THREE ESSAYS IN HEALTH ECONOMICS

BY

PRANAY RAJ PANTA

A thesis

submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Doctor of Philosophy

in Economics

Victoria University of Wellington November 2021

Acknowledgements

I wish to thank many people without whom this dissertation would not have been possible. First and foremost, I extend my sincere gratitude to Prof. Ilan Noy for supervising my research. He has allowed me to be independent while supporting me wholeheartedly in my academic and career pursuits. He has been a great teacher, mentor, and a role model.

I would also like to thank Prof. Don Shin for his extraordinary patience. I am greatly appreciative of him for sharing with me his expertise in economic thinking and econometrics. I have made great advances in my understanding of empirical economic research since my association with him, and for that I am extremely thankful to him.

I thank GS1 NZ for funding my scholarship that allowed me the financial independence to focus on my research. I would also like to thank my dissertation committee for valuable feedback.

I would like to especially thank my parents and my siblings for their unconditional love. They have provided me with immense comfort during the most difficult times in my life. I would also like to thank my past teachers and professors at Caldwell University (USA) and Jacobs University (Germany). They have been a great source of knowledge and wisdom.

I cherish the moments with many people who have been my friends throughout my life. They have all been great people and for that I consider myself extremely lucky. I also thank my PhD colleagues, teachers, and staff at Victoria University of Wellington for the many interesting conversations and academic support.

"If I have seen further, it is by standing on the shoulders of the giants" - Issac Newton

Dedicated to my parents

THREE ESSAYS IN HEALTH ECONOMICS

This dissertation contains three essays on economic analysis of the New Zealand healthcare market particularly relating to publicly funded elective surgeries and the pharmaceutical industry.

In Chapter 1, using administrative data on all patients booked for publicly funded elective surgery in 2014, we investigate whether ethnic disparities exist in waiting time for elective surgeries in New Zealand. Unlike existing studies on the topic, we extensively control for sample selection bias, clinical severity, and regional factors affecting resource supply and healthcare utilization. We find evidence of Maori, Pacific, and Asian patients waiting longer for elective surgery when compared to New Zealand Europeans. Our results indicate that Maori wait on average 3 percent longer for elective surgery, however, they are not consistent across all model specifications. We find that Pacific and Asian patients wait 8 - 9 percent longer for elective surgeries compared to NZE. Our results are considerably stronger for Pacific and Asian compared to Maori across all model specifications. Furthermore, these differences are pervasive across several surgical categories and are as high as 22 percent across angiography procedures. We also find evidence that waiting time in New Zealand is geographically inequitable. Although an increasing number of empirical studies find evidence of inequities in waiting time socioeconomic status, our analysis reveals no such effects once regional differences affecting waiting time are controlled for.

In Chapter 2, using administrative data on patients who have been booked and treated for publicly funded elective surgeries over 2011-2015, we investigate how supply of and demand for elective surgeries respond to waiting times in New Zealand. Unlike existing studies, we endogenize waiting times in a system of three equations, which nests the conventional demand-and-supply model. Each structural equation is over-identified by excluded instruments, and the exclusion restriction is

justified on a priori grounds while over-identifying restrictions are tested. The analysis finds that, in the case of New Zealand, the demand for elective surgery is inelastic, whereas supply is elastic. From a policy perspective, the results suggest that a long-term increase in supply will lead to a permanent reduction in waiting times and thus improving timely access to care.

In Chapter 3, I develop an intuitive model based on the traditional newsvendor framework that enables pharmaceutical wholesalers to efficiently forecast demand and reduce inventory while adjusting for required customer service level (CSL) targets. First, using historical demand data for a major pharmaceutical wholesaler in New Zealand, I present several demand forecasting models and assess their performance in predicting demand, based on several forecast error metrics. I find that artificial neural network (ANN) models are generally stable for demand forecasting and sometimes outperform traditional demand forecasting methods. Second, using the demand forecast as an input, I derive a periodic review inventory model based on the newsvendor inventory framework. Numerical analysis of the proposed model shows that, compared to current practice, the model enables wholesalers to significantly reduce both inventory levels and ordering, while maintaining a very high non-stock-out probability constraint and CSL target close to 100 percent. I conduct separate analyses for both government subsidized low value-high demand drug and nonsubsidized low demand-high value drug and find the results to be consistent across both types of drugs.

Contents

1	Eth	Ethnic disparities in waiting time for publicly funded elective surgery: Evidence from								
	New	Zealand	10							
	1.1	Introduction	10							
	1.2	Institutional Background	12							
	1.3	Literature Review	16							
	1.4	Data	21							
	1.5	Empirical Specification	26							
	1.6	Results	28							
	1.7	Robustness tests	31							
	1.8	Discussion	35							
	1.9	Conclusion	40							
2	Esti	mating the impact of waiting times on supply and demand for elective surgery in								
	New	Zealand	43							
	2.1	Introduction	43							
	2.2	The New Zealand healthcare market for elective surgery	46							
	2.3	Relevant Literature	48							
	2.4	Data and Variables	50							
		2.4.1 Key Variables	52							
		2.4.2 Instrumental Variables	55							

	2.5	Empirical Strategy							
		2.5.1 Waiting time equation	60						
		2.5.2 Supply equation	61						
		2.5.3 Demand equation	62						
	2.6	Empirical Results	63						
	2.7	Robustness tests	66						
	2.8	Conclusion	77						
3	Dem	and forecasting and inventory control for pharmaceutical wholesalers - A numer-							
	ical s	study	80						
	3.1	Introduction	80						
	3.2	Institutional Background	83						
	3.3	Related Literature	85						
	3.4	Mathematical Modelling	88						
		3.4.1 Demand forecasting	88						
		3.4.2 Inventory optimization	91						
	3.5	Experimental Results	96						
		3.5.1 Data	96						
		3.5.2 Demand Pattern Comparison	97						
		3.5.3 Demand Forecasting Results	98						
		3.5.4 Inventory Problem	02						
	3.6	Conclusion	05						
A	Add	itional results	17						
	A.1	Coefficient plots for ethnic gaps by each procedure	17						
	A.2	Coefficient plots for ethnic gaps by each DHB	19						
	A.3	Additional OLS results	20						
	A.4	Econometric diagnostic tests results	22						

List of Tables

1.1	Descriptive Statistics	24
1.2	OLS Results: All categories (pooled) and individual health speciality specific cat-	
	egories	30
1.3	OLS with alternate measures of waiting time and Quantile Regression Results	32
1.4	Ethnic group proportions by DHB (Source: NZ Census data, 2013)	34
2.1	Descriptive Statistics, Unit of Observation - 20 District Health Boards (DHBs)	54
2.2	Estimation of 'Triangular Model' with mean waiting times	64
2.3	Estimation of conventional demand and supply model	67
2.4	Estimation of 'Triangular Model' with alternate Instrumental Variables	69
2.5	Estimation of 'Triangular Model' with median waiting times	70
2.6	Percent pacific as an instrument of waiting time	71
2.7	Percent Asian as an instrument of waiting time	72
2.8	Estimation of 'Triangular Model' for a specific surgical speciality: Orthopaedic	
	surgery	74
2.9	Estimation of 'Triangular model' with DHB fixed effects (20 DHBs)	76
3.1	Total monthly demand forecasts for 2018 (Drug A)	99
3.2	Total monthly demand forecasts for 2018 (Drug B)	99
3.3	Comparison of forecast errors for demand forecasting models	100
3.4	Total sales, orders, and inventory for 2018 (Drug A)	103

3.5	Total sales, orders, and inventory for 2018 (Drug B)
A.1	OLS Results: All ethnicities and Procedure-by-DHB fixed effects
A.2	Econometric diagnostic tests results: Waiting time equations
A.3	Econometric diagnostic tests results: Demand equations
A.4	Econometric diagnostic tests results: Supply equations

List of Figures

1.1	Elective services patient flow	13
1.2	Waiting time differences across procedures and DHBs (Source: NBRS dataset, 2014)	26
2.1	Waiting time vs. difference (quantity demanded - quantity supplied) across DHBs	
	for hip replacement procedure in 2011 and 2015.	55
3.1	Pharmaceutical Supply Chain: An overview	81
3.2	Network of wholesaler warehouses in New Zealand (Source: Grant Thornton NZ,	
	2016)	84
3.3	Feed-forward neural network (Adapted from Hyndman and Athanasopoulous, 2018)	90
3.4	Trends and Seasonality Analysis for monthly drug sales	97
3.5	Monthly Demand forecast comparisons for Drug A and Drug B (Year:2018)	100
3.6	Comparisons of total orders and inventory for 2018 (Actual vs. Forecast) for Drug	
	A and Drug B	104
A.1	Coefficient plots by each procedure: Maori	117
A.2	Coefficient plots by each procedure: Pacific	118
A.3	Coefficient plots by each procedure: Asian	118
A.4	Coefficient plots by each DHB: Maori	119
A.5	Coefficient plots by each DHB: Pacific	119
A.6	Coefficient plots by each DHB: Asian	120

Chapter 1

Ethnic disparities in waiting time for publicly funded elective surgery: Evidence from New Zealand

1.1 Introduction

Universal healthcare is an important feature of many healthcare systems. Such health systems are publicly funded through general taxation and patients do not make a direct financial contribution to healthcare providers at the point of service delivery. In the absence of price, rationing of healthcare demand usually takes the form of waiting time. While the average waiting time in a region is determined by the supply/demand condition of the local health markets, patients experience substantial variation in waiting time depending on their multiple clinical factors. As such, with the severity of the patients' health condition fully controlled for, there should be no direct link between patients' other observable group characteristics (e.g. ethnicity and gender) and their ability to access healthcare. Yet, both empirical and anecdotal evidence suggest that these objectives are seldom achieved.

Although ethnic inequities in healthcare access and outcomes for emergency and life-threatening

clinical conditions has been extensively studied internationally, ethnic disparities in waiting time across publicly funded elective surgeries are less-studied in the context of OECD countries. To do so, we use patient records from 2014 and ask whether waiting times are different across ethnic groups for publicly funded elective surgeries in New Zealand (NZ hereafter). The contributions of our analysis are fourfold. First, we use administrative data on all patients booked for all publicly funded elective surgeries in NZ, while existing studies focus on survey data or a small sample of patients from one surgical procedure and limited geographic area. Second, we adequately control for patient severity by using a unique Clinical Priority Assignment Criteria score (CPAC hereafter) which is used to determine the urgency of the surgery once a patient is accepted into the waiting list. Existing studies either use no controls for patient severity or use inadequate measures such as Diagnosis Related Groups (DRG hereafter) and other broad urgency categories. Controlling for patient severity is important as many healthcare systems require patients to be prioritized based on the severity of their condition (Siciliani & Hurst, 2005a). A failure to control for severity will lead to biased results when there exist differences across ethnic groups in the severity level. Third, due to the unique nature of the NZ healthcare market, any sample selection bias due to hospital choice is likely to be very minimal in our analysis. Patients in NZ do not have a choice of public hospital and are treated in hospitals within their area of residence. These regional hospitals are funded and managed by their respective District Health Boards (DHBs hereafter) and most DHBs operate a single facility for elective surgical care. Therefore, unlike existing studies, the relevant catchment area is not mis-specified in our analysis. Fourth, we adequately control for regional heterogeneity across DHBs, without which estimated effects may be overstated. Geographic regions are likely to differ significantly in terms of healthcare resources supply (number of public hospitals, beds and doctors) and need based healthcare utilisation (regional co-morbidity and health levels). Across health systems in many OECD countries, patients can 'shop around for hospitals' with multiple hospitals as substitutes which makes the catchment populations across regions endogenous and produces biased estimates of other variables affecting waiting time (Gravelle, Smith, & Xavier, 2003).

We use a multivariate regression to estimate the effect of ethnicity on waiting time. Our analysis reveals that waiting time in New Zealand is both geographically and ethnically inequitable. We find evidence of Maori, Pacific, and Asian patients waiting longer for elective surgery when compared to New Zealand Europeans (NZE). These ethnic gaps are stronger for Asian and Pacific patients, who on average wait 8-9 percent longer for surgery. The estimated ethnic gaps for Maori are less strong and indicate that Maori wait on average 3 percent longer for elective surgeries. The waiting time differences are highest across angiography procedures with Pacific patients waiting 22 percent longer for elective surgery. Unlike numerous existing studies, we find no evidence of discrimination by socioeconomic status (SES hereafter). Our findings suggest that despite a national policy framework concerning waiting time for elective surgeries in NZ, access to surgery remains inequitable. Therefore, reducing existing inequities in waiting time must be an important aspect of the NZ government's health policy and agenda.

Our paper is organized as follows. First, in Section 2, we provide some background on the NZ healthcare market relating to publicly funded elective surgery. We then briefly review the relevant existing literature in Section 3 and describe the dataset in Section 4. Section 5 discusses our empirical specification and the results are summarized in Section 6. In Section 7, we perform additional robustness tests to confirm our main results and provide further discussions on the results in Section 8. Finally, we summarize the main findings of our study and conclude the paper in Section 9.

1.2 Institutional Background

New Zealand has a two-tiered health system with provisions for both private and public hospitals performing elective surgery. Unique to NZ is a booking system for surgery which uses the CPAC score¹. Unlike other OECD countries, patients in NZ do not have a choice between specialists or between public hospitals. A possibility of an alternate provider is only available if such provisions

¹A detailed review of the New Zealand priority assessment and the booking system is provided in Dew et al. (2005a).

are deemed necessary by the patients' DHB, or if the patients seek private treatment by paying out of pocket. In NZ, the DHBs are responsible for providing or funding provisions for health services for patients registered in their district. This system prioritises patients by 'clinical need' and 'the patient's ability to benefit from surgery'. This rationing by prioritisation is aimed at achieving equal access to surgery for all citizens and residents. A typical elective services patient flow within the public system, not accounting for any emergency inflows or exits to private systems, is summarised in Figure 1.1.



Figure 1.1: Elective services patient flow

In most cases, once a patient has been referred to a specialist by a GP, he or she must receive a

First Specialist Assessment (FSA) within four months if the hospital believe treatment is necessary and the patient meets the criteria for referral. The hospital service informs the patient with a clinical appointment with a specialist or suggests other course of actions if the patient does not meet the criteria for clinical assessment. During the assessment, the specialist will assess the patient's condition and assign a CPAC score based on clinical factors, patient-reported health levels, and the impact on patient's social needs such as work, income, and independence. Those patients who would benefit from surgery but have a CPAC score below the surgical threshold are placed in an 'active review' status where they must be reassessed within the next six months (earlier if their condition deteriorates). There are targets set by the Ministry of Health (MOH hereafter) for the surgical volume that each DHB is required to manage, and as such there are significant constraints imposed on the DHBs. Nonetheless, the thresholds required to qualify for surgery can be adjusted by the DHBs based on available resources, and consequently be altered to meet the targets set by the MOH. Moreover, surgeons have the capacity to override scores if they feel the score does not adequately reflect the patient's need for elective surgery (Dew et al., 2005a). Hence, multiple factors affect the interplay of supply and demand of healthcare, and consequently waiting time for elective surgeries across administrative regions.

If the patient's CPAC score is above a clinical threshold ², the patient will be determined as needing surgery and will be put into the waiting list. The hospital service will inform the patient that surgery will be provided within the next four months ³. The booking mechanisms may be different across hospitals. In most cases, a few weeks before the date of the scheduled surgery, the hospital clerk will inform the patient of the date of treatment. The patient can then interact with the hospital clerk on issues related to unavailability or transportation, and the hospital clerk can provide advice to reschedule the surgery. If the patient has multiple cancellations or missed appointments, they can be removed from the waiting list.

The patient is likely to have significant interactions with his or her GP before being referred

²The threshold varies across DHBs and across surgical procedures and is calculated by taking into consideration the demand and capacity constraints across each DHB.

³The waiting time limit was six months before January 2015.

to a specialist. There is some evidence that personal beliefs of healthcare professionals affect the outcome for patients. For example, in NZ, Maori are more likely to be discharged from the hospital earlier and are less likely to be referred to a specialist (Ellison-Loschmann & Pearce, 2006; Wilson & Barton, 2012). In a separate study, Derrett et al. (2013) find that the booking system and the CPAC scoring mechanism have been a 'messy reality' and there are problems with 'interrater' reliability of point scoring systems and that the CPAC scoring mechanism is a 'subjective process'. A recent Auditor General's report also has revealed significant issues with appropriate prioritisation, equitable accessibility, and national consistency during the booking process (Derrett et al., 2013).

On a more positive note, there have been significant attempts made by the MOH to address existing issues of equity and consistency within the booking system. There are several policy guidelines that focus on patient flow management and levels of service. The DHB's policies are also specifically focused on reducing waiting times, delivering scheduled treatment, and improving consistency. The government has recently announced that all DHBs will now be replaced with a national health agency, like in the UK, to adopt "a nationwide system approach to address the needs of all New Zealanders".

As there are multiple interactions between the patient and the healthcare provider during the referral and CPAC scoring process, the referral and scoring mechanism may be unfavourable to certain ethnic groups. However, once patients are booked into the waiting list, the patients are likely to only interact with their healthcare provider once the hospital informs them their surgery date. In our analysis, we specifically look at the disparities in waiting time once patients have been booked for surgery. Our assumption is that the CPAC scores are based on medical diagnosis and once the CPAC score has been assigned and patients are booked for surgery, there should be no measurable difference in waiting time for elective surgery across ethnic groups.

1.3 Literature Review

An extensive literature exists both internationally and in New Zealand concerning health inequities across ethnic groups. These studies are mainly related to surgeries on life threatening conditions such as cancer and organ transplantation. A general pattern observed internationally is that indigenous ethnic groups have substantially worse health outcomes and a higher prevalence of chronic diseases (Bramley et al., 2004). Yeates et al. (2009) find that compared to white people, indigenous people in Australia, Canada, New Zealand, and the United States have significantly less likelihood of receiving renal transplantation. In a systematic review of all existing papers concerning healthcare inequality in NZ, Rahiri et al. (2018) report that most studies reveal 'unequal treatment' exists based on ethnicity in NZ across multiple surgical specialities. This is even though Maori have higher rates of preventable diseases (Rahiri et al., 2018). In a separate NZ study, Derrett et al. (2009) find that access to elective surgery is geographically inequitable. Implementing the booking system could be a 'messy reality with unintended consequences' as the regional differences in healthcare service management can strongly impact access to elective surgery (Derrett et al., 2013).

Despite a high prevalence of studies analysing ethnic inequities in healthcare outcomes, surprisingly, ethnic differences in access to elective surgery is less studied. We were able to locate a study from one hospital in the North West of England, in which the authors find no significant impact of ethnicity on waiting time for ophthalmology and orthopaedic procedures (Hacker & Stanistreet, 2004). Despite a relative scarcity of studies related to ethnic differences in waiting time for elective surgery, there is a growing literature analysing the impact of other patient characteristics, mainly SES, on waiting time for elective surgery. Several studies find evidence of inequality in access to healthcare. Siciliani and Verzulli (2009) use survey data from nine European countries and find that higher education significantly correlated with lower waiting times. They also find income effects, although modest as compared to education, where an increase of income of 10,000 Euros reduces waiting times in all countries except Sweden (Siciliani & Verzulli, 2009). Another study by Tinghög et al. (2014) use administrative data from a regional county in Sweden and find that higher disposable income predicted lower waiting times. However, they find no significant effects of ethnicity or gender (Tinghög et al., 2014). A separate study by Sørensen et al. (2009) in Denmark find that people with a higher SES are referred more to a specialist. A study by Kaarboe and Carlsen (2014) in Norway finds that women with lower income and education status are assigned a lower priority for surgery and the direct relationship between education and priority is also true for men below the age of 50. In a study from Germany, Heinrich et al. (2018) find that patients with private health insurance wait less than patients with statutory health insurance for outpatient treatment within the public system. In a separate study in England, Laudicella et al. (2012) find that both education and income have distinct impact on waiting times and patients from least deprived areas wait less than patients from most deprived regions. In the case of Australia, Johar et al. (2013) find large and significant evidence of discrimination by SES and conclude that socioeconomically advantaged patients are highly prioritised. Nevertheless, in a separate study in Australia by Sharma et al. (2013), the authors find that after controlling for sample selection bias due to the presence of the private sector, the gradient between SES and waiting time reduce significantly in size but does not disappear completely. There are several other studies slightly varying in scope with differing conclusions. A comprehensive study by Van Doorslaer et al. (2000) in the US and several countries across Europe find little or no evidence of significant inequality in delivery of healthcare services, although in most of the countries there is a pattern of 'pro-rich inequity' for contacts with physicians. Likewise, Löfvendahl et al. (2005) find no evidence of inequality according to SES in Sweden for access into elective surgery and the differences in access are better explained by hospital conditions. Arnesen et al. (2002) also conclude SES and gender do not explain variations in waiting times in Norway.

In our paper, we are focused mainly on the analysis of ethnic differences in waiting time for elective surgery. SES is largely unobserved when patients are seeking access to elective surgery, and hence, the case for explicit inequity by SES is unclear. Furthermore, existing studies in NZ are more focused in analysis concerning ethnic differences across Maori and NZE patient groups and as such group differences in healthcare access and outcomes are less studied within the context

of Pacific and Asian patient groups. According to census data published by Statistics NZ⁴, Asian ethnic groups are the fastest growing NZ resident population with 15.1 percent of the population identifying as Asian in 2018, up from 11.8 percent in 2013. In that respect, there is a strong likelihood that the Asian population will be the second largest ethnic group in NZ in the upcoming years, as NZE, Maori, and Pacific ethnic groups currently make up 70.2 percent, 16.5 percent, and 8.1 percent of the population respectively. Hence, it is worthwhile to extend discussions on ethnic differences in healthcare access and outcomes in New Zealand to Pacific and Asian ethnic minorities as well. Finally, existing studies in NZ and elsewhere are limited to a very small sample often based on survey data for single procedures (usually cancer treatments) and limited geographic area. We further the analysis in following ways:

(i) Endogeneity of supply factors

Healthcare systems in the UK, Australia and Nordic countries allow patients to choose public hospitals for elective surgery and patients can shop around for hospitals with lower waiting times. This makes multiple hospitals to be substitutes, and as patients care about distance as well waiting times, the catchment populations for hospitals are endogenous (Gravelle, Smith, & Xavier, 2003). Although, controls for supply factors such as area level hospital admissions and number of public hospital beds have been employed in previous studies, failure to allow for correlations between supply factors and unobserved variables affecting hospital utilisation will lead to inconsistent estimates of the effect of the supply variables (Gravelle, Sutton, et al., 2003). This is particularly true when area deprivation profiles are used as a measure of SES as is often the case in existing studies. This is valid only under the assumption that patients rarely cross their area of residence to receive healthcare. When an area level deprivation index is used as a proxy for SES without controlling for supply factors in the catchment area, the area level deprivation index also picks up the unobserved need variables that affect waiting time and overestimate the impact of SES and other variables on waiting time (Gravelle, Sutton, et al., 2003).

⁴https://www.stats.govt.nz/news/new-zealands-population-reflects-growing-diversity.

Hospital choice is not an option in New Zealand and most patients are treated in their residence DHB area. These DHBs vary significantly in terms of the size and funding provisions. Furthermore, most DHBs usually operate a single facility for elective services, and as such DHB dummies adequately pick up unobserved factors affecting area level hospital utilisation (regional co-morbidity and health levels) and other supply factors (such as hospital beds, doctors and efficiency) and produce coefficients on other variables not contaminated by endogeneity (Gravelle, Smith, & Xavier, 2003).

(ii) Control for clinical condition severity

Many health systems prioritize patients according to the severity of their condition as such patients requiring urgent care are treated earlier (Siciliani & Hurst, 2005a). Previous studies use patient characteristics such as age, surgical procedure type, self-reported health levels and patient DRGs as controls for patient severity. These are likely to be imperfect measures. In our study, along with existing controls such as DRGs, we have data on patients CPAC scores which is an explicit measure of the severity of their condition. As such, patient related clinical factors which are major determinants of waiting time for surgery are adequately controlled in our analysis. Furthermore, failure to control for differences across ethnic groups in urgency of health care leads to biased results regarding discrimination against certain ethnic groups in the waiting time. It can be argued that perceived differences across groups can be explained by the fact that certain patient groups are more capable of communicating their healthcare need to their healthcare service providers, and as such have lower waiting times. The use of CPAC scores as an additional control variable also reduces any bias caused by the ability to communicate healthcare need as that would already be reflected in the CPAC score.

As discussed in the previous section, there are existing issues with reliability and consistency of the scoring mechanism. However, in theory, CPAC score is an explicit measure of patient severity. Our assumption is that the CPAC scores are purely based on medical diagnoses. In addition to CPAC scores, we have also included the DRG codes as additional controls for patient severity. DRG codes have been widely used in medical practices around the world as DRG codes group patients according to their diagnoses and presence of other co-morbidity and complications. Many existing studies do not use any controls for patient severity, and we would argue that failure to control for clinical severity will lead to inconsistent estimates of other variables affecting waiting times.

Furthermore, as an additional exercise we also analysed ethnic differences in CPAC scores across patients controlling for the procedure and patient DHB (by including procedure by DHB fixed effects) and found that Maori and Pacific patients in fact have higher CPAC scores on average than NZE patients. The results are reported in Appendix A, Table A.1. We found no statistically significant differences in CPAC scores for Asian patients or by SES. This is consistent with the observation that Maori and Pacific patients have higher co-morbidity than NZE patients and are more likely to be assigned a higher CPAC score. It is possible that the CPAC score maybe biased by the same forces that drive disparities in waiting times, however, our analysis reveals that CPAC scores are likely based on medical diagnosis. Moreover, as patients do not have an interpersonal relationship with their specialist as they do with their GP, we would argue that their likelihood of heavily influencing CPAC scores are minimal. Therefore, when CPAC scores are controlled for along with DRG codes, our hypothesis is that there should be no observable difference between ethnic groups in waiting time for elective surgery.

(iii) Sample selection bias

One feature of healthcare systems is the presence of the private sector. Ideally, studies aiming to analyse equity issues should include responses of all potential entrants into the public healthcare system. There exists a possibility that patients from a higher SES/dominant ethnic groups might subsequently respond to longer waiting times by leaving the system and seeking private care (Sharma et al., 2013). This phenomenon might generate a biased negative gradient between SES and waiting time due to the fact that public hospitals treat poor patients with both high and low waiting times but only rich patients with low waiting times, calling for an analysis of a two-tier health market which include both the public and the private sectors (Sharma et al., 2013). Furthermore, patient heterogeneity with respect to hospital choice and treatment also lead to self-selection bias as patients differ in risk and time preferences for travel and surgical procedures based on their socioeconomic status (Moscelli et al., 2018).

As discussed earlier, self-selection bias due to patient preferences for public hospitals is not an issue as hospital choice is not an option for patients in NZ. More importantly, we perform additional analyses using different measures of waiting time that are free of the public/private selection issue. The first one is a measure of the waiting time calculated as the time between GP referral date and date of first assessment by the specialist. As patients are more likely to exit the queue only after they learn about their expected waiting time after being booked for surgery, any influence of the presence of the private sector is likely to be very minimal in this additional exercise. The second one is the waiting time measured as the time between the patient's entry into the booking system and their *first scheduled* date for surgery instead of their actual date of surgery. As such, the first scheduled date for elective surgery will be free from any bias caused by rescheduling elective surgeries or missed appointments. Our main results are remarkably robust with respect to different measures of waiting time.

1.4 Data

The dataset is based upon variables derived from the National Booking Reporting System (NBRS) dataset and the National Minimum Dataset (NMDS) for 2014⁵, both of which are provided by the

⁵Except for some minor differences in results across individual years and procedures, our results are largely consistent across model specifications for other calendar years, however, for brevity reasons we report results only for 2014 as it was the most recent year for which complete data was available.

NZ MOH⁶. These two datasets can be linked using a unique encrypted patient ID.

The NBRS extract provided to us contains information on all patients booked for elective surgery along with other procedure specific and individual patient specific characteristics. The NBRS dataset contain information on patients age, gender, ethnicity, booked procedures, assigned CPAC scores, area of residence, the agency/DHB funding the treatment, and the dates when the patient was first referred to the specialist along with assessment, booking and surgery dates. Waiting time is measured as the difference between the date of a patient's entry into the booking system and the date of surgery. The NMDS is a record of all inpatient hospital admissions and thus allows us to confirm the patient information derived from the NBRS dataset with increased certainty thus reducing any measurement error. Furthermore, the NMDS dataset also contain additional information on each patient's DRG codes which groups patients based on the hospital event, diagnosis, procedure, and resource consumption needs across each surgical hospital episode.

As with most elective surgery data, individual patient income was not available. Hence, we use the New Zealand Index of Deprivation⁷ (NZDep hereafter) which measures the level of deprivation of residents in each small census area. There are around 2020 area units in NZ with a median population of 2000 people.⁸ The deprivation index is based on nine census variables which include education, income, employment status, household income, access to internet and transportation and other factors of socioeconomic deprivation. The NZDep combines nine variables using theories developed from international deprivation research, using principal component analysis (PCA) and both construct and criterion validation and hence is shown to be a good proxy for individual SES (Salmond & Crampton, 2012). The NZDep index assigns a score of 1 to 10 for each area unit with ten being the most deprived and 1 being the least deprived. We further create five relative deprivation quintiles (1-5, with 5 being most deprived and 1 being least deprived) based on the

⁶A detailed description of the data-sets can be downloaded here: https://www.health.govt.nz/system/files/documents/publications/nbrs_data_dictionary_v4.3.pdf.

⁷A detailed description of the NZdep index and the dataset can be downloaded here:

https://www.otago.ac.nz/wellington/research/hirp/otago020194.html.

⁸Further information on area units can be accessed here:

http://archive.stats.govt.nz/browse_for_stats/population/estimates_and_projections/ area-unit-population-projections.aspx.

distribution of the deprivation index values across domiciles. Such a measure has been previously used in similar studies in Australia and NZ as a proxy for SES (Johar et al. (2013), Sharma et al. (2013), Seneviratne et al. (2015)). Furthermore, the NZDep index has been extensively used in NZ by local governments to allocate healthcare funding and service levels (Salmond & Crampton, 2012). Finally, the NZDep index is also an appropriate control for local barriers to access to primary care, transportation and other enabling social factors which may affect waiting time for elective surgeries.

Because our dataset is based on extensive matching of patient information across two data sets, any inconsistent information across the two sets is excluded in the analysis. We also excluded observations with unusually high (above two years) or missing waiting times. We only include those procedures for which the CPAC score is on an increasing scale of 1-100, in which higher scores represent more severe cases. The dataset only included patients aged 18 years or older. We only include patients across four major ethnic groups: Europeans, Maori, Pacific and Asian. NZ Europeans form the majority of patients receiving elective surgery and as such is the reference group in our analysis. There were a total of 1073 observations which were dropped from the analysis out of which 176 were Middle Eastern, 37 were Latin American and Hispanic, 101 were African, and 759 were patients who were either from 'Other ethnicity', 'refused to answer', 'did not know their ethnicity', or had their 'response unidentifiable'. Out of the 1073 observations, 75 were removed due to missing values for some variables. We re-estimated our analysis by including these additional observations into a fifth 'other' category and found results that are remarkably similar to our main analysis. We attach the results in the Appendix Section in Table A.1. There was a total of 159 unique surgical procedures within 18 broad surgical categories during the study period for which publicly funded elective surgery was provided. In our main analysis, we only used 10 procedures which appear more frequently for which good quality data is available and that also represent diverse healthcare specialities. Data for many booked procedures were not available for all DHBs or were frequently labelled into an 'other' category, and hence were not usable to

make valid comparisons. Our final dataset consists of 35,349 observations.⁹ The variable means for our final dataset are presented in Table 1.1.

Variable Name		Overall	NZ	Maori	Pacific	Asian	SES 1	SES 2	SES 3	SES 4	SES 5
			Euro-								
			pean								
Waiting time (days)		73.82	73.88	76.69	73.49	67.79	69.35	71.41	71.45	75.54	77.34
CPAC Score		63.18	63.95	64.35	56.66	56.08	65.39	64.45	63.25	62.23	62.25
AGE		65.96	67.68	58.61	59.81	62.58	66.25	65.71	66.70	66.80	64.66
Cost Weight		1.45	1.50	1.39	1.18	1.03	1.52	1.51	1.45	1.42	1.38
Gender	Female	0.52	0.51	0.56	0.55	0.54	0.5	0.51	0.51	0.52	0.54
	Male	0.48	0.49	0.44	0.45	0.46	0.5	0.49	0.49	0.48	0.46
Ethnicity	NZ Eu-	0.78	1				0.88	0.85	0.85	0.81	0.61
	ropean										
	Maori	0.11		1			0.04	0.05	0.07	0.11	0.22
	Pacific	0.05			1		0.02	0.02	0.02	0.04	0.11
	Asian	0.06				1	0.06	0.08	0.06	0.05	0.05
Socioeconomic Status	1	0.12	0.14	0.04	0.04	0.13	1				
(SES)											
	2	0.16	0.17	0.07	0.07	0.21		1			
	3	0.20	0.22	0.12	0.08	0.20			1		
	4	0.25	0.26	0.23	0.20	0.20				1	
	5	0.27	0.22	0.54	0.61	25.7					1
Number of Observa- tions(N)		35349	27511	4040	1736	2062	4256	5549	7105	8772	9667

Table 1.1: Descriptive Statistics

Most of the variables presented in Table 1.1 are self-explanatory. The average waiting time for the overall sample was 74 days. Average waiting times are lesser for Pacific (73 days) and Asian (68 days) ethnic groups when compared to NZE patients (74 days). This is partly related to the fact that these groups have lower CPAC scores than NZE. As demonstrated subsequently, once the urgency of patients' health conditions is more fully controlled for by the CPAC score, Asian and Pacific Islanders wait longer than otherwise comparable NZE patients. Patients belonging to the least deprived regions have lower waiting times (69 days) as compared to patients living in the most deprived regions (77 days). In addition, the least deprived regions in New Zealand have a significantly larger share of NZE residents (88 percent) when compared to only 4 percent Maori and 2 percent Pacific residents. However, the most deprived regions have only 66 percent of NZE residents as compared to Maori and Pacific residents who make up almost 22 percent

⁹Our results are similar even in a larger sample of data including all procedures, age group and ethnicity.

and 11 percent share respectively. The 2013 NZ census data¹⁰ also reveals that median personal income is significantly lower for Pacific and Asian ethnicity compared to both NZE and Maori individuals. Our analysis finds that although average waiting time vary by SES, these differences can be adequately explained by regional differences in supply resources and healthcare utilisation which affect waiting time. Average age for the overall sample was 66 years but was considerably lower for Maori (58.61 years) and Pacific (59.81 years) patient groups. Even though our data sample across NZE patients are equally represented by gender, the proportion of female patients are higher amongst Maori, Pacific and Asian ethnic groups. This could be because there are marked differences in life expectancy at birth by gender and ethnic groups in NZ. According to the NZ Ministry of Social Development, life expectancy at birth for non-Maori female and non-Maori males are 83.9 years and 80.3 years respectively, whereas life expectancy for Maori females and Maori males are 77.1 years and 73.0 years respectively. The life expectancy for pacific females is 78.7 years whereas it is only 74.5 years for males. Finally, the 'Cost Weight' variable is a cost variable that is a proxy for the total resource utilisation for each elective surgery performed. In the absence of price, Cost weight provides DHBs a measure of total resource consumption by the elective surgical hospital episode. As resource consumption could be an important determinant of waiting time for elective surgery, we include 'Cost Weight' into our model specifications.

We also observe marked differences in average waiting times across DHBs (not reported in Table 1.1). Urban populations represented by DHBs such as Auckland (63 days) and Capital and Coast (57 days) have considerably lower waiting time compared to DHBs representing rural populations such as Tairawhiti (94 days), Northland (98 days) and Waikato (88 days). Tairawhiti is also the DHB with the highest share of Maori population (almost 50 percent) compared to any other region in NZ. Also, waiting time vary significantly across procedures as well. Cardiac procedures such as angiography have considerably lower waiting times (42 days) as compared to orthopaedic procedures (95 days). Furthermore, these procedure specific differences in waiting time are also consistent across DHBs. To illustrate this, we plot waiting times for two procedures (Angiography

¹⁰Retrieved from: http://archive.stats.govt.nz/Census/2013-census/profile-and-summary -reports/quickstats-income/personal-income-ethnic.

and Hip Replacement) across DHBs represented in the plot in Figure 1.2. Figure 1.2 depicts that waiting time are considerably lower for angiography procedures across all DHBs when compared to waiting times for hip replacement. Furthermore, patients from Auckland wait shorter periods for both procedures, while waiting times are large for Tairawhiti and Waikato across both procedures. These differences reveal that waiting times in NZ are geographically inequitable. Geographic inequality and variations in waiting times across elective procedures in New Zealand have been previously discussed by Derrett et al. (2009) and Derrett et al. (2013) and as such we do not pursue such an analysis in our paper.



Figure 1.2: Waiting time differences across procedures and DHBs (Source: NBRS dataset, 2014)

1.5 Empirical Specification

We use a multivariate linear regression model to estimate the effects of ethnicity on waiting time for elective surgery, with extensive controls for all patient related clinical and non-clinical factors and regional fixed effects. The model specification is as follows:

$$ln(W_{ir}) = \beta_0 + \beta_1 CPAC_i + \beta_2 e_i + \beta_3 s_i + \beta_4 p_i + h_r + \mu_i,$$
(1.1)

The dependent variable $(ln(W_{ir}))'$ is the log of waiting time for patient 'i' in DHB 'r'. As the waiting time variable is skewed towards the right, we transform the waiting time variable using the logarithmic function. 'CPAC' is the individual patient's priority score determining urgency. 'e' is a vector of dummy variables capturing ethnicity and 's' is a vector of dummy variables capturing SES. Assuming that our model specification is adequately specified, inequality exists when $\beta_2 > \beta_2$ 0' and ' $\beta_3 > 0$ '. 'p' is a vector of patient characteristics such as age, gender, cost weight, dummies for DRGs (241 categories), CPAC scoring codes, cost weight code and the calendar month when the patient was booked for surgery. DRGs controls¹¹ for waiting time differences across various procedures and also provide further controls for patient severity along with the cost weight variable. As DHB performance is assessed at the quarterly level, there could be adjustments and seasonal variations in waiting time across different periods within the same calendar year. As relatively fewer patients are added to the booking system and treated within the same month, we need not worry about the simultaneity of waiting time and demand for elective surgery. h_r are DHB dummies controlling for observed and unobserved factors within DHBs that are common to all patients within the same DHB, such as supply resources (beds, doctors, hospital efficiency etc.) and factors affecting regional healthcare utilisation (such as regional co-morbidity and health levels). μ_i is the error term.

The OLS model may be sensitive to the presence of outliers. Moreover, there is a possibility that the relationship between waiting time and ethnicity may vary at different levels of waiting time, and therefore, we also estimate differences in waiting time by ethnicity across the whole distribution of waiting time by employing a quantile regression approach (Sharma et al., 2013). Moreover, there is a possibility that patients with higher waiting times may seek private surgery, and sample selection bias may occur at higher distribution of waiting times (Sharma et al., 2013). Therefore, employing a quantile regression model also allows us to estimate ethnic gaps in waiting time across a lower distribution of the waiting time where the influence of the private sector is likely

¹¹There were a total of 241 unique DRG codes in the sample.

to be low. For an OLS regression model estimating the effects of covariates X on the conditional mean E(W|X) in the model $E(W|X) = X\beta$, where β is the impact of the increasing mean value of the covariates on the increasing mean value of the dependent variable, the quantile regression model for (τ)th conditional quantile estimates the $q\tau(W) = X\beta\tau$, where $\beta\tau$ can be interpreted as the impact of X on the τ th conditional quantile of W given X (Johar et al., 2013).

1.6 Results

Table 1.2 displays the results from estimating Equation 1.1. We find significant effects of ethnicity on waiting time for elective surgery for Pacific and Asian patients across multiple model specifications except orthopaedic surgery. We find stronger results for Asian and Pacific patients compared to Maori as reflected in the higher coefficient estimates. For Maori, we only find significant ethnic gaps across model specification for all categories pooled together and for Angiography procedures only. The first column in Table 1.2 presents OLS estimates for the overall sample pooled for all surgical categories that include General Surgery, Ophthalmology procedures, Orthopaedic procedures and Angiography procedures. Our results show that Maori patients wait 3 percent longer for elective surgery whereas Pacific and Asian patients respectively wait 9 percent and 8 percent longer for elective surgeries. The result for Asian is particularly interesting. In Table 1.1, we observed that the average waiting time is substantially lower for Asian than NZE. The result in Table 1.2 finds that once patient related clinical and non-clinical characteristics are fully controlled for, Asians are waiting much longer than NZE. In the subsequent columns we present OLS estimates with our preferred model specification across data sample stratified by individual surgical categories. We present results over four major surgical categories which include General surgery, Ophthalmology procedures, orthopaedic surgery and Cardiac procedure (angiography). We find that ethnic differences in waiting times are pervasive for Asian and Pacific patients across all surgical categories except Orthopaedic surgery. Although, average waiting times were considerably lower across all DHBs for angiography procedures when compared to other surgical specialities,

ethnic differences in waiting times were highest for angiography with Pacific patients waiting almost 22 percent longer for publicly funded elective surgery and both Maori and Asian patients waiting 9 and 13 percent longer. Across all our estimates for individual surgical categories, we find no effect of SES on waiting time for elective surgery.

Results for other control variables can be summarized as follows: As expected, we find large and significant effects of CPAC scores on waiting time for surgery. The negative direction of the coefficient estimates suggests that more severe patients with higher CPAC score wait less time for elective surgery. We find significant effects of gender only across Ophthalmology procedures with men waiting 4 percent fewer days than women. We find a significant positive effect of age on waiting time for surgery and the effect is highest across Cataract procedures. This result suggests that older patients wait longer for elective surgery. We found negative effects of cost weight; however, the results were only significant for orthopaedic surgery.

Finally, DHB fixed effects were included in all model specifications, and the coefficient estimates for many of them were large and highly significant and were consistent across surgical categories. Moreover, the calendar month fixed effects were also significant suggesting there exists seasonal variation in waiting times for elective surgeries within DHBs.

	All Categories	General Surgery	Ophthalmology Procedures	Orthopaedic Procedures	Angiography (Cardiology)
(Intercent)	4 07***	4 08***	2 76***	6 34***	4 13***
(intercept)	(0.52)	(0.49)	(0.53)	(0.46)	(0.42)
log(CPAC Score)	-0.20***	-0.48***	-0.24***	-0.39***	-0.61***
	(0.01)	(0.05)	(0.02)	(0.09)	(0.06)
REF: Female					
Male	-0.02^{*}	0.02	-0.04***	-0.01	-0.04
	(0.01)	(0.02)	(0.01)	(0.01)	(0.03)
DEE NZE					
REF: NZ European Maori	0.03**	0.04	0.01	0.01	0.09*
Waom	(0.01)	(0.04)	(0.02)	(0.03)	(0.04)
				()	
Pacific	0.09***	0.08^{*}	0.05^{*}	-0.00	0.20***
	(0.02)	(0.04)	(0.02)	(0.05)	(0.06)
Asian	0.08***	0.07*	0.05*	-0.09	0.12**
Asian	(0.02)	(0.03)	(0.02)	(0.06)	(0.05)
	(010_)	(0.02)	(***=)	(0.00)	(0.00)
REF: SES 1					
SES 2	0.01	0.02	-0.01	0.03	0.01
	(0.01)	(0.03)	(0.02)	(0.03)	(0.05)
SES 3	-0.01	0.00	-0.02	0.00	0.04
	(0.01)	(0.03)	(0.02)	(0.03)	(0.04)
SES 4	0.00	0.02	-0.01	-0.01	0.01
	(0.01)	(0.03)	(0.02)	(0.03)	(0.04)
SES 5	0.02	0.01	0.01	0.02	0.02
	(0.01)	(0.03)	(0.02)	(0.03)	(0.04)
log(age)	0.09***	0.07**	0.18***	-0.07	0.11
	(0.02)	(0.03)	(0.04)	(0.04)	(0.08)
Cost Weight	-0.06	-0.00	0.05	-0.06^{*}	-0.05
U	(0.03)	(0.06)	(0.29)	(0.03)	(0.12)
DHB Effects			v	v	\checkmark
DRG Effects	1	1	1	1	1
Calendar Month	1	1	1	1	\checkmark
CDAC Call	1	,	,	,	1
CrAC Codes	V	~	V	V	V
CostWeight Codes	1	1	1	1	1
R ²	0.40	0.45	0.29	0.41	0.34
Num. obs.	35349	9600	12953	7046	2959

Table 1.2: OLS Results: All categories (pooled) and individual health speciality specific categories

1.7 Robustness tests

In this section we conduct additional robustness tests concerning our main results. First, we examine how the OLS results in Table 1.2 are sensitive to the presence of outliers. Moreover, there is a possibility that the relationship between waiting time and our key explanatory variables vary at different levels of waiting times Sharma et al. (2013). Hence, we also estimated quantile regression models at different conditional percentiles of the log of waiting times.

As discussed earlier, sample selection bias could exist due to the presence of the private sector or public hospital choice. There is a possibility that a sample of publicly funded patients in the analysis might include all patients from lower SES with expected high and low waiting time but only patients from higher SES with expected low waiting time (Sharma et al., 2013). To address this concern, we re-estimated our preferred model with an alternate estimate of waiting time measured as the difference between the patient's referral date by the GP and the first specialist assessment date. This would mean that the waiting time measure is independent of the expected waiting time for access to elective surgery and hence the influence of the private sector option is minimal. Furthermore, it could be argued that certain ethnic groups are more likely to miss appointments and hence wait longer for elective surgery. This could lead to biased estimates; however, our dataset allows us to measure waiting time based on the patient's first scheduled date for surgery. This measure of waiting time would be independent of any patient related factors that could potentially affect the final waiting time after being booked into the waiting list.

The results are reported in Table 1.3. and show that our main findings survive these additional exercises. Nevertheless, estimates of ethnic gaps for Maori lose their statistical significance. The results confirm that there exist significant ethnic gaps in waiting time for Pacific and Asian patients whereas we do not see an effect of SES on waiting times. In particular, waiting time is ethnically inequitable in NZ across all quantiles of the waiting time distribution for Asian and Pacific patients.

	OLS Results (Referral Date)	OLS Results (Scheduled Date)	Q10	Q25	Q50	Q75	Q90
(Intercept)	4.03***	4.25***	3.95***	3.19***	4.14***	4.88***	5.15***
	(0.18)	(0.55)	(1.15)	(0.95)	(0.66)	(0.51)	(0.50)
logCPAC Score)	-0.16***	-0.21***	-0.34***	-0.23***	-0.13***	-0.07***	-0.05***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
REF: Female	-0.02	-0.02**	-0.01	-0.01	-0.01	-0.01	-0.02**
Male	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
REF: NZ European	0.03	0.02	0.01	0.03	0.01	0.02*	0.01
Maori	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Pacific	0.08*	0.08***	0.08*	0.11***	0.06***	0.05***	0.04**
	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)
Asian	0.09**	0.08***	0.08**	0.09***	0.08***	0.03*	0.01
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
REF: SES 1	0.03	0.01	0.03	0.04	0.01	0.01	-0.01
SES 2	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
SES 3	-0.01	-0.01	0.02	0.00	-0.01	-0.01	-0.01
	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
SES 4	0.02	-0.00	0.00	0.01	0.00	0.01	-0.01
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
SES 5	0.04	0.01	0.05*	0.04	0.03*	0.01	0.00
	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
log(age)	-0.08**	0.07***	0.11***	0.09***	0.07***	0.05***	0.02
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
costweight		-0.05 (0.03)	-0.17*** (0.05)	-0.03 (0.04)	-0.01 (0.03)	-0.01 (0.02)	-0.06** (0.02)
DHB Effects	1	1	1	1	1	1	1
DRG Effects	×	1	1	1	1	1	1
Procedure Effects	1	×	×	×	×	×	×
Calendar Month	1	1	1	1	1	1	1
CPAC Codes	×	1	1	1	1	1	1
CostWeight Codes	×	1	1	1	1	1	1
Num. obs.	21947	35258	35349	35349	35349	35349	35349

Table 1.3: OLS with alternate measures of waiting time and Quantile Regression Results

 $\frac{2+2+1}{2} < 0.001, \frac{**p}{2} < 0.01, \frac{*p}{2} < 0.05.$ Robust Standard Errors are reported in the parenthe-sis for OLS estimates.

Figure 1.2 reveals that there are significant variations in waiting time across both DHBs and procedures. Moreover, our results in Table 1.2 show that the estimated coefficients on our ethnicity variables vary in magnitudes across procedures. Hence, we also estimate the effects by each DHB and each procedure to investigate if the disparities in waiting time are driven by only a few DHBs and procedures or if these disparities are system wide. Such analyses are also worthwhile because DHBs vary widely in terms of area level hospital utilisation factors such as regional co-morbidities and health levels and other supply factors such as hospital beds, doctors, and supplier quality.

In Appendix A, we present a plot of coefficient estimates with 95 percent confidence intervals for ethnic gaps across Maori, Pacific, and Asian groups by estimating separate regression models for each DHB and each procedure. The plot for coefficient estimates for Maori, Pacific, and Asian by each procedure is presented in Figure A.1, Figure A.2, and Figure A.3 respectively. The plot for coefficient estimates for Maori, Pacific, and Asian by each DHB is presented in Figure A.4, Figure A.5, and Figure A.6 respectively.

The results in Figure A.1, Figure A.2, and Figure A.3 indicate that Maori, Pacific, and Asian ethnic groups wait significantly longer for elective surgery for Angiogram procedures. The effect for Angiogram procedures are both large and significant as indicated in Table 1.2. Moreover, these results are also consistent across Ophthalmology procedures such as Vitrectomy and Cataract surgeries and other procedures such as prostatectomy and breast surgery. The results are less consistent across orthopaedic procedures (hip replacement and knee replacement) and in many occasions the estimated effects are lower for Maori, Pacific, and Asian especially for knee replacement. One possible explanation for such a result has been provided by (Lungu et al., 2019). Joint replacement surgeries are beneficial when they reflect patient preferences rather than physician's discretionary choices and hence, we observe significant variation in waiting time for elective surgery for these procedures (Lungu et al., 2019). This suggests that ethnic gaps may be higher in procedures that reflect physician's discretionary choices.

The results in Figure A.4, Figure A.5, and Figure A.6 show large variations in ethnic gaps across DHBs. Before, interpreting the results, it is also important to take into consideration the

distribution of Maori, Pacific, and Asian ethnic groups across regions to cross-reference the findings with the geographic distribution of ethnic groups across NZ. The proportion of ethnic groups across regions are presented in Table 1.4.

DHB	Proportion	Proportion	Proportion	Proportion
	NZE	Maori	Pacific	Asian
Northland	75.67	32.36	3.21	2.83
Waitemata	71.48	9.32	7.99	18.27
Auckland	57.43	7.70	12.49	28.18
Counties Manukau	48.02	15.46	23.82	23.10
Waikato	76.95	21.73	3.91	7.29
Lakes	70.70	34.58	4.24	5.34
Bay of Plenty	78.36	24.36	2.47	4.77
Tairawhiti	60.82	48.86	3.83	2.42
Taranaki	86.23	17.42	1.63	3.45
Hawke's Bay	77.65	24.40	4.37	3.57
Whanganui	80.57	24.85	3.05	2.65
Midcentral	81.93	18.33	3.72	6.12
Hutt	74.84	16.17	9.23	10.05
Capital and Coast	76.29	10.64	8.37	12.18
Wairarapa	88.40	16.10	2.68	2.13
Nelson Marlborough	90.59	9.43	1.71	3.06
West Coast	91.20	10.54	1.05	2.24
Canterbury	86.23	8.23	2.63	7.46
South Canterbury	92.39	7.20	1.07	2.54
Southern	89.08	9.19	2.06	4.57

Table 1.4: Ethnic group proportions by DHB (Source: NZ Census data, 2013)

Table 1.4 shows that while Asian and Pacific patients are predominantly located near the Auckland (Counties Manukau, Waitemata, and Auckland) and the Wellington region (Capital and Coast, Hutt), there is a greater variation in Maori population across regions mainly within the North Island. The plot coefficients show that there are significant variations across DHBs both in terms of the sign and the magnitude of estimated ethnic gaps. The results are less consistent across smaller DHBs such as West Coast, Lakes, South Canterbury, Wairarapa, and Tairawhiti. On many occasions, the sample included less than 5 observations for Pacific and Asian patients across these DHBs, hence, the coefficient estimates may be largely driven by outliers in the data. As revealed in Table 1.4, Pacific and Asian population are overly represented in Auckland, Counties Manukau, Capital and Coast, and Hutt DHBs. The coefficient estimates for Asian and Pacific patients across these DHBs reveal that Asian and Pacific patients wait longer for surgery compared to NZE. It is important to note that the plot coefficients for these larger DHBs are dwarfed by unusually large coefficients for smaller DHBs in our sample, however, in many cases the coefficient estimates of ethnic gaps across these larger DHBs were both higher and statistically significant. Because we have a large sample of patient records available for Maori across multiple DHBs, the plot coefficient estimates are more stable and indicate that across many DHBs Maori have consistently higher waiting times for elective surgery compared to NZE.

1.8 Discussion

Our main finding is that waiting time in NZ is ethnically inequitable, mainly for Pacific and Asian ethnic groups. The waiting time differences are relatively limited across Maori, however, they are consistently higher on average compared to NZE. Across Pacific and Asian patients, waiting time differences are both large and significant and are pervasive across various surgical procedures.

There may be several reasons why waiting time in NZ is ethnically inequitable. In this section, we discuss some possible mechanisms for existing disparities in waiting time across ethnic groups. Exploring possible mechanisms is also important as policy implications for addressing these disparities may differ based on different underlying causes of these disparities. Existing studies investigating ethnic disparities in waiting time for elective surgery is scarce, and hence possible mechanisms for these disparities are not adequately identified in existing literature. This is because most studies are concerned with analysing the impact of SES on waiting time. We would argue that many existing studies do not account for endogeneity of supply factors and hence the existing disparities in waiting time according to SES may be overstated. In our analysis, we estimated the main results without controlling for supply factors (DHB fixed effects) and found significant effects of SES on waiting time for elective surgery i.e. patients from higher SES have lower waiting times. However, once we controlled for supply factors by including DHB fixed effects, we found no significant effects of SES on waiting times for elective surgery. This suggests that waiting time disparities by SES in existing studies maybe driven by failure to adequately control for supply factors.
As discussed previously, once patient related clinical factors are controlled for, individual patient specific characteristics should not impact waiting times for elective surgery. Despite that, several studies often cite existing inequality in waiting time for elective surgeries. Studies suggest that such disparities can be attributed to several factors such as the care process, health insurance status and social and cultural differences on the part of the healthcare provider (Alrwisan & Eworuke, 2016). Furthermore, other possible mechanisms including clinical uncertainty with physicians operating with varying beliefs about the patient's condition, ethnic diversity, and differences in understanding of the cultural, social, and economic factors impacting patient's health status have also been suggested as potential factors that may cause these healthcare disparities (Nelson, 2002). There may also be other potential mechanisms for existing ethnic disparities in waiting time and we discuss some of these potential causes in this section.

(i) Discrimination by health professionals and schedulers

As discussed previously, personal cultural beliefs of health professionals influence the outcome of interaction with the patients (Nelson, 2002). Evelyn Theunissen (2011) and White et al. (2010) find that indigenous patients find delayed medical interventions compared to non-indigenous people. This is also likely since patients have a much closer interpersonal relationship with their GP and their interactions may affect the likelihood of them being referred to a specialist and the date between referral and their First Specialist assessment. Our results indicate that some of the existing ethnic gaps in waiting time may be caused by discrimination. This is because we subsequently rule out some other possible mechanisms for ethnic gaps in waiting time.

(ii) Different ethnic groups have fewer flexible schedules due to cultural obligations

Although it is likely that scheduling flexibility due to cultural obligations may cause some patients to reschedule the surgery date, it is unlikely that the booking clerks are aware or take explicit considerations of patients' cultural obligations while booking the patients for surgery. Moreover, our dataset allows for us to rule out this possibility. As an additional robustness check, we also estimated ethnic gaps in waiting times measured as the date between assessment and the first scheduled surgery date. Results reported in Table 1.3. that Asian and Pacific patients wait significantly longer for surgery compared to NZE patients. Hence, we would not expect scheduling issues due to cultural obligations as being the likely cause of ethnic gaps in waiting time.

(iii) Exit from the sample by high SES NZE

It is possible that patients from higher SES may skip the booking system and choose to seek private treatment. Nevertheless, patients are not informed of their actual date of surgery beforehand and they only receive a confirmation that a surgery will be provided to them a couple of weeks before the actual date of surgery. Hence, the number of patients from high SES leaving the booking system is likely to be very low. In principle, once two patients receive the same CPAC score, they should have equal waiting times on average. Hence, once we control for CPAC scores, any existing disparities in waiting times for elective surgeries are unlikely to be caused by exit from the sample by high SES NZE. Moreover, our quantile regression results indicate that ethnic gaps remain at all distributions of waiting time. As patients are likely to leave the booking system only if they find that they will be waiting longer for surgery, the impact of high SES leaving the booking system is likely to be very minimal in our analysis.

(iv) Education/Income may be negatively correlated with waiting time

It can be argued that our measure of SES given by area level deprivation index is an imperfect proxy for individual SES and controlling for deprivation index may still lead to biased results if NZE have more education/ income than others within an area. We acknowledge that it would have been ideal to analyse the impact of individual income/education on waiting time for elective surgery, however hospitals do not record individual income while collecting patient information, and therefore, data on patient's individual income were not made available to us. Nevertheless, we would argue that the area level deprivation index captures individual

income for the residents along with several other enabling factors that may impact an individual SES such as household income, education, access to the internet and transportation and hence is a good proxy for SES. Moreover, as we outlined earlier, area level deprivation index has been extensively used in existing literature to control for income.

(v) Language barriers

It is possible that language barriers may have a significant impact on patients' ability to communicate their healthcare need to the specialists. It may also cause delays in filling relevant paperwork. Our results also indicate that ethnic gaps are larger for Asian and Pacific patients across regions where they may be larger percentages of people with language barriers, whereas for Maori patients they are more consistent across regions. We do not have an indicator in our dataset that records a patient's language preferences, hence, we cannot conclusively affirm if language barriers are likely causes of these inequities.

(vi) Correlations within the geographic distribution of specialists and ethnicities

It is possible that significant correlations within geographic distributions of specialists and ethnicities may exist. For example, specialists treating certain conditions may be disproportionately located in areas with lower ethnic minority representation. We also see a huge within DHB across procedure variation in our data. Therefore, to control for the influence of the geographic mismatch between specialists across procedures, we also tested our main results including procedure by DHB fixed effects. The results for this additional exercise are presented in Table A.1 in Appendix A. Our results indicate that, when controlling with procedure by DHB fixed effects, the magnitude of the ethnicity coefficients reduces but does not disappear completely and they are higher by a percentage point for Maori patients. This suggests that some of the existing gaps in waiting time in ethnicity maybe driven by the geographic mismatch of specialists and ethnicities.

Our results indicate that existing disparities in waiting time for elective surgery across ethnic groups may be attributable to preferences of health professionals and other factors such as lan-

guage barriers. As discussed earlier, waiting time are larger across procedures that reflect physicians' preferences and discretionary choices. Moreover, there is some indication of a geographic mismatch of specialists and ethnicities. While estimating multiple model specifications, we find large and significant coefficient estimates for DHB dummies, suggesting that waiting time in New Zealand is geographically inequitable. This phenomenon has been documented previously in the existing literature pertaining to the booking system in New Zealand by Derrett et al. (2009) and Derrett et al. (2013). Those authors argue that despite a national policy framework, significant regional differences in access to surgery persist and there remains large shortfalls in treating all patients requiring surgery (Derrett et al., 2013). Furthermore, some DHBs are more affected than others in terms of affected surgical services, staff shortages, bed shortages, funding and management (Derrett et al., 2013). Despite significant funding injections into the management of waiting lists, there has not been any significant improvements in access to elective surgery (Derrett et al., 2013).

Our results contrast with many existing studies (Siciliani & Verzulli (2009); Tinghög et al. (2014); Sørensen et al. (2009); Kaarboe & Carlsen (2014); Heinrich et al. (2018); Johar et al. (2013); Sharma et al. (2013)) which find evidence of inequality in waiting time by SES. Because we are analysing a different healthcare market in NZ, exact mechanisms of waiting time differences might be different which makes cross-country comparisons difficult. Nonetheless, our extensive controls for sample size, sample selection bias, supply factors and patient severity considerably strengthens the validity of our estimates. We also show that once regional factors are controlled for, any perceived differences in waiting time by SES disappears which suggest that the impact of SES on previous studies may be overestimated. Furthermore, public hospital choice option in healthcare markets in other OECD countries makes hospitals/regions as substitutes for the relevant catchment population (Gravelle, Smith, & Xavier, 2003). It can be argued that our measure of SES given by area level deprivation index is an imperfect proxy for individual SES. However, we would argue that using individual income would mean that other factors such as household income, education status, access to the internet and transportation, and other enabling factors affecting waiting time

would be unobserved and as such the NZ Deprivation index adequately captures SES along with various other factors that may contribute to waiting time differences.

Our results have important policy implications. First, discussions related to improving health outcomes for ethnic minorities in NZ should be extended to both Pacific and Asian ethnic groups as current conversations largely focus on Maori patients. Our results indicate that existing disparities in waiting times are on many occasions even larger for Pacific and Asian ethnic groups. However, our results also indicate that ethnic gaps in the case of Maori, Pacific, and Asian patients maybe driven different factors. It may be worthwhile to study if these differences are due to discrimination by schedulers or due to language barriers and cultural issues that may disproportionately impact patients from certain ethnic groups. This means a targeted policy especially for procedures that reflect physicians' preferences and for DHBs that have higher proportion of ethnic minorities may work well in improving waiting time inequities. Moreover, DHBs vary greatly in average waiting times and there is some evidence of geographic mismatch between health care availability and deprived ethnic minorities. Therefore, increasing access and availability of healthcare resources should be a primary focus for the NZ government while devising health policy.

1.9 Conclusion

Rationing by waiting times for elective surgery is an important feature of many healthcare systems around the world. As such, patients are prioritized on their relative healthcare need rather than their individual characteristics such as ethnicity, gender and their ability to pay. There is a growing amount of empirical literature which reveals that even in the absence of price, there exists significant inequities in access to healthcare. In line with these studies, we analyse the NZ healthcare market for elective surgery to determine mainly if ethnic disparities exist in waiting time. We use a large administrative dataset with extensive controls for severity of the clinical condition and regional factors affecting waiting time.

Although our analysis does not reveal all underlying causes of waiting time differences by eth-

nicity in NZ, we extensively controlled for sample selection bias caused by patient preferences, clinical severity, and other regional factors. Furthermore, Maori, Pacific and Asian ethnic groups are socioeconomically deprived in New Zealand, have poorer health conditions and outcomes and predominantly reside in geographically deprived areas. Therefore, reducing geographic and ethnic inequities in waiting time should be a primary concern for the NZ government when devising its health policy and agenda. Moreover, health policies on elective surgery should be targeted specifically towards improving outcomes for procedures that depend more on physician's preferences rather than patient preferences. Finally, most discussions on health inequities in NZ focuses on differences between Maori and NZE. Our findings, however, present evidence that these discussions should be extended to the Pacific and Asian patients as well. As previously discussed, the census data reveals that Asian ethnic groups are the fastest growing population groups in NZ, and this warrants discussions on issues related to healthcare inequities.

We find evidence that waiting time in NZ is ethnically inequitable especially for Pacific and Asian patients who have significantly unfavourable conditions for access to elective surgery in NZ. The case for inequality in waiting times for Maori patients is less strong. However, we find consistent results for Maori waiting longer for elective surgeries across most procedures and across most DHBs. Furthermore, ethnic differences in waiting times are persistent despite controlling for other enabling factors which may affect waiting time such as access to primary care and transportation. We also find evidence that waiting time in NZ is geographically inequitable. Our analysis finds that once regional factors are controlled for, SES does not impact waiting time for elective surgery refuting various studies who make such claim in other healthcare markets. Although cross-country comparisons are difficult, our analysis reveals that regional factors which may impact resource allocation and healthcare utilisation affect waiting time for the relevant catchment population, and as such there is a possibility that SES inequality in waiting time may be overestimated in studies in other healthcare markets.

Patients in NZ are constrained within their own DHBs and ethnic minorities predominantly reside in geographically deprived regions, hence, existing inequities in access to elective surgery

is highly concerning in NZ. NZ policymakers must recognize that despite a national policy framework, rationing framework and scoring prioritization mechanism, the objectives of universal healthcare system are not fully achieved in NZ.

Chapter 2

Estimating the impact of waiting times on supply and demand for elective surgery in New Zealand

2.1 Introduction

Publicly funded healthcare is typically financed by general taxation, such that the 'receivers' of healthcare services do not make a direct financial contribution to the 'providers' at the point of service delivery. As resources are scarce, a rationing mechanism to manage the demand for such services is usually necessary. Price is traditionally used as a rationing device in many service sectors. In the absence of price, rationing often takes the form of waiting time.

Waiting time is considered a major policy concern in many OECD countries (Siciliani & Hurst, 2005b). Waiting time as a rationing mechanism has significant implications for the interplay between supply and demand for healthcare. The utility derived from the consumption of health services declines significantly the longer the individual must wait for these services (Lindsay & Feigenbaum, 1984; R. M. Martin et al., 2003; S. Martin & Smith, 1999). This may also worsen health outcomes for some individuals (Meerding et al., 2005). There also exists empirical evidence suggesting inequality in both access and utilisation of health services within public healthcare systems (Monstad et al., 2014; Sharma et al., 2013; Tinghög et al., 2014). Longer wait times can induce individuals with a stronger socioeconomic status to seek private treatment. They can achieve this by buying private insurance or paying out of pocket. This can reduce demand for public healthcare. On the supply side, the 'providers' of healthcare might be driven by altruistic motivations or performance target concerns set by their respective health ministries (Cullis et al., 2000; Jones, 2012). Moreover, both geographic variations in available resources and socioeconomic factors within the region affect supply and demand and subsequently waiting time for elective surgery. There exists varying dynamics within health systems across countries that affect waiting times for elective surgery and the subsequent allocation of resources. These system dynamics make the study of the impact of waiting time on supply and demand for elective surgery a worthwhile topic for economic research.

An extensive literature exists on the estimation of supply and demand elasticities with respect to waiting times across other OECD countries, especially within the context of the National Health Service (NHS) in the UK. Nevertheless, our study is the first to estimate supply and demand elasticities within the New Zealand healthcare context. Because countries differ with regard to institutional arrangements, the public–private mix, funding schemes and resource allocation mechanisms, it is worthwhile to take up cross-country studies. In terms of methodology, we construct and analyse a 'triangular model' of local health markets. The model is based on the premise that there exist equilibrium differentials in excess demand for health services (and therefore waiting times) across different regions within a single country, and as such the model consists of three separate equations of waiting time, supply and demand. In contrast, existing studies such as Martin Smith (1999), Riganti et al. (2017), and Stavrunova and Yerokhin (2011) often adopted a two-equation model of demand and supply by imposing a 'market clearing' condition of equality of demand and supply, without specifying a separate equation of waiting time. While Gravelle, Smith, and Xavier (2003) and S. Martin et al. (2007) estimate supply and demand elasticities with separate measures of supply and demand, the empirical strategies used by these papers were considerably different from ours. Among others, waiting times are endogenously determined in our three-equation system, whereas they are exogenous in their two-equation system. Our model includes the behaviour of waiting time as an additional structural equation. Most importantly, our model is flexible enough to include existing models as special cases. Therefore, our study adds to the existing literature by its use of a novel empirical strategy to estimate supply and demand elasticities as well as by investigating the issue in the less-studied healthcare market of New Zealand.

Our triangular model consists of three structural equations. Both demand and supply equations are functions of the waiting time, and the waiting time is affected by the excess demand. Identification of each structural equation requires exclusion of at least one exogenous variable from the equation under consideration. In our set up, the regional mortality rate, supply of public beds, and the population share of Maori, respectively, enter only the demand, supply, and the waiting time equation, making the structural parameters of each equation over-identified. A subsequent section justifies exclusion restrictions on a priori grounds and provides various test results including the Sargan test of over-identifying restrictions, the weak-instrument test, and the Hausman specification test.

From the New Zealand perspective, existing studies on waiting times for elective surgery have focused on the impact of waiting times on health-related quality of life and on the impact of waiting times on equity provisions for surgery (Derrett et al. (1999)), or on the analysis of the scoring prioritisation mechanism (Dew et al. (2005b)). Analyses of supply and demand elasticities are important from a policy perspective because they allow policy makers to understand how supply and demand for elective surgery respond to changes in waiting times. Demand elasticities are important because it informs policymakers if supply side initiatives to reduce waiting times maybe offset by large increases in demand (Riganti et al., 2017). Hence, if demand is elastic, supply side initiatives to reduce waiting times. Hence, estimates of demand and supply elasticities guide policy makers to make decisions on subsequent allocations of resources and to estimate the impact of their decisions on reducing waiting times for publicly funded elective surgery. Therefore, our study is

also of significant relevance to public health policy in New Zealand.

Employing a comprehensive dataset of all patients seeking publicly funded elective surgery in New Zealand during the 2011–2015 period, we find that demand is inelastic to waiting time, whereas supply is elastic. This means that a long-term exogenous increase in supply can lead to a permanent reduction of waiting times and thus improve the timely access to care.

This paper is organized as follows. First, in Section 2, we provide some background on the New Zealand healthcare market relating to publicly funded elective surgeries. We then briefly review the relevant existing literature in Section 3 and describe the dataset and key variables in Section 4. Section 5 discusses our econometric model and empirical estimation strategy, and the results are discussed in Section 6. In Section 7, we perform additional robustness tests to confirm our main findings. Lastly, in Section 8, we summarize the main findings of the paper and conclude with some policy implications.

2.2 The New Zealand healthcare market for elective surgery

New Zealand has a two-tiered health system with provisions for both private and public hospitals performing elective surgery. However, unique to New Zealand is the scoring prioritisation system, also known as the New Zealand Booking System (NZBS), which was introduced in 1998. This system prioritises patients by 'need' as well as the patient's 'ability to benefit' from surgery. This rationing by prioritisation is aimed at achieving equal access to surgery for all citizens and residents. A typical elective services patient flow within the public system, not accounting for any emergency inflows or leakages to private systems, is summarised in Figure 1.1 in Chapter 1.

Once a patient has been referred to a specialist by a GP, he or she must receive a First Specialist Assessment (FSA) within four months of their referral date. If surgery is deemed appropriate, a surgical procedure must be performed within the following four months.¹ These standards are set

¹Before July 2013, the maximum waiting time for receiving elective surgery was six months. In July 2013, the maximum waiting time was reduced to five months, which was further reduced to a maximum of four months from January 2015.

by the Ministry of Health (MOH hereafter), and there are financial penalties for District Health Boards (DHBs hereafter) for failure to meet these targets. Nevertheless, the specialist can decline surgery and return the patient to primary care if the surgery is considered inappropriate or if the priority score for elective surgery is below a certain threshold. The threshold varies across DHBs and surgical procedures and is calculated by considering the demand and capacity constraints across each DHB. Those who would benefit from surgery but have a priority score below the threshold are placed in an 'active review' status and must be reassessed within the subsequent six months (earlier if their condition deteriorates). Furthermore, since the MOH sets targets for the surgical volume that each DHB is required to manage, significant constraints are imposed on DHBs by the MOH, which in turn affects the interplay of supply and demand across administrative regions.

The DHBs provide health and disability services to their area's population. Currently, there are 20 DHBs in New Zealand.². These DHBs are defined at the mesh block level. The DHBs are responsible for providing or funding provisions for health services in their district. The funding and management of hospital facilities for elective surgery is also provisioned by the DHBs in each district. Patients residing in a certain geographic area are assigned to one of the hospitals represented by their DHB for elective services, and the inter-regional movement of staff and patients occurs on a case-by-case basis.

Based on the estimates for 2018 from Stats NZ,³ the population served by the DHBs range from over 500,000 people for larger DHBs such as Waitemata, Auckland and Counties Manukau to less than 50,000 people for smaller DHBs such as Taranaki, Wairarapa and West Coast. The smaller DHBs only operate a single hospital facility. Larger DHBs operate separate facilities for elective surgical services with additional facilities providing acute, diagnostic, mental health, rehabilitation, children's and maternity services. Budgets and resources are centrally managed (the only/largest facility usually serving as the hub) and are not allocated separately for each facility. The majority of the workforce work across multiple sites within each DHB. Hence, in the case of New Zealand,

²There were 21 DHBs before the merger of Otago DHB and Southland DHB into Southern DHB in 2010.

³Data retrieved from :https://www.stats.govt.nz/information-releases/2018-census-nz-stat-tables.

the DHBs are both the 'funder' and the 'provider' of health services. For the purpose of this study, we restricted our analysis to the DHB level because there exist key input and output differences at the facility level both between and within DHBs, thus making comparisons across facilities problematic.

Unlike in the UK, hospital choice is not an option for patients in New Zealand, and the possibility of obtaining an alternative provider is only available if such provisions are deemed necessary by the patient's DHB or if the patient seeks private treatment. The New Zealand healthcare market also differs significantly from those of other OECD countries in terms of the public–private mix. In Australia, almost one-half of the population have some form of private health insurance (PHI), with 47 percent covered by PHI in New South Wales (Stavrunova & Yerokhin, 2011). In the UK, 96 percent of acute beds are provided by public hospitals, while in Italy 81.7 percent are provided by public hospitals (Riganti et al., 2017). According to the New Zealand MOH, more than 35 percent of adults and 28 percent of children are covered by PHI. The proportion of adults with PHI coverage has decreased from 40 percent in 1996-1997 to 35 percent in 2011–2015; for children, PHI coverage has decreased from 31 percent in 1996-1997 to 28 percent in 2012–2015 (Ministry of Health NZ, 2016). Hence, differences in the public–private mix, rationing mechanisms and allocation of resources make the New Zealand healthcare market worthwhile for economic analysis.

2.3 Relevant Literature

Estimating supply and demand elasticities is highly relevant from a policy perspective. A fair number of studies estimate demand and supply elasticities within the context of public healthcare systems. For instance, Lindsay Feigenbaum (1984) develop a theoretical model for waiting times as a rationing mechanism. Their findings suggest that the rate of joining a waiting list is negatively related to the expected delay in supply and to the rate at which demand diminishes over time. Another study by Martin and Smith (1999) examine the rationing feature within the British NHS. Their findings suggest that demand elasticities are low (with estimates ranging around -0.2) and

imply that an exogenous increase in supply (funding additional resources) can have a positive impact on reducing waiting times. A similar study conducted in New South Wales in Australia, however, reported a higher elasticity of demand (- 1.7) as compared to the British NHS studies (Stavrunova & Yerokhin, 2011). The authors suggest that such a high elasticity of demand can be explained by the high number of people with PHI in Australia (Stavrunova & Yerokhin, 2011). Riganti et al. (2017) find that waiting times in Italy had elasticity estimates ranging from -0.15 to - 0.24, which is similar to the findings of the British NHS study by Martin and Smith (1999).

Due to data limitations, these papers assume that the waiting lists in question are in equilibrium, and that the queue length is relatively stable. Because waiting time simultaneously affects supply and demand, these studies used an instrumental variables approach in which several demand and supply shifters act as instruments for waiting times in the supply and demand equations. These previous studies were constrained by data, and typically used the measure of supply (total number of discharges for each surgical procedures) as a dependent variable across both supply and demand equations justified by the 'market clearing' assumption. The works by Martin and Smith (1999) and Stavrunova and Yerokhin (2011) employ cross-sectional data and estimate the demand and supply equations at the 'electoral ward' and 'postal code' level, respectively. On the other hand, Riganti et al. (2017) use a time-series data estimated at the regional level while studying waiting time in Italy. Gravelle, Smith, and Xavier (2003) estimate elasticities with separate measures of supply and demand using hospital data from the UK. These authors use both fixed effects and random effects models to estimate these elasticities. It can be argued that convergence to market equilibrium is slow given assumptions about patient expectations, and the dynamic nature of the market and policy changes (Gravelle, Smith, & Xavier, 2003). Based on a generalisation of a model developed by Gravelle, Smith, and Xavier (2003), S. Martin et al. (2007) also estimate separate empirical models for demand and supply at the provider (hospital) level using ordinary least squares (OLS) methods along with seemingly unrelated regression (SUR) methods to correct for the correlation between the residuals of the OLS estimates.

Since we did not impose a 'market clearing' assumption, we could estimate separate models of

demand and supply. Like the paper by S. Martin et al. (2007), we use different measures of demand (number of patients added to the system) and supply (number of patients treated). We measured quantity supplied as the number of elective surgeries provided at the DHB level varying by surgical procedures and year. We used the number of patients booked for surgery as a measure of demand. In the case of New Zealand, because of the threshold rationing mechanism, using only the number of patients added to the booking system as a measure of demand would lead to misspecification. Because decisions to admit are subject to supply constraints we also included patients placed in the 'active review' status in our measure of demand.

Although some previous studies estimate elasticities with separate measures of demand and supply, our methodology is different, as we employed a 'triangular model' of waiting time, supply and demand to estimate three separate equations. To our knowledge, our study is the first to estimate supply and demand elasticities within the context of the New Zealand elective surgery market employing a novel empirical strategy. Although we do not impose a 'market clearing' condition in our main analysis, as an additional robustness check, we also present the results for supply and demand elasticities imposing the equilibrium condition of quantity supplied (Q^S) equals quantity demanded (Q^D). Finally, our results are robust to alternate model specifications and variable selections, and our model selections are justified by both economic theory and econometric diagnostics tests. Finally, we also explore the effects of PHI and medical doctor availability on supply and demand for elective surgery, which has been less studied in the extant literature. Therefore, our paper makes important contributions to the study of elective surgery healthcare markets.

2.4 Data and Variables

The dataset is based on variables derived from the National Booking Reporting System (NBRS)⁴ data provided by the NZ MOH. The NBRS extract provided to us contains information on all patients booked for elective surgery along with other procedure-specific and individual patient

⁴Data source: https://www.health.govt.nz/system/files/documents/publications/nbrs_data _dictionary_v4.4.pdf.

specific characteristics. The original extract contained data from 2003–2016. However, we were informed that the data from before 2010 and the data for 2016 were incomplete. Moreover, the number of DHBs has been reduced from 21 to 20 since 2010. Hence, for the purpose of data consistency, we restricted our analysis to a study period of five years (2011–2015). We also made use of NZ Health survey data (2011–2015)⁵ for a measure of GP access due to lack of transport and percentage of people within regions with PHI. Finally, we incorporated NZ medical workforce reports (2011–2015)⁶ for a measure of doctors per 100,000 residents across DHBs and mortality collection data⁷ to measure DHB-specific mortality rates.

The unit of observation in our data is 20 DHBs. It has been argued that supply equations are best estimated at the 'provider' level (Jones, 2012). As we established earlier, the notion of 'provider' in the case of New Zealand is more appropriate for the DHBs as patients can only seek treatment within their DHBs. Also, the demand estimates are best estimated at the individual patient level, including the responses of all potential entrants into the system. Due to data constraints, such research is not yet feasible. For the purpose of our research, we defined the DHBs as the relevant 'catchment area', which we would argue is a harmless assumption because the patients residing in a geographic area are usually treated within their own respective DHBs. Previous studies have used postal codes/electoral wards/hospitals as their unit of observation. However, as mentioned by Gravelle, Smith, and Xavier (2003), catchment populations for small area/hospital levels are endogenous, as patients can seek treatment outside their residence/hospital. Hence, supply and demand variables can be mis-specified when estimated at the small area level as done in the previous studies. Previous models have used spatial interaction methods to derive supply and demand shifters at the small area level; however, we do not pursue such an exercise and restrict our analysis to the DHB level. It can be argued that demand responses can be influenced by supply consideration acting at the area level. However, in our demand measure, we also include patients

⁵Data source : https://www.health.govt.nz/nz-health-statistics/national-collections-and -surveys/surveys/new-zealand-health-survey.

⁶Data source: https://www.mcnz.org.nz/about-us/what-we-do/workforce-survey/.

⁷Data source : https://www.health.govt.nz/system/files/documents/publications/mortality _data_dictionary_v1.7.pdf.

placed in an 'active review' status who are essentially patients requiring elective surgery but have been assigned low priority due to supply considerations and are not booked for elective surgery. Hence, we would argue that demand responses are accurately specified in our model.

There was a total of 189 unique surgical procedures within 23 broad surgical categories during the study period for which publicly funded elective surgeries were provided. For the purpose of our study, we used only 15 procedures across six major surgical specialities. The broad surgical categories included general surgery, orthopaedic surgery, gynaecology, urology, otorhinolaryngology (ENT) and ophthalmology. Surgical procedures⁸ for which comprehensive data were available were selected and represented a diverse selection of elective surgical procedures. We chose six major surgical categories because comprehensive data across all DHBs were available for these procedures. Many procedures are not provided by all DHBs across all years and there are several procedures specified as 'Other' which cannot be identified. Moreover, waiting time for our chosen surgical categories and for procedures such as hip replacement, knee replacement, prostatectomy, and cataract surgeries have been extensively studied by many prior researchers such as Riganti et al. (2017), Martin and Smith (1999), and S. Martin et al. (2007). Therefore, we use 15 procedures across six major surgical categories.

2.4.1 Key Variables

After cleaning the data by removing duplicate entries and records with unusually high or missing waiting times, the final data contained a total of 283,897 entry events, i.e. our measure of demand, and 265,221 exit events, i.e. our measure of supply. For each surgical procedure across each DHB for each particular year, we calculated the demand measure as the sum of the total number of certainty approval given for surgery and the number of patients who were in an 'active review' status. The supply measure was the total number of publicly funded treatments provided for each

⁸The surgical procedures include cholecystectomy, rectal prolapse/anorectal surgery, hernia repair, haemorrhoidectomy, breast surgery, myringotomy, hysterectomy, cataract procedures, ankle/foot procedures, shoulder/arm/wrist procedures, hand procedures, carpal tunnel release, prostatectomy, hip replacement and knee replacement.

unique combination of surgical category, year and region. The waiting time was calculated for each patient as the difference between the treatment date and the date certainty was given, i.e. the date on which the patients were advised that they would receive publicly funded treatment. We calculated the waiting time measure as the average of a patient's waiting time for each surgical category, by region and by time. There are several ways to derive the waiting time variables, and different studies have used various approaches. None are expected to be a precise measure of waiting time, however, as indicated by S. Martin et al. (2007)), the average waiting time in our specification is "merely a proxy for the waiting time signal available to consultants and patients that will inform patients' decisions about whether to stay in the health system or to seek treatment in the private sector" (S. Martin et al., 2007). Hence, we exclude waiting time above one year (365 days). Higher waiting times are more likely to reflect data errors or special circumstances. In the initial records, the highest reported waiting time was 4,486 days and total records with waiting time higher than 1 year (365 days) were approximately 1.03 percent of the total records.

Excess demand given by '*Difference*' is the difference between quantity demanded and quantity supplied for 100,000 residents in a DHB for each procedure. Excess demand can be negative when for a particular year the total elective surgery supplied is greater than total elective surgery demanded. This is plausible because there are many elective surgeries procedures performed in the current year which were booked in the previous year. We also use quality measures for DHBs that varies across each DHB and year, but not by procedure. The variable '*AccessGP*' is a measure of the proportion of people within each DHB who lack access to GP services due to lack of transport. The existing literature suggests that transportation barriers have a considerable impact on access to healthcare and lead to poorer management of chronic illness and thus poorer health outcomes (Syed et al., 2013). DHBs are responsible for funding primary healthcare organisations (PHOs) and are hence responsible for funding community based general practices. Therefore, '*AccessGP*' is a proxy variable for both supplier operational efficiency and ease of access for surgery. '*AccessGP*' can also affect demand, as easier access to primary care services might lead to more referrals (Gravelle, Smith, & Xavier, 2003), thereby increasing demand. Thus, '*AccessGP*' appear in all

waiting time, demand and supply equations.

The variable '*MedicalDoctors*' is a measure of the proportion of medical doctors⁹ per 100,000 residents across each DHB for each year. Although, the proportion of doctors might also represent supplier capacity, empirical evidence suggests that medical doctors working in the public and private sectors act as substitutes for elective surgery (Atella & Deb, 2008). Furthermore, we also include the percentage of people with PHI as a proxy variable to represent the availability of private elective surgery. '*MedicalDoctors*' and '*PHI*' both represent complementary/competing services to public hospitals and hence appear in the demand equation. Moreover, these variables also appear in the supply and waiting time equation, as the availability of substitutes might induce public providers to seek private services for urgent and diagnostic elective procedures, thereby impacting supply for elective surgery and waiting times. The effect of PHI prevalence and the regional proportion of medical doctors on supply and demand for elective surgery have been less studied in previous research related to this topic.

The definition of the variables along with their descriptive statistics are summarized in Table 2.1:

Variable Name	Definition	Mean	S.D.	Min.	Max.
Quantity demanded (Q^D)	Number of people seeking elective surgery per 100,000 res-	99.96	97.63	0.60	750
	idents				
Quantity supplied (Q^S)	Number of people receiving elective surgery per 100,000	93.74	91.08	0.40	673.990
	residents				
Difference((Q^D) - (Q^S))	Difference in quantity demanded and quantity supplied per	6.23	24.90	-135.22	259.634
	100,000 residents.				
MeanWait	Average waiting time (Days)	91.56	41.03	14.80	319.90
MedianWait	Median waiting time (Days)	82.66	42.27	13	334
PHI	Percentages of people with PHI	30.28	8.38	13.90	49.90
MaoriProportion	Proportion of residents who identify as Maori per 100,000	17289.31	9759.01	6908.64	45089.69
	residents				
MedicalDoctors	Doctors per 100,000 residents.	256.27	83.32	140	566
MortalityRate	Mortality Rate across regions per 1000 people	7.68	1.48	4.91	10.81
AccessGP	Percentage of people without access to medical services	3.30	1.75	0.40	9.70
	due to lack of transport				
PublicBeds	Public hospital beds per 100,000 residents across regions.	271.45	47.69	181.20	377.79

Table 2.1: Descriptive Statistics	, Unit of Observation	- 20 District Health	Boards (DHBs)
-----------------------------------	-----------------------	----------------------	---------------

⁹These do not include general practitioners (GPs).

In Figure 2.1, we plot waiting time against the 'Difference' between quantity demanded and quantity supplied across DHBs for each year: 2011 and 2015. We chose hip replacement in the graph as it was the most frequently performed elective surgery for which balanced data were available. Figure 3 shows that the average waiting time has declined over the years. Waiting time is considerably lower for larger DHBs¹⁰ such as Auckland and Counties Manukau; whereas smaller and more deprived areas represented by DHBs, such as Whanganui and Wairarapa, have longer waiting times – more importantly, these cross-DHB differences were persistent over time.



Figure 2.1: Waiting time vs. difference (quantity demanded - quantity supplied) across DHBs for hip replacement procedure in 2011 and 2015.

2.4.2 Instrumental Variables

Our triangular model consists of three structural equations. Both demand and supply equations are functions of the waiting time, and the waiting time is affected by the excess demand. Identification of each structural equation requires exclusion of at least one exogenous variable from the equation under consideration. Therefore, we need one unique variable that appears in each equation only so justify the exclusion restrictions. In our set up, the regional mortality rate, supply of public beds, and the population share of Maori, respectively, enter only the demand, supply, and the waiting

¹⁰DHB codes labelled in the graph :https://www.health.govt.nz/nz-health-statistics/ data-references/code-tables/common-code-tables/district-health-board-code-table.

time equation, making the structural parameters of each equation over-identified. The justification of the choice of these instrumental variables are provided in the subsequent paragraph of this section.

(i) Regional Mortality Rates

The demand for elective surgery is mainly determined by the demographic and socioeconomic characteristics of the regional population. Although it is difficult to accurately construct an index for regional socioeconomic status, epidemiological research generally use income, education, occupational class etc. as indicators of socioeconomic status (Galobardes et al., 2006). Existing studies have consistently shown that mortality rate increases as socioeconomic status decreases (Bond Huie et al., 2003). The factors that explain socioeconomic status and mortality include overall poorer health, increased number of comorbid conditions, and lack of access to or underuse of healthcare services (Mackenbach et al., 2008). In our equations, we capture regional access to healthcare by the '*AccessGP*' variable and we use mortality rate across regions standardized by resident population as an exogenous variable appearing only in the demand equation which is a proxy for regional comorbidity/health levels.

Although conditions related to expected demand might influence the government to decide on regional healthcare budget and capacity, these considerations are unlikely to be concerned mainly with elective surgeries. As elective surgeries are non-life-threatening conditions that do not require immediate treatment, budget decisions for elective surgeries are unlikely to be made based on current/previous year's mortality rates. The likelihood of a patient dying while waiting for elective surgery is going to be minimal and in cases where the patient's issues are complicated requiring immediate treatment, these cases are usually treated as emergencies. Therefore, there should be no direct link between mortality rates and waiting times. It can be argued that mortality rates reflect quality of healthcare across regions. Nevertheless, a systematic review of the empirical literature shows very little correlation between quality of care and standardized mortality rates (Lilford & Pronovost, 2010). Multiple empirical studies argue that hospital admissions with preventable deaths are significantly low and differences in quality at an institutional level cannot be supported as quality of care within hospitals are much greater than differences between hospitals (Lilford & Pronovost, 2010). All in all, existing studies are in consensus that utilization of health care is mainly associated with health status of the population and the influence of "predisposing variables (age, sex, and education level) while the effect of enabling variables (income, satisfaction with care and accessibility)" is unclear(Fernández-Olano et al., 2006). As an additional exercise, we also used 'arthritis prevalence' instead of mortality rates as an exogenous demand shifter representing regional health status/comorbidity levels. Although, the relevance of this variable as an instrument for the waiting time in the supply equation was reduced, it is also safely excluded from both the supply and demand equations on a priori grounds. This additional exercise, however, makes little difference to our main results.

(ii) Public Bed Supply

The supply equation is based on the theoretical premise devised by Martin and Smith (1999) in which hospital managers seek to maximize a utility function which depends on waiting time, the efficiency with which resources devoted to elective surgery are used, and resources devoted to non-surgical activity, subject to an overall resource constraint. The solution to the utility maximization problem results in a supply function for elective surgery which is increasing in the amount of total resources available and waiting time (S. Martin & Smith, 1999).

Measurement of the magnitude of DHB level supply of resources is usually constrained due to data availability. Nevertheless, we follow the existing literature on the topic (S. Martin & Smith (1999); Stavrunova & Yerokhin (2011); R. M. Martin et al. (2003); Riganti et al. (2017)) and use the proportion of public beds per 100,000 resident population as a proxy for the budget constraint in the supply function. The *PublicBeds* variable is expected to capture

the effect of public hospital budget constraint on the supply of surgery. Because hospital beds are a binding constraint for publicly funded elective surgery, that is a valid assumption (S. Martin & Smith, 1999). DHBs usually operate within fixed budgets for elective surgery with acute hospital services assigned the first call on hospital resources. DHB managers must consider capacity while aiming to maintain waiting times advised by the MOH. When waiting times targets cannot be met due to capacity/budget constraints, this would result in increased waiting times for patients which is captured in our modelling. Although in our main analysis, we use public bed supply as an exogenous supply shifter, as an additional robustness check, we use the yearly DHB budget variable (standardised by resident population) instead of the public bed supply as an explanatory variable representing supplier capacity in the supply equation. The budget variable is a valid instrument for the waiting time in the demand equation and can be safely excluded from the waiting time equation. The result from this additional exercise are remarkably similar to our main analysis.

(iii) Proportion of Maori

In our model the waiting time is dependent only on the difference between demand and supply of surgery and other exogenous factors affecting waiting time. The variable that enters only the waiting time equation is the proportion of Maori people¹¹ in each region. In a study by Aumua et al. (2018) for paediatric surgical procedures, the authors find that the majority of patients with higher waiting times were Maori and that higher waiting times were not explained by socioeconomic factors affecting health status. Other things being held constant, being a Maori person does not affect the demand for elective surgery nor supply of it, but still can affect waiting time when he/she is discriminated against in the health market. In a systematic review of multiple studies across two decades by Graham and Masters-Awatere (2020), the authors conclude that Maori patients find current health system as being discriminatory. In chapter 1, we also find that Maori are discriminated against and have higher waiting times compared to NZE patients across various surgical

¹¹Due to data limitations, the Maori proportion variable varies only across DHBs.

procedures. However, these effects are even stronger for Asian and Pacific patient groups. Therefore, as an additional robustness test, we also estimate the impact of waiting time on supply and demand for elective surgery using the proportion of Pacific and Asian ethnicities across DHBs as separate instruments of waiting time. As is subsequently shown, our results are consistent across these alternate specifications and indicate that the proportion of Maori, Asian, Pacific ethnicities are all valid instruments for waiting time.

In many healthcare systems across western European countries, and in New Zealand, GPs act as 'gatekeepers' to more specialised care. It can be argued that proportion of ethnic minorities in a region can affect regional demand. A study in Netherlands (Uiters et al. (2006)) with similar healthcare conditions as in New Zealand, found no evidence of the gate keeping role of GPs on healthcare utilisation across ethnic minority groups. We would argue that in cases where regional demand may be affected due to proportion of ethnic minorities, it is mainly due to financial and other barriers to healthcare access, as such the '*AccessGP*' variable in our modelling adequately captures this effect.

Making waiting time endogenous together with supply and demand introduces significant challenges to the selection of appropriate instruments. This is because supply, demand, and waiting time within a healthcare setting are intricately linked. On the demand side, it would have been ideal to find a variable that result only in elective surgery. However, such variable has not been yet suggested in the literature. On the supply side, it would have been ideal to use a capacity-based instrument for elective surgery, such as shocks to physician staffing that impacts only elective surgeries. However, such instruments are not available and consistent with existing literature in the topic we use public beds as a proxy for budget constraints in the supply equation. Therefore, the results presented in this analysis should be interpreted with that caveat in mind.

2.5 Empirical Strategy

The studies related to waiting times for elective surgery include but are not limited to (Cullis & Jones (1986); Goddard et al. (1995); Gravelle, Smith, & Xavier (2003); Lindsay & Feigenbaum (1984); R. M. Martin et al. (2003); S. Martin & Smith (1999); S. Martin et al. (2007)). Our paper both draws influences from all of these sources and extends previous work by adopting a more flexible model for the regional health market.

Our 'triangular' model was built on the observation that different regions experience different waiting times (excess demand for health services) and that cross-regional differences are not a short-term phenomenon. Viewing the cross-regional differences in waiting times as equilibrium differentials, we obtained separate measures of demand and supply and constructed a system of three structural equations. More precisely, three endogenous variables were employed: the demand for surgery, the supply of surgery, and waiting time. Imposing a 'market clearing' condition would reduce the number of endogenous variables to two, where demand for surgery and supply of surgery would be equal at a given waiting time. Below, we specify the three structural equations of supply, demand and waiting time focusing on the exclusion principle.

2.5.1 Waiting time equation

We model waiting time as:

$$W_{irt} = \alpha_0 + \alpha_1 D_{irt} + \alpha_2 Z_r^W + \alpha_2 Y_{rt} + h_i + h_t + \mu_{irt}, \qquad (2.1)$$

where, subscript '*i*' is the surgical procedure, subscript '*r*' is the administrative area (DHB), and subscript '*t*' represents the year. The dependent variable '*W*' is the waiting time (calculated for each combination of surgical procedure (*i*), DHB(*r*), and year(*t*). '*D*' is the quantity difference between measure of demand (Q^D) and measure of supply (Q^S) for elective surgeries per 100,000 residents in each DHB. We hypothesised that waiting time is a function of excess demand for elective surgery, with greater demand increasing the waiting time. As excess demand approaches zero, the waiting time will depend on exogenous variables only, making the model a conventional two-equation system for demand and supply. Put differently, the traditional 'market-clearing' model can be regarded as the limiting case of our 'triangular model'. For the purpose of checking the robustness of the results, however, we also tested the conventional model in the subsequent section and found a similar result in a qualitative sense.

 Z^{W} is the exogenous variable appearing only in the waiting time equation that impacts only the waiting time and does not affect the supply and demand for elective surgery. This exogenous variable appearing only in the waiting time equation is '*MaoriProportion*', which is a measure of the proportion of people with Maori ethnicity groups within each DHB.

'Y' includes variables that appear in all waiting time, supply, and demand equations. These variables are '*PHI*', '*MedicalDoctors*', and '*AccessGP*'. h_i are surgical procedure dummies. h_t are year dummies. μ_{irt} is the error term.

All variables are log transformed and should be interpreted as elasticities.

2.5.2 Supply equation

It can be expected that increased waiting time would induce the DHBs to increase their supply in the form of capacity increases, which can be represented by the size of the medical workforce and the provision of hospital beds available for elective surgery. We model supply of elective surgery as being subject to capacity and budget constraints as follows

$$Q_{irt}^{s} = \beta_0 + \beta_1 W_{irt} + \beta_2 Z_{rt}^{s} + \beta_2 Y_{rt} + h_i + h_t + \mu_{irt}, \qquad (2.2)$$

where, the dependent variable, Q_{irt}^{S} , is the number of elective surgeries provided at each DHB for each procedure and year, and 'W' is the associated waiting time. Z^{S} represents an exogenous supply shifter, which is the supplier capacity given by '*PublicBeds*'. μ_{irt} is the error term.

All variables are log transformed and should be interpreted as elasticities.

2.5.3 Demand equation

It can be expected that increased waiting time would lead to reduced demand, as the utility of elective surgery for patients would decrease the longer, they had to wait for the procedure. The expected reduction in demand can be caused by people switching to alternative forms of treatment or seeking private care. It can also be argued that the provider's service quality would increase the demand for surgery and that additionally, the demand would be higher in regions with higher regional vulnerability and underlying regional population comorbidities.

The subsequent empirical specification can be written as:

$$Q_{irt}^{D} = \gamma_{0} + \gamma_{1} W_{irt} + \gamma_{2} Z_{rt}^{D} + \gamma_{2} Y_{rt} + h_{i} + h_{t} + \mu_{irt}, \qquad (2.3)$$

where, the dependent variable, Q_{irt}^D , is the number of patients seeking elective surgery (calculated for each combination of procedure, time, and DHB), and 'W' is the associated waiting time. Z^D represents an exogenous demand shifter, which is the measure of regional vulnerability and population comorbidity levels given by '*MortalityRate*'. μ_{irt} is the error term.

All variables are log transformed so that the coefficients can be interpreted as elasticities.

Lastly, each equation is estimated by the standard two-stage-least-squares (2SLS hereafter).¹²

¹²In principle, in the first-stage prediction of waiting time, it must be considered that the waiting time is rightcensored and that the NZ Ministry of Health (MOH) requires elective surgery to be performed within a specified time frame (six months during the study period). This calls for a Tobit model. However, we did not pursue this line of inquiry for various reasons. Most importantly, a non-linear prediction of the waiting time makes the correlation between instrumental variables and the actual waiting time arbitrary based on the selected functional form (normal

2.6 Empirical Results

Table 2.2 presents our main results. It is important to note that all variables are log transformed so that the coefficients can be interpreted as elasticities. Hence, the dependent variable throughout the results presented in this analysis are natural logs of quantity demanded, quantity supplied, and waiting time. While we followed convention by reporting the results from both OLS and 2SLS methods, for each equation, a Hausman test strongly rejects the OLS in favour of 2SLS at any conventional significance level. In the 2SLS estimation, we performed several diagnostics tests¹³ to check the validity of the instruments and to test for over-identifying restrictions. The results for the diagnostics tests across multiple model specifications are presented in Appendix A in Table A.2, Table A.3, and Table A.4. We did not find any evidence of misspecification, and the instruments used were strongly correlated with the endogenous regressors.¹⁴

distribution). In addition, there were many cases across all DHBs in which waiting times have significantly exceeded the recommended maximum waiting time period, and we incorporated these cases into our analysis.

¹³We performed weak instruments, Wu- Hausman and Sargan-tests to check the validity of the instruments, to justify the use of the 2SLS estimator, and to test for over-identifying restrictions.

¹⁴Excess demand is also endogenous in the waiting time equation.

	OLS Wait	2SLS Wait	OLS Demand	2SLS Demand	OLS Supply	2SLS Supply
(Intercept)	3.832***	1.620	2.739***	4.967***	1.601**	-7.862***
	(0.363)	(1.058)	(0.492)	(0.732)	(0.646)	(2.130)
Difference $(Q^D - Q^3)$	0.003***	0.038***				
	(0.000)	(0.009)				
log(MeanWait)			0.165***	-0.478**	0 195***	1 477***
log(wealt wait)			(0.041)	(0.152)	(0.045)	(0.272)
			(0.041)	(0.152)	(0.043)	(0.273)
log(MaoriProportion)	0.057**	0.118*				
	(0.024)	(0.061)				
log(MortalityRate)			0.706***	0.898***		
			(0.091)	(0.107)		
log(PublicBeds)					0.728***	1.385***
					(0.093)	(0.179)
log(PHI)	-0.028	0.120	-0.199***	-0.194***	-0.399***	-0.161*
108(1111)	(0.042)	(0.112)	(0.060)	(0.065)	(0.056)	(0.086)
	(0.012)	(0.112)	(0.000)	(0.005)	(0.050)	(0.000)
log(MedicalDoctors)	0.030	0.186**	0.044	0.084	-0.158***	-0.298***
	(0.033)	(0.091)	(0.054)	(0.060)	(0.059)	(0.079)
log(AccessGP)	-0.011	0.003	-0.145^{***}	-0.131***	-0.179***	-0.155***
	(0.016)	(0.039)	(0.024)	(0.026)	(0.025)	(0.031)
Num. obs.	1463	1463	1463	1463	1463	1463
First-Stage F Stat.		10.160		67.399		32.104

Table 2.2: Estimation of 'Triangular Model' with mean waiting times

 $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$, Standard Errors are reported in the parenthesis. Results for year and procedure dummies are omitted.

Focusing on the 2SLS results, it can be concluded that while demand for elective surgery is unresponsive to waiting time, supply is positively responsive to waiting time. The coefficient estimates of (-0.48) in the demand equation suggest that 1 percent increase in waiting time is associated with 0.48 percent reduction in demand, whereas the coefficient estimate of (1.48) for the supply equation implies that 1 percent increase in waiting time is associated with 1.48 percent increase in supply. The estimated coefficient of waiting time (1.48) reveals that the supply of elective surgery is highly elastic with respect to waiting time and is also statistically significant. The estimated coefficient of waiting time in the demand equation in the fourth column (-0.48) demonstrates that the demand for elective surgery is inelastic with respect to waiting time. With regard to the waiting time equation in the second column, the excess demand for elective surgery ($Q^D - Q^S$) is positively associated with waiting time, and the estimated coefficient is also statistically significant, even at the 1 percent level.

The exogenous demand shifter given by '*MortalityRate*' has a positive effect on demand for public surgery, suggesting that vulnerable populations with poorer health levels have a higher demand for elective surgery. The exogenous supply shifter '*PublicBeds*' has an expected positive impact on supply, implying that greater capacity leads to more elective surgery services. The proportion of Maori people entered the waiting time equation positively and significantly, indicating that regions with a higher proportion of Maori people experience longer waiting times.

For the additional control variables that appear in all three equations, we found that a lack of access to GPs ('AccessGP') which represents both ease of access to surgery and supplier efficiency, reduces both demand and supply for elective surgery. Lack of GP access might lead to lower referrals by GPs, thus reducing demand for elective surgery in the public sector. Furthermore, lower supplier operational efficiency leads to a reduction in supply for public surgery. The percentage of people with PHI reduces both supply and demand for elective surgery, suggesting that the availability of private healthcare services can act as both complementary/competing services to elective surgery in the public sector. A possible explanation for this was provided by (Derrett et al. (1999)) who argued that private services tend to fill the gap for public surgery when public provisions are poor, and that good private provisions decrease motivation to develop the public sector. The estimated coefficients on the number of non- GP medical doctors per 100,000 residents 'MedicalDoctors' yielded interesting results. The increased number of medical doctors across DHBs reduced the supply of elective surgery, suggesting that surgeons act as substitutes for public surgery confirming the findings of Atella and Deb (2008) that public and private specialists are substitute sources of medical care. Atella and Deb (2008) argue that when patients derive less utility from the public sector, they may switch to the private sector and this may reduce their willingness to pay taxes for the public sector simultaneously resulting in poorer quality outcomes. Alternately, the increased proportion of doctors causing a reduced supply for surgery, which might be attributable to the fact that most doctors working in the public sector also work for the private sector simultaneously. Hence, there is a possible conflict of interest, and it might be in doctors'

best interest to limit provisions for public surgery (Derrett et al., 1999).¹⁵

The fact that demand is inelastic with respect to waiting time indicates that an increased investment in supply resources in the public healthcare system is likely to reduce waiting times and thus improve access to public surgery in New Zealand. Furthermore, elastic supply suggests that an exogenous increase in demand for elective surgery is unlikely to significantly increase waiting times for elective surgery. The positive coefficient on the '*PublicBeds*' variable and the negative coefficient on the '*MedicalDoctors*' variable suggest that increased investment in increasing medical capacity and infrastructure might improve the supply of elective surgery as compared to increased investments in the medical workforce. The higher prevalence/availability of competing/complementary services reduces both demand and supply by a similar extent,¹⁶, while not affecting waiting times significantly, demonstrating that the prevalence of private health services does not significantly improve outcomes for public patients. It may indeed work to redirect resources from the public sector to the private sector.

2.7 Robustness tests

In this section we discuss various robustness tests to confirm our main result that the supply of elective surgery is elastic with respect to waiting time, whereas demand is not. The description for additional specifications and their results are presented below.

¹⁵The dynamics and the subsequent impact of the private health sector and the proportion of medical doctors are interesting areas for future research.

¹⁶A t-test cannot reject the null hypothesis of equality of the coefficients on log(PHI) between the demand supply equations even at the 10 percent level.

(i) Conventional Demand and Supply Model

Following the majority of studies on this topic, we estimated a 'market clearing' model under the assumption that quantity demanded (Q^D) equals quantity supplied (Q^S) for a given waiting time. Most structural coefficients were consistent with our predictions which are presented in Table 2.3. Compared to the results from our preferred model in Table 2.2, the estimated supply elasticity remained virtually identical even in the restrictive model. While the estimated demand elasticity increased slightly (in absolute value, -0.79), the difference between the two estimated elasticities was still not only statistically significant but also empirically important. Consequently, our main result remains valid after this new exercise, at least in a qualitative sense: The supply of elective surgery is more responsive to waiting time, relative to demand.

	2SLS Demand	2SLS Supply			
(Intercept)	5.681***	-7.862***			
	(0.814)	(2.130)			
log(MeanWait)	-0.789^{***}	1.478***			
	(0.169)	(0.273)			
log(MortalityPate)	1 026***				
log(wortantyRate)	(0.110)				
	(0.119)				
log(PublicBeds)		1.385***			
		(0.179)			
	0 22 4***	0.161*			
log(PHI)	-0.234	-0.101			
	(0.072)	(0.086)			
log(MedicalDoctors)	0.156**	-0.298***			
	(0.066)	(0.079)			
1(ACD)	0 121***	0 155***			
log(AcessGP)	-0.131	-0.155			
	(0.029)	(0.031)			
Num, obs.	1463	1463			
First-Stage F Stat.	134.69	64.06			
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard Errors are					

Table 2.3:	Estimation	of c	conventional	demand	and	suppl	y mod	el

reported in the parenthesis. Results for Year and Procedure

ummies are omitted

(ii) Alternate Instrumental Variables

In order to confirm the validity of our main results, we also experimented with alternate instruments. Due to data constraints, it is often difficult to find valid instruments while estimating supply and demand functions within the healthcare market. Nevertheless, we use arthritis prevalence in the regional population as a proxy for regional comorbidity/ health status instead of the 'MortalityRate' variable. The 'Arthritis' variable enters only in the demand equation. Likewise, we measure and use DHB annual budgets standardized by resident population instead of 'PublicBeds' as a proxy for supplier capacity. 'Budgetpop' enters only in the supply equation. These alternate instruments can be justified on similar a priori grounds as our original instrumental variables described in Section 4. We use the same data as in our main analysis to estimate the structural parameters, however, we have omitted data from 2015 as budget information for DHBs was not available during the data collection stages. Although arthritis prevalence was significantly correlated with waiting time, the strength of the instrument was reduced as evident by a low First-Stage F Stat. (4.34). Across all equations the Hausman Specification tests strongly reject OLS in favour of 2SLS and Sargan tests indicate that the over-identifying restrictions are valid. Again, our results are similar to our main findings and confirm that while supply for healthcare is elastic (2.91), the demand is not (-0.28). The results are presented below in Table 2.4.

	2SLS Wait	2SLS Demand	2SLS Supply
(Intercept)	1.204	6.472***	-6.173
	(1.385)	(0.866)	(4.947)
diff	0.041***		
	(0.011)		
log(MeanWait)		-0.290^{*}	2.918***
		(0.175)	(1.119)
log(MaoriProportion)	0.121		
	(0.076)		
log(Arthritis)		0.286***	
		(0.053)	
log(BudgetPop)			2.497***
			(0.792)
log(PHI)	0.145	-0.379***	-0.164
<u>č</u>	(0.127)	(0.059)	(0.159)
log(MedicalDoctors)	0.240^{*}	0.003	-0.949***
Ū.	(0.125)	(0.065)	(0.315)
	. ,		
log(AccessGP)	0.010	-0.142***	-0.139***
	(0.045)	(0.026)	(0.052)
Num. obs.	1171	1171	1171
First Stage F Stat.	6.71	46.98	4.35
**** = 1 0 01 *** = 1 0 05 *= 1 0	10 Chardend Emer		

Table 2.4: Estimation of 'Triangular Model' with alternate Instrumental Variables

***p < 0.01, **p < 0.05, *p < 0.10. Standard Errors are reported in the parenthesis. Results for Year and Procedure dummies are omitted

(iii) Alternate waiting time measure (Median waiting times)

It can be argued that mean waiting time is generally sensitive to outliers; hence, we performed the analysis again using median waiting times. We reverted to our preferred model for this purpose. The results reported in Table 2.5 are virtually identical to those reported in Table 2.2 in a qualitative sense.

	2SLS Wait	2SLS Demand	2SLS Supply			
(Intercept)	1.663	5.051***	-8.479***			
	(1.056)	(0.774)	(2.518)			
Difference $(Q^D - Q^S)$	0.038***					
	(0.008)					
log(MedianWait)		-0.476***	1.530***			
		(0.156)	(0.318)			
log(MaoriProportion)	0.123**					
	(0.061)					
log(MortalityRate)		0.889***				
		(0.109)				
log(PublicBeds)			1.422***			
			(0.208)			
log(DIII)	0.162	0 190***	0 206**			
log(F111)	(0.112)	-0.180	-0.200			
	(0.112)	(0.067)	(0.091)			
log(MedicalDoctors)	0 104	0.044	-0.180**			
log(inedicale octors)	(0.091)	(0.061)	(0.083)			
	(0.091)	(0.001)	(0.005)			
log(AccessGP)	-0.003	-0.134***	-0.147***			
Č.	(0.039)	(0.027)	(0.036)			
Num. obs.	1463	1463	1463			
First-Stage F Stat.	10.16	45.326	19.871			
$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.05$	*** $p < 0.01, ** p < 0.05, * p < 0.10$. Standard Errors are					

Table 2.5: Estimation of 'Triangular Model' with median waiting times.

reported in the parenthesis. Results for Year and Procedure

dummies are omitted

(iv) Percent Asian and Percent Pacific as separate instruments of waiting time

In chapter 1, we find that Asian and Pacific patients have significantly higher waiting times for elective surgeries compared to NZE and Maori, so we test our main results using proportion of Asian and Pacific ethnicities across DHBs as separate instruments of waiting time. Our results from this additional exercise are remarkably similar to our main results.

	OLS Wait	2SLS Wait	OLS Demand	2SLS Demand	OLS Supply	2SLS Supply
(Intercept)	4.71***	3.38***	2.74***	4.95***	1.60^{*}	-7.95***
· •	(0.18)	(0.56)	(0.49)	(0.73)	(0.65)	(2.13)
Difference($O^D - O^S$)	0.00***	0.04***				
	(0.00)	(0.01)				
	(0000)	(010-)				
log(MeanWait)			0.17***	-0.47**	0.20***	1 49***
log(lifeall (fait)			(0.04)	(0.15)	(0.05)	(0.27)
			(0.04)	(0.15)	(0.05)	(0.27)
log(PacificProportion)	_0.05***	_0 10**				
log(1 active 10portion)	-0.03	-0.10				
	(0.01)	(0.04)				
log(MortelityPeta)			0 71***	0.00***		
log(wortantyKate)			(0.00)	(0.11)		
			(0.09)	(0.11)		
$1 \cdot \cdot (\mathbf{D} \cdot 1 \cdot 1^{*} \cdot \mathbf{D} \cdot 1 \cdot 1)$					0 72***	1 20***
log(PublicBeds)					0.73	1.39***
					(0.09)	(0.18)
log(PHI)	-0.05	0.07	-0.20***	-0.19**	-0.40***	-0.16
	(0.03)	(0.09)	(0.06)	(0.06)	(0.06)	(0.09)
log(MedicalDoctors)	0.05	0.23*	0.04	0.08	-0.16**	-0.30***
	(0.03)	(0.10)	(0.05)	(0.06)	(0.06)	(0.08)
log(AccessGP)	0.03*	0.08	-0.15^{***}	-0.13***	-0.18^{***}	-0.16***
	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.03)
Num. obs.	1463	1463	1463	1463	1463	1463
First-Stage F Stat.		10.160		67.65		32.34

Table 2.6: Percent pacific as an instrument of waiting time

***p < 0.01, **p < 0.05, *p < 0.10, Standard Errors are reported in the paranthesis. Results for year and procedure dummies are omitted.
	OLS Wait	2SLS Wait	OLS Demand	2SLS Demand	OLS Supply	2SLS Supply
(Intercept)	4.51***	3.04***	2.74***	4.97***	1.60*	-7.81***
	(0.18)	(0.60)	(0.49)	(0.73)	(0.65)	(2.13)
		. ,		. ,		. ,
Difference($O^D - O^S$)	0.00***	0.04***				
(£ £)	(0.00)	(0.01)				
log(MeanWait)	(0.00)	(0.01)	0.17***	-0.48**	0 20***	1 47***
log(incul (ful))			(0.04)	(0.15)	(0.05)	(0.27)
			(0.04)	(0.15)	(0.05)	(0.27)
log(AsianProportion)	_0.00***	-0.08*				
log(Asiani loportion)	(0.02)	(0.05)				
	(0.02)	(0.05)				
log(MortalityPata)			0.71***	0.00***		
log(montainty Kate)			0.71	(0.11)		
			(0.09)	(0.11)		
log(DublioDada)					0 72***	1 20***
log(PublicBeds)					0.75	(0.18)
					(0.09)	(0.18)
le «(DIII)	0.06	0.12	0.20***	0 10**	0.40***	0.16
log(PHI)	0.06	0.13	-0.20	-0.19	-0.40	-0.16
	(0.04)	(0.12)	(0.06)	(0.06)	(0.06)	(0.09)
	0.00*	0.05*	0.04	0.00	0.1(**	0.20***
log(MedicalDoctors)	0.08	0.25	0.04	0.08	-0.16	-0.30***
	(0.03)	(0.10)	(0.05)	(0.06)	(0.06)	(0.08)
	0.02	0.05	0.15***	0.12***	0.10***	0.1/***
log(AccessGP)	0.03	0.05	-0.15	-0.13	-0.18	-0.16***
	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.03)
Num. obs.	1463	1463	1463	1463	1463	1463
First-Stage F Stat.		10.160		67.31		32.02

Table 2.7: Percent Asian as an instrument of waiting time

***p < 0.01, **p < 0.05, *p < 0.10, Standard Errors are reported in the paranthesis. Results for year and procedure dummies are omitted.

Our results in Table 2.6 and Table 2.7 are similar to our main findings and confirm that demand for elective surgeries are inelastic whereas supply is elastic. The econometric diagnostic tests also strongly indicate that both proportion Pacific and proportion Asian are valid instruments of waiting time as indicated by the F-statistic of over 10 in Table 2.6 and Table 2.7. The associated p-values for the Hausmann test and the Sargan test are presented in the Appendix Section in Table A.2, Table A.3, and Table A.4.

The results reported in Table 2.6 and Table 2.7 show that on average increased proportion of Pacific and Asian patients in an area lead to lower waiting times. Based on the findings on Chapter 1, we would expect the proportion of Pacific and Asian to be associated with longer waiting times. Although, this may appear inconsistent, it is consistent with the observation in Table 1.1, that waiting times are lower on average for Pacific and Asian patients. It is likely that Asian and Pacific patients may seek elective surgery only after their condition has significantly worsened. Hence, only after we control for patient severity by including CPAC scores, we find Asian and Pacific patients waiting longer for elective surgery. The difference in the sign of the coefficients for our Maori and Asian/Pacific variables could also be because observed relationships are different for within DHBs and across DHBs specifications. Moreover, these relationships may differ at individual and aggregate levels of analysis. As observed in Table 1.4, proportion of Maori exhibit a greater variation across DHBs whereas individuals from Pacific and Asian backgrounds are mainly concentrated across four DHBs around the Auckland and the Wellington region. Moreover, average waiting times are consistently higher for Maori across all DHBs. Hence, our choice of Maori variable as a valid source of exogenous variation of waiting time appears justified.

(v) Single surgical speciality (Orthopaedic surgery)

We also estimated our preferred model for procedures within a single surgical speciality, i.e. orthopaedic surgery. We chose orthopaedic surgery because good-quality data across all DHBs were available for orthopaedic surgery as compared to other specialities. The results from Table 2.8 confirm our main finding that the supply is very elastic, whereas the demand is inelastic¹⁷.

	2SLS Wait	2SLS Demand	2SLS Supply
(Intercept)	0.067	1.532	-17.378***
	(1.450)	(0.957)	(3.391)
Difference $(Q^D - Q^S)$	0.072***		
	(0.015)		
log(MeanWait)		-0.273*	2.104***
		(0.163)	(0.366)
log(MaoriProportion)	0.153*		
	(0.079)		
log(MortalityRate)		1.234***	
		(0.150)	
log(PublicBeds)			2.240***
			(0.321)
log(PHI)	0.304**	-0.330***	-0.256*
	(0.153)	(0.096)	(0.135)
log(MedicalDoctors)	0.287**	0.344***	-0.221*
	(0.124)	(0.085)	(0.130)
log(AccessGP)	0.055	-0.259***	-0.158***
	(0.057)	(0.037)	(0.054)
Num. obs.	1133	1133	1133
First-Stage F Stat.	11.727	80.731	28.102

Table 2.8: Estimation of 'Triangular Model' for a specific surgical speciality: Orthopaedic surgery

***p < 0.01, **p < 0.05, *p < 0.10. Standard Errors are reported in the parenthesis. Results for Year and Procedure dummies are omitted

¹⁷Despite some changes in the magnitudes of the estimated coefficients, the estimated signs remain the same between Table 2.2 and 2.8.

(vi) DHB Fixed Effects

Lastly, we also estimated our 'triangular model' with DHB fixed effects. To specify the supply, demand and waiting time equation, we aggregated the administrative data (2011–2014) for individual patients across DHBs at the quarterly level. We used 10 surgical procedures for which good-quality data were available. We employed regional, year and surgical procedure dummies to account for regional differences, policy changes and differences in clinical attitudes. Due to the regional fixed effects, the demand and supply shifters used in our main analysis appear as weak instruments, as regional variables do not vary significantly over time within DHBs. Hence, we use the quarterly expense reports of DHBs, and used total outsourced services expenses given by 'OutsourcedServicesExpenses' and the proportion of accident-related cases given by 'AccidentProportion' as exogenous demand shifters. For the supply equation we use the proportion of emergency cases given by 'EmergencyProportion' and total medical personnel expenses given by 'MedicalPersonnelExpenses' as exogenous supply shifters. The exogenous variable given by 'MaoriProportion' appearing only in the waiting time equation is the proportion of total surgical episodes with Maori patients. The supply and demand shifters and 'MaoriProportion' vary by DHB, year, and quarter but not by surgical procedures. We hypothesize that quarterly performance reviews affect the supply for elective surgery as they impact the utility function of supply managers who make supply decisions. The quarterly dummies also enter into the waiting time equation, but do not appear in the demand equation; the quarterly dummies are also one of the instruments used when estimating the demand equation.

The conclusion from the results outlined in Table 2.9 is that even when DHB fixed effects are controlled for, demand is inelastic while supply is elastic with respect to waiting time, reconfirming our main result. The increase in outsourced service expenses and proportion of accident cases both increase the demand for elective surgery. The explanation for this may be that increased expenses on outsourcing services are related to DHBs seeking private

	2SLS Wait	2SLS Demand	2SIS Supply
(Intercept)	3.836***	4.694***	-2.577
	(0.489)	(0.501)	(1.592)
Difference $(Q^D - Q^S)$	0.038**		
	(0.018)		
le r(Maar Wait)		0 220***	1 0 4 0 ***
log(wean wait)		-0.558	1.242
		(0.111)	(0.310)
MaoriProportion	2.240		
in a control of the c	(1.490)		
	(11.50)		
OutsourcedServicesExpenses		5.222***	
L.		(0.809)	
AccidentProportion		5.971***	
		(1.990)	
MedicalPersonnelExpenses			3.837**
			(1.876)
EmergencyProportion			-1.319**
			(0.654)
Num also	2125	2125	2125
INUIII. ODS.	5125	5125	5125
First-Stage F Stat.	1.345	27.880	11.956

Table 2.9: Estimation of 'Triangular model' with DHB fixed effects (20 DHBs)

***p < 0.01, **p < 0.05, *p < 0.10. Standard Errors are reported in the parenthesis. Results for Year, Procedure,

DHB dummies are omitted

providers for specialist assessment/relevant diagnosis for elective surgery, thus increasing the number of admissions to the booking system. Because everyone in New Zealand is covered for accidents by a 'no-fault scheme' by the Accident Compensation Corporation (ACC), the proportion of accidents can increase the demand for elective surgery as those procedures covered by the ACC tend to have shorter waiting times. In contrast, the proportion of emergency patients can decrease the supply of elective surgery, as emergency cases might consume medical resources intended for non-emergency elective surgery cases. Furthermore, increased medical personnel expenses representing supplier capacity can increase the supply for elective surgery. The proportion of Maori patients variable, which only appears in the waiting time equation, is associated with higher waiting times, although the coefficients are not significant (t-value 1.50, p-value 0.13). The reported estimates with DHB fixed effects pass the diagnostic tests for over-identifying restrictions and instrument relevance in the demand and supply equations; however, in the waiting time equation, the instrument's relevance cannot be established, as indicated by a low first-stage F-statistic. Furthermore, in some specifications with alternate procedures, our demand equation does not pass the over-identifying restrictions test. However, in alternate specifications, the signs and magnitudes of the coefficients remain similar. Hence, these results should be interpreted with that caveat in mind and with caution.

2.8 Conclusion

In this chapter, we estimated separate models for waiting time, supply and demand for elective surgeries. By adopting a 'triangular' model in which we estimated each equation separately, we found evidence that supply and demand for elective surgery does respond to waiting time signals in the public sector. Our results are consistent with economic theories and demonstrate a negative impact of waiting time on demand and a positive impact of waiting time on supply. In our main analysis, our demand estimate was -0.48, suggesting that demand is inelastic with respect to waiting time, whereas supply is elastic, with an elasticity coefficient of 1.48. An elasticity estimate below 1, as in the case of demand elasticity, suggests that demand is inelastic with respect to waiting times. This means supply side initiatives will have a positive impact on reducing waiting time for elective surgeries and the gains in reduced waiting times will not be offset by large increases in demand (Riganti et al., 2017). Moreover, the fact that supply is elastic suggests that an exogenous shock in demand will not significantly increase waiting time for elective surgeries.

Our demand estimates were higher than those of previous studies from the UK and Italy, which reported demand elasticities of -0.21. Estimates from Australia, however, yielded evidence of elastic demand. The explanation for the slightly higher demand elasticity reported in New Zealand could be attributable to the difference in the public–private mix, with New Zealand having a relatively higher proportion of people with private insurance as compared to the UK and Italy but a smaller proportion as compared to New South Wales in Australia. The difference could also be

due to differences in estimation procedures or because elasticities differ within equilibrium and disequilibrium.

In our paper, we made no assumption of equilibrium. Most studies in the UK, Australia and Italy have estimated demand and supply equations with a single measure of utilisation (typically supply) with an equilibrium assumption. Furthermore, we adopted a novel empirical specification. Finally, we also explored the dynamics related to ease of access to primary care, proportion of medical doctors and prevalence of PHI on waiting time, demand, and supply for elective surgery. Our results are robust to alternate specifications and consistent with economic theories.

Our results have important policy implications, indicating that an increase in supply for elective services will reduce waiting times and can thus potentially improve access to public surgery. Furthermore, waiting lists are likely to remain stable, and any exogenous increase in demand is unlikely to cause significant increases in waiting times. Supplier capacity is positively associated with supply for elective surgery. Therefore, increased investments by the MOH in surgical capacity might lead to better outcomes and a subsequent reduction in waiting times. Ease of access to primary healthcare services also negatively affects both demand and supply for elective surgery. Therefore, improved conditions for GP and primary healthcare access across more vulnerable populations must be an important consideration in government policy and agenda. Finally, we found that a higher proportion of medical doctors and the availability of private healthcare services act as both complementary and competing services to the public healthcare system; however, they might not necessarily improve waiting times for patients.

Finally, the usual caveats to empirical work must be applied when interpreting our results, which must be subject to the validation of the assumptions we made and of the degree our selected variables represent true supply and demand conditions. Although we incorporated the dynamics of private healthcare services into our model, there is definitely room to incorporate additional data. The impact of PHI and medical doctors on supply and demand for elective surgeries is an important topic of empirical work and deserves its own separate analysis. We leave such an extension to future work. Finally, the health economics literature pertaining to New Zealand is limited. We

therefore hope that this chapter helps to address the need for more careful econometric analysis of the New Zealand healthcare market and contributes to the wider healthcare literature, especially that relating to publicly funded elective surgeries.

Chapter 3

Demand forecasting and inventory control for pharmaceutical wholesalers - A numerical study

3.1 Introduction

Demand forecasting and inventory control are integral functions of business process management especially relating to supply chain processes. In these processes, there is an inherent trade-off between customer demand satisfaction and inventory costs. Customer demand can always be satisfied if a sufficiently large inventory is maintained. However, maintaining larger inventory entails problems such as tied up capital, inventory write-downs, and reduced profits. In this regard, inventory decisions based on accurate sales forecasts allow companies to not only increase their profits but also enables them to respond to complexities with greater flexibility. This is particularly relevant to the pharmaceutical industry where some medicines have a short shelf life and maintaining higher product quality is paramount because it directly impacts public health and pharmaceutical wholesalers must consistently maintain a high customer service level (CSL), close to 100 percent, while simultaneously limiting inventory costs (Uthayakumar & Priyan, 2013). This means they usually operate on relatively thin margins.

A typical configuration of a pharmaceutical supply chain (PSC) can be summarized in Figure 3.1. The activities of a typical PSC involve the flow and transfer of medicines from manufacturers to patients. The role of the pharmaceutical wholesaler in a PSC is to source medicines from suppliers and to deliver them to pharmacies and hospitals. In the NZ context, upon which this study is based, pharmaceutical wholesalers face several constraints due to regulations on both the supply and demand for medicines. These regulations, explained in subsequent sections of this paper, mean that wholesalers face significant constraints on ordering from suppliers while simultaneously having to meet an expectation from pharmacies to maintain an almost 100 percent CSL.





There is an extensive literature on demand forecasting and inventory control in industries such as electronics, food, retail, and clothing. However, very few studies focus on the pharmaceutical industry. This is mainly due to strict regulations and confidentiality surrounding medicines which makes data availability limited. In this study, I use historical monthly sales and inventory data (from January 2014 to December 2018) for a major pharmaceutical wholesaler in NZ to forecast demands for two types of drugs; one is a government subsidized low-value high-demand drug and the other a non-subsidized high-value low-demand drug. The demand forecasts are performed us-

ing traditional time series forecasting methods such as ARIMA and exponential smoothing and the results are compared with non-linear and non-parametric Artificial Neural Network (ANN) models (Neural Network Autoregression (NNAR), Multilayer Perceptron (MLP), and Extreme Learning Machine (ELM)). The model with the lowest forecast error is selected to predict monthly demand and the demand forecast is then used as inputs in an inventory control model. The inventory model is a version of the periodic review model based on the traditional newsvendor¹ problem framework which is derived in the subsequent section of this paper. The inventory model allows us to determine the optimal inventory levels while adjusting for desired CSL targets. The results of the inventory control model are then compared to the current practice stock-on-hand data for the pharmaceutical wholesaler.

The numerical analysis of the mathematical models reveals important findings. First, the study finds that ANN models are stable for demand forecasting and outperform traditional linear demand forecasting models on some occasions. Second, the inventory control model also reveals that, compared to current practice, the proposed model enables wholesalers to reduce inventory levels (by up to 80 percent) while maintaining a very high non-stock-out probability constraint and CSL target close to 100 percent. Furthermore, the proposed periodic inventory control model is based on the premise that ordering of medicines is done once every period (monthly), hence, the model allows wholesalers to reduce ordering costs by limiting the frequency of orders.

Demand forecasting using ANN methods and optimal inventory stocking within a pharmaceutical wholesaler setting has been previously less studied. The existing papers are limited to single medical products across limited geographic area for a small-scale wholesaler. In this analysis, I use a comprehensive dataset for a major wholesaler in NZ for both subsidized and non-subsidized drugs. Moreover, to the best of my knowledge, the newsvendor periodic review model framework has not been previously applied to a pharmaceutical wholesaler setting. Hence, the chapter makes important academic contributions by comparing various methods for demand forecasting and applying the traditional newsvendor problem framework to a pharmaceutical wholesaler's inventory

¹A newsvendor model is a fundamental stochastic inventory model characterized by fixed prices and uncertain demand for a perishable product (newspapers) (Snyder & Shen, 2019).

problem.

The rest of the chapter is organized as follows. Section 2 provides a description of the current state of PSC in NZ and its unique implications for pharmaceutical wholesalers. Section 3 reviews the relevant literature and in Section 4, the mathematical models for demand forecasting and inventory control are presented. Section 5 presents the numerical analysis of the mathematical approach to predict sales and determine optimal inventory levels. Finally, Section 6 concludes with the summary of this study along with important findings, achievements, and directions for future work.

3.2 Institutional Background

The NZ pharmaceutical supply chain and market structure is characterized by an interaction between several private and public entities. The private entities comprise the traditional supply chain network agents which include manufacturers, wholesalers, logistic providers, and pharmacies. In NZ, two public entities, the District Health Boards (DHB) and the Pharmaceutical Management Agency (PHARMAC) have a significant impact on business operations of the PSC network agents. The DHBs aim to provide community pharmacy services to their patients by establishing contracts with over 1000 pharmacies in NZ via the Community Pharmacies Service Agreement (Grant Thornton NZ, 2016). The agreement requires pharmacies to maintain a very high CSL in which 90 percent of prescriptions are to be filled within 1 hour, 99 percent to be filled within the end of the next business day and 100 percent within two business days (Grant Thornton NZ, 2016). The pharmacies in turn make orders from wholesalers who run a network of warehouses to meet medicine demand from the pharmacies. There are currently 3 major wholesalers who operate 17 warehouses across NZ. As shown in Figure 3.2, the wholesalers are ProPharma, CDC, and PWL whereas HCL is the 3PL logistics provider. The network of warehouses for these three wholesalers are presented in Figure 3.2 (Grant Thornton NZ, 2016).



Figure 3.2: Network of wholesaler warehouses in New Zealand (Source: Grant Thornton NZ, 2016)

These wholesalers order from various suppliers to meet demand from the pharmacies. One unique aspect of the NZ pharmaceutical market is the role of PHARMAC. PHARMAC maintains a 'Schedule' of drugs which consists of over 4500 medicines which are funded by PHARMAC and is available to the public at no or minimal cost. Almost 90 percent of medicines in the Schedule use single sourcing through a single supplier. Through negotiated contracts with manufacturers, PHARMAC is able to lower wholesale prices in return for guaranteed market share to the man-

ufacturers. PHARMAC contracts also require suppliers to hold at least 3-month stock and have ongoing supply to support it. Suppliers provide both Schedule and Non-schedule drugs to the wholesalers.

Some of the major cost factors for wholesalers include fixed and overhead costs of CPSA operations: stock finance, overhead management, logistics costs, and pick and pack labour (Sapere Research Group, 2015). According to a Deloitte report, a wholesaler's average margin is 3.5 percent (net discounts) on top of the drug price negotiated by PHARMAC (Sapere Research Group, 2015). According to the same report, increased inflation has increased supply-chain costs which is a source of considerable uncertainty to wholesalers and pharmacies (Sapere Research Group, 2015). Therefore, it is paramount that wholesalers are capable of forecasting demand to precise levels and optimize inventory to increase margins and maintain efficient operations.

3.3 Related Literature

Demand forecasting and inventory control is a topic of major interest to both academic researchers and industry practitioners. There has been an extensive body of research into these topics in the field of supply chain management in the last 50 years (Syntetos et al., 2009). Demand forecasting and effective inventory management remain significant challenges to manufacturers, retailers, and wholesalers across all major industries. A systematic review of related literature on demand forecasting and inventory control can be found across many industries such as fashion (Nenni et al., 2013), telecommunications (Fildes & Kumar, 2002), and energy (Suganthi & Samuel, 2012). In the field of health care, research into demand forecasting and inventory management has been growing in recent years (Uthayakumar & Priyan, 2013). As outlined earlier, an important feature of the pharmaceutical industry is the expected CSL of 100 percent due to product supply and quality being directly related to patient health (Uthayakumar & Priyan, 2013).

Demand forecasting is mainly categorized into linear and non-linear methods. The choice of the appropriate method usually depends on the nature of the data and the modelling of the problem (Khalil Zadeh et al., 2014). Since the original work by Brown and Holt in the 1950s, exponential smoothing methods are still widely used demand forecasting techniques in business today (Hyndman et al., 2008). Linear methods such as Auto Regressive Integrated Moving Average (ARIMA) models are also popular method to forecast demand across various industries (Box et al., 2011). Recent advances in computer applications and machine learning algorithms, such as Artificial Neural Network (ANN) have led to a greater use of these techniques in demand forecasting (Kochak & Sharma, 2015). ANN techniques also suffer from their own set of setbacks, especially the difficulty of identifying important forecasting variables (Lu et al., 2012). Nevertheless, ANN methods have been proven to be an efficient way to forecast non-linear demand and are increasingly used compared to conventional statistical methods (Carbonneau et al., 2008). Various comparative studies have been carried out to judge effectiveness of ANN models to traditional ARIMA and exponential smoothing techniques. Although there are a few studies who report better performance using traditional linear methods (Heravi et al., 2004), the majority of recent studies largely favour ANN techniques to produce better demand forecasts. In an extensive review of the literature on the use of ANN across several forecasting domains, Zhang (2004) found that in one third of the cases, ANN were equivalent to linear methods in terms of performance and in about two-thirds of the cases, ANN techniques performed better. Consequently, in recent years, some studies report good evidence supporting a case for developing a hybrid ARIMA-ANN model, as shown by Zhang (2003), which can outperform both separate methodologies.

Similar to demand forecasting, inventory problems are also extensively studied in literature. Keeping inventory (stock) for future sale is an important business function. Inventory models can be classified as either deterministic where the demand for a time period is known, or stochastic where the demand is a random variable with a known probability distribution (Snyder & Shen, 2019). Furthermore, these inventory models can also be classified based on the way the inventory is reviewed, either continuously or periodically. In a continuous review model, an order is placed as soon as inventory falls to a certain level, whereas in a periodic review model, orders are placed at discrete points in time (Snyder & Shen, 2019). Finally, factors such as product shortages, lead

time, and set-up cost are important variables that can be considered while deriving appropriate inventory models. A detailed review of various inventory model theories and their application can be found in Snyder and Shen (2019).

There are several studies varying in scope and methodologies which are relevant to this paper. In a paper by Merkuryeva et al. (2019), the authors use a 'Simple Moving Average' technique and multiple regression methods to study demand forecasting in the pharmaceutical industry. In another paper, Ribeiro et al. (2017) study pharmaceutical sales prediction in a distribution company in Portugal using data mining methods. In another paper for a pharmaceutical distribution company in Iran, the authors perform a comparative study of linear and non-linear methods and find that the hybrid methods provide better forecasts while predicting future medicine sales (Khalil Zadeh et al., 2014). One of the first attempts to apply inventory theory in the health sector was developed by Satir and Cengiz (1987), in which the authors used a stochastic periodic review model to study stock-out and budgets. Little and Coughlan (2008) developed a constraint-based model for determining stock levels in a storage location with constraints allowing for criticality of medicines. In a separate study, Kelle et al. (2012) modelled optimal allocation of medicines based on ordering and holding costs with a service level constraint. Finally, there are several simulation studies such as Vila-Parrish et al. (2008) and Dellaert and van de Poel (1996) that consider both multi-product and multi-period cases within the pharmaceutical supply chain.

This paper provides a more comprehensive analysis of demand forecasting and inventory control strategies for a pharmaceutical wholesaler in NZ using good quality data. I consider multiple products across multiple periods and provide an extensive comparative analysis of the performance of multiple demand forecasting techniques. Furthermore, to my knowledge, this is the first time that a newsvendor model framework has been applied to inventory problems in the context of a pharmaceutical wholesaler. Hence, this study presents several important contributions to the extant literature on topics related to the pharmaceutical supply chain especially concerning pharmaceutical wholesalers.

3.4 Mathematical Modelling

3.4.1 Demand forecasting

A pharmaceutical wholesaler's ability to reduce inventory and ordering costs is dependent on its capacity to accurately forecast demand for medicines. Several forecasting techniques are widely applied across various industries and the choice of the appropriate method is based on the nature of the product, sales patterns, and data availability. In this study, the forecasting of demand is based on historical pharmaceutical sales data for the wholesaler, and therefore, this research concentrates only on the time series domain. Based on the nature of the pharmaceutical sales data and frequency with which these popular forecasting techniques are used in the existing literature, this chapter focuses on three demand forecasting methods which are described below.

1. Exponential smoothing (Holt - Winter's (HW) method)

In exponential smoothing methods, forecasts are weighted averages of past observations with weights decaying exponentially as observations get older (Hyndman & Athanasopoulos, 2018). Several exponential smoothing methods to forecast demand have been suggested. If the sales patterns exhibit no seasonality and trend, a simple exponential forecasting technique can be used. However, in our analysis we use the Holt - Winter's (HW) additive exponential smoothing method derived by Holt and Winter (1960) which considers both seasonality and trend in the data (Hyndman & Athanasopoulos, 2018).

There are three components to the HW method - level, trend, and seasonality (Goodwin et al., 2010). Level refers to current level of sales/demand, trend refers to any increasing or decreasing values in the series, and seasonality refers to any short-term cycle in the series (Goodwin et al., 2010). Based on historical data, the HW method estimates 'smoothing constants' for each of these components by taking a weighted average of the previous estimates of each component's values (Goodwin et al., 2010).

2. Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) models are also widely used in forecasting demand. There are also three components to the ARIMA method - Autoregression, Moving Average (MA), and differencing (I). Autoregression implies that future values are predicted based on past values, where past values of the variable are used as explanatory variables in a regression like model (Hyndman & Athanasopoulos, 2018). MA uses a regression like model to establish dependency relationship between past errors which are used as explanatory variables (Hyndman & Athanasopoulos, 2018). Finally, 'I' implies that the data have replaced by a differencing process to make the data stationary (Hyndman & Athanasopoulos, 2018). The ARIMA model is based on three parameters, 'p, d, q' where p is the AR order, d is the degree of differencing, and q is the MA order (Hyndman & Athanasopoulos, 2018).

The full model as given by (Hyndman & Athanasopoulos, 2018) can be expressed with the equation below:

$$y'_{t} = c + \phi_{1}y'_{t-1} + \dots + \phi_{p}y'_{t-p} + \theta_{1}\varepsilon_{t-1} + \dots + \theta_{q}\varepsilon_{t-q} + \varepsilon_{t}, \qquad (3.1)$$

where, y'_t is the differenced series and the predictor variables on the right-hand side contains both the lagged values of the variable of interest and lagged errors (Hyndman & Athanasopoulos, 2018).

3. Artificial Neural Network (ANN)

In recent years, ANN methods have been widely used in solving forecasting problems. ANN models are organized in layers of nodes where the model trains itself from the data (Hyndman & Athanasopoulos, 2018). Generally, ANN models are composed of three layers-input layer, hidden layer, and output layer and these layers consists of nodes that mimic neurons in the brain, where the 'nodes' perform calculations and feed these calculations to the output layers that produces the forecast value (Hyndman & Athanasopoulos, 2018). Hence, such type of models is also called feed-forward neural network models (Hyndman & Athana-

sopoulos, 2018). The calculations attached to these nodes are learned from the data. A visual representation of a feed-forward neural network is provided in Figure 3.3.



Figure 3.3: Feed-forward neural network (Adapted from Hyndman and Athanasopoulous, 2018)

In this chapter, I use three types of feed-forward neural network models namely- Neural Network Autoregression (NNAR), Extreme Learning Machine (ELM), and Multi-layer Perceptron (MLP). In the case of NNAR models, past observation or lagged values of the data are used as inputs in the input layer (Hyndman & Athanasopoulos, 2018). In the case of MLP, the model trains itself from the data and determines the output weights for the hidden layer through backpropagation (Durairaj & Revathi, 2015). All of these methods are similar because they are a class of feed-forward neural network models but are different on how the coefficients and weights get assigned to different input and output layers. ANN methods are essentially black box approaches and it is difficult to provide a satisfactory explanation of their behaviour (Benítez et al., 1997). Nevertheless, these models are generally able to make accurate predictions and hence are widely applied in forecasting.

There are several differences between traditional time-series forecasting methods such as ARIMA/HW and more modern ANN methods. One of the distinct advantages of the ANN models is that it is non-linear and hence makes no a priori assumptions about the nature of the data (Khashei & Bijari, 2010). Moreover, the models are data driven and are flexible to adapt to the data and can uncover unseen patterns within the data (Khashei & Bijari, 2010). However, ANN techniques also suffer from their own set of setbacks, especially the

difficulty of identifying important forecasting variables (Lu et al., 2012). A detailed review on the theory and application of the various types of multilayer feed-forward neural networks can be found in Hyndman and Athanasopoulos (2018), Huang et al. (2006), and Gardner and Dorling (1998).

Before applying the forecasts as inputs for the inventory control problem, it is necessary to conduct a comparative assessment on the performances of the above-mentioned forecasting models. This can be analysed by calculating various forecast error metrics. The measurements used in this study are Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), which are shown in equations below:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_{i} - Y_{i})^{2}}$$
(3.2)

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{Y}_i - Y_i}{\hat{Y}_i} \right|$$
(3.3)

where ' \hat{Y}_i ' and ' Y_i ' represent predicted and actual values at time 'i' respectively.

3.4.2 Inventory optimization

In this paper, the theoretical model I use to determine optimal inventory level is based on the newsvendor framework whose mathematical analysis originated with Arrow et al. (1951). The subsequent development of the theoretical model and some of the notations used are also influenced by the work of Snyder and Shen (2019).

Consider a pharmaceutical wholesaler who orders a single product from a single supplier and sells the product during a single time period. It is assumed that the wholesaler faces stochastic demand from a single pharmacy with a known demand distribution. Although the wholesaler sells multiple medicines over multiple periods to multiple pharmacies, the multi-product/multi-period models can be thought of as multiple independent copies of the single-period model presented in this analysis. This assumption reduces the proposed model to a periodic review model based

on the newsvendor problem framework (Snyder & Shen, 2019). The assumption is valid because pharmaceutical wholesalers in NZ face significant upstream and downstream constraints for both ordering and meeting pharmacy demand. As described earlier, the wholesaler can only place an order once every period and because the pharmacies must maintain a very high CSL, the unmet demand from the pharmacies is likely to result in lost sales to the wholesaler. Therefore, the operations for subsequent periods are not linked, justifying the development of the single-period model in this analysis.

The original newsvendor framework assumes that excess inventory is not carried over to the next period. A detailed explanation of periodic review model with zero fixed costs and derivations on setting the optimal base stock policy can be found in (Snyder & Shen, 2019). I apply the traditional newsvendor framework to the pharmaceutical wholesaler inventory problem using mathematical analysis given by Snyder and Shen (2019) for setting optimal base stock.

In each time period, we observe the current inventory position and place an order whose size is sufficient to bring the inventory level (IL) up to the base stock level (S). Therefore, if IL \leq S, then the wholesalers orders up to S, and when IL > S, then the wholesaler leaves the IL where it is. S is a constant and it is later shown that the base stock policy (i.e. ordering up to S) is optimal when the fixed costs (K) are zero. Although in practice K is typically positive, because the wholesaler sells multiple products and faces significant constraints on ordering, and thus be indifferent towards placing an order every period. This means the fixed cost is constant at every ordering period and can be ignored. The model also assumes that the lead time (L) is zero, i.e. the order is delivered as soon as it is ordered and the discount factor (γ) is equal to 1 and there is no discounting such that\$1 spent (or received) in period t+1 is equivalent to \$1 in period t. The assumption of no lead time is reasonable for a pharmaceutical wholesaler in NZ. This is because wholesalers in NZ must order from their manufacturers at specific periods of time (usually two weeks in advance) through a 3PL logistics service provider. Hence, pharmaceutical wholesalers in NZ are aware in advance when the order would be delivered. Therefore, for our modelling purposes, the assumption of no lead time is justified.

The sequence of events in each period t is as follows.

- 1. The inventory level (IL) at period t is observed.
- 2. Quantity (Q) is ordered at the start of the period according to the base-stock policy, which is optimal such that,

$$Q_t = \begin{cases} S - IL, & if IL \le S \\ 0, & if IL > S \end{cases}$$
(3.4)

- 3. A random demand (d_t) is observed and is satisfied from inventory.
- 4. Holding costs (h) is assessed in the case of excess inventory and stock-out costs (c) is assessed if the demand is not satisfied. Any excess inventory is carried over to the next period, whereas, unfulfilled demand are lost sales and are not ordered again. The costs given by 'h' and 'c' are also referred to as overage and underage costs in the academic literature and are usually represented by c_0 and c_u respectively by many prior researchers. It is assumed that other costs such as purchasing and processing are already included in 'h' and that its negative is included in 'c'.

Therefore, if $d_t > Q_t$, then $c_u = c$, and if $d_t < Q_t$, then $c_o = h$.

In this set-up the goal is to determine 'S' and to minimize the expected cost in the single period for the pharmaceutical wholesaler. Therefore, if the wholesaler orders up to 'S', the cost function Y(S) is given as:

$$Y(S) = Expected overage cost + Expected underage cost.$$
 (3.5)

This can be subsequently written using the following notation:

$$Y(S) = c_o E[S - d] + c_u E[d - S]$$
(3.6)

Equation 3.6 represents the cost function of the wholesaler where the wholesaler incurs an overage cost when S > d and an underage cost when d > S. We can also represent these values as integrals in the following form, where *d* is the random demand.

$$Y(S) = c_0 \int_0^S (S - d)g(d) \, dd + c_u \int_S^\infty (d - S)g(d) \, dd \tag{3.7}$$

In order to minimize Y(S), its first derivative is set to 0, i.e $\frac{dY(S)}{dS} = 0$

Then according to Leibniz rule, we get,

$$\frac{dY(S)}{dS} = c_0 F(S) - c_u (1 - F(S))$$
(3.8)

The second derivative of Equation 3.8 is going to be $\frac{d^2Y(S)}{dS} = (c_o + c_u)F(S)$ which is ≥ 0 . Hence, Y(S) is convex and has an optimal solution which is given as:

$$\frac{dY(S)}{dS} = c_0 F(S) - c_u (1 - F(S)) = 0$$
(3.9)

Therefore,

$$F(S) = \left(\frac{c_u}{c_u + c_o}\right) \tag{3.10}$$

Equation 3.10 is the critical fractile which helps determine the order size and its implicit formulation originated with Arrow (1951) (Snyder & Shen, 2019). The critical fractile in this case also represents the percentage of the time the pharmacies' demand is satisfied by the wholesaler. The intuition here is that as the stock-out costs / underage costs given by c_u is larger relative to the inventory holding cost / overage costs given by c_o , the critical fractile approaches to 1 i.e. the percentage of the time demand ought to be satisfied by the wholesaler approaches to 100 percent. In principle, 'S', which is the optimal base stock is the minimum inventory required to satisfy a level of demand which satisfies the wholesaler's targeted service levels. The minimum service level the wholesaler wishes to satisfy is based on the preferences of the wholesaler and the associated costs with keeping too much inventory or too little inventory. Therefore, based on the critical fractile and the expected stock-out costs and inventory holding costs, the pharmaceutical wholesaler can decide on the optimal stock levels while simultaneously adjusting for desired CSL targets.

In the case of a normally distributed demand i.e. $D \sim N(\mu, \sigma^2)$ with pdf f and cdf F, the optimal base stock policy is given by (Snyder & Shen, 2019).

$$F(S) = \left(\frac{c_u}{c_u + c_o}\right) \iff \Phi\left(\frac{S - \mu}{\sigma}\right) = \left(\frac{c_u}{c_u + c_o}\right)$$
(3.11)

Hence, we can solve for S. The cdf F given in Equation 3.11 for optimal base stock (S) is given by Φ and its inverse appears in Equation 3.12 while solving for S (Snyder & Shen, 2019).

$$S = \mu + \sigma \phi^{-1} \left(\frac{c_u}{c_u + c_o} \right) \tag{3.12}$$

In our set-up, the mean and the standard deviation is not known. Therefore, instead of the mean, the input values are forecast using the demand forecasting techniques explained in Section 3.4.1. Therefore, optimal base stock (S) as given in Equation 3.12 takes the form as given in Equation 3.13 and a detailed explanation for this can also be found in Nahmias and Olsen (2020).

Optimal base stock (S) =
$$\mu_t + \sigma_{e,t} \phi^{-1} \left(\frac{c_u}{c_u + c_o} \right)$$
 (3.13)

where μ_t is the forecast for the demand in each period t and $\sigma_{e,t}$ is the standard deviation of the forecast error (Snyder & Shen, 2019).

3.5 Experimental Results

3.5.1 Data

This study uses monthly sales and stock on hand (inventory) data from January 2014 to December 2018 (total 60 data points) for a NZ pharmaceutical wholesaler. Although data was available for multiple medicines, only two drugs are used for comparative assessment in this study. For the purpose of this study, the drugs are called Drug A and Drug B. Drug A is a government subsidized low value high demand drug used to treat osteoporosis symptoms. Drug B is a non-subsidized low demand drug which is relatively more expensive than Drug A and is used to treat Type 2 diabetes by lowering blood sugar levels in diabetic patients. Both drugs are supplied by a single supplier to the wholesaler. Although the company operates multiple warehouses across NZ that source to multiple pharmacies; the data used in this study is the aggregate sales data for all pharmacies across all warehouses for the wholesaler in NZ. The first 48 data points (80 % of the total sample) is used as the training sample, and the remaining 12 data points (20 % of the sample) is used as the testing sample for measuring demand forecast accuracy and evaluating the inventory control problem. Ideally, in order to generate a more robust forecast, a larger sample of data is preferred. However, such data were not available through the wholesaler. Therefore, the results of this analysis should be interpreted with that caveat in mind. Nevertheless, our results show that the techniques employed in this chapter are stable for demand forecasting.

The study uses demand forecasting techniques explained in Section 3.4.1 to forecast demand for both drugs A and B and evaluates the performance of the forecasts based on two criteria: root mean squared error (RMSE) and mean absolute percentage error (MAPE). Both RMSE and MAPE evaluate the deviation of the demand forecast from the actual sales values, and hence a lower measure of these metrics implies greater accuracy. The forecasts with lowest RMSE and MAPE are then used as inputs for the inventory problem and optimal inventory levels are calculated and compared with current practice.

3.5.2 Demand Pattern Comparison

As different drugs vary on their usage, price and other characteristics, the sales of these drugs exhibit different demand patterns. For the purpose of this study, the choice of respective drugs was specifically motivated by price, subsidy status, and volume of demand. To allow for easy comparison, Figure 3.4 presented below illustrates both trend and seasonality patterns for both drug A and drug B. Plot A and Plot C graph monthly sales for January 2014 to December 2018 for drugs A and B respectively. These plots provide an indication of trends. Simultaneously, Plot B and Plot D graph total sales by month for 2014 to 2018 for these drugs and illustrate seasonality.



Figure 3.4: Trends and Seasonality Analysis for monthly drug sales

As evident from Figure 3.4 (Plots A and B), Drug A exhibits both trend and seasonality. The total number of monthly sales have continuously decreased from 2014 to 2018 and monthly sales have peaked in the month of May (except for 2016) and towards the end of the year (except for 2018). However, as illustrated in Plots C and D in Figure 3.4, the total monthly sales for Drug B is relatively low compared to Drug A and the demand appears to be consistent throughout the years without any observable patterns of trend and seasonality.

3.5.3 Demand Forecasting Results

Table 3.1 presents the demand forecasts for Drug A for HW, ARIMA, NNAR, ELM, and MLP forecasting techniques and Table 3.2 reports demand forecasts for Drug B. Table 3.3 presents the numerical values of RMSE and MAPE indicators to assess forecast accuracy. Graphical illustration of the forecast results for Drug A and Drug B are presented in Figure 3.5. The ARIMA forecast results are automatically obtained in 'R' using an algorithm which combines unit root tests, minimisation of the AICc and MLE to obtain an ARIMA model. Similarly, the optimal number of lags as inputs for the NNAR model is also selected automatically according to the AIC.

Month	Actual	HW	ARIMA	NNAR	ELM	MLP
	Demand	Forecast	Forecast	Forecast	Forecast	Forecast
January	9642	8110	8774	10031	9199	8737
February	9678	8853	9232	9979	10386	10137
March	9942	9461	9755	10374	9929	9755
April	10328	8720	9034	9925	10179	10307
May	11029	9754	10391	10976	10670	11142
June	9449	9405	9838	10260	9946	10250
July	10297	8920	8845	9978	10104	9938
August	10318	8981	9694	10167	10373	11236
September	9156	8719	9172	9883	10120	10831
October	9924	8415	8930	10016	9937	10283
November	9968	9038	9736	10174	10767	11929
December	9761	9735	9980	10228	10280	11739

Table 3.1: Total monthly demand forecasts for 2018 (Drug A)

Table 3.2: Total monthly demand forecasts for 2018 (Drug B)

Month	Actual	HW	ARIMA	NNAR	ELM	MLP
	Demand	Forecast	Forecast	Forecast	Forecast	Forecast
January	117	132	168	159	135	116
February	128	146	158	157	141	137
March	136	144	167	156	140	128
April	166	164	148	129	138	133
May	132	137	161	160	142	147
June	152	157	158	157	140	125
July	140	137	156	156	141	125
August	151	133	167	158	142	125
September	139	142	156	156	141	130
October	123	134	163	157	142	116
November	88	163	166	157	142	133
December	102	159	162	157	142	104



Figure 3.5: Monthly Demand forecast comparisons for Drug A and Drug B (Year:2018)

Table 3.3: Comparison of forecast errors for demand forecasting models

Product Type	Foregest orrer matrics	Forecasting Method					
i touuct type	Forecast error metrics	HW	ARIMA	NNAR	ELM	MLP	
Drug A	RMSE	1098.82	753.27	425.53	497.73	1055.99	
	MAPE	9.40	6.07	3.73	4.03	8.33	
Drug B	RMSE	28.82	38.37	35.13	23.36	20.87	
	MAPE	17.27	28.39	25.76	15.49	13.04	

The analysis of the results tables show that the predicted sales values obtained through ANN techniques (NNAR, ELM, and MLP) are in general closer to the real sales values when compared to the results obtained from HW and ARIMA methods. For Drug A, based on the forecast accuracy measures obtained in Table 3.3, it can be observed that all techniques report good performance and accuracy with forecast errors below 10 percent. The forecast errors measure the average percentage

difference between the forecasts predicted by the model and the actual observed values of demand. NNAR and ELM models have considerably lower RMSE values and report forecast accuracy errors of less than 5 percent, which is generally considered to be a very high precision level according to industry standards and the academic literature. It can therefore be concluded that for medicines with both trend and seasonality component and when past sales inform future sales, both traditional methods and ANN techniques can be used for demand forecasting.

For Drug B, the forecast results are relatively poorer compared to forecasts for Drug A. This is mainly because the drug demand does not exhibit any linear behaviour, seasonality, and trend. This makes the usage of lagged values of the time series as inputs in the models ineffective. Hence, both ARIMA and NNAR models that include lagged values of the time series in their calculations perform significantly more weakly than other forecasting techniques. However, even for Drug B, ANN techniques (MLP and ELM) have considerably lower RMSE and MAPE values compared to other forecasting methods. Although the differences are relatively small, ANN methods outperform traditional exponential smoothing and ARIMA method in demand forecasting for both types of drugs. We can conclude that ANN techniques (NNAR in the case of drug A and MLP in the case of drug B) are stable for forecasting drug consumption. Based on the forecast accuracy measures reported in Table 3.3, forecasts obtained from the NNAR method and forecasts obtained from MLP method were used as input values in the inventory control problem for Drug A and Drug B, respectively. The numerical analysis and the results of the inventory control problem are discussed in the subsequent section.

It can be observed that when the demand pattern exhibits some linear behaviour and trend as in the case of drug A, all forecasting techniques generate stable forecasts. In the case of drug B, where there is no observable pattern in demand, the forecast accuracy is less stable across all methods. As discussed earlier, ANN methods are non-parametric and non-linear forecasting methods and are generally considered black box approaches where it is difficult to know the underlying structures of the functions of the model. Nevertheless, they have been widely applied in demand forecasting due to their ability to generate accurate predictions, and our analysis also confirm that the ANN methods generally deliver stable forecasts.

3.5.4 Inventory Problem

The numerical analysis of the inventory problem for the pharmaceutical wholesaler is based on the theoretical framework presented in Section 3.4.2. The presented theoretical model is a periodic review model in which orders are placed at discrete points in time i.e. once every month. Although the model is simple, it captures all the trade-offs involved in the decision-making process for setting optimal inventory. The inventory holding costs and penalty costs for inability to satisfy demand are likely to be the most important inputs for supply chain managers while deciding on inventory levels and the fact that wholesalers in NZ are only able to order from the suppliers at fixed periods, the presented model captures all essential elements of the NZ pharmaceutical wholesale sector. As explained earlier, the intuition of the inventory model is that as the stock-out costs / underage costs given by ' c_u ' becomes larger relative to the inventory holding cost / overage costs given by ' c_o ' , the critical fractile approaches to 1 or equivalently the percentage of the time demand ought to be satisfied by the wholesaler approaches 100 percent. A more realistic solution to the inventory problem would be to solve it with actual values of c_u and c_o , however, such data is not available. Hence, the inventory problem is instead solved based on the assumption that the critical fractile is 0.999 or that it is extremely costly for the wholesaler to stock-out, that is, I assume the wholesaler maintains a probability of no stock-outs of 0.999 or a CSL level close to 100 percent. Therefore, the optimal stock levels for drug A and drug B are each given by the following expression:

Optimal base stock (S) =
$$\mu_t + \sigma_{e,t}\phi^{-1}(0.99)$$
 (3.14)

where, $\mu_t + \sigma_{e,t}$ are demand forecast and standard deviation of the forecast error for Drug A and Drug B, respectively. The critical value ' $\phi^{-1}(0.999) = 3.09$.

Table 3.4 and Table 3.5 present results for total sales, orders, and inventory levels based on the forecast while simultaneously comparing the results with current practice for drug A and drug B

respectively. Graphical illustration of the results (actual vs forecast) for both drugs are presented in Figure 3.5.

Month	Actual	Sales	Actual	Order	Actual	Inventory
	Sales	Forecast	Orders	Forecast	Inventory	Forecast
January	9642	10031	8630	2908	7308	1586
February	9678	9979	9767	9590	7397	1498
March	9942	10374	10180	10073	7635	1629
April	10328	9925	9598	9493	6905	795
May	11029	10976	10465	11379	6341	1145
June	9449	10260	9788	10312	6680	2008
July	10297	9978	9963	9167	6346	879
August	10318	10167	10751	10487	6779	1047
September	9156	9883	9810	10034	7433	1925
October	9924	10016	8062	9289	5571	1290
November	9968	10174	10753	10081	6356	1403
December	9761	10228	9377	10022	5972	1665

Table 3.4: Total sales, orders, and inventory for 2018 (Drug A)

Table 3.5: Total sales, orders, and inventory for 2018 (Drug B)

Month	Actual	Sales	Actual	Order	Actual	Inventory
	Sales	Forecast	Orders	Forecast	Inventory	Forecast
January	117	116	118	41	141	64
February	128	137	120	139	133	75
March	136	128	132	119	129	58
April	166	133	148	141	111	33
May	132	147	166	180	145	81
June	152	125	108	110	101	39
July	140	125	157	151	118	50
August	151	125	156	141	123	40
September	139	130	156	155	140	56
October	123	116	99	125	116	58
November	88	133	127	140	155	111
December	102	104	54	59	107	68



Figure 3.6: Comparisons of total orders and inventory for 2018 (Actual vs. Forecast) for Drug A and Drug B

The analysis of the results table and the graphical illustration in Figure 3.6 show that the presented model can successfully help reduce inventory levels despite maintaining a very high nonstock-out probability and CSL of 99.99 percent. Although current practice is efficient in fulfilling demand without stocking out, it comes at a cost of carrying large inventories.

There are several possible reasons for persistent levels of overstock by wholesalers. Due to confidential nature of pharmaceutical data, the dataset provided to me was aggregate levels of stock by the pharmaceutical wholesaler across multiple warehouses in NZ, hence individual warehouses may not carry such large inventories. Moreover, the drugs are generic and have a relatively longer shelf life. Therefore, the associated costs of holding such large inventory may be minimal compared to the costs of running out. Hence, incentives for pharmaceutical wholesalers to reduce inventory may be low.

In the case of Drug A, it can be observed that demand can be forecast precisely, therefore, there is significant scope to reduce inventory levels by up to 80 percent as evident from Table 3.4. Despite

the fact there is less precision in terms of forecast accuracy for Drug B, numerical analysis of the presented inventory model in Table 3.5 reveal that inventories can be reduced by up to 50 percent. These are both significant reductions when compared to current practice. The inventory model presented can also help reduce ordering; however, these reductions are not significant (less than 5 percent), at least in terms of total orders. Nevertheless, the fact that the presented inventory model is a periodic review model implies that ordering costs can be reduced by limiting the frequency of placing an order. Therefore, it can be safely concluded that the presented inventory model is stable for forecasting optimal inventory levels and can yield better performance when compared to current practice.

It is important to note that although ANN methods generate stable forecasts, one is not consistently performing better than the other methods suggested in this analysis. As suggested by Zhang (2003), there is an increasing trend in academic literature to use a hybrid ARIMA-ANN models to predict demand where these hybrid methods can potentially take advantages of multiple approaches and these have been shown to be very effective in generating more accurate predictions. Our results confirm that such an approach is likely to be very effective in generating better predictions.

3.6 Conclusion

This chapter presents an integrated framework for demand forecasting and inventory planning for a pharmaceutical wholesaler in New Zealand. As outlined earlier in the chapter, pharmaceutical products have a short shelf life and maintaining high product quality is vital because it directly impacts public health. Furthermore, regulatory pressures and supply and demand conditions significantly increase complexities in decision making for PSC network agents. Therefore, accurate demand forecasting for medicines and subsequent inventory planning is vital to maintain efficient operations and to increase profits.

The numerical analysis presented here makes a comparative assessment of several demand

forecasting techniques such as exponential smoothing, ARIMA, and ANN methods (NNETAR, MLP, ELM). The forecasting results are obtained by analysing real pharmaceutical sales data (for both subsidized and non-subsidized medicines) from a major wholesaler in NZ, and the performance of individual techniques are assessed based on various forecast error metrics. Consistent with most comparative studies on demand forecasting, the study finds that ANN methods are generally stable for demand forecasting and sometimes perform better than both exponential smoothing and ARIMA methods. However, when future predictions are based on lagged values of past sales, both traditional methods for demand forecasting and ANN methods are equally stable for generating good forecasts. Nevertheless, the MLP method seems to produce consistently better forecasts across both products. Subsequently, using the obtained demand forecasts as inputs, a periodic review inventory model is derived based on the newsvendor framework which is then applied to forecast optimal inventory levels. The results are then compared with current practice using actual stock-on-hand data from the wholesaler. The analysis finds that significant reduction in both inventory and ordering can be achieved by adopting the proposed periodic review inventory model.

There are several ways in which the scope of this study can be extended for further investigation. This study presents a framework for demand forecasting and inventory planning for pharmaceutical wholesalers only. Therefore, future studies could incorporate both suppliers and pharmacies and model the entire PSC network with information sharing to allow for complex interactions between the PSC network agents. Moreover, several assumptions such as zero set-up cost, zero lead times, and periodic review reduce this model to a simple framework. Future studies can extend the existing model to include variable set-up costs and lead times. Finally, there is an increasing interest in developing a hybrid ARIMA-ANN model to predict future demand and there is good evidence to support the claim that these hybrid methods can potentially take advantages of both approaches and produce results that can outperform both separate methodologies (Zhang, 2003).

The pharmaceutical industry is usually characterized by high complexity and because it is concerned with human health, CSL cannot be compromised. This usually leads to a practice of carrying huge inventories to ensure 100 percent CSL. The numerical analysis finds that for generic drugs with observed patterns of seasonality and trend, demand can be predicted with significantly high precision. Even for drugs that exhibit no observed patterns, managers can still benefit from using ANN forecasting techniques that yield relatively efficient sales predictions. The proposed inventory model also shows that inventory reductions can be achieved without compromising CSL while simultaneously limiting ordering. The proposed inventory model is an intuitive framework that can be adjusted based on desired CSL targets and several cost factors affecting inventory decisions. The fact that industry practitioners have access to large amounts of data and cost information implies that they can confidently engage in demand forecasting and inventory planning that can lead to significantly better outcomes and higher profits.
References

- Alrwisan, A., & Eworuke, E. (2016). Are discrepancies in waiting time for chest pain at emergency departments between african americans and whites improving over time? *The Journal of emergency medicine*, *50*(2), 349–355.
- Arnesen, K. E., Erikssen, J., & Stavem, K. (2002). Gender and socioeconomic status as determinants of waiting time for inpatient surgery in a system with implicit queue management. *Health policy*, 62(3), 329–341.
- Arrow, K. J., Harris, T., & Marschak, J. (1951). Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, 250–272.
- Atella, V., & Deb, P. (2008). Are primary care physicians, public and private sector specialists substitutes or complements? evidence from a simultaneous equations model for count data. *Journal of Health Economics*, 27(3), 770–785.
- Aumua, C., Krishna, S., Samarakkody, U., & de Lore, D. (2018). Racial disparity in an outreach pediatric surgical service. *Journal of pediatric surgery*, *53*(5), 1069–1072.
- Benítez, J. M., Castro, J. L., & Requena, I. (1997). Are artificial neural networks black boxes? *IEEE Transactions on neural networks*, 8(5), 1156–1164.
- Bond Huie, S. A., Krueger, P. M., Rogers, R. G., & Hummer, R. A. (2003). Wealth, race, and mortality. *Social Science Quarterly*, 84(3), 667–684.

- Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2011). *Time series analysis: forecasting and control* (Vol. 734). John Wiley & Sons.
- Bramley, D., Hebert, P., Jackson, R., & Chassin, M. (2004). Indigenous disparities in diseasespecific mortality, a cross-country comparison: New zealand, australia, canada, and the united states. *The New Zealand Medical Journal (Online)*, *117*(1207).
- Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140–1154.
- Cullis, J. G., & Jones, P. R. (1986). Rationing by waiting lists: an implication. *The American economic review*, 76(1), 250–256.
- Cullis, J. G., Jones, P. R., & Propper, C. (2000). Waiting lists and medical treatment: analysis and policies. *Handbook of health economics*, *1*, 1201–1249.
- Derrett, S., Bevin, T. H., Herbison, P., & Paul, C. (2009). Access to elective surgery in new zealand: considering equity and the private and public mix. *The International journal of health planning and management*, 24(2), 147–160.
- Derrett, S., Cousins, K., & Gauld, R. (2013). A messy reality: an analysis of new zealand's elective surgery scoring system via media sources, 2000–2006. *The International Journal of Health Planning and Management*, 28(1), 48–62.
- Derrett, S., Paul, C., & Morris, J. M. (1999). Waiting for elective surgery: effects on health-related quality of life. *International Journal for Quality in Health Care*, *11*(1), 47–57.
- Dew, K., Cumming, J., McLeod, D., Morgan, S., McKinlay, E., Dowell, A., & Love, T. (2005a). Explicit rationing of elective services: implementing the new zealand reforms. *Health Policy*, 74(1), 1–12.

- Dew, K., Cumming, J., McLeod, D., Morgan, S., McKinlay, E., Dowell, A., & Love, T. (2005b). Explicit rationing of elective services: implementing the new zealand reforms. *Health Policy*, 74(1), 1–12.
- Durairaj, M., & Revathi, V. (2015). Prediction of heart disease using back propagation mlp algorithm. *International Journal of Scientific & Technology Research*, 4(8), 235–239.
- Ellison-Loschmann, L., & Pearce, N. (2006). Improving access to health care among new zealand's maori population. *American journal of public health*, *96*(4), 612–617.
- Evelyn Theunissen, K. (2011). The nurse's role in improving health disparities experienced by the indigenous māori of new zealand. *Contemporary Nurse*, *39*(2), 281–286.
- Fernández-Olano, C., Hidalgo, J. L.-T., Cerdá-Díaz, R., Requena-Gallego, M., Sánchez-Castaño, C., Urbistondo-Cascales, L., & Otero-Puime, A. (2006). Factors associated with health care utilization by the elderly in a public health care system. *Health policy*, 75(2), 131–139.
- Fildes, R., & Kumar, V. (2002). Telecommunications demand forecasting—a review. *International journal of forecasting*, *18*(4), 489–522.
- Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Smith, G. D. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health*, 60(1), 7–12.
- Goddard, J., Malek, M., & Tavakoli, M. (1995). An economic model of the market for hospital treatment for non-urgent conditions. *Health Economics*, *4*(1), 41–55.
- Goodwin, P., et al. (2010). The holt-winters approach to exponential smoothing: 50 years old and going strong. *Foresight*, *19*(19), 30–33.
- Grant Thornton NZ. (2016). True costs of the pharmaceutical supply chain.
- Gravelle, H., Smith, P., & Xavier, A. (2003). Performance signals in the public sector: the case of health care. *Oxford Economic Papers*, 55(1), 81–103.

- Gravelle, H., Sutton, M., Morris, S., Windmeijer, F., Leyland, A., Dibben, C., & Muirhead, M. (2003). Modelling supply and demand influences on the use of health care: implications for deriving a needs-based capitation formula. *Health economics*, *12*(12), 985–1004.
- Hacker, J., & Stanistreet, D. (2004). Equity in waiting times for two surgical specialties: a case study at a hospital in the north west of england. *Journal of Public Health*, 26(1), 56–60.
- Heinrich, N., Wübker, A., & Wuckel, C. (2018). Waiting times for outpatient treatment in germany: New experimental evidence from primary data. *Jahrbücher für Nationalökonomie und Statistik*, 238(5), 375–394.
- Heravi, S., Osborn, D. R., & Birchenhall, C. (2004). Linear versus neural network forecasts for european industrial production series. *International Journal of Forecasting*, 20(3), 435–446.
- Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3), 489–501.
- Hyndman, R., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.
- Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media.
- Johar, M., Jones, G., Keane, M. P., Savage, E., & Stavrunova, O. (2013). Discrimination in a universal health system: explaining socioeconomic waiting time gaps. *Journal of health economics*, 32(1), 181–194.
- Jones, A. M. (2012). The elgar companion to health economics. Edward Elgar Publishing.
- Kaarboe, O., & Carlsen, F. (2014). Waiting times and socioeconomic status. evidence from norway. *Health economics*, *23*(1), 93–107.
- Kelle, P., Woosley, J., & Schneider, H. (2012). Pharmaceutical supply chain specifics and inventory solutions for a hospital case. *Operations Research for Health Care*, *1*(2-3), 54–63.

- Khalil Zadeh, N., Sepehri, M. M., & Farvaresh, H. (2014). Intelligent sales prediction for pharmaceutical distribution companies: A data mining based approach. *Mathematical Problems in Engineering*, 2014.
- Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*, *37*(1), 479–489.
- Kochak, A., & Sharma, S. (2015). Demand forecasting using neural network for supply chain management. *International journal of mechanical engineering and robotics research*, 4(1), 96– 104.
- Laudicella, M., Siciliani, L., & Cookson, R. (2012). Waiting times and socioeconomic status: evidence from england. *Social science & medicine*, 74(9), 1331–1341.
- Lilford, R., & Pronovost, P. (2010). Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *Bmj*, *340*.
- Lindsay, C. M., & Feigenbaum, B. (1984). Rationing by waiting lists. *The American Economic Review*, 74(3), 404–417.
- Löfvendahl, S., Eckerlund, I., Hansagi, H., Malmqvist, B., Resch, S., & Hanning, M. (2005).
 Waiting for orthopaedic surgery: factors associated with waiting times and patients' opinion. *International Journal for Quality in Health Care*, *17*(2), 133–140.
- Lu, C.-J., Lee, T.-S., & Lian, C.-M. (2012). Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks. *Decision Support Systems*, 54(1), 584–596.
- Lungu, D. A., Ruggieri, T. G., & Nuti, S. (2019). Decision making tools for managing waiting times and treatment rates in elective surgery. *BMC health services research*, *19*(1), 1–9.

- Mackenbach, J. P., Stirbu, I., Roskam, A.-J. R., Schaap, M. M., Menvielle, G., Leinsalu, M., & Kunst, A. E. (2008). Socioeconomic inequalities in health in 22 european countries. *New England journal of medicine*, 358(23), 2468–2481.
- Martin, R. M., Sterne, J. A., Gunnell, D., Ebrahim, S., Smith, G. D., & Frankel, S. (2003). Nhs waiting lists and evidence of national or local failure: analysis of health service data. *Bmj*, 326(7382), 188.
- Martin, S., Rice, N., Jacobs, R., & Smith, P. (2007). The market for elective surgery: joint estimation of supply and demand. *Journal of health economics*, 26(2), 263–285.
- Martin, S., & Smith, P. C. (1999). Rationing by waiting lists: an empirical investigation. *Journal of Public Economics*, *71*(1), 141–164.
- Meerding, W.-J., IJzelenberg, W., Koopmanschap, M., Severens, J. L., & Burdorf, A. (2005). Health problems lead to considerable productivity loss at work among workers with high physical load jobs. *Journal of clinical epidemiology*, 58(5), 517–523.
- Merkuryeva, G., Valberga, A., & Smirnov, A. (2019). Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science*, *149*, 3–10.
- Monstad, K., Engesæter, L. B., & Espehaug, B. (2014). Waiting time and socioeconomic status—an individual-level analysis. *Health Economics*, 23(4), 446–461.
- Moscelli, G., Siciliani, L., Gutacker, N., & Cookson, R. (2018). Socioeconomic inequality of access to healthcare: Does choice explain the gradient? *Journal of Health Economics*, 57, 290–314.
- Nelson, A. (2002). Unequal treatment: confronting racial and ethnic disparities in health care. *Journal of the national medical association*, *94*(8), 666.
- Nenni, M. E., Giustiniano, L., & Pirolo, L. (2013). Demand forecasting in the fashion industry: a review. *International Journal of Engineering Business Management*, *5*, 37.

- Rahiri, J.-L., Alexander, Z., Harwood, M., Koea, J., & Hill, A. G. (2018). Systematic review of disparities in surgical care for maori in new zealand. *ANZ Journal of Surgery*, 88(7-8), 683–689.
- Ribeiro, A., Seruca, I., & Durão, N. (2017). Improving organizational decision support: Detection of outliers and sales prediction for a pharmaceutical distribution company. *Procedia computer science*, 121, 282–290.
- Riganti, A., Siciliani, L., & Fiorio, C. V. (2017). The effect of waiting times on demand and supply for elective surgery: Evidence from italy. *Health economics*, *26*, 92–105.
- Salmond, C. E., & Crampton, P. (2012). Development of new zealand's deprivation index (nzdep) and its uptake as a national policy tool. *Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique*, S7–S11.
- Sapere Research Group. (2015). Strategic think piece on pharmaceutical margins.
- Seneviratne, S., Campbell, I., Scott, N., Coles, C., & Lawrenson, R. (2015). Treatment delay for maori women with breast cancer in new zealand. *Ethnicity & health*, 20(2), 178–193.
- Sharma, A., Siciliani, L., & Harris, A. (2013). Waiting times and socioeconomic status: does sample selection matter? *Economic Modelling*, *33*, 659–667.
- Siciliani, L., & Hurst, J. (2005a). Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 oecd countries. *Health policy*, 72(2), 201–215.
- Siciliani, L., & Hurst, J. (2005b). Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 oecd countries. *Health policy*, 72(2), 201–215.
- Siciliani, L., & Verzulli, R. (2009). Waiting times and socioeconomic status among elderly europeans: evidence from share. *Health Economics*, *18*(11), 1295–1306.
- Snyder, L. V., & Shen, Z.-J. M. (2019). *Fundamentals of supply chain theory*. Wiley Online Library.

- Sørensen, T. H., Olsen, K. R., & Vedsted, P. (2009). Association between general practice referral rates and patients' socioeconomic status and access to specialised health care: A populationbased nationwide study. *Health policy*, 92(2-3), 180–186.
- Stavrunova, O., & Yerokhin, O. (2011). An equilibrium model of waiting times for elective surgery in nsw public hospitals. *Economic Record*, 87(278), 384–398.
- Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting—a review. *Renewable and sustainable energy reviews*, *16*(2), 1223–1240.
- Syed, S. T., Gerber, B. S., & Sharp, L. K. (2013). Traveling towards disease: transportation barriers to health care access. *Journal of community health*, *38*(5), 976–993.
- Syntetos, A. A., Boylan, J. E., & Disney, S. M. (2009). Forecasting for inventory planning: a 50-year review. *Journal of the Operational Research Society*, *60*(sup1), S149–S160.
- Tinghög, G., Andersson, D., Tinghög, P., & Lyttkens, C. H. (2014). Horizontal inequality in rationing by waiting lists. *International Journal of Health Services*, *44*(1), 169–184.
- Uiters, E., Deville, W. L., Foets, M., & Groenewegen, P. P. (2006). Use of health care services by ethnic minorities in the netherlands: do patterns differ? *The European Journal of Public Health*, 16(4), 388–393.
- Uthayakumar, R., & Priyan, S. (2013). Pharmaceutical supply chain and inventory management strategies: Optimization for a pharmaceutical company and a hospital. *Operations Research for Health Care*, *2*(3), 52–64.
- Van Doorslaer, E., Wagstaff, A., Van der Burg, H., Christiansen, T., De Graeve, D., Duchesne, I., ... others (2000). Equity in the delivery of health care in europe and the us. *Journal of health* economics, 19(5), 553–583.

- Vila-Parrish, A. R., Ivy, J. S., & King, R. E. (2008). A simulation-based approach for inventory modeling of perishable pharmaceuticals. In 2008 winter simulation conference (pp. 1532– 1538).
- White, H., Walsh, W., Brown, A., Riddell, T., Tonkin, A., Jeremy, R., ... Kritharides, L. (2010).
 Rheumatic heart disease in indigenous populations. *Heart, Lung and Circulation*, 19(5-6), 273–281.
- Wilson, D., & Barton, P. (2012). Indigenous hospital experiences: A new zealand case study. *Journal of clinical nursing*, 21(15-16), 2316–2326.
- Yeates, K. E., Cass, A., Sequist, T. D., McDonald, S. P., Jardine, M. J., Trpeski, L., & Ayanian, J. Z. (2009). Indigenous people in australia, canada, new zealand and the united states are less likely to receive renal transplantation. *Kidney international*, 76(6), 659–664.
- Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, *50*, 159–175.
- Zhang, G. P. (2004). Neural networks in business forecasting. IGI global.

Appendix A

Additional results

A.1 Coefficient plots for ethnic gaps by each procedure



Figure A.1: Coefficient plots by each procedure: Maori



Figure A.2: Coefficient plots by each procedure: Pacific



Figure A.3: Coefficient plots by each procedure: Asian

A.2 Coefficient plots for ethnic gaps by each DHB



Figure A.4: Coefficient plots by each DHB: Maori



Figure A.5: Coefficient plots by each DHB: Pacific



Figure A.6: Coefficient plots by each DHB: Asian

A.3 Additional OLS results

Table A.1: OLS Results: All ethnicities and Procedure-by-DHB fixed effects

	All ethnicities	Procedure by DHB FE	Dependent Variable: CPAC scores
(Intercept)	4.19***	4.51***	3.74***
	(0.67)	(0.66)	(0.03)
log(CPAC Score)	-0.20***	-0 34***	
log(CIAC Score)	(0.01)	(0.02)	
REF: Female			
Male	-0.02*	-0.02**	0.00
	(0.01)	(0.01)	(0.00)
REF: NZ European			
Maori	0.03**	0.04**	0.01*
	(0.01)	(0.01)	(0.00)
D :C	0.00***	0.05**	0.01*
Pacific	$(0.08)^{-0.02}$	0.05**	0.01*
	(0.02)	(0.02)	(0.01)
Asian	0.08***	0.04^{*}	-0.00
	(0.02)	(0.02)	(0.01)
Other	0.02		
	(0.02)		
REF: SES 1			
SES 2	0.01	0.00	0.00
	(0.01)	(0.01)	(0.00)
SEC 2	0.01	0.00	0.00
3E3 3	(0.01)	(0.01)	(0.00)
	(0.01)	(0.01)	(0.00)
SES 4	-0.00	-0.00	0.01
	(0.01)	(0.01)	(0.00)
SEC 5	0.02	0.02	0.00
3E3 J	(0.02)	(0.02)	(0.00)
	(0.01)	(0.01)	(0.00)
log(age)	0.08***	0.07***	0.01
	(0.02)	(0.02)	(0.01)
Coat Waisht	0.07*	0.04	0.04***
Cost weight	-0.07	-0.04	(0.00)
	(0.05)	(0.05)	(0.00)
DHB Effects	1	X	×
DDGEG	,		~
DRG Effects	1	X	×
Procedure by DHB Effects	×	1	1
		-	-
Calendar Month	1	1	\checkmark
	,	,	~
CPAC Codes	1	\checkmark	×
CostWeight Codes	1	1	1
	2	2	-
R ²	0.40	0.44	0.82
Num. obs.	36347	35349	35349

EXAMPLE 1005. Store p < 0.01, p < 0.01, p < 0.05. Standard Errors are reported in the parenthesis. Results for DHB, DRG, calendar month, and CPAC codes dummies are omitted.

A.4 Econometric diagnostic tests results

Regression Models	Corresponding Table	Weak- Instrument Test (F-stat)	Hausmann Test (p-values)	Sargan Test (p-values)
Main Results	2.2	10.16	0.00	0.06
Alternate IVs	2.4	6.71	0.00	0.99
Median Waiting Times	2.5	10.16	0.00	0.08
Percent Asian	2.6	9.20	0.00	0.89
Percent Pacific	2.7	10.22	0.00	0.34
Orthopaedic Surgery	2.8	11.72	0.00	0.30
DHB Fixed Effects	2.9	1.35	0.00	0.49

Table A.2: Econometric diagnostic tests results: Waiting time equations

Table A.3: Econometric diagnostic tests results: Demand equations

Regression Models	Corresponding	Weak-	Hausmann Test	Sargan Test
	Table	Instrument	(p-values)	(p-values)
		Test (F-stat)		
Main Results	2.2	67.40	0.00	0.71
Equilibrium Model	2.3	67.40	0.00	0.69
Alternate IVs	2.4	46.98	0.00	0.51
Median Waiting Times	2.5	45.32	0.00	0.76
Percent Asian	2.6	67.31	0.00	0.25
Percent Pacific	2.7	67.65	0.00	0.70
Orthopaedic Surgery	2.8	80.71	0.00	0.65
DHB Fixed Effects	2.9	27.88	0.00	0.09

Table A.4: Econometric diagnostic tests results: Supply equations

Regression Models	Corresponding	Weak-	Hausmann Test	Sargan Test
	Table	Instrument	(p-values)	(p-values)
		Test (F-stat)		
Main Results	2.2	10.16	0.00	0.06
Equilibrium Model	2.3	10.16	0.00	0.06
Alternate IVs	2.4	4.35	0.00	0.24
Median Waiting Times	2.5	19.88	0.00	0.83
Percent Asian	2.6	32.02	0.00	0.10
Percent Pacific	2.7	32.34	0.00	0.70
Orthopaedic Surgery	2.8	28.10	0.00	0.27
DHB Fixed Effects	2.9	11.92	0.00	0.25