = 0 ;
end
`/* Check indicators over all claims in each CCI group */`
for *i = 1 to 17* **do**
  if *any indicator has 1 for CCI group i* **then**
    total_GRP(i) = 1 ;
  else
    total_GRP(i) = 0;
  end
end
`/* If both diabetes flags are 1, keep the highest severity */`
if *sum(total_GRP(11),total_GRP(10))=2* **then**
  total_GRP(10) = 0 ;
end
`/* If both liver disease flags are 1, keep the highest severity */`
if *sum(total_GRP(9),total_GRP(15))=2* **then**
  total_GRP(9) = 0 ;
end
`/* If both cancer flags are 1, keep the highest severity */`
if *sum(total_GRP(14),total_GRP(16))=2* **then**
  total_GRP(14) = 0 ;
end
`/* use Charlson weights to calculate a weighted score */`
wgtcc = sum(of total_GRP(1)-total_GRP(10)) + total_GRP(11)*2 + total_GRP(12)*2 + total_GRP(13)*2 + total_GRP(14)*2 + total_GRP(15)*3 + total_GRP(16)*6 + total_GRP(17)*6;

Appendix C

Small Area Variance Analysis (SAVA)

The small area variance analysis (SAVA) method is used when we compare the utilization rates across regions in Ontario. This Appendix provides an in-depth elaboration on how we calculate the age-standardized utilization rates across regions, and the formulas we use to derive different SAVA statistics.

First, to calculate the age-standardized utilization rates in each census division, we aggregate age into 10-year age groups (0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70+), and calculate the utilization rate in each age group stratum, using the population in each census division and each age group from RPDB as the rate denominator. Next, we use the entire population in RPDB as the standard population, and calculate the proportion (weight) of population in each age group using the entire population. Finally, for each census division, we derive the direct age-standardized rate by applying the age-group specific rates to the standard population. The age-standardized rate is a weighted average with weights taken from the standard population, and reflects the rate if the population had the same age distribution as the standard.

Next, we derive the extreme quotient (EQ) and coefficient of variation (CV) using the direct age-standardized rates for each i-th Healthcare Area, denoted by DSR_i, for $i = 1, ..., 49$. The EQ is expressed as [44]

$$EQ = \frac{\max(DSR_i)}{\min(DSR_i)} \tag{C.0.1}$$

The CV is expressed as [44]

$$CV = \frac{\sqrt{\sum(DSR_i - \overline{DSR})^2/(49-1)}}{\sum DSR_i/49} \quad \text{with} \quad \overline{DSR} = \frac{\sum DSR_i}{49} \tag{C.0.2}$$

The remaining two statistics, the Chi-squared statistic and systematic component of variance (SCV), use the observed and expected cases per census division. For the Chi-squared test, the null hypothesis is that there was no significant variation across regions, and is tested using a χ^2 test with k-1 degrees of freedom (k = number of areas analysed, that is 49) [112]. The χ^2 value is expressed as [112]

$$\chi^2 = \sum \frac{(y_i - e_i)^2}{e_i} \approx \chi^2_{k-1} \tag{C.0.3}$$

where y_i is the observed of MRI scans and e_i is the expected number of MRI sans in each area, based on age-specific MRI utilization rates for the whole population. The age-specific utilization rates for the whole population are calculated using the population in RPDB as the standard population.

Lastly, the SCV is another measurement of variation across regions [112], and is expressed as [44]

$$SCV = \frac{1}{k}(\sum \frac{(y_i - e_i)^2}{e_i} - \sum \frac{1}{e_i}) \tag{C.0.4}$$

SCV > 3 is considered significant variation across regions, SCV up to 10 is considered high variation and SCV > 10 is considered very high variation [4].

Appendix D

Descriptive Statistics of Wait Times, 2008-2017

D.1 Priority 1 (target of 1 day)

Year	Mean (stdev)	25%	50%	75%	90%	% within target
2008	2.45 (19.56)	0	0	1	3	84.24
2009	1.42 (10.88)	0	0	1	2	85.40
2010	1.73 (14.26)	0	0	1	2	87.24
2011	1.51 (12.52)	0	0	1	2	88.59
2012	1.01 (8.09)	0	0	0	1	90.90
2013	1.42 (15.28)	0	0	0	1	92.78
2014	1.32 (12.89)	0	0	0	1	92.71
2015	1.25 (8.49)	0	0	0	1	92.11
2016	0.86 (5.50)	0	0	1	1	91.30
2017	0.60 (4.36)	0	0	0	1	93.90

D.2 Priority 2 (target of 2 days)

Year	Mean (stdev)	25%	50%	75%	90%	% within target
2008	10.04 (26.55)	0	2	8	28	55.91
2009	5.81 (21.11)	0	1	4	12	66.74
2010	4.50 (16.50)	0	1	3	9	69.91
2011	4.01 (16.05)	0	1	3	7	72.39
2012	2.93 (14.09)	0	1	2	5	77.20
2013	2.32 (10.09)	0	1	2	4	78.68
2014	2.28 (9.56)	0	1	2	5	78.68
2015	2.46 (9.81)	0	1	2	5	77.59
2016	2.37 (10.47)	0	1	2	4	80.69
2017	2.07 (10.28)	0	1	2	4	81.20

D.3 Priority 3 (target of 10 days)

Year	Mean (stdev)	25%	50%	75%	90%	% within target
2008	32.68 (50.32)	6	16	36	79	36.83
2009	32.38 (52.63)	6	15	36	81	39.93
2010	24.39 (41.45)	5	12	27	57	47.46
2011	26.63 (42.49)	6	13	32	65	45.29
2012	24.53 (46.91)	5	11	24	51	48.98
2013	18.55 (33.85)	5	10	20	38	52.27
2014	21.38 (33.78)	5	12	26	48	46.35
2015	23.68 (37.38)	5	11	27	62	47.69
2016	22.18 (39.31)	4	9	23	56	53.01
2017	22.32 (39.70)	5	10	23	56	53.60

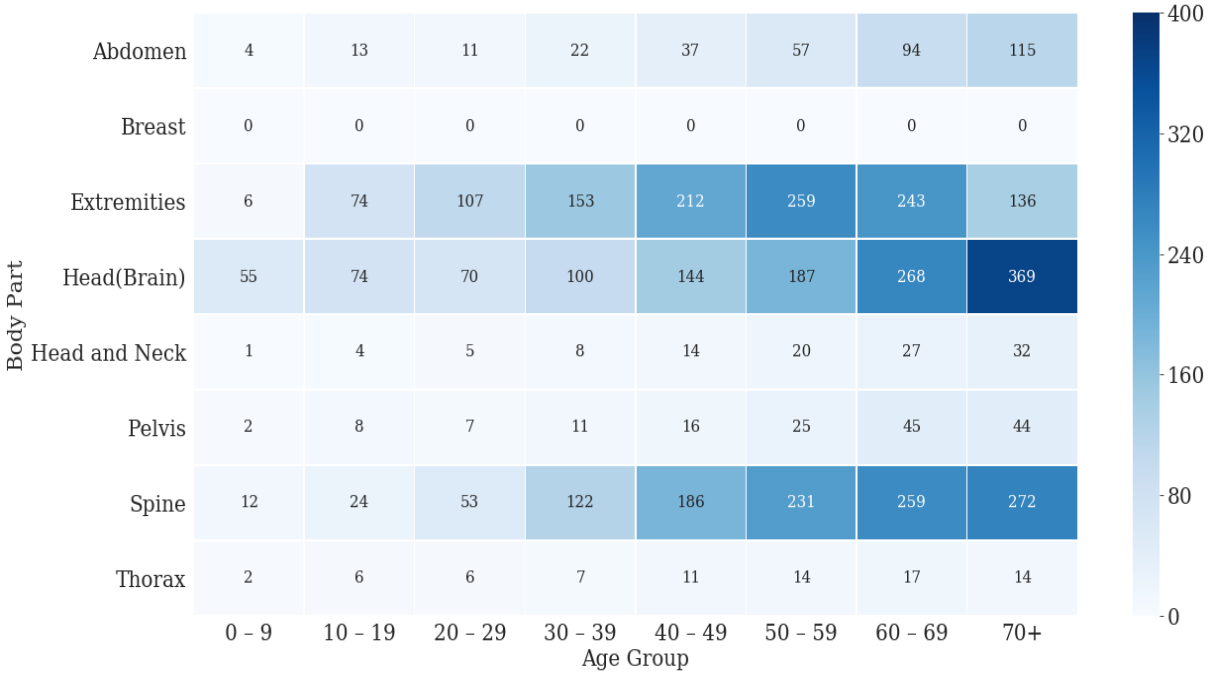
D.4 Priority 4 (target of 28 days)

Year	Mean (stdev)	25%	50%	75%	90%	% within target
2008	56.99 (64.48)	20	39	70	116	36.99
2009	61.86 (63.12)	22	49	81	120	31.58
2010	64.17 (62.40)	23	49	87	133	31.46
2011	58.46 (62.56)	21	44	74	116	33.48
2012	53.11 (63.93)	17	36	66	106	40.38
2013	44.32 (58.46)	16	30	50	84	47.87
2014	51.66 (59.90)	22	37	64	92	36.32
2015	59.73 (62.14)	24	46	77	111	30.48
2016	63.38 (66.60)	24	46	81	132	30.93
2017	63.91 (66.82)	25	48	77	130	29.58

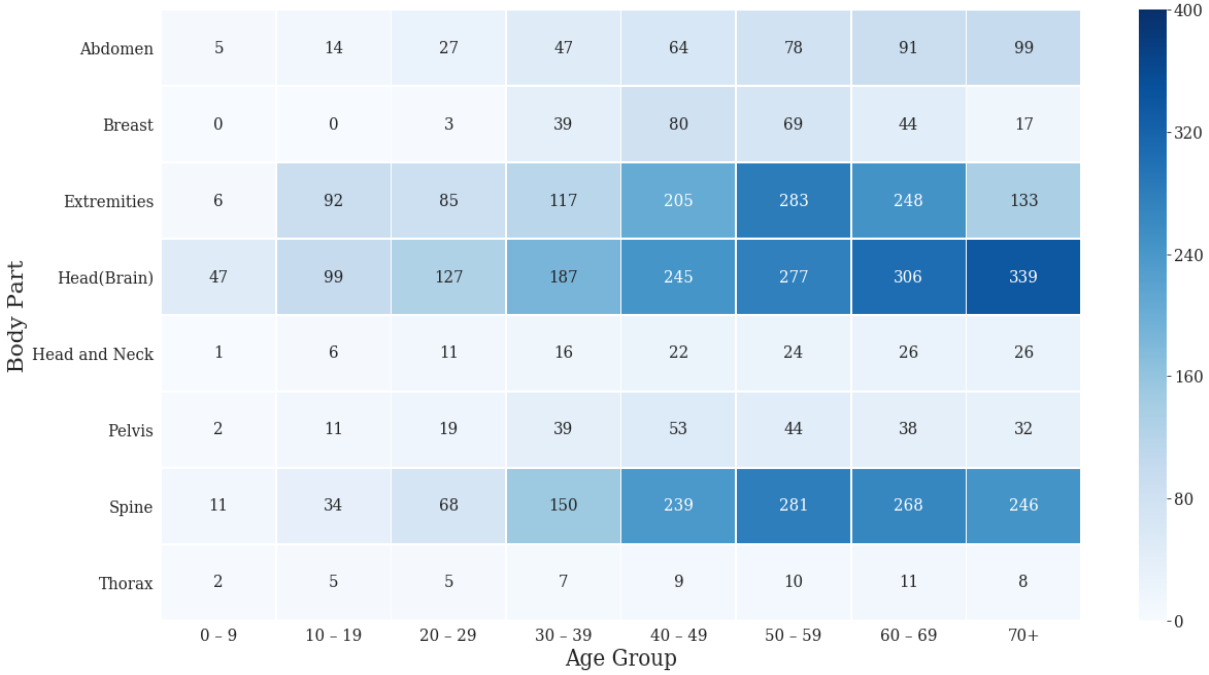
Appendix E

Age- and Sex-specific MRI Utilization Rate by body parts, 2017

E.1 Male

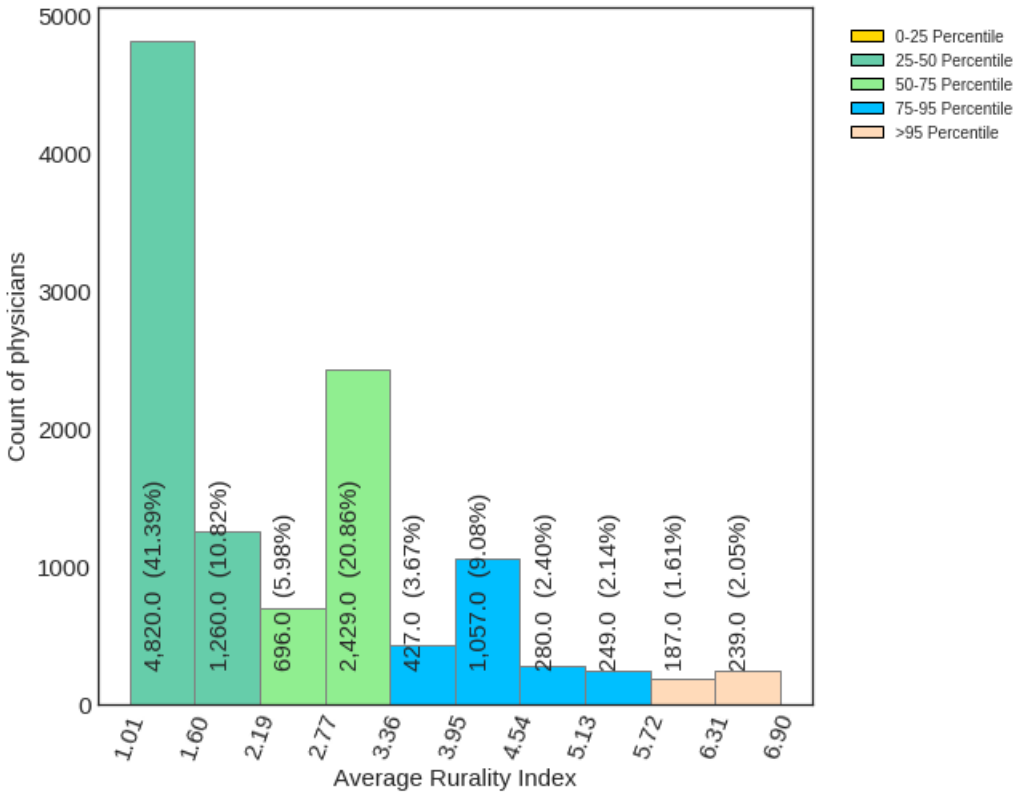


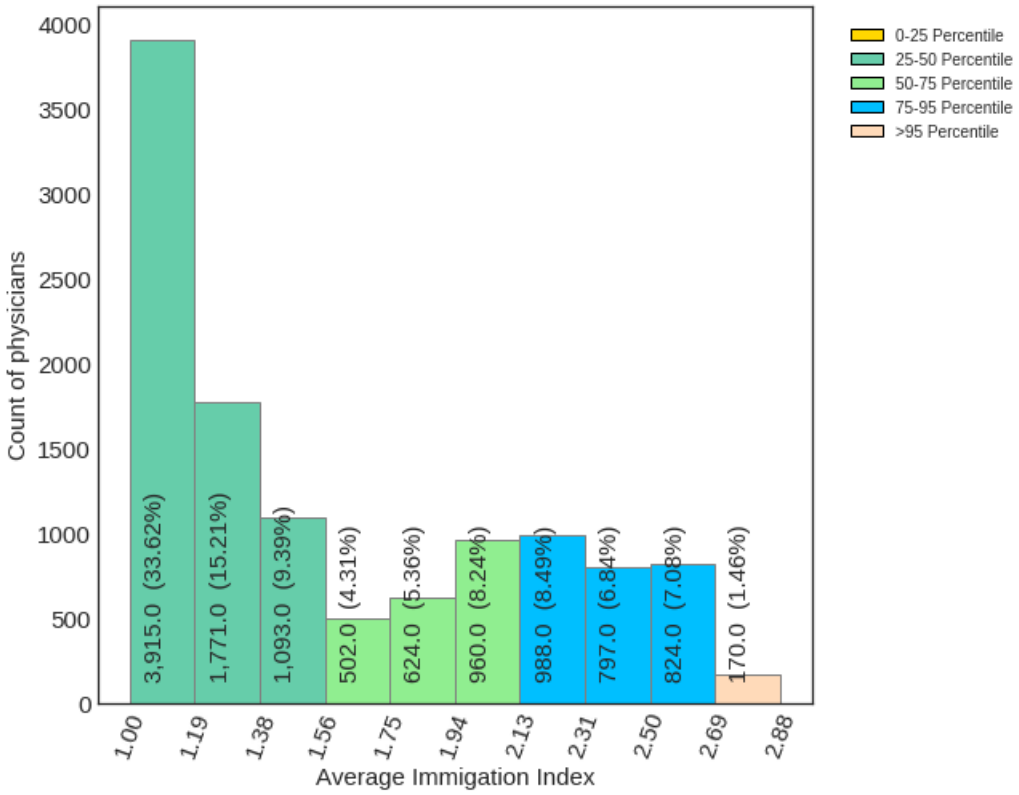
E.2 Female

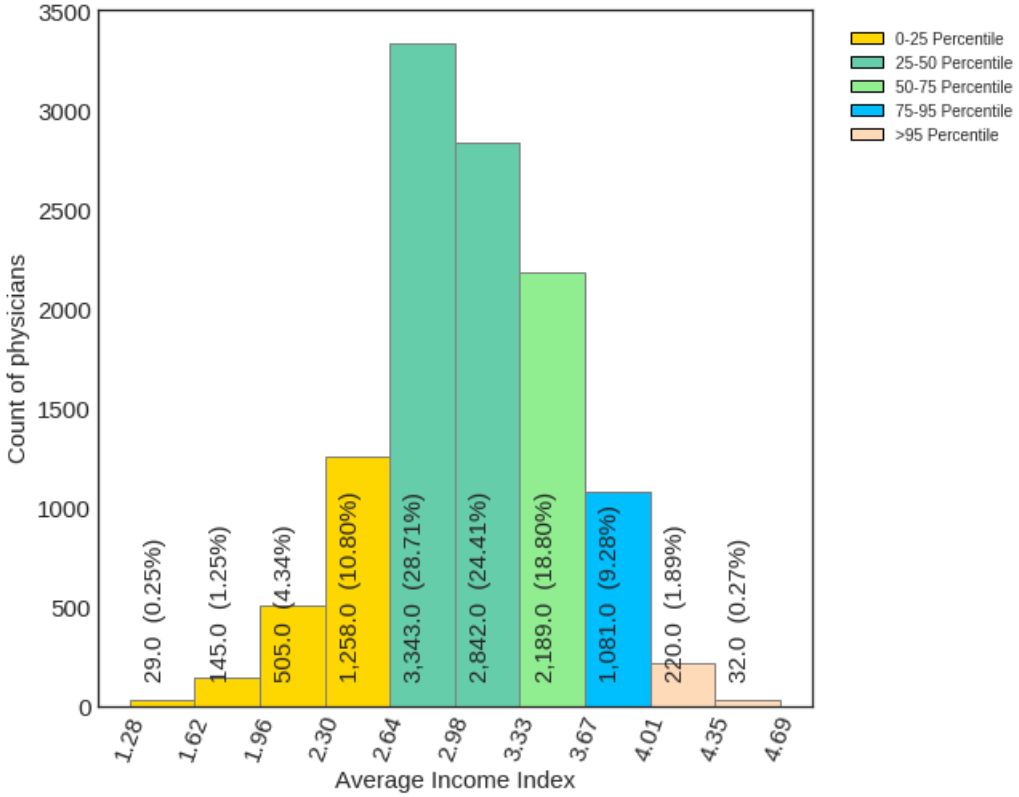

Body Part
Abdomen 5 14 27 47 64 78 91 99
Breast 0 0 3 39 80 69 44 17
Extremities 6 92 85 117 205 283 248 133
Head(Brain) 47 99 127 187 245 277 306 339
Head and Neck 1 6 11 16 22 24 26 26
Pelvis 2 11 19 39 53 44 38 32
Spine 11 34 68 150 239 281 268 246
Thorax 2 5 5 7 9 10 11 8
0 – 9
10 – 19
20 – 29
30 – 39
40 – 49
50 – 59
60 – 69
70+
Age Group
400
320
240
160
80
0

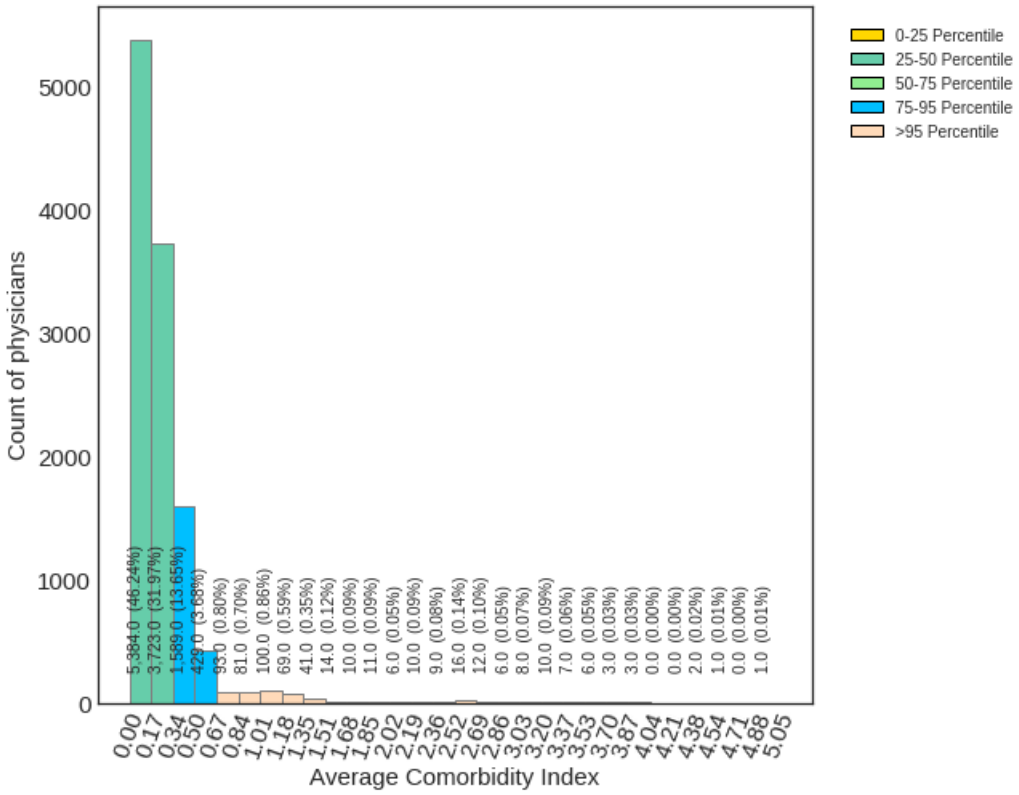
Appendix F

Histograms of Average Index for Rurality, Immigration, Income, and Comorbidity




0-25 Percentile
25-50 Percentile
50-75 Percentile
75-95 Percentile
>95 Percentile
Count of physicians
Average Immigation Index
3,915.0 (33.62%)
1,771.0 (15.21%)
1,093.0 (9.39%)
502.0 (4.31%)
624.0 (5.36%)
960.0 (8.24%)
988.0 (8.49%)
797.0 (6.84%)
824.0 (7.08%)
170.0 (1.46%)


0-25 Percentile
25-50 Percentile
50-75 Percentile
75-95 Percentile
>95 Percentile
Count of physicians
Average Income Index
29.0 (0.25%)
145.0 (1.25%)
505.0 (4.34%)
1,258.0 (10.80%)
3,343.0 (28.71%)
2,842.0 (24.41%)
2,189.0 (18.80%)
1,081.0 (9.28%)
220.0 (1.89%)
32.0 (0.27%)


0-25 Percentile
25-50 Percentile
50-75 Percentile
75-95 Percentile
>95 Percentile
Count of physicians
0
1000
2000
3000
4000
5000
5,384.0 (46.24%)
3,723.0 (31.97%)
1,589.0 (13.65%)
429.0 (3.68%)
93.0 (0.80%)
81.0 (0.70%)
100.0 (0.86%)
69.0 (0.59%)
41.0 (0.35%)
14.0 (0.12%)
10.0 (0.09%)
11.0 (0.09%)
6.0 (0.05%)
10.0 (0.09%)
9.0 (0.08%)
16.0 (0.14%)
12.0 (0.10%)
6.0 (0.05%)
8.0 (0.07%)
10.0 (0.09%)
7.0 (0.06%)
6.0 (0.05%)
3.0 (0.03%)
3.0 (0.03%)
0.0 (0.00%)
0.0 (0.00%)
2.0 (0.02%)
1.0 (0.01%)
0.0 (0.00%)
1.0 (0.01%)
0.00
0.17
0.34
0.50
0.67
0.84
1.01
1.18
1.35
1.51
1.68
1.85
2.02
2.19
2.36
2.52
2.69
2.86
3.03
3.20
3.37
3.53
3.70
3.87
4.04
4.21
4.38
4.54
4.71
4.88
5.05
Average Comorbidity Index

Appendix G

Physician Practice Patterns Analysis

At the first stage of our physician profiling analysis, for all three selected tests, we perform risk adjustment only on age and sex distributions of the physician's case mix, and investigate the differences in practice patterns across tester groups. For each test, we retrieve the top 50 billing codes in OHIP claimed by higher testers, and compare them with those claimed by typical and lower testers. Our finding shows similar practice patterns for brain MRIs across tester levels. On the other hand, for extremities and spine MRIs, high-ordering physicians were associated with more sports medicine Focused Practice Assessment (FPA) practices. Table G presents for each test, the difference of sports medicine FPA practices across tester groups, in terms of claims per physician (total billing claims divided by total number of physicians within the group).

Table G.1: Difference in sports medicine FPA practices across tester groups

Billing Code	Explanation	Claims per physician		
		Lower	Typical	Higher
Extremities				
A917	Sports Medicine Focused Practice Assessment (FPA)	0.21	0.54	52.29
Spine				
A917	Sports Medicine Focused Practice Assessment (FPA)	3.11	8.78	27.28

Appendix H

Outputs for Risk Adjustment Models

H.1 Extremities

Table H.1: Risk adjustment model output, extremities

	Rate ratio (95% CI) [1]	
	Non sports medicine practitioners	Sports medicine practitioners
Intercept	5.8 (5.68-5.92)	24.79 (19.75-31.11)
Patient sex		
Male (%)	0.91 (0.89-0.93)	1.22 (1.02-1.47)
Patient age (%)		
0-19	1.45 (1.4-1.5)	4.83 (3.07-7.59)
20-39	1.28 (1.22-1.34)	8.55 (4.73-15.45)
40-59	1.52 (1.48-1.56)	3.89 (2.8-5.41)
60-79	2.08 (1.96-2.2)	10.15 (5.21-19.74)
80+ (reference)	1	1
Sports medicine engagement (%)	-	1.51 (1.33,1.72)

[1] All risk ratio estimates are for 1 stdev increase in the parameter of interest

H.2 Spine

Table H.2: Risk adjustment model output, spine

	Rate ratio (95% CI)	
	Non sports medicine practitioners	Sports medicine practitioners
Intercept	6.1 (5.98-6.21)	11.61 (8.79-15.34)
Patient sex		
Male (%)	0.9 (0.89-0.92)	0.98 (0.79-1.22)
Patient age (%)		
0-19	1.14 (1.1-1.17)	1.19 (0.69-2.02)
20-39	0.96 (0.92-1)	1.22 (0.6-2.49)
40-59	1.34 (1.31-1.37)	1.29 (0.87-1.92)
60-79	1.55 (1.48-1.63)	1.67 (0.76-3.66)
80+ (reference)	1	1
Sports medicine engagement (%)	-	1.53 (1.3,1.81)

H.3 Brain

Table H.3: Risk adjustment model output, brain

	Rate ratio (95% CI)
Intercept	4.07 (3.97,4.18)
Patient sex	
Male (%)	0.83 (0.81-0.85)
Patient age (%)	
0-19	1.16 (1.1-1.22)
20-39	1.05 (0.98-1.12)
40-59	1.23 (1.19-1.28)
60-79	1.64 (1.52-1.76)
80+ (reference)	1

Bibliography

[1] Rajan Agarwal, Meredith Bergey, Seema Sonnad, Howard Butowsky, Mythreyi Bhargavan, and Michael H Bleshman. Inpatient CT and MRI utilization: trends in the academic hospital setting. *Journal of the American College of Radiology*, 7(12):949–955, 2010.

[2] Alan Agresti. *Categorical data analysis*, volume 482. John Wiley & Sons, 2003.

[3] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

[4] John Appleby, Veena Raleigh, Francesca Frosini, Gwyn Bevan, Haiyan Gao, and Tom Lyscom. Variations in health care. *The good, the bad and the inexplicable. London: The King's Fund*, 2011.

[5] Michael Armbrust, Reynold S Xin, Cheng Lian, Yin Huai, Davies Liu, Joseph K Bradley, Xiangrui Meng, Tomer Kaftan, Michael J Franklin, Ali Ghodsi, et al. Spark sql: Relational data processing in spark. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pages 1383–1394, 2015.

[6] Gordon Arnett and David C Hadorm. Steering committee of the western canada waiting list. developing priority criteria for hip and knee replacement: results from the western canada waiting list project. *Can J Surg*, 46:290–6, 2003.

[7] Elizabeth J Atkinson, Cynthia S Crowson, Rachel A Pedersen, and Terry M Therneau. Poisson models for person-years and expected rates. *Mayo Foundation Tech Report# 81*, 2008.

[8] Alex Bekker. Spark vs. hadoop mapreduce: Which big data framework to choose. URL https://www.scnsoft.com/blog/spark-vs-hadoop-mapreduce.

[9] Douglas G Bonett. Confidence interval for a coefficient of quartile variation. *Computational Statistics & Data Analysis*, 50(11):2953–2957, 2006.

[10] BrainFacts.org. Technologies that peer inside your head. URL https://www.brainfacts.org/in-the-lab/tools-and-techniques/2014/brain-scans-technologies-that-peer-inside-your-headl.

[11] Emmalin Buajitti, Tristan Watson, Todd Norwood, Kathy Kornas, Catherine Bornbaum, David Henry, and Laura C Rosella. Regional variation of premature mortality in ontario, canada: a spatial analysis. *Population health metrics*, 17(1):9, 2019.

[12] Suzanne C Byrne, Brendan Barrett, and Rick Bhatia. The impact of diagnostic imaging wait times on the prognosis of lung cancer. *Canadian Association of Radiologists' Journal*, 66(1):53–57, 2015.

[13] Statistics Canada. Postal code conversion file plus (pccf+) reference guide.

[14] Right Care. Right care, 2020. URL https://rightcarealliance.org/.

[15] Benjamin TB Chan and Peter C Austin. Patient, physician, and community factors affecting referrals to specialists in ontario, canada: a population-based, multi-level modelling approach. *Medical Care*, pages 500–511, 2003.

[16] Mary E Charlson, Peter Pompei, Kathy L Ales, and C Ronald MacKenzie. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*, 40(5):373–383, 1987.

[17] Ran-Chou Chen, Herng-Ching Lin, Dachen Chu, Tom Chen, Sheng-Tzu Hung, and Nai-Wen Kuo. Physicians' characteristics associated with repeat use of computed tomography and magnetic resonance imaging. *Journal of the Formosan Medical Association*, 110(9):587–592, 2011.

[18] Ran-Chou Chen, Dachen Chu, Herng-Ching Lin, Tom Chen, Sheng-Tzu Hung, and Nai-Wen Kuo. Association of hospital characteristics and diagnosis with the repeat use of CT and MRI: a nationwide population-based study in an Asian country. *American Journal of Roentgenology*, 198 (4):858–865, 2012.

[19] Deena J Chisolm, Ann Scheck McAlearney, Sofia Veneris, David Fisher, Melissa Holtzlander, and Karen S McCoy. The role of computerized order sets in pediatric inpatient asthma treatment. *Pediatric Allergy and Immunology*, 17(3):199–206, 2006.

[20] Camilo Cid, Randall P Ellis, Verónica Vargas, Juergen Wasem, and Lorena Prieto. Global risk-adjusted payment models. In *World Scientific Handbook of Global Health Economics and Public Policy: Volume 1: The Economics of Health and Health Systems*, pages 311–362. World Scientific, 2016.

[21] Ottawa: CIHI. Medical imaging exams in canada double in six years, 2010. URL http://www.cihi.ca/CIHI-extportal/internet/en/Document/types+of+care/specialized+servi.

[22] Janine Clarke. Difficulty accessing health care services in canada, 2016.

[23] Yves Croissant. *mlogit: Multinomial Logit Models*, 2020. URL https://CRAN.R-project.org/package=mlogit. R package version 1.1-0.

[24] Salvador de Mateo and Enrique Regidor. Standardisation or modelling of mortality rates. *Journal of Epidemiology & Community Health*, 50(6):681–682, 1996.

[25] Michelle L Degelman and Katya M Herman. Immigrant status and having a regular medical doctor among canadian adults. *Canadian Journal of Public Health*, 107(1):e75–e80, 2016.

[26] Elizabeth R DeLong, Eric D Peterson, David M DeLong, Lawrence H Muhlbaier, Suzanne Hackett, and Daniel B Mark. Comparing risk-adjustment methods for provider profiling. *Statistics in Medicine*, 16(23):2645–2664, 1997.

[27] Sandor Demeter, Martin Reed, Lisa Lix, Leonard MacWilliam, and William D Leslie. Socioeconomic status and the utilization of diagnostic imaging in an urban setting. *CMAJ*, 173(10): 1173–1177, 2005.

[28] Richard A Deyo, Daniel C Cherkin, and Marcia A Ciol. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology*, 45(6):613–619, 1992.

[29] Paula Diehr. Encyclopedia of biostatistics. 2005.

[30] Paula Diehr, Kevin Cain, Frederick Connell, and Ernest Volinn. What is too much variation? the null hypothesis in small-area analysis. *Health services research*, 24(6):741, 1990.

[31] Jens Dittrich and Jorge-Arnulfo Quiané-Ruiz. Efficient big data processing in hadoop mapreduce. *Proceedings of the VLDB Endowment*, 5(12):2014–2015, 2012.

[32] Spark Documentation. Cluster mode overview. URL https://spark.apache.org/docs/1.1.0/cluster-overview.html.

[33] Isabel dos Santos Silva. *Cancer epidemiology: principles and methods - Chapter 14: Dealing with confounding in the analysis*. IARC, 1999.

[34] John M Eisenberg. Measuring quality: are we ready to compare the quality of care among physician groups? *Annals of Internal Medicine*, 136(2):153–154, 2002.

[35] Nada Elgendy and Ahmed Elragal. Big data analytics: a literature review paper. In *Industrial Conference on Data Mining*, pages 214–227. Springer, 2014.

[36] Derek J Emery, Alan J Forster, Kaveh G Shojania, Stephanie Magnan, Michelle Tubman, and Thomas E Feasby. Management of MRI wait lists in Canada. *Healthcare Policy*, 4(3):76, 2009.

[37] Derek J Emery, Kaveh G Shojania, Alan J Forster, Naghmeh Mojaverian, and Thomas E Feasby. Overuse of magnetic resonance imaging. *JAMA internal medicine*, 173(9):823–825, 2013.

[38] Benjamin Fine, Susan E Schultz, Lawrence White, and David Henry. Impact of restricting diagnostic imaging reimbursement for uncomplicated low back pain in ontario: a population-based interrupted time series analysis. *CMAJ open*, 5(4):E760, 2017.

[39] A Stewart Fotheringham and David WS Wong. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A: Economy and Space*, 23(7):1025–1044, 1991.

[40] William Gardner, Edward P Mulvey, and Esther C Shaw. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, 118(3):392, 1995.

[41] Joseph E Gonzalez, Reynold S Xin, Ankur Dave, Daniel Crankshaw, Michael J Franklin, and Ion Stoica. Graphx: Graph processing in a distributed dataflow framework. In *11th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 14)*, pages 599–613, 2014.

[42] Stephen F Hall, Colleen Webber, Patti A Groome, Christopher M Booth, Paul Nguyen, and Yvonne DeWit. Do doctors who order more routine medical tests diagnose more cancers? a population-based study from Ontario Canada. *Cancer Medicine*, 8(2):850–859, 2019.

[43] William R Hendee, Gary J Becker, James P Borgstede, Jennifer Bosma, William J Casarella, Beth A Erickson, C Douglas Maynard, James H Thrall, and Paul E Wallner. Addressing overutilization in medical imaging. *Radiology*, 257(1):240–245, 2010.

[44] Berta Ibáñez, Julián Librero, Enrique Bernal-Delgado, Salvador Peiró, Beatriz González López-Valcarcel, Natalia Martínez, and Felipe Aizpuru. Is there much variation in variation? revisiting statistics of small area variation in health services research. *BMC health services research*, 9(1): 60, 2009.

[45] KJ Jager, C Zoccali, A Macleod, and FW Dekker. Confounding: what it is and how to deal with it. *Kidney international*, 73(3):256–260, 2008.

[46] D Jason Busse, Paul E Alexander, Amane Abdul-Razzak, John J Riva, D John Dufton, Madison Zhang, Markus Faulhaber, CFPC Rachel Couban, MISt Gordon H Guyatt, Y Raja Rampersaud, et al. Appropriateness of spinal imaging use in canada. 2013.

[47] Mahtab Karami. Clinical decision support systems and medical imaging. *Radiology Management*, 37(2):25, 2015.

[48] Vaquar Khan. Difference between scaling horizontally and vertically. URL https://github.com/vaquarkhan/vaquarkhan/wiki/Difference-between-scaling-horizontally-and-vertically.

[49] Ramin Khorasani. Computerized physician order entry and decision support: improving the quality of care. *Radiographics*, 21(4):1015–1018, 2001.

[50] Michael Kirby. Review of ontario's wait time information system, 2007. URL http://www.health.gov.on.ca/en/common/ministry/publications/reports/kirby_rep/kirby_rep.pdf.

[51] Wayne R Kubick. Big data, information and meaning. *Applied Clinical Trials*, 21(2):26, 2012.

[52] Roz Diane Lasker, David W Shapiro, and Anthony M Tucker. Realizing the potential of practice pattern profiling. *Inquiry*, pages 287–297, 1992.

[53] Andreas Laupacis, Institute for Clinical Evaluative Sciences in Ontario, and Raymond Przybysz. Access to MRI in Ontario: Addressing the information gap. july 2003. *Institute for Clinical Evaluative Sciences*, 2003.

[54] David C Levin, Vijay M Rao, and Laurence Parker. Physician orders contribute to high-tech imaging slowdown. *Health Affairs*, 29(1):189–195, 2010.

[55] Wendy Levinson and Tai Huynh. Engaging physicians and patients in conversations about unnecessary tests and procedures: Choosing Wisely Canada. *CMAJ*, 186(5):325–326, 2014.

[56] Wendy Levinson, Marjon Kallewaard, R Sacha Bhatia, Daniel Wolfson, Sam Shortt, and Eve A Kerr. 'Choosing Wisely': a growing international campaign. *BMJ Qual Saf*, 24(2):167–174, 2015.

[57] Steven Lewis, Morris L Barer, Claudia Sanmartin, Sam Sheps, Samuel ED Shortt, and Paul W McDonald. Ending waiting-list mismanagement: principles and practice. *CMAJ*, 162(9):1297–1300, 2000.

[58] Chao-Jui Li, Yuan-Jhen Syue, Tsung-Cheng Tsai, Kuan-Han Wu, Chien-Hung Lee, and Yan-Ren Lin. The impact of emergency physician seniority on clinical efficiency, emergency department resource use, patient outcomes, and disposition accuracy. *Medicine*, 95(6), 2016.

[59] Clare Liddy, Jatinderpreet Singh, Ryan Kelly, Simone Dahrouge, Monica Taljaard, and Jamie Younger. What is the impact of primary care model type on specialist referral rates? a cross-sectional study. *BMC family practice*, 15(1):22, 2014.

[60] Nicholas T Longford. A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74(4):817–827, 1987.

[61] ICE Magazine. MRI Market Continues Upward Trend, 2018. URL https://theicecommunity.com/mri-market-continues-upward-trend.

[62] Rupa Mahanti. Scalability rules: 50 principles for scaling web site. *Software Quality Professional*, 14(1):35–48, 2011.

[63] Ian C Marschner. glm2: Fitting generalized linear models with convergence problems. *The R Journal*, 3:12–15, 2011.

[64] Peter McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127, 1980.

[65] Peter McCullagh. *Generalized linear models*. Routledge, 2018.

[66] I McCurdie. Imaging in sport and exercise medicine:"a sports physician's outlook and needs". *The British journal of radiology*, 85(1016):1198–1200, 2012.

[67] Paul McDonald. *Waiting Lists and Waiting Times for Health Care in Canada: More Management, More Money?: Summary Report*. Health Canada, 1998.

[68] Thomas G McGuire and Richard C Van Kleef. Risk sharing. In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets*, pages 105–131. Elsevier, 2018.

[69] Mechanical and University of Toronto Industrial Engineering. How to use Netu: The morLAB cluster. 2012.

[70] Ateev Mehrotra, Rachel O Reid, John L Adams, Mark W Friedberg, Elizabeth A McGlynn, and Peter S Hussey. Physicians with the least experience have higher cost profiles than do physicians with the most experience. *Health Affairs*, 31(11):2453–2463, 2012.

[71] Marty Putyra Michael Watts, Susan Newell. Ontario taking next steps to integrate health care system. URL https://www.osler.com/en/resources/regulations/2019/ontario-taking-next-steps-to-integrate-health-care-system.

[72] Saam Morshed, Paul Tornetta III, and Mohit Bhandari. Analysis of observational studies: a guide to understanding statistical methods. *JBJS*, 91(Supplement_3):50–60, 2009.

[73] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384, 1972.

[74] Sabina Nuti and Milena Vainieri. Managing waiting times in diagnostic medical imaging. *BMJ open*, 2(6):e001255, 2012.

[75] Ministry of Health and Long-Term Care. MRI and CT Scanning Services, 2018. URL http://www.auditor.on.ca/en/content/annualreports/arreports/en18/v1_308en18.pdf.

[76] Ontario Ministry of Health and Long-Term Care. Amendments to the schedule of benefits to reflect evidence-informed care. toronto, 2012. URL www.health.gov.on.ca/en/pro/programs/ecfa/action/primary/lb_sob.aspx.

[77] Ontario Ministry of Health and Long-Term Care. Schedule of benifits. physician services under the health insurance act, 2016.

[78] Ontario Ministry of Health and Long-Term Care. GP focused practice designation policy and program overview, 2016. URL https://cs.oma.org/Member/omamail/MailItems/PCFA/GP%20Focus%20Application%20June%202016.pdf.

[79] University of Manitoba. Charlson Comorbidity Index. URL http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?printer=Y&conceptID=1098.

[80] Health Quality Ontario. Wait times for diagnostic imaging, 2020. URL https://www.hqontario.ca/System-Performance/Wait-Times-for-Diagnostic-Imaging.

[81] D Wayne Osgood. Poisson-based regression analysis of aggregate crime rates. *Journal of quantitative criminology*, 16(1):21–43, 2000.

[82] Vinay Parmar, L Thompson, and Hifz Aniq. Comparison of referrals for lumbar spine magnetic resonance imaging from physiotherapists, primary care and secondary care: how should referral pathways be optimised? *Physiotherapy*, 101(1):82–87, 2015.

[83] B Paxton, M Lungren, S Jung, P Kranz, and R Kilani. A case study in lumbar spine MRI and physician self-referral of imaging. In *Radiological Society of North America 2011 Scientific Assembly and Annual Meeting*, 2011.

[84] Aviva Petrie and Caroline Sabin. *Medical statistics at a glance*. John Wiley & Sons, 2019.

[85] John T Pohlman and Dennis W Leitner. A comparison of ordinary least squares and logistic regression. *Ohio Journal of Science*, 2003.

[86] Luciano M Prevedello, Ali S Raja, Richard D Zane, Aaron Sodickson, Stuart Lipsitz, Louise Schneider, Richard Hanson, Srinivasan Mukundan, and Ramin Khorasani. Variation in use of head computed tomography by emergency physicians. *The American Journal of Medicine*, 125(4):356–364, 2012.

[87] Hude Quan, Vijaya Sundararajan, Patricia Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, L Duncan Saunders, Cynthia A Beck, Thomas E Feasby, and William A Ghali. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care*, pages 1130–1139, 2005.

[88] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL https://www.R-project.org/.

[89] Envision Radiology. X-Ray vs. CT vs. MRI. URL https://www.envrad.com/difference-between-x-ray-ct-scan-and-mri.

[90] RadiologyInfo.org. Magnetic Resonance Imaging (MRI) - Body. URL https://www.radiologyinfo.org/en/info.cfm.

[91] Richard J Rossi. *Mathematical Statistics : An Introduction to Likelihood Based Inference*. John Wiley & Sons, 2018.

[92] Jeffrey M Rothschild, R Khorasani, and DW Bates. Guidelines and decision support help improve image utilization. *Diagnostic Imaging*, 22(11):95, 2000.

[93] Vikas Saini, Shannon Brownlee, Adam G Elshaug, Paul Glasziou, and Iona Heath. Addressing overuse and underuse around the world. *The Lancet*, 390(10090):105–107, 2017.

[94] Claudia Sanmartin, Samuel ED Shortt, Morris L Barer, Sam Sheps, Steven Lewis, and Paul W McDonald. Waiting for medical services in Canada: lots of heat, but little light. *CMAJ*, 162(9): 1305–1310, 2000.

[95] Meir H Scheinfeld, Jee-Young Moon, Michele J Fagan, Reubin Davoudzadeh, Dan Wang, and Benjamin H Taragin. MRI usage in a pediatric emergency department: an analysis of usage and usage trends over 5 years. *Pediatric Radiology*, 47(3):327–332, 2017.

[96] Gideon Schwarz et al. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[97] Maninder Singh Setia, Amelie Quesnel-Vallee, Michal Abrahamowicz, Pierre Tousignant, and John Lynch. Access to health-care in canadian immigrants: a longitudinal study of the national population health survey. *Health & Social Care in the Community*, 19(1):70–79, 2011.

[98] Dilpreet Singh and Chandan K Reddy. A survey on platforms for big data analytics. *Journal of Big Data*, 2(1):8, 2015.

[99] Jon Starkweather and Amanda Kay Moske. Multinomial logistic regression. *Consulted page at September 10th: http://www. unt. edu/rss/class/Jon/Benchmarks/MLR_JDS_Aug2011. pdf*, 29: 2825–2830, 2011.

[100] The Statesman. Delhi government to provide free MRI scan. URL https://www.thestatesman.com/cities/delhi-government-to-provide-free-mri-scan-1481031172.html.

[101] Jane Stoever and Leslie Champlin. New policy, dual residency programs support fps who provide emergency care. *The Annals of Family Medicine*, 4(4):375–376, 2006.

[102] Muhammad Syafrudin, Norma Latif Fitriyani, Donglai Li, Ganjar Alfian, Jongtae Rhee, and Yong-Shin Kang. An open source-based real-time data processing architecture framework for manufacturing sustainability. *Sustainability*, 9(11):2139, 2017.

[103] Talented. What is mapreduce? URL https://www.talend.com/resources/what-is-mapreduce/.

[104] TechTarget. big data analytics, 2019. URL https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics.

[105] Medical News Today. What to know about MRI scans. URL https://www.medicalnewstoday.com/articles/146309#what-is-an-mri-scan.

[106] Juliana Tolles and William J Meurer. Logistic regression: relating patient characteristics to outcomes. *Jama*, 316(5):533–534, 2016.

[107] ENC Tong, ACA Clements, MA Haynes, MA Jones, AP Morton, and M Whitby. Improved hospital-level risk adjustment for surveillance of healthcare-associated bloodstream infections: a retrospective cohort study. *BMC Infectious Diseases*, 9(1):145, 2009.

[108] Giovanni Tripepi, Kitty J Jager, Friedo W Dekker, and Carmine Zoccali. Stratification for confounding–part 2: direct and indirect standardization. *Nephron Clinical Practice*, 116(4):c322–c325, 2010.

[109] Edwin JR van Beek and Eric A Hoffman. Functional imaging: CT and MRI. *Clinics in Chest Medicine*, 29(1):195–216, 2008.

[110] William N Venables and Brian D Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. URL http://www.stats.ox.ac.uk/pub/MASS4. ISBN 0-387-95457-0.

[111] H Gilbert Welch, Mark E Miller, and W Pete Welch. Physician profiling–an analysis of inpatient practice patterns in Florida and Oregon. *New England Journal of Medicine*, 330(9):607–612, 1994.

[112] Ragnar Westerling. Components of small area variation in death rates: a method applied to data from Sweden. *Journal of Epidemiology & Community Health*, 49(2):214–221, 1995.

[113] Richard Williams. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*, 6(1):58–82, 2006.

[114] JJ You, DA Alter, K Iron, PM Slaughter, A Kopp, R Przybysz, D Thiruchelvam, L Devore, and A Laupacis. Diagnostic services in Ontario: descriptive analysis and jurisdictional review. *ICES Investigative Report. Toronto, Canada: Institute for Clinical Evaluative Sciences*, 2007.

[115] John J You, Wendy Levinson, and Andreas Laupacis. Attitudes of family physicians, specialists and radiologists about the use of computed tomography and magnetic resonance imaging in ontario. *Healthcare Policy*, 5(1):54, 2009.

[116] Gary J Young, Stephen Flaherty, E David Zepeda, Koenraad J Mortele, and John L Griffith. Effects of physician experience, specialty training, and self-referral on inappropriate diagnostic imaging. *Journal of general internal medicine*, pages 1–7, 2020.

[117] Tjalling J Ypma. Historical development of the Newton–Raphson method. *SIAM review*, 37(4): 531–551, 1995.

[118] Matei Zaharia, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauly, Michael J Franklin, Scott Shenker, and Ion Stoica. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In *Presented as part of the 9th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 12)*, pages 15–28, 2012.

[119] Matei Zaharia, Tathagata Das, Haoyuan Li, Timothy Hunter, Scott Shenker, and Ion Stoica. Discretized streams: Fault-tolerant streaming computation at scale. In *Proceedings of the twenty-fourth ACM symposium on operating systems principles*, pages 423–438, 2013.

[120] Matei Zaharia, Reynold S Xin, Patrick Wendell, Tathagata Das, Michael Armbrust, Ankur Dave, Xiangrui Meng, Josh Rosen, Shivaram Venkataraman, Michael J Franklin, et al. Apache spark: a unified engine for big data processing. *Communications of the ACM*, 59(11):56–65, 2016.