Running Some Tests: Essays on Doctors, Nurses and Hospital Health Care

Sophie Sam Joyce

A Dissertation
Presented to the Faculty
of The University of Auckland
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Economics
Adviser: Associate Professor Rhema Vaithianathian

March 2013

Abstract

This dissertation consists of three essays on the economics of health-care delivery in hospitals. The first two essays estimate the impact of doctor-patient demographic concordance on a doctor's decision making for diagnostic resources and medical treatments. Demographic concordance occurs when a doctor and patient have the same ethnic group and/or gender. The third essay estimates a relationship between ward-level nursing hours and a patient's health outcome. These three papers use detailed data obtained by me from a hospital. Ethics approval to use this data was granted from the New Zealand Health and Disability Council (Reference number: NTX/11/EXP/029).

The first essay estimates a relationship between gender and/or ethnic concordance between a doctor and patient and the amount of diagnostic tests ordered during a hospital stay. Diagnostic test orders have increased in many developed countries. For example, in the United States the cost of 'unnecessary' diagnostic tests and procedures has recently been estimated at between USD 200 to 250 billion per annum (Berwick and Hackbarth, 2012; Thompson, 2011). Therefore, ways to reduce unnecessary diagnostic test ordering is of interest to health policy makers. I test whether doctors order higher or lower amounts of diagnostic tests when they treat patients with the same demographic features, relative to when they treat patients with no shared demographic features. I find a statistically significant reduction in laboratory and radiology tests when a doctor treats a patient of the same gender and/or ethnicity relative to demographically discordant patients. Assuming demographic concordance variables are exogenous, I suggest two reasons for a reduction in diagnostic test orders. The first reason is an information gain in demographically concordant consultations. Because information on a patient's health status comes from doctor-patient consultation and diagnostic tests, a reduction in diagnostic test orders suggests a doctor has obtained better quality information from consultation. An improved consultation

could include gains in communication, and/or the physical exam. The information gain hypothesis only holds if preferences for the amount of laboratory and radiology tests do not change in demographically concordant relative to discordant pairs. The second reason for a reduction in test orders is if demographically concordant patients and/or doctors prefer to order fewer diagnostic tests, for example by choosing a less aggressive treatment plan. Unlike the United States, I do not expect litigation and insurance arrangements to explain a reduction in diagnostic tests, because health care is publicly provided and doctors are not at a personal risk of litigation in the hospital.

The second essay estimates a relationship between doctor-patient ethnic concordance and a women's likelihood of having an emergency caesarean procedure. Studies have documented variation in caesarean procedure rates across ethnicity groupings within a country (Rumball-Smith, 2009; von Katterfeld et al., 2011; Vangen et al., 2000; Getahun et al., 2009). This paper makes a novel contribution to literature explaining ethnic-based variation in caesarean rates by investigating the effect of provider-patient ethnic concordance or discordance on the decision to have an emergency caesarean. An emergency caesarean is decided after a women has gone into labour, and women who receive a planned caesarean are excluded from my sample. Differences in the unobserved health status of women in ethnically concordant and discordant groups is therefore not expected to explain my results, because all women in the patient sample have been considered physically able to undergo a natural birth. I use the three largest casemanager ethnicity groupings in my data; European, Indian and Asian. I find that Asian women with an Asian casemanager are on average 6% ($p = .0001$) less likely to have an emergency caesarean compared to an Asian women treated by a European or Indian casemanager[1]. Ethnic concordance for European and Indian patients is statistically insignificant. I suggest three explanations for why Asian women are less likely to receive an emergency caesarean when treated by an

[1]A casemanager could be a midwife, staff nurse or doctor that is primarily responsible for a patient's care in hospital.

Asian casemanager. These are a reduction in maternal distress, clinical uncertainty and/or patient preference for an emergency caesarean. Two primary reasons for ethnic-based variation in caesarean rates are differences across ethnicity groupings in unobserved patient health characteristics and preferences for caesarean procedures. My result suggests that ethnicity-specific health characteristics and preferences do not fully explain a higher caesarean rate for Asian women in New Zealand.

The third essay estimates a relationship between ward-level hours of nursing staff and a patient's health outcome. Patient health outcomes are mortality and length of ward stay. There is a large body of empirical literature on the relationship between hospital nurse-to-patient ratios and patient health outcomes (for reviews see; Lang et al. (2004); Kane et al. (2007)). This paper contributes to this empirical literature in three ways: by using a detailed nursing staff dataset, using a novel instrumental variable for nursing hours and by considering the separate effect of nursing and patient hours in a ward on a patient's health outcome. My instrumental variable is the amount of sick and bereavement leave taken by nurses on a ward. Initially, there is a statistically significant positive relationship between nursing hours on a ward and a patient's likelihood of mortality. After instrumentation, nursing hours on a ward has a negative, but statistically insignificant, effect on the likelihood of mortality. A patient's length of stay is modeled with a competing risk survival model. Discharge home is the main outcome. Competing risks are transfer to another health-care facility and in-hospital mortality. My main result is that cumulative exposure to higher patient hours on a ward is associated with a longer hospital stay in 16 out of 20 wards. An explanation for this is that increased demand by other patients on fixed hospital resources, such as medical equipment and doctor and nurse time, lowers the ability to deliver timely hospital health care. As a result, patients stay in hospital longer to receive the health care they need. This information could be

useful for hospital administrators, because it suggests improving patient flow through a hospital during high demand times could reduce the average length of stay.

Contents

List of Tables

List of Figures

Chapter 1

Diagnostic tests and Provider-Patient Demographic Concordance

1.1 Introduction

The overuse of diagnostic tests and procedures has received considerable attention in recent years, particularly in the United States where the cost of unnecessary diagnostic tests and procedures has been estimated at USD 200 to 250 billion (Berwick and Hackbarth, 2012; Thompson, 2011). The concern over wasteful diagnostic test orders is also demonstrated in a recent initiative by the American College of Physicians to publish lists of diagnostic tests and procedures considered unnecessary in a given clinical situation (Cassel and Guest, 2012).

In addition to contributing to wasteful medical spending, diagnostic tests can be harmful and can put patients at risk of unnecessary medical treatment (Korenstein et al., 2012; Volk and Ubel, 2011). Firstly, radiology tests expose patients to harmful radiation. Secondly, diagnostic tests can return a false positive and a patient can undergo treatment for a condition they do not have. A related problem is a test revealing a clinical abnormality which may not develop into a threat to patient health if left untreated. This generates further medical treatment that is not warranted by a patient's true clinical need. Sutton (2011, pg.1600) characterised a doctor's diagnostic situation thus; 'Sifting through and assigning importance to vast quantities of unfiltered information is a defining skill of our [medical] profession. One of the unintended yet unavoidable consequences of increased diagnostic information is increased false positives.'

This paper estimates a relationship between doctor-patient gender and/or ethnic concordance and the amount of laboratory and radiology resources used in hospital treatment. Several papers have suggested that there are improvements in doctor and/or patient satisfaction with a consultation when a doctor and patient are demographically concordant (LaVeist and Nuru-Jeter, 2002; Street Jr et al., 2008; Sandhu et al., 2009). One outcome of an improved consultation could be fewer diagnostic tests and procedures. This could occur through improvements in doctor-patient

trust, communication and other factors that encourage more efficient resource use. On the other hand, demographically concordant doctor-patient pairs could result in higher resource use if doctors tend to provide more medical treatment to patients of their own demographic group, relative to different demographic groups.

I use administrative data from a large hospital. This data records each laboratory and radiology test ordered during a patient's hospital stay. The data also records a primary doctor for each inpatient event[1]. To determine whether demographic concordance occurred, doctors' demographic characteristics are obtained from a Human Resources department. The study period is 2004 to 2011 and there are approximately 225,354 inpatient events treated by 324 unique doctors in the regression analysis.

I find a statistically significant reduction in laboratory and radiology tests ordered during a hospital stay when a doctor and patient have the same gender and/or ethnicity, relative to when a doctor treats a patient with no demographic concordance (on gender and/or ethnicity).

A doctor treating a patient of the same gender (or ethnic group), relative to a patient with no demographic concordance, is associated with an on average reduction in the total *cost* of laboratory tests of approximately \$4.5 (or \$4.6). Gender (ethnic) concordance is associated with an on average reduction in total laboratory *test* orders of approximately .7 (.8). Gender (ethnic) concordance is associated with an on average reduction in the likelihood of prescribing one or more radiology tests of 1.1% (2.5%). This paper also finds a greater reduction in diagnostic resources when patients and doctors have the same ethnic group *and* gender, relative to when they have only gender or ethnic concordance.

The main feature of my empirical strategy is a fixed-effect on doctor and Major Diagnostic Category. I therefore estimate the relationship between demographic con-

[1]A primary doctor is defined by the hospital as the main doctor responsible for a patient's hospital stay. I use primary doctors to determine whether demographic concordance occurred for an inpatient event. This is discussed further in Section 1.4.7

cordance and diagnostic resources only within patients a doctor treats. Secondly, I assume doctor-patient demographic concordance is in practice exogenous. However, if informal sorting of patients to doctors occurs - because of demographic characteristics - this would likely generate a positive bias in the amount of diagnostic resources. Informal sorting would therefore be associated with a bias that is opposite to the negative relationship in my data. I discuss the assignment of doctors to patients during a hospital stay in Section 1.4.7.

I suggest two explanations for a reduction in diagnostic resources when doctors and patients are demographically concordant.

The first explanation is improved information from a consultation when a doctor and patient are demographically concordant. Doctors obtain information from two sources when diagnosing and treating patients. The first source is information from the consultation process. This includes information from communication and a physical exam. The second source is information from diagnostic tests. Because I observe a reduction in diagnostic tests, this suggests more information on a patient's health condition has been obtained from the consultation.

This explanation is supported by empirical studies on the quality of communication when doctors and patients are demographically concordant on gender and ethnicity (Street Jr et al., 2007; Van Ryn and Burke, 2000; Gordon et al., 2006). These papers use data collected from surveys after a consultation has occurred, or researchers code observations from video-taped consultations. The quality of a consultation has been measured by; its length, the content of questioning, body language, and satisfaction rated by doctors and patients. These papers have suggested that there is improved consultation quality in demographically concordant pairs, though sample sizes tend to be small. Because my paper uses changes in diagnostic resources to infer consultation quality, it is a unique contribution to the literature on consultation quality in demographically concordant doctor-patient pairs.

Balsa and McGuire (2001, 2003) model improved information in ethnically concordant relative to discordant doctor-patient consultations. In their model, doctors interpret a noisier signal of a patient's health condition when ethnically discordant. A doctor may then over- or under-treat a patient, because the quality of information on a patient's health condition is poor. A related area of research is on the revealed preferences patients have for the demographic characteristics of their General Practitioner (GP). For example, Godager (2012) finds that both male and female citizens in Norway prefer a GP of the same gender. They use revealed preferences of citizens that were asked to rank their most preferred GP.

A second explanation for a negative relationship between demographic concordance and diagnostic resources is that preferences for the amount of laboratory and radiology tests could change when a doctor and patient are demographically concordant. For example, doctors and/or patients may prefer a less invasive treatment plan in demographically concordant pairs. This explanation would also be consistent with fewer diagnostic tests. I do however control for treatment decisions which should mitigate some of the differences in treatment plans across demographically concordant and discordant pairs. My paper is unable to distinguish between improved consultation and preference change explanations for a reduction in diagnostic resources in demographically concordant pairs.

Two reasons for the overuse of diagnostic resources in the United States are litigation and insurance arrangements. Both of these explanations could be associated with fewer diagnostic tests ordered in demographically concordant doctor-patient pairs. For example, a doctor treating a patient of the same demographic characteristics could be less concerned about patient litigation and/or is more likely to provide medical care according to patient need rather than reimbursement incentives. My data is from a hospital in New Zealand, where health care is publicly provided and doctors are not at a personal risk of litigation. Therefore, litigation and insurance arrangements

are not expected to explain a reduction in diagnostic test orders in demographically concordant doctor-patient pairs. Lastly, I assume doctors are not prejudiced towards a patient of the same demographic group. That is, I assume doctors do not intentionally under-treat a patient of the *same* demographic characteristics because of a dislike for that demographic group.

My significant negative finding suggests that improvements in the doctor-patient relationship, for example through shared demographic characteristics, could reduce the use of diagnostic tests.

The next section outlines the theoretical and empirical literature on doctor and patient demographic characteristics and health care delivery. Later sections explain the data source and empirical strategy. Results and conclusion are in Section 1.6 and 1.7.

1.2 Background

1.2.1 Theory on doctor-patient demographic characteristics and clinical consultation

Balsa and McGuire (2003) model three ways a patient's ethnic group might enter into a doctor's decision making about medical treatment. The three explanations are prejudice, clinical uncertainty and stereotypes. These three explanations might explain the ethnic-based variation in medical treatment observed in the United States.

To set up their model, Balsa and McGuire use a baseline situation which leads to a fair and efficient allocation of medical treatment. In this model, there are two patient types; white and black. Black and white populations have a distribution of illness severity. This is normally distributed with a population mean and variance. Doctors have complete information on the severity of black and white patients in the baseline model. Patients will receive a medical treatment if their illness severity (i.e. expected

benefit from treatment) exceeds a threshold. This threshold is decided by doctors maximising the population's expected net benefit, given information on a patient's severity of illness. Because doctors have *complete information* on the severity of black and white patients, doctors select treatment if a patient's true health condition exceeds the threshold. The allocation is fair because black and white patients are treated according to their true illness severity, and not their race or ethnic group. The allocation is also efficient because no other patient can be made better off, without making another patient worse off[2].

The first explanation for ethnic-based health disparities are prejudices or biases doctors hold against an ethnic group. This involves a doctor treating a patient from an ethnic group with a lower regard than a patient from another ethnic group. This enters into the model as a transaction cost (e.g. psychological cost), based on Becker's specification of discrimination in the labour market. A transaction cost, for white doctors providing treatment to black patients, *raises the threshold* required for black patients to be treated. The allocation is unfair because black patients are under-treated relative to white patients, given their clinical need. Whether the allocation is efficient depends on the legitimacy of doctor's 'psychological cost'. Efficiency is concerned with the total welfare of society. Treatment for black patients may not be increased without being offset by the psychological cost to doctors and/or making white patients worse off. If this 'psychological cost' is not a legitimate preference, the allocation with prejudiced doctors would be inefficient.

The second explanation for ethnic disparities in medical treatment is clinical uncertainty. Balsa and McGuire separate clinical uncertainty into two mechanisms; (1) a miscommunication and (2) a rational profiling model. In the miscommunication model, black patients emit a *noisy signal* of their condition to white doctors. There

[2]A fair allocation is when every individual's treatment decision is independent of their race or ethnic group. An efficient allocation is defined as no other allocation that makes another individual better off without making another individual worse off.

is no noise in a white patient's signal of their condition. Doctors are Bayesian decision makers; the noisier a signal the greater weight doctors will put on racial priors instead of a patient's signal of their condition. Racial priors are the population means for illness severity. Doctors revert to these population means when deciding treatment for black patients. Black patients with a true severity below the threshold have a positive probability of treatment (and therefore harm, because treatment is excessive relative to clinical need). There is a risk of of non-treatment for black patients with a true severity above the threshold. This allocation is unfair because black and white patients with the same level of severity could be treated differently. The allocation is efficient because the doctor uses all information available to maximise the expected benefit of both black and white patients.

In the rational profiling model, doctors observe a noisy signal for both black and white patients. However, the distribution of the signal error is the same for black and white patients. The distribution of illness severity for black and white populations have *different means*, and the doctor is aware of this. Different means could come from differences in biological characteristics across populations. For example, a lower rate of metabolism for depression drugs in black than white patients could result in drugs administered in lower doses to black patients. The doctor will then set a lower treatment threshold for the group with a higher mean severity. In the profiling model, compared to the miscommunication model, white patients also have a risk of being mistreated. In addition, the risk of being over or under treated for patients depends on the relative population means of black and white groups. If the mean of black illness severity is lower then white, black patients have a greater risk of undertreatment than the risk of overtreatment. If the mean illness severity for black patients is greater than white, then black patients have a greater risk of overtreatment than the risk of undertreatment. The allocation of treatments in the rational profiling model is unfair because black and white patients with the same severity level are treated

differently. The allocation is however efficient: using a common threshold for both groups would generate greater misdiagnoses, because information on mean differences across the groups is excluded.

The last source for ethnic treatment disparities are stereotypes. Stereotypes are beliefs doctors have about the health-related behaviors of minority patients. These beliefs are inaccurate compared to beliefs in rational profiling. Doctors observe a patient's true severity accurately (no noise), and they value the net benefit of treatment for both groups equally (no bias). Balsa and McGuire show how a belief held by doctors that black patients are less likely to comply with treatment recommendations can generate a self-fulfilling equilibria. There is also a cost to putting in effort for doctors and patients. Effort by doctors is in observing and following up whether patients comply. Effort by patients is in complying with treatment. The net benefit from treatment therefore depends on both the doctor (to observe and follow up) and patient effort (to comply). These efforts can take high or low values. The doctor will maximise the patient's net benefit minus their own cost and benefit. There are two equilibria. Firstly, for doctors to observe a patient (high doctor effort), and patients to comply (high patient effort). Secondly, for doctors not to observe (low doctor effort), and patients not to comply (low patient effort). This equilibria can result in black patients recieving fewer treatments. If white doctors believe that black patients do not cooperate, this can generate self-fulfilling beliefs because doctors do not want to put in the effort to follow up and black patients have less incentive to comply (coordination failure). This allocation is unfair, because black and white patients are treated differently given the same illness severity. It is also inefficient because black patients and white doctors could reach a higher equilibrium of both cooperating.

Balsa and McGuire's model describes how a patient's ethnic group could enter into a doctor's decision making. Their model refers to medical treatment decisions, whereas I investigate diagnostic resource decisions. I discuss at the end of the next

section how a relationship between doctor-patient ethnic concordance and the amount of diagnostic resources relates to Balsa and McGuire's three explanations for ethnic-based disparities in medical treatments.

The 'principal-agent' theory in economics has also been applied to the doctor-patient consultation by Mooney and Ryan (1993). Their principal-agent model maximises the utility functions of doctors and patients in the doctor-patient encounter. In contrast to traditional principal-agent theory, the utility functions are interdependent when a doctor (agent) is treating a principal (patient). The principal is characterised by having less information than the agent (on medical diagnoses and treatments). Because utility functions are interdependent in health, the maximisation problem requires the doctor - who acts on behalf of the patient in discussing and administering medical treatments - to observe the patient's utility function. The contents of the patient's utility function have been discussed by Mooney and Ryan (1993). Mainly, the utility function contains the patient's health condition and preferences for medical treatment. The information problem characterising the agency situation is separated into; the extent to which the patient transfers information on their health and preferences, the extent to which the doctor transfers information on diagnoses and treatment options, and the exchange between doctor and patient in selecting a treatment that matches the patient's health needs and preferences (Scott and Vick, 1999, pg. 114). This information exchange characterises the maximisation problem in the doctor-patient relationship. Increased similarity between a doctor and patient may facilitate both parties to exchange more information, and for the doctor to act on behalf of a patient's utility function, suggesting an improvement in the consultation.

Empirical tests for statistical discrimination

My research question relates to empirical tests for statistical discrimination in health care, because improved doctor-patient communication is a plausible explanation for

a reduction in diagnostic tests when a doctor and patient have the same demographic characteristics. Empirical tests for statistical discrimination test for the effect of information from doctor-patient communication on treatment decisions. Balsa et al. (2005) and Chandra and Staiger (2010) test for statistical discrimination as an explanation for medical treatment disparities across ethnicity groupings. Grytten et al. (2011) test for statistical discrimination as an explanation for why highly-educated women are more likely to undergo a caesarean procedure in Sweden.

Balsa et al. (2005) empirically test for their two proposed mechanisms for statistical discrimination - miscommunication and rational profiling - in the diagnosis of hypertension, diabetes, and depression among black and white Americans. In the rational profiling model, doctors use prior beliefs about a patient's racial group when deciding medical condition and treatment. In the miscommunication model, doctors are less able to understand and interpret a patient's signal because they belong to a different ethnic group. The authors use data collected from a survey filled in by doctors and patients after a consultation.

To test for rational profiling, Balsa and McGuire construct priors for how likely it is for an age-gender-race combination to have a diagnosis of diabetes, hypertension, or depression from epidemiological data. They then look at the significance of these 'priors' on the diagnosis choice. If it is unlikely for a combination of age-gender-race to have a diagnosis, and a patient fitting that group is less likely to receive a diagnosis, then this is evidence of rational profiling - because doctors are using prior information when assigning a diagnosis. They find evidence of this behavior in the diagnosis of hypertension, diabetes, and depression. To test for miscommunication, the authors use an interaction term composed of a measure of a patient's health signal and race. This tests if patients from different ethnicity groupings send different signals when consulting doctors. The measure of a patient's signal of their condition was obtained

by surveying patients on how they communicated their symptoms. They find evidence for this in diagnosing depression in black patients.

Chandra and Staiger (2010) test for statistical discrimination and prejudice as two competing explanations for ethnic-based disparities in treatment. They theoretically model the doctor's medical treatment decision process. They then compare what their model implies for medical decisions to data on medical procedures. Specifically, they investigate if under-treatment of black patients is due to prejudice or statistical discrimination. In their model, doctors select a cardiac revascularisation procedure for an Acute Myocardial Infarction (AMI) patient if they anticipate their *marginal benefit* of treatment is above a threshold level.

The authors use differences in patient outcomes to infer whether a patient group has a higher or lower marginal benefit from treatment. The authors argue that, controlling for the propensity to be treated, prejudice against minority groups would show itself as a higher marginal benefit after treatment for minority patients. This is because minority patients receive fewer procedures, but show greater benefit from such procedures. On the other hand, statistical discrimination would imply that patient outcomes (marginal benefits from treatment) are equalized across majority and minority groups, despite black and white populations having differing rates of cardiac procedures. Different rates of procedures across populations would suggest doctors take into account differences in the population mean severity (i.e. racial priors about benefit from treatment) into the decision process. The authors control for the severity of a patient through the Charlson score. The authors do not find evidence for prejudice-based explanations, because black patients exhibit *lower* benefits (i.e. higher mortality rates) when treated with a cardiac procedure.

A shortcoming of Chandra and Staiger's argument is that doctors may choose to provide treatment to more severe patients, and these patients may have poor health outcomes. The authors argue that because minority patients have poor health

outcomes, this cannot be due to prejudice. Prejudice would however be consistent with poor health outcomes if doctors require a *higher severity* threshold for minority patients to be treated compared to majority populations.

Grytten et al. (2011) distinguish between statistical discrimination and agency theory as competing explanations for why highly educated women are more likely to receive caesarean procedures in Sweden. The authors argue that statistical discrimination theory implies more highly educated women ('expert patients') are better able to communicate their symptoms and preference for delivery methods than non-expert patients. They argue this will result in higher rates of caesarean section procedures for expert than non-expert patients. Agency theory assumes doctors have an incentive to avoid caesarean section procedures, because of the cost to hospitals. In this case, a doctor is more able to persuade a non-expert patient than an expert patient to forgo a caesarean. Agency theory is therefore also consistent with higher rates of caesarean section procedures for expert patients.

Increased use of diagnostic technology over time is used to distinguish between statistical discrimination and agency theory. Under agency theory, diagnostic technology would not reduce the gap between expert and non-expert patients in their rate of caesarean section procedures. This is because non-expert patients are unable to interpret and use the information from diagnostic devices. In this case, doctors are still able to persuade non-expert patients not to undergo a caesarean. If statistical discrimination is behind differences in expert and non-expert rates of caesareans, then the authors expect the gap between expert and non-expert mothers *to decline over time*. This is because a doctor relies less on patient signals of their condition (which advantages expert patients) and more on diagnostic technology. They find that disparities in caesarean rates for highly and low educated women reduced from 1967 to 2005, which suggests to the authors that statistical discrimination explains some of

the reason why more highly educated women are more likely to receive a caesarean procedure.

There are shortcomings to Grytten et al.'s argument. Grytten et al. relies on changes in diagnostic technology as the *sole* reason for a reduction in the gap of caesarean procedures between expert and non-expert mothers. It may be the case that non-expert patients are more aware of caesarean procedures now compared to the past. For example, caesarean may have become more fashionable for all citizens as countries have industrialised. Furthermore, the authors do not distinguish between emergency caesarean and planned caesarean procedures. These treatment decisions are made in different environments and diagnostic technology may play different roles in each decision. For example, because emergency caesareans are decided in labour, there may be little time to use diagnostic devices compared to ultrasound devices used during pregnancy. In this case, doctors would still rely on an expert patient's communication during labour, but communication would have less impact on planned caesarean decisions.

Testing for the presence of statistical discrimination in treatment decisions is difficult. Firstly, clinical uncertainty is hard to measure. In this paper, I use the amount of diagnostic test resources used during hospital treatment. Because diagnostic tests only provide information on a patient's health status, the number of tests ordered is indicative of the level of clinical uncertainty. This is because a doctor will gain information from the verbal and physical consultation and order their diagnostic tests based on this information. If there is more clinical uncertainty they will order a larger number of tests in order to arrive at a diagnosis and treatment plan. I therefore argue that the amount of diagnostic tests ordered is a good indication of clinical uncertainty in a consultation. Based on this argument we would expect from Balsa and McGuire's model that a doctor would order *fewer* diagnostic tests when demographically concordant. This is because patients have improved communication with a doctor when

ethnically concordant. A similar dynamic could be expected for when a doctor and patient are concordant on gender as well.

However, the second issue with measuring statistical discrimination is that it is difficult to show that clinical uncertainty is a *unique explanation* for medical treatment variation across groups of patients. For example, in Grytten et al.'s paper a change in caesarean procedures over time could be explained by changes in patient preferences for caesarean procedures, rather than the introduction of diagnostic technology. In this paper, I cannot exclude the explanation that *preferences* for the amount of diagnostic tests (by either doctor or patient) could be related to demographic concordance. That is, a patient or doctor may prefer to use fewer diagnostic resources when demographically concordant. I do however include variables for medical treatment decisions to control for variation in medical treatment decisions.

1.2.2 Empirical papers on doctor-patient demographic concordance and health outcomes

There are a number of empirical papers on the impact of doctor-patient demographic concordance on health care delivery. Health care outcomes have included; referral rates for cardiac operations, patient satisfaction with a consultation, length of consultation, doctor's treatment decisions, patient compliance with treatment, and so on. The effect of doctor-patient demographic concordance on the amount of diagnostic tests ordered is, to my knowledge, novel.

Consultation quality

A number of papers use survey data collected from patients and doctors after a consultation to study the relationship between demographic concordance (gender and ethnicity) and measures for consultation quality. The following reviews the main

papers in this literature, with an emphasis on how researchers obtained their measure of consultation quality.

Cooper-Patrick et al. (1999) investigate the effect of doctor-patient racial and gender concordance on participatory decision making (PDM) in a consultation. They survey 1816 adult patients after a consultation with a doctor at a primary care practice. Thirty two primary care practices in Washington DC, United States were involved in the study during the November 1996 to June 1998 period. The outcome (dependent variable) is a patient's percieved measure of how participatory the decision-making process was. This measure was developed by Kaplan et al. (1995), and is calculated from three questions patients are asked relating to how participatory they thought the consultation was. These are; '(1) If there were a choice between treatments, how often would this doctor ask you to help make the decision? (2) How often does this doctor give you some control over your treatment? and (3) How often does this doctor ask you to take some of the responsibility for your treatment?' (Cooper-Patrick et al., 1999, pg. 584). These are answered on a 0 (never) - 4 (very often) scale and are combined and converted to a 100 point scale for the dependent variable. Control variables include the patient's age, gender, education, marital status, self-reported health (5-point scale, poor to excellent), and how many years a doctor has been a patient's primary care doctor. They found that patients with racially concordant doctors were statistically significantly more likely to have a higher PDM measure. They also found that patients with a female doctor were on average associated with a higher PDM, but that gender concordance was not statistically significantly related to the PDM measure.

LaVeist and Nuru-Jeter (2002) look at the effect of doctor-patient racial concordance on patient satisfaction of care. Their measure of patient satisfaction is composed of answers to questions from 1 (poor) to 4 (excellent). Patients are asked to rate their doctor's performance based on the following five questions; '(1) provid-

ing good health care, (2) treating you with dignity, (3) making sure you understand what you've been told, (4) listening to your health problems, and (5) being accessible by phone or in-person' (LaVeist and Nuru-Jeter, 2002, pg. 297). These items were averaged resulting in an outcome ranging from 1 - 4 of how satisfied a patient was with their consultation. They find that patients with a doctor who is race concordant reported higher satisfaction with their doctor than patients who were not race concordant.

Street Jr et al. (2008) studied doctor-patient demographic concordance and communication in the consultation. They recorded consultations for 270 patients with 29 physicians. Patients completed a survey before and after a consultation. Consultations were recorded and researchers coded a consulation for observations on a doctor's informativeness, supportiveness and partnership building, as well as the patient's willingness and ability to communicate. They found some evidence of racial differences in communciation e.g. Asian doctors were more likely to percieve Black patients as significantly less effective communicators than White or Black doctors. In addition, black physicians felt that black patients were more satisfied with care received than Asian doctors, but there was no difference for black patients with white doctors.

Sandhu et al. (2009) reviews papers on gender dyads in doctor-patient relationships. Consultation outcomes in their review included; talk content, behaviour and non-verbal communication and the consultation length. Consultation length was the most studied outcome for gender dyad studies. The authors find that the majority of studies have shown some effect of gender concordance on consultation outcomes. The review also investigates differences in gender concordance, for example, female-female matches tended to involve longer consultations and contain more biomedical content than male-male matches. The authors also note that the evidence base is small.

Lastly, Stepanikova et al. (2006) find white patients treated by white doctors are less likely to report medical errors than if they are treated by a non-white doctor. For minority patients, the race of the doctor does not affect the probability of a patient reporting a medical error.

Medical treatment decisions

A smaller number of empirical papers study the effect of demographic, particularly racial, concordance on medical treatment decisions. The number of papers is small in part because administrative datasets rarely have information on the doctor's demographic characteristics. King et al. (2004) looks at the effect of racial concordance between black and white doctors and patients on the time to receiving protease inhibitors in patients with HIV. They find black patients received protease inhibitors significantly later when they were treated by a white doctor than a black doctor.

Chen et al. (2001) tested for a significant interaction in the race (black and white) of doctor and patient in the use of cardiac catheterisation following an AMI event. The interaction term is composed of dummy variables for a patient and doctor's ethnic group. They find that black patients had lower rates of cardiac catheterisation, and that this was *regardless of the race of the doctor*. The authors however do not control for a doctor's skill level in their interaction model. For example, whether a doctor held a specialist or junior position in the hospital. They include a control variable for the doctor's medical specialty; internal medicine, cardiology, other internal-medicine sub-specialty, and family practice. A dummy variable for the race of the doctor would however control for some of the unobserved differences in skill level across black and white doctors.

General Practitioner selection

In the economics literature, Godager (2012) model patient selection of GPs as a function of the age difference and gender concordance between a GP and patient. The authors use data collected from patient surveys in Oslo, Norway. The revealed preference for GPs is obtained from responses by inhabitants to rank their three most preferred GPs in descending order. The authors obtain a representative sample by selecting respondents in proportion to the characteristics of the population. They used a sample of 15,000 patients with 437 unique GPs. They find that patients from both genders prefer a GP of the same gender. They also find that higher educated women prefer a GP of a matching gender than a less educated woman. They also find that patients prefer a GP that is older rather than younger relative to their own age.

1.3 Data

In this section I firstly overview my data source. I then discuss how I obtained demographic information for doctors. I lastly summarise my laboratory and radiology data.

1.3.1 Source of data

Data was obtained from a large public hospital in New Zealand. I used three departments within the hospital to collect data. This resulted in a dataset with information on doctors' demographic characteristics, and the diagnostic test orders for each inpatient event. My dataset has the advantage of detailed information on diagnostic resources and doctor characteristics, but is limited to one hospital.

My dataset covers inpatient events between 2004 and 2011. This time period is used because information on diagnostic test orders was only recorded from the 2003

financial year onwards. The initial year of diagnostic test collection could be less reliable, because the system was newly implemented, hence I use inpatient events from 2004 to 2011.

The first source of data is electronic inpatient event records. An inpatient event is recorded when a patient spends more than three hours in hospital. Diagnostic coders working for the hospital will look at each inpatient admission and code an event if it is longer than three hours.

Inpatient event data contains information on the name of a patient's primary doctor. I use a patient's primary doctor to determine whether a patient was treated by a doctor with the same demographic features[3].

Information on a doctor's gender and ethnicity is obtained from the hospital's Human Resources (HR) department. Demographic information is self-reported by doctors on employment forms. Demographic information was obtained from HR for all nursing and medical staff that received payment from the hospital.

HR data is merged with inpatient data by matching on doctors' names. Details are in Section 1.3.2. Matching had to be undertaken on names because there is no unique identifier for employees used by both inpatient data services and HR. HR only collected demographic information from 2005 onwards. As a result, there is a possibility I do not observe a doctor that left employment before 2005. There is only one year of data between 2004 and 2005, and therefore the number of doctors in this position is expected to be small.

The third source of data is laboratory and radiology tests. This is obtained from a patient costing system that records tests ordered during a patient's hospital stay. Diagnostic tests are linked to each inpatient event by an identification code.

[3]A patient's primary doctor is referred to as a 'casemanager' in inpatient data. This is because a patient's casemanager may not be a doctor. For example, they might be a nurse or midwife. This paper uses the term doctor, instead of casemanager, for ease of understanding. It's a realistic assumption that most diagnostic test orders in a hospital would come from a doctor. Casemanagers are not always identified in the data as a doctor or nurse, so it is not possible to consider these casemanager groups separately.

1.3.2 Matching doctor demographic information to inpatient data

The main data source is inpatient event data. This records the name of the primary doctor[4]. To obtain demographic information on doctors, HR data was merged with inpatient data by matching the names of doctors. Ethics approval was obtained for this procedure (New Zealand Health and Disability Council, reference number: NTX/11/EXP/029).

HR data has information on the first, middle and last names of employees. Inpatient data has variable fields for a doctor's last name, first name, preferred given name and middle name. Unfortunately, these variable fields for a doctor's name are not always populated in inpatient event data. At worst, there is only information on the last name and first initials of the doctor. The matching process therefore worked sequentially by firstly merging with the most information on names. Patients that are not merged are then attempted to be merged requiring less information.

The first merge was on the full first, middle, and last name. The second merge used the full first and last name and third letter. The third merge used the full first and last name. The last merge was on the first letter(s) and full last name. For this set of merges, each unique merge was checked manually and flagged if the match did not look correct. A match is flagged incorrect if both datasets had information on first names and these did not match. Or, medical area (e.g. maternal, cardiology) and designation (e.g. medical officer or nursing staff) indicated in HR records did not match the type of patients treated by that doctor in inpatient data. After this matching process, a crosswalk was constructed to link inpatient data with HR information. I also check that my results are robust when estimated with doctors matched only on their full first and last name.

[4]I use the term 'doctor' henceforth to refer to the primary doctor for inpatient events.

I am able to match 417,119 eligible inpatient events with a doctor in HR records during the 2004-2011 time period. The number of eligible inpatient events during the sample period that I am unable to match is 114,070. Eligible inpatient events restrict patients to being aged 5 years or over, not entering hospital for an elective procedure, not being discharged from a rehabilitation facility, and lastly not having a primary doctor that is an 'AED Consultant' (i.e. doctor is not personally identified). Details for how this eligible patient population is selected is in Section 1.4.4.

Table 1.27 (in Appendix 1.E) has summary statistics of inpatient events for patients matched with HR data. Table 1.28 has summary statistics for the inpatient population that is not matched with HR data. Of all inpatient events not matched with a HR record, 30% are for MDC 14 which comprises mostly birth events (Table 1.28). It is quite likely the primary provider is employed by another organisation (e.g. a private obstetrician or midwife), but enters hospital to provide care when a women goes into labour.

There are several reasons a doctor in inpatient data is not matched to HR data. These include; HR not having a record of a doctor (e.g. if they are hired by another health care provider but work at the hospital), the doctor is a team of people, doctor was identified as a 'consultant', doctors recorded a different version of their name (e.g. Tony compared to Anthony), and misspelling in either inpatient or HR records of an employee's name.

Furthermore, of doctors that are matched to HR data, there is a sizable portion that did not complete information on their ethnicity and/or gender. Doctors are asked by HR to fill out forms with their age, ethnicity and gender and this is then entered into HR records. Some employees also entered an 'other' ethnic category, which resulted in an inability to determine ethnic concordance. Ethnicity and gender information is also self-reported by employees and is not checked for accuracy by another party.

Inpatient events that involve a doctor for which we do not have ethnic and gender information are excluded from my sample. Table 1.13 in Appendix 1.A has summary statistics for inpatient events where the doctor has gender and ethnicity information. There are 275,695 inpatient events with information from HR on doctor's ethnicity and gender.

Table 1.1 has demographic information for doctors that have information on gender and ethnicity. I was able to merge a total of 480 doctors, of which 472 have information on gender, 325 have information on ethnicity and 460 have information on age. There are 324 doctors with information on both gender and ethnicity. This population of doctors is used to estimate the diagnostic resource and demographic concordance relationship. Of the 324 doctors, 82.4% are European, 7.4% are Asian, and 4.9% are Indian. Other doctor ethnicity groupings are smaller. Of 324 doctors, 59.3% are male. The average number of patients per doctor in the study is 850 with a median of 418.

The number of doctors with less than 20 observations is 66, and 26 of these are in MDC 14 which corresponds to the Childbirth and Pregnancy medical category. These caregivers are quite likely midwives or obstetricians who enter hospital only when their patient gives birth. Another 15 are in MDC 3 which is the Ear, Nose and Throat category and involves a large share of prearranged surgical operations. Casemanagers in these roles (e.g. midwives and surgeons) could work for other employers, or on a part-time basis with the hospital. The number of patients per doctor is relevant to my empirical strategy. This is because I require sufficient numbers of patients within doctors to estimate coefficients in a fixed-effect model.

1.3.3 Laboratory and Radiology data

The quantity and cost of laboratory and radiology tests for an inpatient event are obtained from a patient costing system used by the hospital. All laboratory or ra-

Table 1.1: Doctor characteristics

	N	Mean	Sd	Min	Max
European doc	324	0.824	0.381	0	1
Asian doc	324	0.074	0.262	0	1
Indian doc	324	0.049	0.217	0	1
Age	319	43.944	9.433	22	71
Male	324	0.593	0.492	0	1
N patients	324	850.917	1199.508	1	8287
N merged	480				
N w gender	472				
N w ethnicity	325				
N w age	460				
N w <20 patients	66				
N w <20 patients in MDC 14	26				
N w <20 patients in MDC 3	15				

Notes: N is number of doctors with HR information on ethnicity and gender. Scalar statistics at bottom of table are for number of doctors merged, and merged with ethnicity, gender or age information.

diology tests a doctor orders during an inpatient event are recorded and attached a cost. In the following, I discuss how costs are assigned to tests, how diagnostic data is collected, and features of the laboratory and radiology cost data.

Costing of diagnostic data

Costs for each type of laboratory and radiology test are calculated internally by the organisation. Tests are given a cost by dividing overheads for the organisation (equipment, labour etc.) into the expected costs for a laboratory or radiology test.

The cost for each type of laboratory test can change over financial years. This change is mostly small, because the expected resource use for a test type is relatively stable. To avoid changes in test costs over financial years, the average cost for each type of test during the entire study period is used. Costing data for the financial year of 2002-2003 and 2003-2004 did not contain information on the type of test ordered. If a test was missing information on the type, the original cost assigned to the test was used. A dummy variable for each financial year would also capture changes in test costs over time in the organisation. Patient events that had missing information

on the type of test were also tagged and excluded from the sample as a robustness check.

Radiology data has an 'other' test type category. This cost varies across individuals. This therefore prevented using one average cost for each test type in radiology data. In any case, I do not use radiology costs, this decision is discussed in Section 1.4.3.

Collection of diagnostic data

An employee at the hospital is responsible for the maintenance of laboratory and radiology data. Because laboratory and radiology tests are recorded electronically and monitored by an employee for accuracy, missing observations of tests for an inpatient event are expected to be rare, or at least not systematic.

Approximately 20% of hospital inpatient events have no recorded laboratory tests. This may raise concern about missing observations of laboratory test orders. To investigate the completeness of laboratory data, patient characteristics for inpatient events that have zero laboratory tests are examined. I would, for example, not expect patients staying in hospital for a considerable amount of time to have no laboratory tests.

Table 1.30 in Appendix 1.G summarises inpatient event variables for patients with no laboratory tests. The 95th percentile of patients have a length of hospital stay that is less than two days. This suggests that patients with no laboratory tests comprise mostly short hospital stays. The maximum hospital stay (LOS) is 131 days. All patients with a LOS above the 95th percentile are rehabilitation patients. These individuals require long term care for non-acute conditions.

In addition, 38% (15%) of inpatients with no laboratory test are in MDC 2 (14) which is the medical category for eye related conditions (birth events). These medical

conditions are expected to be associated with fewer laboratory and radiology text orders.

Summary statistics therefore indicate that most inpatient events with no laboratory tests are the type of events we might expect to have no diagnostic tests, rather than instances of missing observations on diagnostic test orders.

The next section summarises the features of the laboratory and radiology data. These datasets are considered accurate and there are no significant concerns about missing observations in these datasets.

Laboratory data

Table 1.2 summarises the laboratory cost dataset used in my study. There are 1,261 unique laboratory test types for the population of inpatients in the study. The average test cost is $98 and the median cost is $24.47. There are over 11 million entries in the laboratory cost data. There are 361,734 entries missing information on the specific type of laboratory test. These are mostly from the 03/04 period (354,392). There are only 567 entries out of the total 11 million observations that have a zero or no allocated cost. Laboratory data is therefore fairly complete given the size and scope of the information it contains.

In laboratory cost data, a labour component is added for the collection of blood, this is separately identified and dropped from the data so that only costs for laboratory tests are used to measure laboratory resource use. Table 1.3 lists the top laboratory test types and the proportion of total tests that are of that type. The most common laboratory test is a Full Blood Count with a cost of $7.76.

Radiology data

Table 1.5 summarises the radiology cost dataset. There are 245 unique radiology tests. All entries missing information on the type of test are from the 2003-2004

Table 1.2: Cost information for each type of Laboratory test

	N	Mean	Sd	Max	P50	P90
Cost	1,296	93.26	184.03	2,811.95	34.47	240.61
N data entries	1,179,7102					
N entries missing type	361,734					
N entries missing type 03/04	354,392					
N entries missing type 04/05	6,921					
N entries zero cost	567					

Notes: N=1296 is number of unique laboratory tests. Scalar statistics are the number of laboratory dataset entries that are missing the type of laboratory test, or have a zero cost.

Table 1.3: List of the most common laboratory test types

Laboratory type	Cost	Proportion of test
ALT	1.97	.03
AST	1.98	.03
Albumin	1.86	.03
Alk Phos	1.97	.03
Bilirubin Total	2.07	.03
Creatinine	1.86	.08
Full Blood Count	7.74	.09
Film examination	16.58	.03
GGT	1.93	.03
Glucose	1.86	.04
Potassium	1.81	.08
Sodium	1.81	.08
Urea	2	.07

Notes: Column (1) is the laboratory test type, Column (2) is the cost in the data, Column (3) is the proportion of test type in the data.

financial year. There are only 188 entries of 804,770 observations of radiology tests that have a cost of zero.

Table 1.4 lists the most popular types of radiology tests. The most popular radiology item is a chest scan (on plain film), 30% of all radiology items are in this category with an estimated cost of $60.

Table 1.4: List of most common radiology test types

Radiology type	Cost	Proportion total test
Abdomen	90.7	.05
CT Chest, Abdo, Pelv	428.52	.04
CT Head	276.49	.05
CT Other	135.19	.06
Chest	61.87	.31
Gynae Ultrasound	81.68	.02
Mobiles	76.86	.06
Obstetric Ultrasound	71.62	.02
Pelvis and Hips	60.45	.03
Screening	229.22	.02
Tib, Fib, Ankle	59.74	.02
US Abdomen, Pelvis	122.01	.02
US Vascular	311.53	.02

Notes: Column (1) is the Radiology test type, Column (2) is the cost in the data, Column (3) is the proportion of test type in the data.

Table 1.5: Cost information for each type of Radiology test

	N	Mean	Sd	Max	P50	P90
Cost	245	487.92	716.55	5,961.14	294.48	908.43
N data entries	804,770					
N entries missing type	2,077					
N entries missing type 03/04	2,077					
N entries missing type 04/05	0					
N entries zero cost	188					

Notes: N=254 is number of radiology test types.

1.4 Empirical strategy

This section outlines my empirical strategy, starting with my baseline econometric model. I then discuss the laboratory and radiology resource outcomes, study sample, variables for demographic concordance and other explanatory variables in the model. I then discuss the assignment of patients to doctors, and accordingly whether demographic concordance variables are exogenous in the baseline econometric model. A discussion of alternative approaches to estimating the relationship between demographic concordance and diagnostic resources closes the empirical strategy section.

1.4.1 Econometric model

The main feature of my estimation approach is a combined fixed-effect on doctor and Major Diagnostic Category (MDC). A doctor-MDC fixed-effect estimates the effect of demographic concordance, relative to no demographic concordance, within each doctor's decision-making environment. This is because patients tend to present at hospital with symptoms originating in a biological area or organ system. Major Diagnostic Categories correspond to a single organ system that is the origin of disease. For example, circulatory or respiratory conditions. This is assigned by diagnostic coders after an inpatient event. Within this diagnostic situation, each doctor decides diagnosis and treatment using information from training, experience, patient consultation, and diagnostic tests. A fixed-effect on doctors therefore controls for heterogeneity across doctors in their patient case-mix, experience, training and other idiosyncratic test-ordering behaviours. The econometric model is indexed by inpatient event i, inpatient event $i's$ doctor j, and inpatient event $i's$ MDC, m. The baseline model is:

$$\begin{aligned} diag\_resource_{ijm} &= \alpha + \beta \mathbf{X}_i + \beta_1 gender\_concordance_i + \beta_2 ethnic\_concordance_i \\ &\quad + \beta_3 gender\_and\_ethnic\_concordance_i + \alpha_{jm} + \varepsilon_{ijm} \end{aligned} \tag{1.1}$$

$diag\_resource_{ijm}$ is the amount of inpatient event $i's$ laboratory or radiology resources, these diagnostic measures are discussed in Section 1.4.2 and 1.4.3. X_i is a set of control variables for inpatient event i, these are discussed in Section 1.4.6.

Demographic concordance is estimated by a set of dummy variables for each type of demographic concordance; ethnicity only, gender only and ethnicity plus gender. The base category is no demographic concordance, these variables are discussed in Section 1.4.5.

Baseline models are estimated with Ordinary Least Squares, though non-linear models are also estimated when diagnostic outcomes are discrete (see Sections 1.4.2 and 1.4.3).

My fixed-effect method estimates the difference in mean diagnostic outcomes for demographically concordant relative to discordant patients within each doctor and MDC combination, after controlling for explanatory variables in vector X. In model estimation, standard errors are clustered on the doctor and MDC.

I discuss whether demographic concordance variables are exogenous in the baseline model in Section 1.4.7. With exogenous demographic concordance variables, $\beta_1, \beta_2, \beta_3$ estimates the effect of ethnicity, gender, and ethnicity plus gender concordance, relative to no demographic concordance, on the amount of diagnostic resources within patients with the same MDC and doctor. Unobserved sorting of patients to doctors, because of demographic characteristics, is theoretically possible in the hospital. I however argue (in Section 1.4.7) that any plausible relationship between demographic concordance variables and the error term (ε_{ijm}) would result in a positive bias in the estimate of demographic concordance variables, and accordingly is opposite to the negative relationship I find.

1.4.2 Laboratory test outcome and model specification

I use two different laboratory test measures (outcomes), these are the total cost and quantity of laboratory tests ordered during $i's$ inpatient event. The total laboratory cost is the summation of laboratory test costs that were ordered during inpatient event i. Likewise, the quantity of laboratory tests is the total number of laboratory tests ordered during i's hospital stay. Table 1.6 summarises the laboratory (and radiology) outcomes for the eligible inpatient population.

Table 1.6: Diagnostic resource outcomes

	N	Mean	Sd	Max	P10	P50	P90	P95
Cost of lab.	275,697	206.3	472.3	3,0658.6	0.0	88.2	456.7	760.8
Ln lab cost	224,396	4.8	1.1	10.3	3.4	4.8	6.3	6.8
Quantity lab.	275,697	22.9	47.9	2511.0	0.0	11.0	52.0	85.0
Cost of rad.	275,697	319.6	1,266.0	114,668.5	0.0	55.4	752.2	1296.6
Ln rad. cost	156,567	5.4	1.2	11.6	4.0	5.2	7.1	7.6
Quantity rad.	275,697	1.5	2.8	164.0	0.0	1.0	4.0	6.0

Notes: Summary of diagnostic resource outcomes for the study sample.

Laboratory cost

I estimate the effect of demographic concordance variables for both raw and natural logarithm-transformed laboratory *cost* outcomes.

The advantage of a log-transformed laboratory cost, is that the impact of large cost values on coefficient estimates are reduced. From Table 1.6, the 95th percentile of laboratory costs is $760.80 and the maximum value is $30,658.60. This shows that there is a small number of inpatient events with very large laboratory cost values. Large values affect the estimation of the mean function using OLS, and their impact on mean estimation is reduced when log-transformed.

The second advantage of log-transformed laboratory cost, is that the distribution is more normal. Figure 1.2 plots the kernel density of the log-transformed laboratory cost. This figure shows a more normally distributed cost than the raw cost outcome, shown in Figure 1.1. Ordinary Least Squares assumes that error terms are normally distributed to enable efficient estimation of model coefficients. A more normally distributed outcome is more likely to have normally distributed error terms and therefore satisfy the efficiency condition for OLS estimation.

However, disadvantages of log transformations are that patients with zero laboratory cost (no laboratory tests) are excluded from the sample, and that model coefficients cannot be interpreted on the cost scale. Re-transformation of coefficient estimates to the cost scale, such as Duan's smearing procedure, is problematic for heteroskedastic data.

The advantage of raw laboratory cost outcomes are that coefficient estimates can be interpreted on the cost scale. Figure 1.1 plots the kernel density distribution of the raw laboratory cost. The 95th percentile of costs are excluded. There is a high density at zero, and a dip before increasing. The dip is due to a small number of patients who receive $1-3 of tests.

A disadvantage of the raw cost outcome is that outliers are larger, and large cost values can inflate coefficient estimates in the OLS model. Because of this, I exclude inpatient events with laboratory costs in the 95th percentile.

Laboratory cost models are estimated with Ordinary Least Squares regression. There is a large literature on models for health care expenditure data (Jones, 2011; Hill and Miller, 2010). These models are aimed to addressing the positive and right skewed nature of expenditure data. Jones (2011) and Hill and Miller (2010) compare estimation procedures. Both papers find OLS on raw costs (as opposed to log-transformed costs) is one of the best performing estimation methods. They assess the performance of models by comparing the Mean Squared Error of coefficient estimates. Other estimation methods that are similar to OLS in performance are; Extended Estimating Equations, and the Generalised Linear Model (Jones, 2011).

Laboratory quantity

I also estimate the effect of demographic concordance on the number of laboratory tests ordered during a patient's hospital stay. Figure 1.3 plots the frequency histogram for the quantity of laboratory tests ordered during each patient's hospital stay. Nearly 20% of all inpatient admissions have no laboratory test (to check for laboratory data accuracy, summary statistics for these inpatient events have been discussed in Section 1.3.3).

The quantity of laboratory tests is estimated with fixed-effect OLS and a fixed-effect count data model. Count data models are non-linear and designed for positive

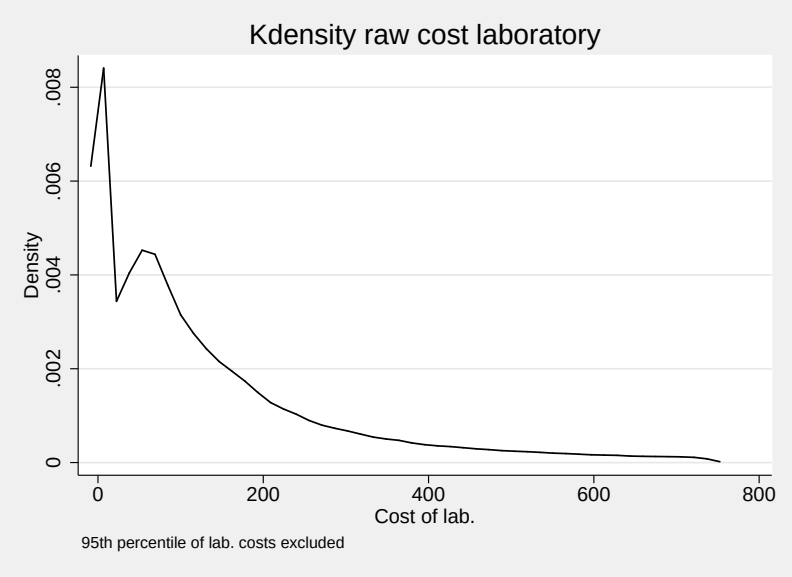

Figure 1.1: Laboratory cost raw scale

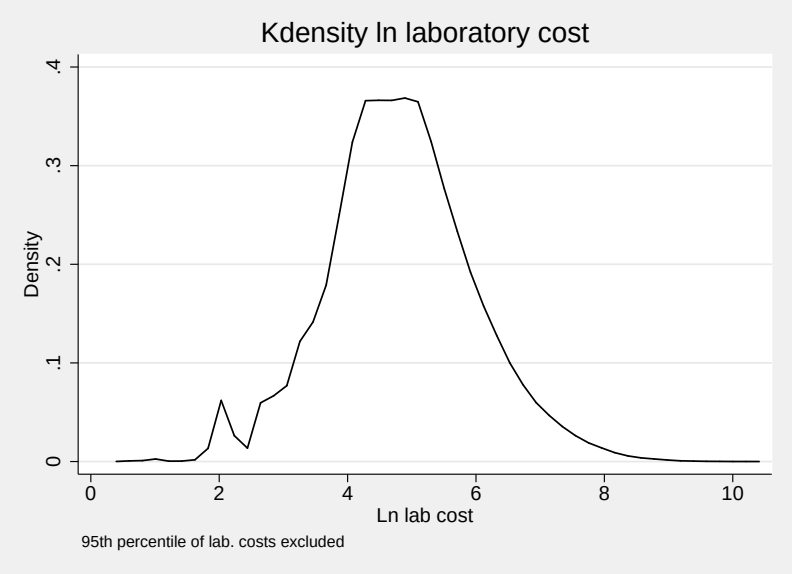

Figure 1.2: Laboratory cost ln scale

integer outcomes. I estimate both a Negative Binomial and Poisson model with bootstrapped standard errors. Bootstrapping standard errors should mitigate some of the problems associated with unobserved heterogeneity, and coefficient estimation, in non-linear models (Wooldridge, 2002, pg. 470).

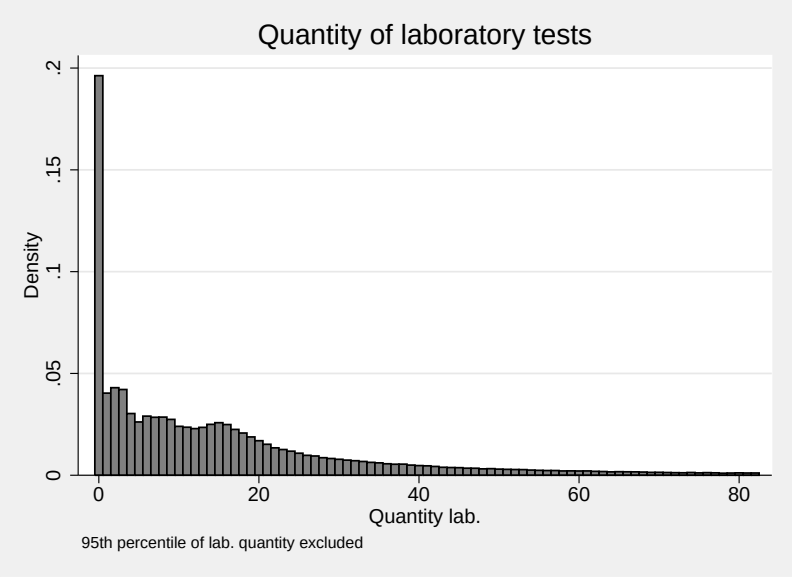

Figure 1.3: Laboratory quantity

1.4.3 Radiology test outcome and model specification

My radiology outcome is a binary variable for whether a patient received one or more radiology tests during their hospital stay. A positive incidence of a radiology test is informative compared to a positive incidence of a laboratory test. This is because 45% of patients do not receive a radiology scan, and the variation in the number of radiology tests ordered per inpatient event is low ($standard\_deviation = 1.2$). In contrast, most patients receive a laboratory test, and there is greater variation in the number of laboratory tests ordered ($standard\_deviation = 47.9$).

Figure 1.6 plots the frequency of radiology tests per inpatient event. Compared to laboratory tests (Figure 1.3), the number of tests per inpatient event is low. The percentage of inpatient events with less than 6 radiology tests is 95% and the mean number of radiology tests is 1.5.

The distribution of radiology *costs*, even in the log scale, is highly non-normal and has a thick right-tail. Figure 1.4 plots the kernel density distribution of radiology cost. There is a high density of costs at zero and a long right tail. Figure 1.5 plots the log-transformed radiology costs, this distribution is also highly non-normal, with

a spike around $40 and a thick right-tailed distribution. The higher density at $40 is due to a large proportion of chest and other plain-film scans in this cost bracket. Some radiology tests are significantly more expensive (e.g. MRI), which contributes to a thick right-tailed distribution. Fitting both the raw and log-transformed cost distribution with OLS is problematic, because assumptions of normally distributed error terms is unrealistic.

The binary radiology outcome is estimated with a fixed-effect linear probability model (OLS)[5]

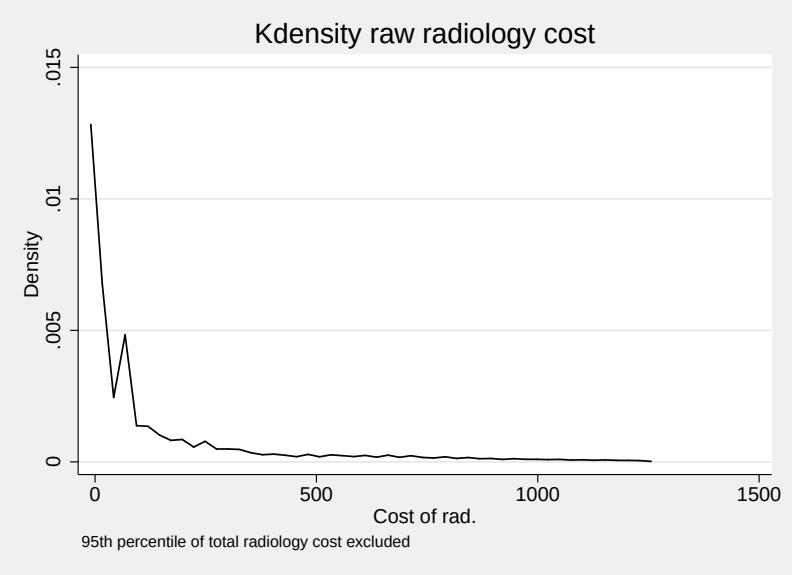

Figure 1.4: Radiology cost in raw scale

[5]I attempted to estimate the binary outcome with a fixed-effect logit model in *Stata 11*. With current computing capabilities I was unable to estimate the fixed-effect logit. Removing the fixed-effect component from a non-linear model is a computationally intensive process, which might explain why I was unable to compute coefficient estimates.

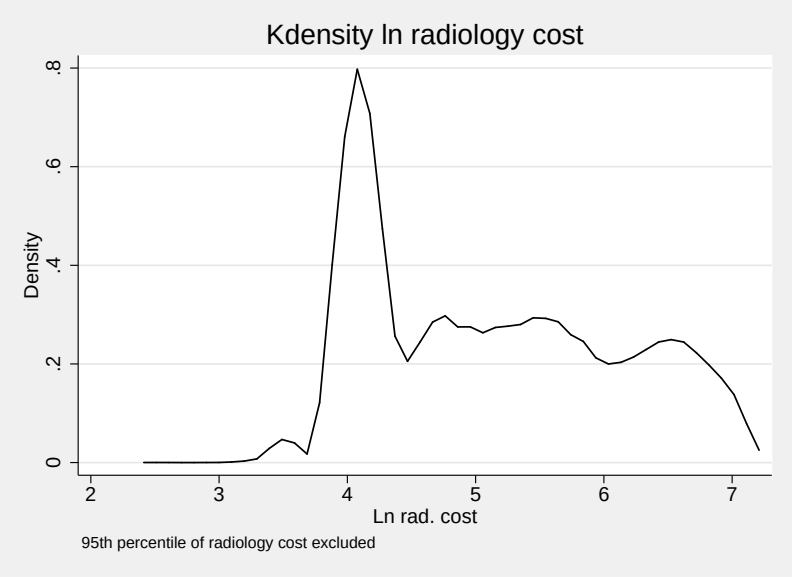

Figure 1.5: Radiology cost in ln scale

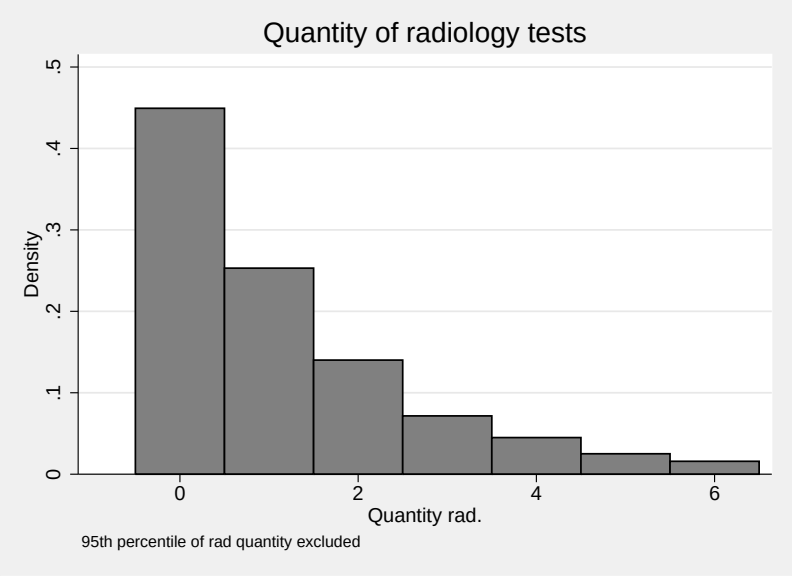

Figure 1.6: Radiology quantity

1.4.4 Study Sample

I exclude some types of inpatient events from my hospital data to obtain a more homogeneous study sample.

I exclude inpatients who (1) enter hospital from a waiting list (i.e. for an elective procedure), (2) are under 5 years of age and (3) are discharged from a rehabilitation unit.

The above patient restrictions select a more homogeneous sample of inpatient events. Specifically, I study inpatient events where there is some diagnostic uncertainty (i.e. excluding 'waitlist' patients), not an intended long term physical or mental rehabilitation process, and encounters involving doctor-patient communication (i.e. patients greater than 5 years of age)[6]. Robustness tests also confirm that the results do not change with the inclusion of these patients.

The remaining eligible inpatient events have a variety of medical conditions. These are summarised in Table 1.33, Appendix 1.I. This lists the proportion of total eligible patients for each Major Diagnostic Category. The top MDCs are Pregnancy, Childbirth and Puerperium (MDC 14, 16% of all inpatients), Circulatory System (MDC 5, 14%), Respiratory System (MDC 4, 9%), Digestive System (MDC 6, 9%), and Kidney and Urinary Tract (MDC 11, 6%). Within each MDC, the names and proportion of admissions for the top Diagnostic Related Groups are listed. For example, 19% of admissions in the Digestive MDC, have a DRG associated with treatment for abdominal pain, and 10% have treatment for a non-severe appendectomy. This table helps to illustrate that when patients arrive with a biological origin of illness, the attending doctor will have to decide if the diagnosis is more severe (e.g. DRG of appendicitis or AMI) or less severe (e.g. DRG of 'abdominal pain' or 'chest pain').

I also investigate estimating my baseline model on patients only in the top MDCs for hospital admissions, (excluding Pregnancy, Childbirth And Puerperium.) These MDCs are Circulatory, Respiratory, Kidney and Urinary tract, and Digestive condi-

[6]Waitlist patients enter hospital for a specific procedure that has been arranged prior to admission. Remaining inpatient admission types in the sample are acute and arranged admissions. Arranged admissions include patients that develop an acute condition but entry into hospital has been arranged by a GP. For Rehabilitation units, patients tend to have different treatment plans than acute care wards. Typically, patients have a high LOS and low diagnostic test orders. Lastly, I want to consider inpatient events involving doctor-patient communication, so I exclude patients under 5 years of age.

tions. Estimation across patients within MDCs selects a more homogeneous patient sample to estimate demographic concordance variables on.

1.4.5 Demographic concordance variables

Each type of doctor-patient demographic concordance is separated into mutually exclusive categories. The base category is no demographic concordance. Dummy variables are included for ethnicity only, gender only, and combined gender and ethnicity concordance.

Ethnic concordance occurs when a doctor and patient belong to the same ethnic group. Patients and doctors are associated with one of seven main ethnicity groupings; European, Maori, Pacific Peoples, Asian, Indian, Middle Eastern, and Latin American/other. These ethnicity groupings are broad and include different languages and countries. For example, Asian ethnic group includes Chinese, Korean, and Japanese ethnicity groupings. Some doctors enter more detailed information on their ethnic group, whereas other doctors enter broader ethnicity groupings. It is therefore not possible to determine more detailed (e.g. Chinese separately to Asian) ethnic concordance for all doctors.

HR data also has information on the age of doctors, a second set of demographic concordance variables is constructed using this information. Age concordance occurs if a patient has an age within five years above or below a doctor's age. Further analysis in the results section is conducted with age concordance.

The proportion of inpatient events with ethnic, gender, ethnic plus gender, and no demographic concordance is in Table 1.7. Of inpatient events, 25.2% have ethnic only concordance, 24.8% have gender only concordance and 31% have only ethnic plus gender concordance. The lowest proportion of inpatient events are in the base-category of no demographic concordance (19%). Similarly, when an age concordance variable

is included in the model, the proportion of inpatient events with no demographic concordance is reduced to 15.7%.

Table 1.7: Proportion of inpatient events in each demographic concordance category

	Proportion
Gender and ethnic concordance:	
Gender only	0.252
Ethnic only	0.248
Gender and ethnic	0.310
No demographic concordance (base cat.)	0.190
Gender, ethnic and age concordance:	
Ethnic only	0.217
Gender only	0.203
Age only	0.033
Gender and ethnic	0.262
Gender and age	0.049
Age and ethnic	0.031
Gender, age and ethnic	0.048
No demographic concordance (base cat.)	0.157
N patient events	275,697

Notes: Column (2) is the proportion of patients in each demographic concordance category in the eligible patient population.

Types of demographic concordance

Demographic concordance variables are a composite of different types of gender and ethnic pairs, for example Male-Male, European-European and so on. Table 1.8 summarises the proportion of each type of pair within a demographic concordance category (e.g. the proportion of gender-only concordance that is male).

The percentage of inpatient events with ethnicity-only concordance that are European is 97.2%. The percentage of inpatient events with gender and ethnicity concordance that are European is 71.8% (European and male) plus 24% (European and female).

The high number of European ethnic concordance is a limitation of my data. In particular, it is difficult to estimate the separate effect of a patient's ethnic group and ethnic concordance on diagnostic test ordering. A patient's ethnic group is separately controlled for to remove variation in diagnostic test ordering due to a patient's ethnic

group. Patient ethnic (or gender) groups may be related to diagnostic tests if distributions of health statistics differ across ethnicity groupings. For example, obesity and smoking is more prevalent in Maori and Pacific peoples, these health behaviours could be associated with higher amounts of diagnostic tests. Maori and Pacific people are also more likely to not have a doctor of the same ethnicity. The lack of variation in the data to estimate ethnic concordance, when patient ethnic group is controlled for, is further discussed in my result section.

Of inpatient events with gender-only concordance, 57% are male. There is therefore greater gender variation to estimate the separate effect of a patient's gender group, and gender concordance, on the amount of diagnostic resources.

Table 1.8: Proportion of each type of demographic concordance

	N	Proportion
Gender concordance only:		
Male	69,524	0.584
Ethnicity concordance only:		
European	68,332	0.972
Asian	68,332	0.009
Indian	68,332	0.013
Maori	68,332	0.004
Pacific	68,332	0.003
Other ethnicity	68,332	0.000
Gender plus ethnicity concordance:		
European and male	85,461	0.718
European and female	85,461	0.240
Minority ethnicity and male	85,461	0.009
Minority ethnicity and female	85,461	0.032
N patient events	275,697	

Notes: N is the number of patients in each demographic concordance category. Column (2) is the proportion of patients for each type of demographic concordance e.g. Male is Male doctor-Male patient.

Comparisons across demographic types of doctors

My empirical method estimates the *average* demographic concordance effect across different types of demographic pairs. For example, gender-only concordance is an average of male and female gender concordant pairs. The effect of demographic

concordance on diagnostic resources might differ across these types of pairs. For example, if male doctors are better communicators on average than female doctors, then demographic concordance may have little effect on a male doctor's diagnostic test ordering, because male doctors obtain adequate information from communication initially.

A problem with comparing demographic concordance across doctor gender and ethnicity groupings is that patient case-mix is expected to differ across demographic groups of doctors. Tables 1.32 and 1.31 in Appendix 1.H present summary statistics for the characteristics of male and female doctors in the data. Female doctors are less likely to provide a surgical procedure; 11% of female doctors have provided a surgical procedure compared to 19% for male doctors. Female doctors are also more likely to work in the MDC (14) associated with childbirth; 51% of all female doctors work in MDC 14 relative to 4.5% for males doctors. As a result, it is reasonable to expect medical area and skill level to differ across gender and ethnicity groupings of doctors.

Skill level and medical area could be related to the relationship between demographic concordance and diagnostic resources. For example, the impact of demographic concordance may be less important for complex illnesses, because test ordering could be driven by a patient's severity of condition, rather than communication. I'm therefore not able to distinguish between differences in *behavioural features* (e.g. communication abilities) and *case-mix* when explaining a difference in the relationship between demographic concordance and diagnostic resources across gender and ethnicity groupings of doctors[7].

[7]Even if we considered male and female doctors within a MDC, there could still be differences in the unobserved level of experience and skill level across doctor gender and ethnicity groupings, which could affect the diagnostic resource relationship.

1.4.6 Other explanatory variables

All explanatory variables are listed with summary statistics in Table 1.13, Appendix 1.A. I separate inpatient explanatory variables into; patient characteristics, timing of admission, type of hospital admission and clinical variables. These are discussed next.

Explanatory variables for patient characteristics are; a male dummy variable, age at admission, socioeconomic deprivation scale of neighbourhood, and patient ethnic group dummy variables. The socioeconomic deprivation scale ranges from 1 to 10, an increasing scale is associated with higher levels of deprivation in a patient's residential neighbourhood. Patient ethnic group dummy variables are; Maori, Pacific peoples, Asian, Indian, Middle Eastern, and Latin American/other. European ethnic group is the base category.

Explanatory variables for patient characteristics control for differences across patients in their demographic and socioeconomic features. Demographic and socioeconomic features could be related to diagnostic test orders through their association with comorbid conditions. For example, patients from lower socioeconomic areas tend to have higher rates of smoking and obesity, which could be related to increased diagnostic test orders because these health behaviours are associated with higher chronicity of health condition.

Explanatory variables for the timing of an inpatient's admission are *dummy* variables for; day of week a patient was admitted, year of admission, and timing of entry to hospital (dummy variables for admission between 12pm-5pm, 5pm-12am, overnight admission and morning admission is base category). Timing of admission could be related to diagnostic test orders if changes in laboratory opening and closing hours (e.g. during a day and over weekends) affect diagnostic test ordering. In addition, there might be long term changes in hospital policy affecting diagnostic test orders. These are captured by dummy variables for year of admission.

Explanatory variables for the type of inpatient admission are *dummy* variables for; transfer from another hospital, entry to hospital through the Accident and Emergency Department (AED), acute admission (relative to base category of arranged admission), an intended day patient admission, accident as cause of hospital admission, if a patient had one or more previous hospital admissions within the last 60 days, and if district health board for inpatient i is that of the hospital (variable 'Home DHB').

The following discusses how each type of hospital admission could be related to diagnostic test orders. Firstly, a transfer indicates an inpatient event has been referred from another health-care facility. This could be associated with fewer diagnostic test orders, because transfer patients are more likely to have previous diagnostic test orders on their patient history file. However, if more severe patients are more likely to be transferred, this could mitigate a negative association.

Secondly, entry to hospital through the AED would control for routine diagnostic test orders when patients present with a symptom. For example, doctors tend to order a fairly standard set of tests when patients arrive at hospital with chest or abdominal pain.

Thirdly, acute admissions could be related to higher test orders, because there is a lower likelihood of diagnostic test orders prior to entering hospital compared to arranged hospital admissions.

Fourthly, an intended day patient would also be associated with lower test orders, because it is more likely there is less diagnostic uncertainty when a patient enters hospital for a prearranged treatment.

Fifthly, patients that enter hospital where an accident has been indicated as the cause of entry would be associated with decreased diagnostic uncertainty and therefore diagnostic tests, because the cause of illness is known to the patient and doctor.

Sixthly, patients who have had one or more inpatient event within the last 60 days would likely be associated with fewer diagnostic test orders, because these patients will have diagnostic test results in their patient history file. I specify a 60-day window instead of a smaller (30-day) window to ensure all patients that have had contact with the hospital within a reasonable time-frame are indicated.

Lastly, district health boards are regional providers of health care. Each citizen is associated with a health board depending on their area of residence. Patients can be treated by health boards other than their residential one, payment is arranged across health boards for providing care. This could be related to test ordering if test orders were conducted in another hospital of the patient's DHB.

Explanatory variables for a patient's clinical condition for an inpatient event i are; length of hospital stay in days, the number of diagnoses, a dummy variable for a surgical theatre event, and dummy variables for each of Charlson's comorbid conditions.

I use the ICD-10-AM diagnostic codes for Charslon's comorbid conditions that have been converted from ICD-9 by Sundararajan et al. (2004). Charslon comorbid conditions control for specific diagnoses that are associated with severe or complex health conditions (these are listed in Table 1.13, Appendix 1.A). An explanation of Charlon's comorbid conditions is also in Sundararajan et al. (2004).

I do not include variables for every inpatient events' primary diagnosis. This is because individual diagnosis codes are too numerous to include in the model (there are over 20,000 ICD-10-AM diagnostic codes). In addition, I am interested in estimating the effect of demographic concordance on the amount of diagnostic test orders required to diagnose and treat a patient. The specific diagnosis arrived at through a doctor's decision-making process is therefore not a necessary control variable in my model - unless demographic concordance is related to unobserved patient severity, I discuss this possibility in the next section.

Lastly, in a robustness check, I include dummy variables for primary diagnosis for patients with (1) Kidney plus Liver and (2) Digestive conditions. Demographic concordance variables remain statistically significant. This suggest that results are not sensitive to differences in the unobserved medical diagnosis of patients in demographically concordant relative to discordant pairs.

Variables for theatre event (surgical procedure) and the length of hospital stay control for the impact of medical treatment decisions on the amount of diagnostic tests ordered. If a doctor's treatment decision is affected by demographic concordance, and treatment decisions are related to the amount of diagnostic tests ordered, any relationship between concordance and test ordering could be due to treatment decisions, rather than communication between a doctor and patient. Controlling for medical treatments received during an inpatient event accounts for variation in test ordering due to medical treatment decisions, rather than the amount of diagnostic resources required to understand and monitor a patient's health status.

1.4.7 Primary doctors and assignment of patients to doctors

This section discusses how doctors are assigned patients and how this could potentially affect the estimation of unbiased coefficients for demographic concordance variables.

I use the primary doctor, recorded in inpatient event data, to determine whether demographic concordance occurred for an inpatient event. The primary doctor is defined by data collections as the main doctor responsible for an inpatient within the hospital unit where the patient spent the majority of their hospital stay. For example, Cardiology, Respiratory, and General Medicine are all hospital units.

A patient can be treated by doctors other than their primary doctor. This could occur either within a hospital unit, or across hospital units during an inpatient event. For example, a patient that enters hospital through AED and is transferred to the

Cardiology unit would have a primary doctor that is associated with the Cardiology unit, but would have been treated by an AED consultant as well.

In this paper I associate all diagnostic test orders during an inpatient event to the primary doctor. Because patients can be treated by multiple doctors, understanding the initial *assignment* and subsequent *transfer* of patients to doctors is important to ensuring demographic concordance variables are exogenous to the error term (ε_{ijm}) in the diagnostic resource equation 1.1.

There are two *components* of the *error term* (ε_{ijm}) that could be *correlated* with demographic concordance variables. These are; (1) *noise* in the measure of diagnostic resource use (caused by doctors, other than the primary one, ordering tests) and (2) patient's unobserved severity of illness. I need to ensure these two components - noise and unobserved patient severity - are unrelated to whether demographic concordance occurs. I discuss each of these scenarios next.

The first possible endogenous relationship is if demographic concordance variables are related to the *noise* in the measurement of diagnostic resources ordered by the primary doctor. I associate all diagnostic test orders during an inpatient event to the primary doctor. Because a patient can be treated by multiple doctors, this generates noise in the measurement of diagnostic resources. I use the total test orders during a hospital stay because test decisions made by previous doctors is relevant information for a primary doctor's test ordering. For example, a primary doctor may not order a specific type of test, because it has been ordered by a previous doctor, and they can see the laboratory results on a patient's file. In addition, it is not possible to determine the ordering doctor for a diagnostic test in my data. I am therefore not able to determine if tests were ordered by another doctor than the primary one.

The second possible endogenous relationship is if demographic concordance is related to unobserved patient severity or complexity of illness. For example, if a more

severe or complex patient is more likely to be transferred to a doctor of the same demographic characteristics.

In this paper, to enable accurate estimates of demographic concordance variables, I argue demographic concordance is in theory exogenous, and if demographic sorting does occur informally and/or occasionally it would be associated with increased diagnostic resource use. This would bias the estimate of coefficients on demographic concordance variables upward and opposite to the (negative) relationship I find in my data. I explain my reasoning in the following paragraphs.

Firstly, the assignment of patients to doctors is in theory exogenous to ε_{ijm}. There is no formal hospital policy of sorting patients to doctors that have the same demographic characteristics. When patients arrive at hospital with an acute condition they are assigned a doctor from a pool of doctors on-call in a medical specialty for that day. A patient may already have an assigned doctor from a previous hospital stay, or through arrangement with a GP. Doctors in a hospital unit (e.g. respiratory or cardiology) takes turns being on-call. When a doctor is not on-call they are attending to their caseload. In practice, patients are not able to select their doctor when they arrive at hospital. After a patient has been assigned a doctor, they can be transferred to another doctor. A transfer could occur because of doctor shift changes, constraints on doctor's time (i.e. current caseload), a patient transfers to another medical specialty, another doctor has more experience or specialty in a medical area, or if a patient and/or doctor requests a transfer for another reason. It is therefore possible that a doctor or patient requests a transfer because of demographic characteristics.

Secondly, given patients can be transferred to other doctors, this paper is then concerned about *informal* sorting of patients to doctors of the same demographic characteristics. Here I argue that if a patient is transferred *because* of doctor and/or patient demographic characteristics, it is more plausible that patients are transferred

to a doctor of the *same* demographic features rather than a doctor of different demographic features.

The first endogenous relationship is if diagnostic test orders by other doctors are related to demographic concordance variables. I argue that sorting because of demographic characteristics only occurs to achieve demographic concordance. This endogenous transfer pattern is associated with a patient being treated by more doctors, and accordingly a patient would have higher, rather than lower, amount of diagnostic test orders. I therefore argue that is is only plausible that there is a positive relationship between noise and demographic concordance variables. A positive relationship is opposite to the negative relationship between concordance variables and diagnostic resources that I find in my data. Therefore any endogenous relationship between demographic concordance and noise in diagnostic resources would likely bias the coefficient estimate on demographic concordance in the opposite (positive) direction to the one I find.

The second source for an endogenous relationship is if patients are more or less likely to be transferred - because of demographic characteristics - depending on the level of their unobserved severity of illness. If unobserved severity and demographic concordance variables are related, I argue that it is more plausible that they would be positively related. That is, a patient of increased severity is more likely to be transferred to a doctor of the same demographic characteristics rather than a doctor of different demographic characteristics. Because more severe or complex patients are associated with higher diagnostic resource use, this would bias the estimate of demographic concordance coefficients in the opposite direction to the one I find.

It is important to remember that I am only concerned about transfers *because of* demographic characteristics that are related to a patient's unobserved severity level. A more plausible transfer, related to a patient's unobserved severity, is for more severe patients to be transferred to a more experienced doctor. This would not affect the

estimation of demographic concordance variables, because differences across doctors in the average severity of their caseload is controlled for by a doctor fixed-effect.

A related issue is if decisions on test ordering are influenced by doctors not listed in a patient's file, for example a supervising and junior doctor could consult over patient diagnosis and treatment. Given doctors' time is constrained and investing in a patient's diagnosis and treatment is particularly time consuming, it is plausible that the primary doctor is largely responsible for the diagnosis and treatment of patients during their hospital stay. We can therefore expect diagnostic test orders to be driven in *large part* by the primary doctor, with some inpatient events involving diagnostic test orders by previous doctors. In addition, a supervisory relationship for a junior doctor will apply to all patients a junior doctor treats. A fixed-effect on doctors should therefore remove the common effect of a supervisory relationship on doctors' decisions. This also applies if a doctor is more or less likely to consult with other doctors. A doctor fixed-effect removes the average effect of a doctor's unique consultation process on diagnostic test orders.

Another concern with diagnostic data collected in hospital is that a patient may enter hospital with prior diagnostic tests ordered outside of hospital. My data is unable to observe tests ordered for a patient outside of hospital. It is assumed that previous test orders are unrelated to demographic concordance variables. A dummy variable for previous hospital admissions within 60 days prior to a hospital event is expected to control for prior diagnostic test orders in a patient's history file.

A common transfer of doctors occurs when a patient arrives at the AED and is transferred to another department within the hospital. Patients arriving at an emergency department tend to get a standard range of tests given a presentation of symptoms (e.g. chest pain). A dummy variable for AED entry controls for test ordering associated with this entry method.

1.4.8 Method discussion

This paper uses a fixed-effect method to estimate concordance effects within doctors and MDCs. A shortcoming of a fixed-effect method is that variables for doctor characteristics cannot be included in the regression. For example, I cannot include variables for a male or European doctor and consequently investigate how these variables are related to diagnostic test ordering. In addition, because only variation within doctors is used to estimate the effect of variables on the outcome, there is less variation to identify coefficients than if there was no fixed-effect.

An alternative approach to estimating demographic concordance effects is to include separate dummy variables for a patient and doctor's ethnic and gender group, and use an interaction term to estimate doctor-patient ethnic and gender concordance on diagnostic resources. A shortcoming of a model with interaction terms is that patients are compared across doctors, as well as within doctors. Estimates for demographic concordance coefficients could therefore pick up the effect of differences in skill level across doctors on diagnostic resources. For example, if demographic concordance is more likely to occur in high cost medical specialties, this could bias the estimate of demographic concordance coefficients upwards. This is because a large portion of observations where demographic concordance occurs will involve doctors that use high amounts of diagnostic resources on average.

Because laboratory costs are continuous, it is possible to estimate the effect of demographic concordance at different percentiles of the laboratory cost distribution using quantile regression. Quantile regression estimates coefficients at different parts of the distribution of y, for example, the 25th or 75th percentile of costs. Quantile regression was developed for models without fixed-effects. Galvao (2011) has developed methods to estimate fixed-effect quantile models, but computation uses grid search and is not possible, given current computing capabilities, on models with more than three variables. If quantile regression is implemented without fixed-effects, problems

associated with comparing patients across doctors (i.e. no fixed-effect) may be minimised, because concordant and discordant patients are compared within a range of laboratory costs. That is, quantile regression will estimate the effect of concordance at low and high values of the laboratory cost distribution. Quantile regression is also less sensitive to outliers and heteroskedasticity of cost data, because coefficients are estimated in a limited range of laboratory cost outcomes. Quantile regression can also only be estimated on the log-transformed laboratory cost. This is because the quantile regression theory does not apply to truncated data (e.g. raw laboratory cost) or discrete outcomes.

1.5 Descriptive statistics

This section presents kernel density distributions for the logarithm of laboratory costs for demographically and non-demographically concordant inpatient events. These are raw distributions; there are no controls for the effect of medical condition, doctor, patient ethnicity or gender on test ordering.

Figures 1.7 to 1.9 plots the kernel density of log-transformed laboratory costs[8]. In all kernel density plots, demographically *discordant* inpatient events have a slightly higher density at higher laboratory costs compared to demographically *concordant* inpatient events. Gender concordant groups have the largest difference in distributions, compared to ethnically concordant groups. This is also reflected in my regression results. In addition, the right-ward shift of the laboratory cost distribution for demographically discordant events is the most pronounced for combined gender and ethnic concordance, relative to patients with only gender or ethnic concordance. These observations are consistent with my regression results.

[8]Bandwidth for kernel density estimation is the same across all models.

The difference in kernel density distributions for demographically concordant and discordant patients is however small. This is further reflected in the small coefficient estimates for demographic concordance variables.

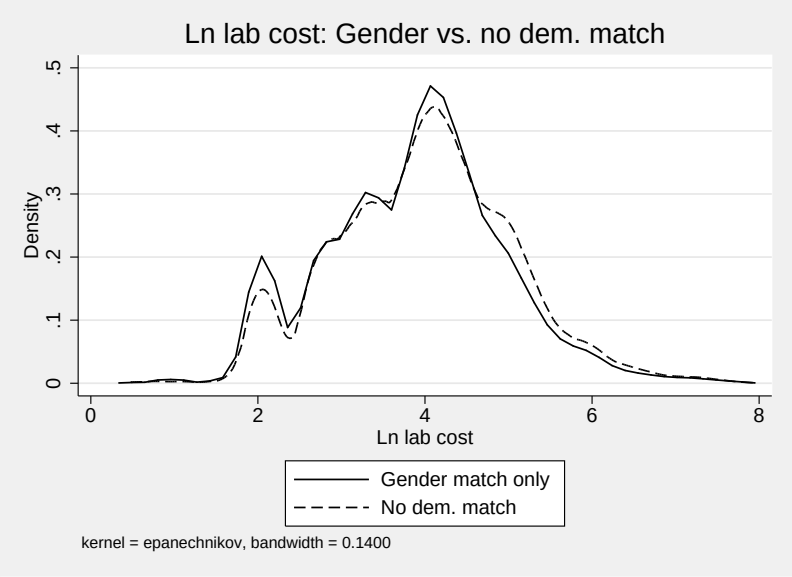

Figure 1.7: Ln laboratory cost by gender match for HR matched data

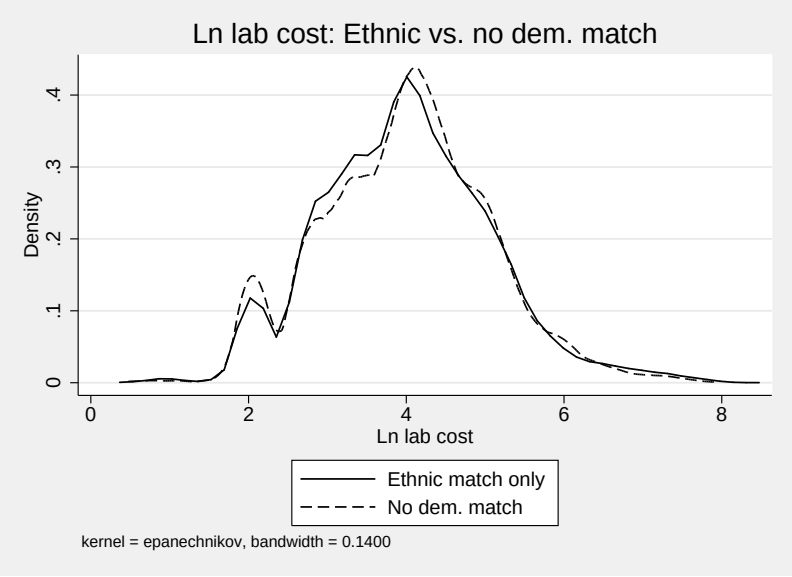

Figure 1.8: Ln laboratory cost by ethnicity match for HR matched data

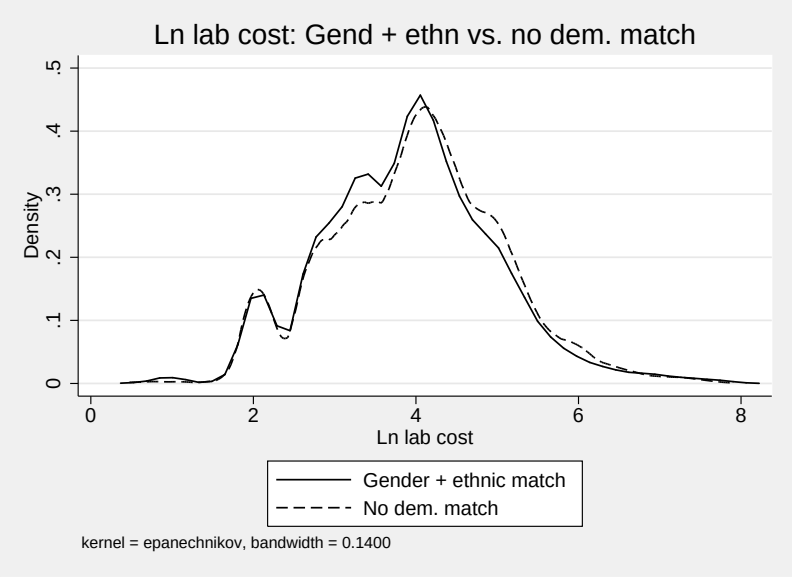

Figure 1.9: Ln laboratory cost by gender and ethnicity match for HR matched data

1.6 Results

Main results from the baseline model for laboratory cost (log-transformed and raw), laboratory quantity and radiology test outcomes are in Tables 1.9, 1.10, 1.11 and 1.12 respectively. In these tables, there are two models for each diagnostic outcome; a fixed-effect on doctors only (Column (1)) and combined doctor plus MDC fixed-effect (Column (2)). The number of groups increases significantly when a fixed-effect is included on doctor and MDC. A fixed-effect on doctors-only is therefore included in baseline tables, because there is more variation within doctor groups to estimate model coefficients. All baseline models are estimated with OLS.

All models are estimated with standard errors clustered on the doctor. Clustering errors is the robust estimation method for standard errors in fixed-effect panel data models (Stock and Watson, 2008).

Further regression tables for laboratory cost outcomes are in Appendix 1.B, including a full table of coefficient values. Robustness checks for laboratory cost outomes are in Appendix 1.C[9].

I discuss the results for gender-only, ethnic-only and gender plus ethnic concordance variables in separate sections. I firstly discuss results from the baseline model and then results from robustness tests, for each type of demographic concordance.

Table 1.9: Laboratory raw cost outcome

	(1) doc FE	(2) doc+mdc FE
Gender only concordance	-4.534^{***}	-4.265^{***}
	(-3.46)	(-3.41)
	[0.00]	[0.00]
Ethnic only concordance	-4.694^{*}	-2.717
	(-2.09)	(-1.39)
	[0.04]	[0.17]
Gender and ethnic concordance	-6.085^{*}	-3.527
	(-2.44)	(-1.59)
	[0.02]	[0.11]
Observations	225344	225344
Within R^2	0.354	0.334
Numb. groups	290	3044
Min group size	2	1
Max group size	7964	6996
Avg group size	777.0	74.0

Notes: Outcome: raw cost of laboratory tests. Sample: 95th percentile of raw laboratory is dropped. Model: linear regression, Column (1) has fixed-effect on doctor, Column (2) has fixed-effect on doctor and MDC. Standard errors clustered on doctor. Robust t-statistics (p-values) in parentheses (brackets). ***p<.01; **p<.05; *p<.10.

1.6.1 Gender only concordance

Gender concordance (β_2) is negative and statistically significant in all baseline models (Tables 1.9, 1.10, 1.11 and 1.12). A doctor treating a patient of the same gender (only) relative to a patient with no demographic concordance is associated with an, on average, reduction in; total laboratory costs of \$4.5, .767 of a laboratory test, or a

[9]I also estimate robustness tests for laboratory quantity and radiology outcomes but these are ommited from this paper for brevity.

Table 1.10: Laboratory log-transformed cost outcome

	(1) doc FE	(2) doc+mdc FE
Gender only concordance	-0.032^{**}	-0.035^{***}
	(-2.77)	(-3.36)
	[0.01]	[0.00]
Ethnic only concordance	-0.031^{*}	-0.025
	(-2.12)	(-1.64)
	[0.04]	[0.10]
Gender and ethnic concordance	-0.032^{*}	-0.027
	(-2.01)	(-1.64)
	[0.05]	[0.10]
Observations	191473	193650
Within R^2	0.326	0.334
Numb. groups	289	2993
Min group size	1	1
Max group size	4572	5123
Avg group size	662.5	64.7

Notes: Outcome: log-transformed laboratory cost. Sample: 99th percentile log-transformed costs excluded. Model: linear regression, Column (1) has fixed-effect on doctor, Column (2) has fixed-effect on doctor and MDC. Standard errors clustered on doctor. Robust t-statistics (p-values) in parentheses (brackets). ***p<.01; **p<.05; *p<.10.

Table 1.11: Laboratory quantity outcome

	(1) doc FE $b/t/p$	(2) doc + mdc FE $b/t/p$
Gender only concordance	-0.782^{**}	-0.656^{**}
	(-3.27)	(-2.96)
	[0.00]	[0.00]
Ethnic only concordance	-0.852^{*}	-0.700^{*}
	(-2.46)	(-2.24)
	[0.01]	[0.03]
Gender and ethnic concordance	-1.004^{**}	-0.741^{*}
	(-2.63)	(-2.12)
	[0.01]	[0.04]
Observations	236559	236559
Within R^2	0.517	0.500
Numb. groups	291	3703
Min group size	2	1
Max group size	8220	7185
Avg group size	812.9	63.9

Notes: Outcome: quantity of laboratory test. Sample: eligible study sample. Model: linear regression, Column (1) has fixed-effect on doctor, Column (2) has fixed-effect on doctor and MDC. Standard errors clustered on doctor. Robust t-statistics (p-values) in parentheses (brackets). ***p<.01; **p<.05; *p<.10.

Table 1.12: Radiology outcome

	(1) doc FE $b/t/p$	(2) doc + mdc FE $b/t/p$
Gender only concordance	-0.011^{*} (-2.02) [0.04]	-0.012^{*} (-2.47) [0.01]
Ethnic only concordance	-0.025^{***} (-3.69) [0.00]	-0.024^{***} (-3.59) [0.00]
Gender and ethnic concordance	-0.037^{***} (-4.87) [0.00]	-0.036^{***} (-4.88) [0.00]
Observations	238916	238916
Within R^2	0.171	0.124
Numb. groups	291	3714
Min group size	2	1
Max group size	8251	7210
Avg group size	821.0	64.3

Notes: Outcome: =1 if radiology test greater than or equal to one, =0 if no radiology test. Sample: eligible study sample. Model: linear regression, Column (1) has fixed-effect on doctor, Column (2) has fixed-effect on doctor and MDC. Standard errors clustered on doctor. Robust t-statistics (p-values) in parentheses (brackets). ***p<.01; **p<.05; *p<.10.

1.1% reduction in the likelihood of having one or more radiology test. The coefficient is statistically significant at the 1% level for laboratory cost and quantity outcomes, and at the 10% level for the radiology outcome. This result is robust to estimation with fixed-effects on doctors (Columns (1)) and combined doctor and MDC (Columns (2)).

In the regression table for the raw laboratory cost outcome, with a fixed-effect on doctor and MDC, the one-sided Wald test for the null hypothesis of $\beta_2 >= 0$ with alternative hypothesis $\beta_2 < 0$, has a p-value of .005. I therefore reject the null hypothesis in favour of $\beta_2 < 0$. This p-value is constructed from halving the (two-sided) p-value included in Table 1.9. One-sided Wald tests of $\beta_2 < 0$ for other diagnostic resource outcomes can similarly be constructed from halving p-values included in Tables 1.10, 1.11 and 1.12.

The coefficients in log-transformed laboratory cost models can be interpreted as semi-elasticities. A doctor treating a patient of the same gender relative to no demo-

graphic concordance is associated with an on average reduction in the total laboratory costs of 3.2%. This estimate is conditional on the population of patients that received a laboratory test. This estimate increases slightly to a 3.5% reduction in laboratory costs when there is a fixed-effect on combined doctor and MDC (Table 1.10).

Robustness of gender concordance results

Tables for robustness checks on *laboratory cost* outcomes are in Appendices 1.C and 1.D. I firstly discuss robustness checks for laboratory *cost* outcomes before turning to results from a non-linear model for laboratory *quantity* outcomes.

The first robustness check removes the fixed-effect on doctors (Column (1), Table 1.19, Appendix 1.C). This test investigates if demographic concordance is robust to including heterogeneity across doctors in their test-ordering behaviour. With no fixed-effect, gender concordance is statistically significant at the 1% level and is estimated at a reduction of $8.1 in total laboratory costs. Therefore, removing a fixed-effect *increases* the estimate of β_2.

The second series of robustness tests investigates outlying inpatient events (observations). Identifying outlying inpatient events is important because demographic concordance coefficients are estimated by comparing mean diagnostic outcomes, after controlling for explanatory variables. Comparing mean outcomes can be sensitive to outlying observations. For example, if one group contained a large laboratory cost value, this would raise the mean considerably for that group relative to the comparison group. We therefore want to estimate the coefficient for demographic concordance on inpatient events that are representative of a typical clinical encounter, rather than coefficient estimates that are influenced by a small number of 'rare' cases. Hospital admissions also involve a great deal of heterogeneity, because patients can enter for vastly different clinical conditions.

Linear regression (i.e. minimising the squared sum of residuals) is especially sensitive to outlying observations (compared to quantile or least median of squares regression). Outliers can include vertical outliers (outlying from predicted y values), bad leverage points (outlying in x and y) and good leverage points (outlying in x but close to y). The effect of vertical outliers on mean estimation is addressed in my paper by dropping observations with raw costs in the 95th percentile of the cost distribution. Bad leverage points are also expected in our data because of significant heterogeneity across patients in unobserved severity of illness[10].

I explore outlying observations in my data by implementing a robust regression method. Robust regression method weights vertical outliers to have less influence on mean estimation, and excludes observations with bad leverage points from the population. Bad leverage points are identified by a Cook's distance statistic larger than 2. An example of an observation with a bad leverage point would be a patient that had unusual values of diagnostic tests given explanatory variables. For example, a patient with a high LOS and low diagnostic test orders. The robust regression method firstly estimates the model and identifies outliers. This model is then re-run by weighting outlying values so they have less influence on mean estimation. This method is available for linear regression *without* fixed-effects. It is not clear how to identify outliers after first de-meaning observations in a fixed-effect method. I therefore estimate a robust regression model on my data without fixed-effects (Column (2), Table 1.19). Results show gender concordance is associated with a larger reduction in laboratory costs than in the baseline model. Gender concordance is associated with a $5.28 reduction in total laboratory costs, compared to a patient with no demographic concordance. The model fit has improved, because outlying observations are given less

[10]A further kind of outlier in panel data are block outliers. These occur if outlying observations are across a block of consecutive time periods. This is however unlikely in our data because there is no real time dimension, and we expect the arrival of 'extreme' (e.g. high cost or unusual) cases to be spread out over a doctor's caseload.

weight in the regression. The estimated R^2 is .678 and all demographic concordance variables are statistically significant at the 1% level.

The impact of vertical outliers on the robustness of mean estimation is also checked by dropping different percentiles of the raw laboratory cost distribution (Table 1.20). These models have a fixed-effect on the doctor. Column (1) includes the whole population, gender concordance is statistically insignificant in this model. At the 99, 95 and 90 percentile, gender concordance effects are statistically significant at the 1% level. This suggests that estimates of gender concordance are not robust to including a small number of extreme vertical outliers in laboratory cost data. This could occur if, for example, a small number of complex medical cases involve gender concordance. This would increase the group mean in the gender concordant relative to discordant group.

The third series of robustness checks estimates demographic concordance on a subsample of patients with the same diagnostic category. One concern with estimating concordance effects across different medical conditions is that the effect of, say, gender on lab test ordering may be related to medical area. Gender may therefore not be adequately controlled for when estimated over the whole population of hospital admissions. For example, a male patient may be associated with higher laboratory costs in cardiology but lower laboratory costs in, for example, general medicine[11].

I firstly estimate the laboratory cost model on patients in any one of the top four MDCs. These are; Respiratory, Circulatory, Kidney plus Urinary, and Digestive. This robustness test restricts the variety of medical conditions in the population, so there is less heterogeneity in medical treatment across patients in the population. Results are in Table 1.21. Column (1) and (2) have a fixed-effect on doctor and doctor-MDC respectively. Gender concordance coefficients are statistically significant at the 1% level. On average, a doctor treating a patient of the same gender relative to

[11] A model interacting gender and ethnic group with MDC dummy variables did not change my main results. I do not present this table for reasons of brevity.

no demographic concordance is associated with a $5.17 reduction in total laboratory cost ordering. The coefficient of $5.17 is slightly larger than the estimated gender concordance coefficient across *all* hospital admissions ($4.5). Column (3) in Table 1.21 estimates concordance effects with a fixed-effect on doctor and DRG. DRG groups determine payments hospitals receive for medical treatment. This controls to a higher degree for the severity of patients and medical treatment decisions. The number of groups increases significantly (to N=43,848), and the size of groups is small at an average of 9.3. The gender concordance coefficient is statistically significant at the 1% level, with an estimated reduction in total laboratory costs of $3.11.

I also estimate my baseline model on patients within each MDC group in the eligible inpatient population. Results for each MDC on the log-transformed laboratory cost are in Tables 1.23 to 1.26 in Appendix 1.D. This allows me to investigate if there are larger coefficients for demographic concordance in some medical areas over others. For example, I would expect there to be larger coefficients for demographic concordance in MDCs that involve greater diagnostic uncertainty than less diagnostic uncertainty. There may be less clinical uncertainty for external health conditions upon presentation at hospital, and therefore demographic concordance may have a reduced impact on diagnostic test ordering, (because there is less potential to gain from improved communication in a consultation.) This would occur for illnesses located internally (e.g. Digestive and Kidney and Urinary) relative to external illnesses (e.g. skin conditions and ENT).

I find that the largest estimated reduction in total laboratory costs arising from gender concordance is for patients in the Kidney and Urinary MDC (6.6% reduction in laboratory costs). Patients in this MDC group comprise a large share of kidney and liver stones. These are internal illnesses for which diagnostic testing is required to confirm the incidence and severity of the illness. MDCs with low numbers of

observations and/or external illnesses (skin, eyes, ENT) do not always have negative or statistically significant relationships.

The last series of robustness checks used a dataset that matched doctors with HR data *and* a manual match process. A manual match process was used to associate a gender and ethnic group using the first and last name of doctors in inpatient data. All patients that are not matched with HR data, or are matched with HR data but do not have information on doctor ethnicity or gender, are manually associated a gender and/or ethnic group. Details of the process used to associate a doctor and ethnic group is in Appendix 1.F. I associate a doctor's gender and ethnic group based on the likelihood of a doctor's name belonging to a gender or ethnic group. For example, a doctor with the name Michael would be associated a male gender group, because there is a high likelihood this name and gender occurs together. This process provides a second dataset of HR and manually matched doctors, thereby providing a larger number of inpatient observations to estimate the baseline model.

This second dataset also allows me to investigate sampling bias. I am not able to observe the gender and ethnicity characteristics of doctors that are not matched with HR data. I am therefore not able to test if there is an over-representation of certain kinds of doctors in the HR matched data relative to the wider hospital population[12]. Sampling bias may occur if certain types of doctor gender or ethnicity groupings are less (or more) likely to record their demographic characteristics in HR

[12]Sampling bias may not be a particularly problematic for this study, because there is a random element in whether I can match demographic information to a doctor. All nurse and medical staff employees were obtained from HR systems. These are all staff members that receive some form of payment from the organisation. Inability to match a doctor is either caused by random differences between HR and inpatient data (e.g. misspelling or only initials for casemanager in inpatient datasets) or because a patient is treated by a doctor that is not employed by the health board. In addition, other situations where we cannot match with HR data involve cases where a doctor is not personally identified, for example a'consultant' or a doctor is part of a team. Table 1.28 provides the percentage of patients that are not matched with HR data that have a consultant term in the primary casemanager's name (11%). These situations are at least identifiable and provide a reason for why an inpatient event is not matched.

forms. A manually matched dataset is an attempt to check that results are robust when estimated on a wider population of doctors in the hospital.

Regression results for patients that have a casemanager that is HR *and* manually matched are in Table 1.17 and Table 1.18 in Appendix 1.C. There are now 407,225 observations in the sample compared to 224,656 for the raw cost outcome. In this dataset, the estimated coefficient for gender concordance is smaller. A doctor treating a patient of the same gender relative to no demographic concordance is associated with a reduction in the total cost of laboratory tests of 2.8% or a $3.5 reduction in total laboratory cost. The gender concordance variable is statistically significant at the 1% level.

There is a caveat on interpretation of results from manually matched inpatient events, because a doctor's ethnic or gender group could be incorrectly associated. I observe a larger number of doctors associated with a European ethnic group after manual matching, compared to HR data. Doctors from Maori and South African ethnicity groupings may be more likely to be associated with a European ethnic group using last names, this may lead to an over-representation of the European ethnic group.

Lastly, I estimate a fixed-effect non-linear model for laboratory quantity outcomes. I implement a fixed-effect Poisson and Negative Binomial model, both of these models have bootstrapped standard errors to deal with heterogeneity across the patient population. Results are in Table 1.16. The coefficient on gender concordance is negative and statistically significant at the 5% level. The marginal effect is estimated at a reduction in the total number of laboratory test orders of .057. This indicates that the effect of gender concordance on the number of laboratory tests ordered is robust to estimation in both a linear and non-linear fixed-effect model.

Overall, robustness tests indicate the coefficient for gender concordance is robust to estimation on different subsets of the data and different model specifications.

1.6.2 Ethnic only concordance

We now turn to interpreting the results for the ethnic concordance variable. When there is a *fixed-effect on doctors only*, the ethnic concordance coefficient (β_1) is statistically significant at the 10% level for laboratory cost and quantity outcomes, and at the 1% level for the radiology outcome. A doctor treating a patient of the same ethnicity (only) relative to a patient with no demographic concordance is associated with an, on average, reduction in total laboratory costs of $4.6, or .88 of a laboratory test, or a 2.4% reduction in the likelihood of having one or more radiology test (Column (1) in Tables 1.9, 1.10, 1.11 and 1.12).

When there is a *fixed-effect on doctor-MDC*, (β_1) remains negative and statistically significant for laboratory quantity and radiology outcomes, but is statistically insignificant for laboratory cost outcomes (Column (2) in Tables 1.9, 1.10, 1.11 and 1.12).

Robustness of ethnicity concordance results

The coefficient for ethnic concordance is not robust to (1) including a fixed-effect on Doctor and MDC for laboratory cost outcomes (Table 1.9) and (2) estimation with a non-linear model for laboratory quantity outcomes (Table 1.16).

Ethnic concordance may be weakly identified in my paper because of a lack of variation in doctor ethnicity groupings. Given that a high proportion of ethnic concordance is European, this could result in less variation in the data to identify the coefficients for a patient's ethnic group and ethnic concordance separately.

Removing a fixed-effect increases the statistical significance of the ethnic concordance variable. The ethnic concordance variable is statistically significant in the robust regression model, and linear regression model without fixed-effects (Table 1.19). The effect of ethnic concordance on reducing the total laboratory cost in the robust regression method is estimated at $5.25.

For the second dataset of manually matched doctor demographic characteristics, the ethnic concordance variable is statistically significant (Table 1.17). This might occur because there are more observations of doctors and hence greater variation in doctor ethnicity to estimate ethnic concordance.

The magnitude of the ethnic concordance variable also varies across MDCs. The coefficient on the ethnic concordance variable is largest in Digestive (MDC 6) and Kidney and Urinary Tract (MDC 11) conditions. This is a similar pattern to a large coefficient on the gender concordance variable for MDCs associated with internal illnesses compared to external illnesses.

Overall, ethnic concordance is less robust than gender concordance. The lack of statistical significance may be contributed to by a lack of variation in doctor ethnicity groupings compared to gender groups.

1.6.3 Ethnicity and gender concordance

When there is a *fixed-effect on doctors only*, the gender and ethnic concordance coefficient (β_3) is statistically significant at the 10% level for laboratory cost and quantity outcomes, and at the 1% level for radiology outcome. A doctor treating a patient with the same gender and ethnic group relative to no demographic concordance, is associated with an, on average, reduction of $6 in total laboratory costs, or 1 laboratory test, or 3.7% reduction in the likelihood of having one or more radiology tests.

When there is a *fixed-effect on doctor and MDC*, the gender and ethnic concordance variable is not statistically significant in the laboratory cost model.

In all baseline models, $\beta_3 > \beta_1$ and $\beta_3 > \beta_2$. The Wald test for the null hypothesis of equality for $\beta_3 = \beta_2$ and $\beta_3 = \beta_1$ and alternative hypothesis $\beta_3 \neq \beta_2$ and $\beta_3 \neq \beta_1$ is implemented for baseline models. For the quantity of laboratory tests, the null hypothesis $\beta_3 = \beta_2$ ($\beta_3 = \beta_1$) is unable to be rejected, the p-value is 0.4675 ($p = 0.4999$). For the raw laboratory cost outcome, the null hypothesis $\beta_3 = \beta_2$ ($\beta_3 = \beta_1$)

is also unable to be rejected for raw laboratory cost outcome, the p-value is 0.2733 ($p = 0.4772$). For the radiology outcome, the null hypothesis $\beta_3 = \beta_2$ ($\beta_3 = \beta_1$) is rejected with p-value of $p = 0.0104$ ($p = 0.0007$). Therefore, gender and ethnicity concordance has a greater effect on reducing the likelihood of one or more radiology test compared to demographic discordance.

Robustness of ethnicity and gender concordance results

The robustness of gender plus ethnic concordance follows a similar pattern to the robustness of the ethnic concordance variable. Specifically, gender and ethnic concordance is not robust (statistically significant) to including a fixed-effect on Doctor and MDC for laboratory cost outcomes. In addition, gender and ethnic concordance is not statistically significant in non-linear laboratory quantity models (Table 1.16).

Gender and ethnic concordance is statistically significant in models without a fixed-effect (Table 1.19), and the second dataset of manually matched doctor characteristics (Table 1.17). Both of these robustness tests increase variation in the data to estimate the independent effect of gender and ethnic matching. Gender and ethnic concordance is predominately European and therefore there are similar concerns about the ability of my data to identify the separate effect of gender and ethnic concordance from a patient's ethnic group. It is for this reason that I focus on the robustness of gender concordance results in this paper.

Age, ethnicity and gender concordance

HR data also contains information on the age of a doctor. Demographic concordance variables for age, gender and ethnic concordance are estimated for patients matched with HR data (Table 1.15, Appendix 1.B). This includes concordance variables for; age-only, age and gender, age and ethnicity, and combined age, ethnicity and gender concordance. All demographic concordance coefficients have a negative sign. Age

has no statistically significant effect on its own, but is statistically significant when combined with both gender and ethnic group. When a patient and doctor have all three demographic characteristics in common, I observe the largest on average reduction in total laboratory costs from any type of demographic concordance considered previously. The combination of gender, similar age range and ethnic concordance is associated with a reduction in the total laboratory costs of $7.18, compared to doctors treating a patient with no demographic concordance.

1.7 Conclusion and discussion

This paper aimed to estimate a relationship between demographic concordance and the amount of diagnostic resources ordered during hospital treatment. Demographic concordance between a doctor and patient is a relatively understudied area in health economics. This is most likely a result of the lack of data on doctors' demographic information in administrative datasets.

My first main result is a statistically significant reduction in laboratory and radiology tests when a doctor and patient have the same gender and/or ethnic group, relative to when a doctor treats a patient with no demographic concordance.

The effect of gender concordance on reducing diagnostic resources is more robust than ethnic concordance. My robustness checks investigate inpatient outliers. I find significant cost outliers and drop the 95th percentile of costs when estimating my model. I also estimate demographic concordance with no fixed-effect, and a fixed-effect on doctor only. With no fixed-effect, coefficient estimates for all types of demographic concordance are larger and statistically significant at the 1% level.

I also estimate the baseline model using different medical populations in hospital. In general, I find demographic concordance coefficients are larger when estimated on

patients with internal rather than external illnesses. However, the statistical strength of this relationship is low, largely because population sizes are small in some MDCs.

The last series of robustness checks used a dataset that associated a gender and ethnic group based on a doctor's first and last name. I find that both gender and ethnic concordance are statistically significant for the baseline model with log-transformed laboratory costs.

My second main result is a greater reduction in diagnostic resources when patients and doctors have the same ethnic *and* gender group, relative to when they have *only* gender or ethnic concordance. This is also reflected in the largest estimated reduction for all diagnostic resource measures for combined ethnic, age and gender concordance.

The estimated size of coefficients for demographic concordance variables is however small, and therefore the relationship between demographic concordance and diagnostic resources is difficult to observe in descriptive statistics.

The main strengths of my empirical method are; a fixed-effect on doctor and MDC, different measures for diagnostic resources, and the ability to exclude an endogenous relationship when explaining my results.

Firstly, a fixed-effect on doctors and MDC situates the estimation of coefficients for demographic concordance in each doctor's decision-making environment. I therefore use differences in diagnostic resources for demographically concordant and discordant patients, within a doctor and medical area, to estimate the impact of concordance on diagnostic test orders.

Secondly, I use four different measures of diagnostic outcomes; laboratory cost, log-transformed laboratory cost, laboratory quantity, and the likelihood a patient has one or more radiology tests during a hospital stay. I observe a negative relationship across all four diagnostic outcomes. This supports the credibility of my result that demographic concordance reduces the diagnostic resources required to treat a patient in hospital.

Thirdly, I argue that demographic concordance variables are either exogenous, or if they are endogenous the expected relationship would bias the coefficient on demographic concordance in the opposite direction to the one I find. Hospital policy assigns doctors based on availability at the time of patient arrival. As a policy, assignment is independent of demographic characteristics. Demographic concordance is therefore in practice exogenous in my baseline model. However, patients can be transferred across doctors, and this could occur *informally* for demographic reasons. If sorting were to occur, for demographic reasons, it would be more likely to occur in the direction to achieve demographic concordance rather than sorting with the intention to achieve a demographic discordance. In the former case, this would result in *higher* amounts of diagnostic resources in demographically concordant pairs - because patients would have been treated by more than one doctor, and this would be associated with greater rather than fewer diagnostic test orders.

Lastly, the hospital is also an appropriate setting to investigate the quality of doctor-patient consultation in demographically concordant pairs. This is because diagnoses and treatments are decided using information at hand in an acute setting. Consultations with General Practitioners often involve a history of doctor-patient interactions, which will affect diagnostic test ordering and treatment decisions, in addition to information obtained from the current consultation.

The limitations of my study are a high proportion of doctors that have a European ethnic group. A high proportion of European doctors reduces variation in the data to identify the effect of ethnic concordance separately from the effect of a patient's ethnic group on diagnostic test ordering. Despite this, ethnic concordance is statistically significant at the 10% level in the doctor fixed-effect model across all diagnostic resources.

A second limitation of my data is the inability to observe the identity of the doctor ordering a diagnostic test in hospital. Given the collaborative nature of some hospital

treatment, there will likely be test orders for some patients that are not due to the 'primary' doctor identified in inpatient data. A large population size increases my confidence that we are estimating some kind of average doctor-patient relationship effect, even if some patients may be treated by more than one doctor and will therefore have 'noisier' diagnostic resource measures than patients treated by only one doctor. Importantly, I don't expect the noise in diagnostic resource measures (generated by multiple ordering doctors) to bias estimates of concordance variables in the negative direction. In addition, doctors can observe the test orders of previous doctors and this is relevant information for how many tests they will order. It is therefore important to use *all* tests ordered during a patient's hospital stay, rather than just a subset of diagnostic test orders.

In this paper, I also discussed reasons for a reduction in diagnostic resources in demographically concordant doctor-patient pairs.

Firstly, because my data is from a New Zealand hospital, it is not expected that concerns about patient litigation or insurance arrangements would explain a reduction in diagnostic resources. Patient litigation or insurance arrangements are two prominent explanations for the over-use of diagnostic resources in the United States. The hospital in my study is publicly funded and there is less risk of personal litigation for doctors.

The first explanation for a reduction in diagnostic resources is an information gain on a patient's health status when doctors and patients are demographically concordant. It is assumed doctors obtain the information they need to decide treatments and monitor health status during a hospital stay. Information can come from the consultation or diagnostic tests. If there is more uncertainty in the consultation, doctors will order more tests to make up for the lack of information, vice versa if doctors are more certain of their information then they will require less information from tests to diagnose and treat patients. The last source of information on a pa-

tient's health condition is from a doctor's individual experience, communication style, specialty and so on; this forms their idiosyncratic test-ordering behaviour. A fixed-effect on doctors controls for idiosyncratic test ordering. If a doctor requires fewer tests on average when treating demographically concordant relative to discordant patients, this is consistent with increased information on a patient's health status from a consultation.

There could be a number of different mechanisms behind an information gain in demographically concordant consultations. These include; better quality communication; doctors' private information gain when patients belong to the same demographic group (e.g. doctors believe a condition is more or less likely in patients from their own demographic group); or a doctor is more trusting of patient information in demographically concordant pairs (i.e. not biased). For example, a doctor may be less trusting of a patient's symptom reports when they are demographically discordant and this could motivate a doctor to order additional diagnostic tests, because they perceive the accuracy of patient information to be poor. These doctor-patient consultation scenarios are all consistent with increased information arising from the consultation process. Balsa and McGuire (2003) also do not stipulate the mechanisms behind their miscommunication model. They do however suggest that miscommunication could originate in prejudice and/or stereotypes; for example, if a doctor is less willing to invest in understanding patient symptom reports from minority patients.

A second explanation for a reduction in diagnostic resources, is that preferences for treatment styles change in demographically concordant relative to discordant pairs. For example, a doctor or patient may prefer a less invasive treatment style when demographically concordant. I am unable to distinguish between an information gain and preference change explanation for my result.

The information gain hypotheses is arguably more plausible given my results, and the literature documenting communication gains in demographically concordant con-

sultations (LaVeist and Nuru-Jeter, 2002; Street Jr et al., 2008; Sandhu et al., 2009). Firstly, I find a significant negative relationship across all three diagnostic resource measures. Secondly, demographic concordance effects are also larger for two or more demographic characteristics. This is consistent with an even greater improvement in doctor-patient communication with increased demographic similarity. Thirdly, demographic concordance is largest for internal clinical conditions such as kidney and digestive conditions. Demographic concordance is statistically insignificant for medical conditions where the cause of hospital admission is externally located. There is plausibly less clinical uncertainty with external rather than internal illnesses. Lastly, this paper also controls for the length of stay and theatre event treatment decisions. This is intended to control for variation in test ordering driven by differences in treatment decisions across patients in demographically concordant and discordant pairs. If a less invasive treatment style were to explain a significant negative relationship, then treatment decision variables would capture much of this explanation which would result in statistically insignificant demographic concordance variables.

An extension of this research could be to investigate how using patient preferences for their doctor might affect the efficiency of health care delivery in a hospital. Previous studies have shown that patients tend to prefer GPs with the same demographic characteristics (particularly for gender) (Godager, 2012). Asking patients if they have a preference for doctor demographic features could identify patients that have difficulty communicating with doctors from a different demographic group. Patients who have no preference could be assigned an available doctor. However, matching all patients with a doctor of the same demographic could have adverse consequences. In the long run, doctors would not learn how to communicate in different situations and the impact of demographic discordance on increasing diagnostic resources could be even greater.

There has been increased interest in reducing the cost associated with unnecessary diagnostic tests and procedures, particularly in the United States. In addition to the benefits of reducing wasteful medical spending, reducing the number of diagnostic tests required to treat a patient could also improve a doctor's ability to make decisions. Increased information places burdens on doctors to identify a patient's condition(s), and its severity and complexity. Imperfect information on a patient's health status has long been recognised as a source of market failure in health care delivery (Arrow, 1963). Arrow (1963) considered uncertainty central to the understanding of health care markets; 'the special economic problems of medical care can be explained as adaptations to the existence of uncertainty in the incidence of disease and the efficacy of treatment' (Arrow, 1963, pg. 941). If a doctor is able to gain more information or develop a rapport with a patient that identifies adequate and satisfactory health care requiring fewer resources, then policy makers could benefit from further study into how the doctor-patient relationship affects medical resource decisions.

1.A Explanatory variables

Table 1.13: Summary of study variables for HR matched patients

	Mean	*Sd*	*Min*	*Max*	*P*50
Demographic variables:					
Gender only concordance	0.20	0.40	0.00	1.00	0.00
Ethnic only concordance	0.28	0.45	0.00	1.00	0.00
Gender and ethnic concordance	0.31	0.46	0.00	1.00	0.00
Patient characteristics:					
Age	52.09	23.05	5.00	105.00	53.00
Male	0.49	0.50	0.00	1.00	0.00
Deprivation scale	5.85	2.94	0.00	10.00	6.00
Pacific pat	0.13	0.33	0.00	1.00	0.00
African pat	0.01	0.07	0.00	1.00	0.00
Maori pat	0.09	0.29	0.00	1.00	0.00
Middleeastern pat	0.01	0.10	0.00	1.00	0.00
Other ethnic pat	0.02	0.15	0.00	1.00	0.00
Asian pat	0.07	0.25	0.00	1.00	0.00
Indian pat	0.05	0.21	0.00	1.00	0.00
Admission timing:					
After 5pm entry	0.25	0.43	0.00	1.00	0.00
After 12pm entry	0.32	0.47	0.00	1.00	0.00
Overnight admiss.	0.07	0.25	0.00	1.00	0.00
Monday admit	0.16	0.37	0.00	1.00	0.00
Tuesday admit	0.16	0.37	0.00	1.00	0.00
Wednesday admit	0.17	0.37	0.00	1.00	0.00
Thursday admit	0.18	0.38	0.00	1.00	0.00
Friday admit	0.15	0.36	0.00	1.00	0.00
Saturday admit	0.09	0.29	0.00	1.00	0.00
2005 admit	0.12	0.33	0.00	1.00	0.00
2006 admit	0.12	0.32	0.00	1.00	0.00
2007 admit	0.13	0.34	0.00	1.00	0.00
2008 admit	0.13	0.34	0.00	1.00	0.00
2009 admit	0.13	0.34	0.00	1.00	0.00
2010 admit	0.14	0.34	0.00	1.00	0.00
2011 admit	0.12	0.33	0.00	1.00	0.00
Admission type:					
Transfer	0.13	0.33	0.00	1.00	0.00
AED entry	0.34	0.47	0.00	1.00	0.00
Prev admiss 60days	0.33	0.47	0.00	1.00	0.00
Accident	0.12	0.32	0.00	1.00	0.00
Day patient	0.13	0.33	0.00	1.00	0.00
Home DHB	0.63	0.48	0.00	1.00	1.00
Acute admiss.	0.73	0.45	0.00	1.00	1.00
Clinical variables:					
LOS	3.29	5.44	0.00	176.00	1.00
Theatre event	0.16	0.36	0.00	1.00	0.00
Diagnosis count	4.42	3.45	1.00	59.00	4.00

Myocardial Infarct.	0.05	0.21	0.00	1.00	0.00
Congestive Heart F	0.04	0.20	0.00	1.00	0.00
Periphral Vascular Dis	0.02	0.15	0.00	1.00	0.00
Cerebrovascular Dis	0.04	0.19	0.00	1.00	0.00
Dementia	0.02	0.14	0.00	1.00	0.00
Chronic Pulmonary D	0.04	0.20	0.00	1.00	0.00
Rheumatic Disease	0.01	0.07	0.00	1.00	0.00
Peptic Ulcer Disease	0.00	0.06	0.00	1.00	0.00
Mild Liver Disease	0.01	0.11	0.00	1.00	0.00
Diabetes w/o complic.	0.04	0.21	0.00	1.00	0.00
Diabetes w complic.	0.07	0.25	0.00	1.00	0.00
Paraplegia + Hemiplegia	0.02	0.14	0.00	1.00	0.00
Renal Disease	0.07	0.25	0.00	1.00	0.00
Cancer	0.10	0.31	0.00	1.00	0.00
Liver Disease	0.00	0.05	0.00	1.00	0.00
Metastatic Carcinoma	0.04	0.21	0.00	1.00	0.00
AIDS/HIV	0.00	0.04	0.00	1.00	0.00
N	225344				

Notes: Summary of explanatory variables for the study sample.

1.B Results: baseline models for laboratory cost outcome

Table 1.14: Laboratory costs

	(1) doc FE	(2) doc+mdc FE
Gender only concordance	-4.534^{***}	-4.265^{***}
	(−3.46)	(−3.41)
	[0.00]	[0.00]
Ethnic only concordance	-4.694^{*}	−2.717
	(−2.09)	(−1.39)
	[0.04]	[0.17]
Gender and ethnic concordance	-6.085^{*}	−3.527
	(−2.44)	(−1.59)
	[0.02]	[0.11]
Controls:		
Age	−0.054	−0.017
	(−0.88)	(−0.29)
	[0.38]	[0.77]
Male	−2.232	−2.058
	(−1.89)	(−1.80)
	[0.06]	[0.07]
Deprivation scale	0.094	0.091
	(0.85)	(0.85)

	[0.39]	[0.40]
After 5pm entry	−2.010	−2.359
	(−0.58)	(−0.74)
	[0.57]	[0.46]
After 12pm entry	1.183	1.199
	(0.63)	(0.68)
	[0.53]	[0.50]
Overnight admiss.	6.131**	6.091***
	(3.27)	(3.43)
	[0.00]	[0.00]
Transfer	−15.799***	−13.939***
	(−3.98)	(−4.27)
	[0.00]	[0.00]
LOS	12.229***	12.422***
	(12.86)	(12.92)
	[0.00]	[0.00]
Theatre event	28.447***	28.135***
	(5.03)	(4.84)
	[0.00]	[0.00]
AED entry	−5.793*	−5.807**
	(−2.59)	(−2.68)
	[0.01]	[0.01]
Prev admiss 60days	−18.804***	−16.939***
	(−7.70)	(−7.95)
	[0.00]	[0.00]
Accident	−49.892***	−45.888***
	(−23.41)	(−23.34)
	[0.00]	[0.00]
Day patient	−37.499***	−38.669***
	(−3.87)	(−4.39)
	[0.00]	[0.00]
Home DHB	−1.500	−1.020
	(−0.88)	(−0.73)
	[0.38]	[0.47]
Diagnosis count	10.522***	10.349***
	(20.49)	(20.03)
	[0.00]	[0.00]
Pacific pat	0.817	2.868
	(0.38)	(1.50)
	[0.71]	[0.13]
African pat	10.250**	8.294*
	(2.62)	(2.51)
	[0.01]	[0.01]
Maori patient	−2.805	−0.105
	(−1.08)	(−0.05)

	[0.28]	[0.96]
Middleeastern pat	−0.196	−0.580
	(−0.05)	(−0.15)
	[0.96]	[0.88]
Other ethnic pat	−5.394*	−3.737
	(−2.04)	(−1.59)
	[0.04]	[0.11]
Asian pat	2.160	2.442
	(1.03)	(1.21)
	[0.30]	[0.23]
Indian pat	1.982	3.644
	(0.94)	(1.95)
	[0.35]	[0.05]
Monday admit	−3.646*	−3.430*
	(−2.02)	(−2.14)
	[0.04]	[0.03]
Tuesday admit	0.798	0.372
	(0.39)	(0.22)
	[0.70]	[0.83]
Wednesday admit	−4.842**	−4.270*
	(−2.61)	(−2.52)
	[0.01]	[0.01]
Thursday admit	−6.280**	−4.964**
	(−3.27)	(−2.99)
	[0.00]	[0.00]
Friday admit	−5.389***	−5.458***
	(−3.67)	(−4.39)
	[0.00]	[0.00]
Saturday admit	−4.989***	−4.891***
	(−3.75)	(−3.88)
	[0.00]	[0.00]
2005	−15.108***	−15.942***
	(−4.48)	(−4.88)
	[0.00]	[0.00]
2006	−21.152***	−21.100***
	(−5.84)	(−5.76)
	[0.00]	[0.00]
2007	−28.445***	−30.147***
	(−7.34)	(−8.43)
	[0.00]	[0.00]
2008	−13.023**	−14.810**
	(−2.63)	(−3.24)
	[0.01]	[0.00]
2009	−6.801	−8.644
	(−1.27)	(−1.78)
	[0.21]	[0.08]

2010	7.023	6.316
	(1.25)	(1.27)
	[0.21]	[0.21]
2011	−2.266	−2.976
	(−0.40)	(−0.60)
	[0.69]	[0.55]
Myocardial Infarct.	6.687*	3.892
	(2.16)	(1.53)
	[0.03]	[0.13]
Congestive Heart F	25.328***	25.139***
	(8.61)	(8.34)
	[0.00]	[0.00]
Periphral Vascular Dis	0.918	−2.058
	(0.28)	(−0.55)
	[0.78]	[0.58]
Cerebrovascular Dis	−17.043*	−12.223
	(−2.54)	(−1.84)
	[0.01]	[0.07]
Dementia	−8.749	−7.013
	(−1.96)	(−1.66)
	[0.05]	[0.10]
Chronic Pulmonary D	−8.318	−7.552
	(−0.83)	(−0.86)
	[0.41]	[0.39]
Rheumatic Disease	22.760**	21.263***
	(3.18)	(4.06)
	[0.00]	[0.00]
Peptic Ulcer Disease	45.739***	44.602***
	(7.70)	(7.80)
	[0.00]	[0.00]
Mild Liver Disease	12.295*	14.434*
	(2.14)	(2.52)
	[0.03]	[0.01]
Diabetes w/o complic.	−2.838	−4.607*
	(−1.29)	(−2.26)
	[0.20]	[0.02]
Diabetes w complic.	−30.000***	−30.094***
	(−14.81)	(−14.80)
	[0.00]	[0.00]
Paraplegia + Hemiplegia	−31.517***	−33.211***
	(−7.62)	(−8.45)
	[0.00]	[0.00]
Renal Disease	15.978***	16.969***
	(6.89)	(8.79)
	[0.00]	[0.00]

Cancer	17.389*	19.876***
	(2.38)	(3.92)
	[0.02]	[0.00]
Liver Disease	49.894***	51.350***
	(6.72)	(7.40)
	[0.00]	[0.00]
Metastatic Carcinoma	−27.260***	−23.746***
	(−7.08)	(−7.45)
	[0.00]	[0.00]
AIDS/HIV	−3.867	5.941
	(−0.36)	(0.48)
	[0.72]	[0.63]
Observations	225344	225344
Within R^2	0.354	0.334
Numb. groups	290	3044
Min group size	2	1
Max group size	7964	6996
Avg group size	777.0	74.0

Notes: Outcome: raw cost of laboratory tests. Sample: 95th percentile of raw laboratory is dropped. Model: linear regression, Column (1) has fixed-effect on doctor, Column (2) has fixed-effect on doctor and MDC. Standard errors clustered on doctor. Robust t-statistics (p-values) in parentheses (brackets). ***p<.01; **p<.05; *p<.10.

Table 1.15: Laboratory cost: age concordance

	(1) doc FE	(2) doc+mdc FE
Ethnic only concordance	−4.289	−2.222
	(−1.80)	(−1.09)
	[0.07]	[0.28]
Gender only concordance	−4.426***	−4.196***
	(−3.43)	(−3.37)
	[0.00]	[0.00]
Age only concordance	1.647	1.794
	(0.92)	(1.08)
	[0.36]	[0.28]
Gender and ethnic concordance	−5.645*	−2.974
	(−2.18)	(−1.30)
	[0.03]	[0.19]
Gender and age concordance	−3.829	−3.184
	(−1.51)	(−1.38)
	[0.13]	[0.17]
Age and ethnic concordance	−4.745	−3.052
	(−1.81)	(−1.30)
	[0.07]	[0.20]
Gender, age and ethnic concordance	−7.180**	−5.085*
	(−2.65)	(−2.16)
	[0.01]	[0.03]
Observations	224656	224656
Within R^2	0.354	0.335
Numb. groups	284	2997
Min group size	2	1
Max group size	7964	6996
Avg group size	791.0	75.0

Notes: Outcome: raw cost of laboratory tests. Sample: 95th percentile of raw laboratory is dropped. Model: linear regression, Column (1) has fixed-effect on doctor, Column (2) has fixed-effect on doctor and MDC. Standard errors clustered on doctor. Robust t-statistics (p-values) in parentheses (brackets). ***p<.01; **p<.05; *p<.10.

Table 1.16: Laboratory quantity: non-linear regression

	(1) poisson	(2) neg. bin.
Gender only concordance	−0.042**	−0.057**
	(−2.83)	(−3.01)
	[0.00]	[0.00]
Ethnic only concordance	−0.021	−0.021
	(−0.92)	(−0.91)
	[0.36]	[0.36]
Gender and ethnic concordance	−0.025	−0.038
	(−1.10)	(−1.42)
	[0.27]	[0.15]
Observations	236534	236534

Notes: Outcome: quantity of laboratory tests. Sample: eligible study sample. Model: Column (1) fixed-effect Poisson with bootstrapped standard errors, Column (2) fixed-effect Negative Binomial with bootstrapped standard errors. Standard errors clustered on doctor. Robust t-statistics (p-values) in parentheses (brackets). ***p<.01; **p<.05; *p<.10.

1.C Results: robustness checks for laboratory cost outcome

Table 1.17: Log-transformed laboratory cost: manually matched population

	(1) doc FE	(2) doc + mdc FE
Gender only concordance	−0.028**	−0.031***
	(−3.17)	(−3.58)
Ethnic only concordance	−0.024*	−0.034**
	(−2.05)	(−2.72)
Gender and ethnic concordance	−0.024*	−0.036**
	(−1.99)	(−2.85)
Observations	342836	346857
Within R^2	0.320	0.329
Numb. groups	638	638
Min group size	1	1
Max group size	4572	4572
Avg group size	537.4	543.7

Notes: Outcome: log-transformed laboratory cost. Sample: manually and HR matched doctors, 99th percentile of ln(cost) dropped. Model: linear regression, Column (1) has fixed-effect on doctor, Column (2) has fixed-effect on doctor and MDC. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 1.18: Laboratory cost: manually matched population

	(1) doc FE	(2) doc + mdc FE
Gender only concordance	−3.562***	−3.369***
	(−3.80)	(−4.04)
Ethnic only concordance	−3.093	−2.172
	(−1.85)	(−1.43)
Gender and ethnic concordance	−3.686*	−2.425
	(−2.00)	(−1.49)
Observations	407225	407225
Within R^2	0.351	0.331
Numb. groups	648	7352
Min group size	1	1
Max group size	7964	6996
Avg group size	628.4	55.4

Notes: Outcome: raw laboratory cost. Sample: manually and HR matched doctors, 95th percentile of ln(cost) dropped. Model: linear regression, Column (1) has fixed-effect on doctor, Column (2) has fixed-effect on doctor and MDC. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 1.19: Laboratory cost: no fixed-effect

	(1) reg	(2) ln reg	(3) robust reg
Gender only concordance	−8.102***	−0.058**	−5.287***
	(−3.78)	(−3.15)	(−9.35)
Ethnic only concordance	−10.855*	−0.074*	−5.252***
	(−2.40)	(−2.41)	(−7.18)
Gender and ethnic concordance	−14.119**	−0.083**	−7.576***
	(−2.85)	(−2.61)	(−9.53)
Observations	224656	179426	224656
R^2	0.411	0.313	0.631

Notes: Outcome: raw and log-transformed laboratory cost. Sample: eligible study sample, 95th percentile of cost distribution excluded. Model: linear regression, Column (1) has no fixed-effect on doctor, Column (2) has no fixed-effect on doctor and log-transformed laboratory cost outcome, Column (3) uses a robust regression routine in *Stata 11* to identify outliers. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 1.20: Laboratory cost: addressing outliers

	(1) All	(2) 99 percent	(3) 90 percent
Gender only concordance	−3.265	−5.660*	−2.615**
	(−0.99)	(−2.39)	(−2.92)
Ethnic only concordance	−14.012**	−10.513*	−2.732
	(−2.75)	(−2.54)	(−1.88)
Gender and ethnic concordance	−11.298*	−12.511**	−4.269**
	(−2.02)	(−2.78)	(−2.70)
Observations	238916	236206	212212
Within R^2	0.414	0.392	0.303
Numb. groups	291	291	290
Min group size	2	1	2
Max group size	8251	8199	7763
Avg group size	821.0	811.7	731.8

Notes: Outcome: raw laboratory cost. Sample: eligible study sample, Column (1) has all eligible study sample, Column (2) has the 99th percentile of laboratory cost distribution excluded, Column (3) has 90th percentile of cost distribution excluded. Model: doctor fixed-effect linear regression. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 1.21: Laboratory cost: restricted medical conditions

	(1) Top MDC	(2) Top MDC fe	(3) drg fe
Gender only concordance	−4.549**	−4.166*	−2.962**
	(−2.87)	(−2.46)	(−2.64)
Ethnic only concordance	−5.988	−5.490	−1.392
	(−1.59)	(−1.57)	(−0.91)
Gender and ethnic concordance	−6.934	−6.078	−2.743
	(−1.72)	(−1.61)	(−1.50)
Observations	97212	97212	225337
Within R^2	0.433	0.431	0.231
Numb. groups	264	878	22531
Min group size	1	1	1
Max group size	7115	6996	5723
Avg group size	368.2	110.7	10.0

Notes: Outcome: raw laboratory cost. Sample: eligible study sample, Column (1) and (2) have all patients within the top four MDC; Respiratory, Digestive, Kidney and Liver, and Cardiology, Column (3) has eligible study sample. Model: linear regression, Column (1) has fixed-effect on doctor, Column (2) has fixed-effect on doctor and MDC, Column (3) has fixed-effect on doctor and DRG. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 1.22: Laboratory cost: kidney and digestive MDC

	(1) Kidney	(2) Digestive
Gender only concordance	−13.114 (−1.27)	−5.920 (−1.29)
Ethnic only concordance	−70.443* (−2.60)	−9.843* (−2.30)
Gender and ethnic concordance	−67.740* (−2.34)	−6.546 (−1.19)
Ureter stone	−15.144** (−2.93)	
Kidney stone	−8.870* (−2.26)	
Hydro calc	−17.493* (−2.37)	
Gastro hem		15.226 (1.82)
IBS		15.342 (1.38)
Obstruction		−45.677*** (−5.83)
Abd pain		−3.183 (−0.42)
Gastro en		−6.707 (−1.18)
Appendicitis		−31.229*** (−4.20)
Abd hernia		−71.063*** (−5.00)
Observations	14628	22982
Within R^2	0.459	0.599
Numb. groups	155	176
Min group size	1	1
Max group size	1452	1732
Avg group size	94.4	130.6

Notes: Outcome: raw laboratory cost. Sample: eligible study sample, Column (1) has patients in the Kidney and Liver MDC, Column (2) has patients in the Digestive MDC. Model: doctor fixed-effect linear regression, additional explanatory variables for the main diagnosis groups within each MDC. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

1.D Results: laboratory cost by MDC

Table 1.23: Log-transformed laboratory cost: MDC 1,3,4

	(1) nervous	(2) ENT	(3) respiratory
Gender only concordance	−0.035	0.064	−0.053
	(−1.18)	(1.05)	(−1.45)
Ethnic only concordance	−0.089	0.048	−0.002
	(−1.76)	(1.01)	(−0.07)
Gender and ethnic concordance	−0.078	0.157*	−0.039
	(−1.46)	(2.42)	(−0.87)
Observations	13161	5434	17220
Within R^2	0.362	0.297	0.420
Numb. groups	234	192	219
Min group size	1	1	1
Max group size	1209	486	3139
Avg group size	56.2	28.3	78.6

Notes: Outcome: log-transformed laboratory cost. Sample: eligible study sample, Column (1) and (2) and (3) have patients in MDC 1, 3, and 4 respectively. Model: doctor fixed-effect linear regression. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 1.24: Log-transformed laboratory cost: MDC 5,6,8

	(1) circulatory	(2) digestive	(3) musculosk
Gender only concordance	0.008	−0.043*	−0.057*
	(0.59)	(−2.14)	(−2.13)
Ethnic only concordance	−0.010	−0.017	−0.003
	(−0.44)	(−0.67)	(−0.03)
Gender and ethnic concordance	−0.004	−0.015	−0.006
	(−0.16)	(−0.47)	(−0.07)
Observations	31017	22426	13600
Within R^2	0.428	0.406	0.466
Numb. groups	231	221	225
Min group size	1	1	1
Max group size	1811	1704	1384
Avg group size	134.3	101.5	60.4

Notes: Outcome: log-transformed laboratory cost. Sample: eligible study sample, Column (1) and (2) and (3) have patients in MDC 5, 6, and 8 respectively. Model: doctor fixed-effect linear regression. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 1.25: Log-transformed laboratory cost: MDC 9,10,11

	(1) skin	(2) endocrine	(3) kidney
Gender only concordance	0.004	0.007	−0.066**
	(0.12)	(0.18)	(−2.70)
Ethnic only concordance	−0.009	0.043	−0.179
	(−0.25)	(0.38)	(−1.56)
Gender and ethnic concordance	0.002	0.047	−0.192
	(0.04)	(0.41)	(−1.65)
Observations	7926	2962	14111
Within R^2	0.299	0.377	0.385
Numb. groups	216	195	192
Min group size	1	1	1
Max group size	442	214	1433
Avg group size	36.7	15.2	73.5

Notes: Outcome: log-transformed laboratory cost. Sample: eligible study sample, Column (1) and (2) and (3) have patients in MDC 9, 10, and 11 respectively. Model: doctor fixed-effect linear regression. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 1.26: Log-transformed laboratory cost: MDC 14,16,18

	(1) pregnancy	(2) blood dis.	(3) infectious
Gender only concordance	−0.684***	−0.029	−0.003
	(−7.82)	(−0.50)	(−0.09)
Ethnic only concordance	−0.082*	0.038	0.074
	(−1.99)	(0.28)	(1.39)
Gender and ethnic concordance	0.684***	−0.041	0.124**
	(7.82)	(−0.31)	(2.63)
Observations	11242	1959	5352
Within R^2	0.211	0.383	0.384
Numb. groups	82	175	223
Min group size	1	1	1
Max group size	2997	142	302
Avg group size	137.1	11.2	24.0

Notes: Outcome: log-transformed laboratory cost. Sample: eligible study sample, Column (1) and (2) and (3) have patients in MDC 14, 16, and 18 respectively. Model: doctor fixed-effect linear regression. Standard errors clustered on doctor. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

1.E Information on patients matched with HR data and not matched with HR data

Table 1.27: Summary statistics for HR merged population

	Mean	*Sd*	*Min*	*Max*	*P*50
Age	49.50	22.76	5.00	107.00	48.00
Male	0.43	0.49	0.00	1.00	0.00
Deprivation scale	5.91	2.92	0.00	10.00	6.00
After 5pm entry	0.26	0.44	0.00	1.00	0.00
After 12pm entry	0.32	0.46	0.00	1.00	0.00
Overnight admiss.	0.08	0.27	0.00	1.00	0.00
Transfer	0.12	0.32	0.00	1.00	0.00
LOS	3.96	7.24	0.00	364.00	1.00
Theatre event	0.19	0.39	0.00	1.00	0.00
AED entry	0.32	0.47	0.00	1.00	0.00
Acute admiss.	0.73	0.45	0.00	1.00	1.00
Prev admiss 60days	0.33	0.47	0.00	1.00	0.00
Accident	0.11	0.32	0.00	1.00	0.00
Day patient	0.11	0.31	0.00	1.00	0.00
Home DHB	0.65	0.48	0.00	1.00	1.00
Diagnosis count	4.79	3.92	1.00	65.00	4.00
Pacific pat	0.13	0.34	0.00	1.00	0.00
African pat	0.01	0.08	0.00	1.00	0.00
Maori patient	0.09	0.29	0.00	1.00	0.00
Middleeastern pat	0.01	0.10	0.00	1.00	0.00
Other ethnic pat	0.02	0.15	0.00	1.00	0.00
Asian pat	0.08	0.27	0.00	1.00	0.00
Indian pat	0.05	0.22	0.00	1.00	0.00
European pat	0.60	0.49	0.00	1.00	1.00
MDC 0	0.00	0.01	0.00	1.00	0.00
MDC 2	0.04	0.19	0.00	1.00	0.00
MDC 3	0.02	0.15	0.00	1.00	0.00
MDC 4	0.08	0.27	0.00	1.00	0.00
MDC 5	0.14	0.35	0.00	1.00	0.00
MDC 6	0.10	0.30	0.00	1.00	0.00
MDC 7	0.02	0.14	0.00	1.00	0.00
MDC 8	0.08	0.27	0.00	1.00	0.00
MDC 9	0.04	0.18	0.00	1.00	0.00
MDC 10	0.01	0.12	0.00	1.00	0.00
MDC 11	0.06	0.23	0.00	1.00	0.00
MDC 12	0.01	0.10	0.00	1.00	0.00
MDC 13	0.02	0.13	0.00	1.00	0.00
MDC 14	0.17	0.37	0.00	1.00	0.00
MDC 15	0.02	0.15	0.00	1.00	0.00
MDC 17	0.03	0.16	0.00	1.00	0.00
MDC 18	0.02	0.15	0.00	1.00	0.00
MDC 19	0.00	0.06	0.00	1.00	0.00
MDC 20	0.00	0.04	0.00	1.00	0.00
MDC 21	0.03	0.17	0.00	1.00	0.00

MDC 22	0.00	0.01	0.00	1.00	0.00
MDC 23	0.05	0.22	0.00	1.00	0.00
Myocardial Infarct.	0.04	0.20	0.00	1.00	0.00
Congestive Heart F	0.04	0.20	0.00	1.00	0.00
Periphral Vascular Dis	0.02	0.15	0.00	1.00	0.00
Cerebrovascular Dis	0.03	0.18	0.00	1.00	0.00
Dementia	0.02	0.13	0.00	1.00	0.00
Chronic Pulmonary D	0.04	0.19	0.00	1.00	0.00
Rheumatic Disease	0.01	0.07	0.00	1.00	0.00
Peptic Ulcer Disease	0.00	0.07	0.00	1.00	0.00
Mild Liver Disease	0.02	0.13	0.00	1.00	0.00
Diabetes w/o complic.	0.04	0.20	0.00	1.00	0.00
Diabetes w complic.	0.07	0.25	0.00	1.00	0.00
Paraplegia + Hemiplegia	0.02	0.13	0.00	1.00	0.00
Renal Disease	0.07	0.25	0.00	1.00	0.00
Cancer	0.10	0.30	0.00	1.00	0.00
Liver Disease	0.00	0.07	0.00	1.00	0.00
Metastatic Carcinoma	0.04	0.20	0.00	1.00	0.00
AIDS/HIV	0.00	0.04	0.00	1.00	0.00
N	417123				

Notes: Summary statistics for patients that have a casemanager that is merged with HR data, this includes patient with a casemanager that does not have information on gender and/or ethnic group.

Table 1.28: Summary statistics for non-HR merged population

	$Mean$	Sd	Min	Max	$P50$
Consultant	0.11	0.31	0.00	1.00	0.00
Control variables:					
Age	35.56	23.96	5.00	104.00	32.00
Male	0.37	0.48	0.00	1.00	0.00
Deprivation scale	5.76	2.94	0.00	10.00	6.00
After 5pm entry	0.24	0.43	0.00	1.00	0.00
After 12pm entry	0.26	0.44	0.00	1.00	0.00
Overnight admiss.	0.08	0.28	0.00	1.00	0.00
Transfer	0.07	0.26	0.00	1.00	0.00
LOS	2.82	5.85	0.00	369.00	1.00
Theatre event	0.22	0.42	0.00	1.00	0.00
AED entry	0.19	0.39	0.00	1.00	0.00
Acute admiss.	0.58	0.49	0.00	1.00	1.00
Prev admiss 60days	0.34	0.47	0.00	1.00	0.00
Accident	0.08	0.27	0.00	1.00	0.00
Day patient	0.18	0.38	0.00	1.00	0.00
Home DHB	0.55	0.50	0.00	1.00	1.00
Diagnosis count	4.05	3.37	1.00	63.00	3.00
Pacific pat	0.13	0.34	0.00	1.00	0.00
African pat	0.01	0.08	0.00	1.00	0.00
Maori patient	0.10	0.31	0.00	1.00	0.00
Middleeastern pat	0.01	0.10	0.00	1.00	0.00
Other ethnic pat	0.02	0.13	0.00	1.00	0.00
Asian pat	0.09	0.29	0.00	1.00	0.00

Indian pat	0.05	0.21	0.00	1.00	0.00
European pat	0.59	0.49	0.00	1.00	1.00
MDC 0	0.00	0.01	0.00	1.00	0.00
MDC 2	0.03	0.17	0.00	1.00	0.00
MDC 3	0.04	0.19	0.00	1.00	0.00
MDC 4	0.08	0.27	0.00	1.00	0.00
MDC 5	0.10	0.29	0.00	1.00	0.00
MDC 6	0.07	0.25	0.00	1.00	0.00
MDC 7	0.01	0.09	0.00	1.00	0.00
MDC 8	0.07	0.26	0.00	1.00	0.00
MDC 9	0.04	0.19	0.00	1.00	0.00
MDC 10	0.02	0.15	0.00	1.00	0.00
MDC 11	0.04	0.19	0.00	1.00	0.00
MDC 12	0.00	0.06	0.00	1.00	0.00
MDC 13	0.01	0.08	0.00	1.00	0.00
MDC 14	0.29	0.45	0.00	1.00	0.00
MDC 15	0.03	0.17	0.00	1.00	0.00
MDC 17	0.06	0.24	0.00	1.00	0.00
MDC 18	0.02	0.14	0.00	1.00	0.00
MDC 19	0.00	0.06	0.00	1.00	0.00
MDC 20	0.00	0.03	0.00	1.00	0.00
MDC 21	0.03	0.16	0.00	1.00	0.00
MDC 22	0.00	0.01	0.00	1.00	0.00
MDC 23	0.02	0.13	0.00	1.00	0.00
Myocardial Infarct.	0.03	0.16	0.00	1.00	0.00
Congestive Heart F	0.03	0.16	0.00	1.00	0.00
Periphral Vascular Dis	0.01	0.12	0.00	1.00	0.00
Cerebrovascular Dis	0.02	0.13	0.00	1.00	0.00
Dementia	0.01	0.10	0.00	1.00	0.00
Chronic Pulmonary D	0.05	0.21	0.00	1.00	0.00
Rheumatic Disease	0.01	0.09	0.00	1.00	0.00
Peptic Ulcer Disease	0.00	0.05	0.00	1.00	0.00
Mild Liver Disease	0.01	0.09	0.00	1.00	0.00
Diabetes w/o complic.	0.03	0.18	0.00	1.00	0.00
Diabetes w complic.	0.04	0.19	0.00	1.00	0.00
Paraplegia + Hemiplegia	0.01	0.10	0.00	1.00	0.00
Renal Disease	0.05	0.21	0.00	1.00	0.00
Cancer	0.13	0.34	0.00	1.00	0.00
Liver Disease	0.00	0.05	0.00	1.00	0.00
Metastatic Carcinoma	0.04	0.18	0.00	1.00	0.00
AIDS/HIV	0.00	0.04	0.00	1.00	0.00
N	114070				

Notes: Summary statistics for patients that have a casemanager that is not merged with HR data.

1.F Manually associating gender and ethnic group for doctors

A portion of inpatient records have a doctor that is unable to be matched with HR information. A manual match process was then undertaken using information on the first and lastname of doctors in inpatient data to associate a gender and ethnic group. Manual matching on doctors names provided a secondary dataset to explore the robustness of my results, this is discussed in the results section. Main results use data that is matched using information from HR records only.

To achieve more demographic matches, and investigate sampling bias if certain ethnicity groupings or genders are less likely to complete doctor information forms, a manual match process was undertaken to associate a gender using casemanager's first name and an ethnic group using the ethnic origin of the lastname[13]. Casemanagers that are not able to be identified are those with an initial for a firstname, generic consultant (e.g. AED consultant) or a team of people. Gender neutral first names were also not matched. This dataset allows me to check the robustness of results. Main results are estimated on patients matched with HR records.

Table 1.29 has details for the number of casemanagers matched after HR and manual matching. There are 941 unique doctor names in the inpatient data that are matched with gender and ethnicity information. There is a high number of casemanagers because the same individual can have different names in inpatient data. For example, the same individual can enter as Mr. X or Dr. X. There is no uniform pattern for this. Separate fields contain information on first names, preferred names,

[13]Fiscella and Fremont (2006) review published articles using information on last names to associate an ethnic group for health services research, they find it can perform well and is a useful tool when information on an ethnic group is lacking. They find it is poor for predicting African American ethnicity groupings in the United States. A similar problem might occur in our data for South African and Maori individuals incorrectly associated with a European ethnic group. Using first names for gender matching is less commonly used. Cassidy et al. (1999) study the phonology of English names and gender.

middle names, last name but this information also varies across inpatient names for the same individual. That is, some casemanagers in inpatient data only have initials, or are missing a third name or preferred name. It is therefore difficult to identify individuals with different inpatient doctor names, because we do not have complete information on first, middle and lastname. There might be some advantages to using the doctor name as it appears in inpatient data as the unique identifier. It sometimes refers to a medical situation (e.g. on call), or area of the hospital. When using fixed effects on casemanagers, comparisons are more likely to be made across patients that are treated in clinically similar situations, in addition to the same individual. A disadvantage is smaller group sizes within casemanagers and therefore less precise estimates. The percentage of European, Indian and Asian casemanagers is similar to HR data with 83.2%, 6.7%, and 5.6% respectively. Though there is an increase in European doctors which might raise concern of inaccurately associating a European ethnic group to a patient of a different ethnic group. The average number of patients per doctor name is smaller at 539. The percentage of male casemanagers is also similar, though there is an increase in matching for maternity admissions. This is most likely due to increased matching of midwives.

To associate a gender and ethnic group based on a doctor's name, on-line service where used to confirm the likelihood of a firstname being male or female and the ethnic origin of a lastname. For gender matching, *www.firstnamesex.com* which gives the proportion of male and female names for a given first name using 1990 US Census data. This website did not have comprehensive information on gender for firstnames from minority ethnicity groupings, such as Asian, Indian, Pacific and Maori, therefore *www.gpeters.com* was used as well. This uses information from the internet to estimate how likely a firstname refers to a male or female sex. Situations where the gender of a name is not clear is not matched.

Surnames where associated with European, Asian, Indian, Pacific, Maori and Middle Eastern ethnicity groupings. African ethnicities are difficult to identify because lastnames often have a European or dutch heritage. Maori casemanagers may also be incorrectly identified as European, because there is a longer history of anglicising surnames and intermarriage. This might result in an overpopulation of European casemanagers. *www.geneology.com* provide information on the ethnic group composition of a surname using immigration data to the U.S.

Table 1.29: Doctor characteristics for manually matched population

	N	Mean	Sd	Min	Max
European doc	941	0.832	0.374	0	1
Asian doc	941	0.067	0.250	0	1
Indian doc	941	0.056	0.231	0	1
Male	941	0.552	0.498	0	1
N patients	941	539.555	825.677	1	8287

Notes: N=941 is number of doctors. Summary statistics for doctors that are merged with HR data and manually associated a gender and ethnic group.

1.G Summary statistics for patients with zero laboratory tests

Table 1.30: Study variables for matched population with lab costs of zero

	$Mean$	Sd	Min	Max	$P50$
Transfer	0.11	0.31	0.00	1.00	0.00
LOS	0.55	1.63	0.00	131.00	0.00
Theatre event	0.12	0.33	0.00	1.00	0.00
AED entry	0.10	0.30	0.00	1.00	0.00
Acute admiss.	0.42	0.49	0.00	1.00	0.00
Prev admiss 60days	0.40	0.49	0.00	1.00	0.00
Accident	0.15	0.36	0.00	1.00	0.00
Day patient	0.34	0.47	0.00	1.00	0.00
Own DHB	0.44	0.50	0.00	1.00	0.00
Diagnosis count	2.81	1.91	1.00	36.00	2.00
Admit 2000	0.00	0.00	0.00	0.00	0.00
Admit 2001	0.00	0.00	0.00	0.00	0.00
Admit 2002	0.00	0.00	0.00	0.00	0.00
Admit 2003	0.06	0.23	0.00	1.00	0.00

Admit 2004	0.11	0.32	0.00	1.00	0.00
Admit 2005	0.13	0.34	0.00	1.00	0.00
Admit 2006	0.14	0.35	0.00	1.00	0.00
Admit 2007	0.15	0.36	0.00	1.00	0.00
Admit 2008	0.13	0.34	0.00	1.00	0.00
Admit 2009	0.12	0.33	0.00	1.00	0.00
Admit 2010	0.11	0.31	0.00	1.00	0.00
Admit 2011	0.05	0.21	0.00	1.00	0.00
MDC 0	0.00	0.00	0.00	1.00	0.00
MDC 2	0.17	0.38	0.00	1.00	0.00
MDC 3	0.03	0.16	0.00	1.00	0.00
MDC 4	0.10	0.30	0.00	1.00	0.00
MDC 5	0.08	0.27	0.00	1.00	0.00
MDC 6	0.03	0.16	0.00	1.00	0.00
MDC 7	0.01	0.10	0.00	1.00	0.00
MDC 8	0.09	0.28	0.00	1.00	0.00
MDC 9	0.02	0.13	0.00	1.00	0.00
MDC 10	0.01	0.10	0.00	1.00	0.00
MDC 11	0.05	0.21	0.00	1.00	0.00
MDC 12	0.01	0.12	0.00	1.00	0.00
MDC 13	0.01	0.12	0.00	1.00	0.00
MDC 14	0.15	0.35	0.00	1.00	0.00
MDC 15	0.03	0.18	0.00	1.00	0.00
MDC 17	0.08	0.26	0.00	1.00	0.00
MDC 18	0.00	0.06	0.00	1.00	0.00
MDC 19	0.01	0.09	0.00	1.00	0.00
MDC 20	0.00	0.02	0.00	1.00	0.00
MDC 21	0.03	0.17	0.00	1.00	0.00
MDC 22	0.00	0.01	0.00	1.00	0.00
MDC 23	0.05	0.22	0.00	1.00	0.00
N	51301				

Notes: Summary statistics for patients that no laboratory tests during their inpatient event.

1.H Male and female doctor characteristics

Table 1.31: Male doctor characteristics

	N	Mean	Sd	Min	Max
European doc	234	0.850	0.357	0	1
Asian doc	234	0.068	0.253	0	1
Indian doc	234	0.051	0.221	0	1
MDC 1	234	0.081	0.187	0	1
MDC 3	234	0.091	0.234	0	1
MDC 4	234	0.079	0.138	0	1
MDC 5	234	0.193	0.289	0	1
MDC 6	234	0.088	0.142	0	.7190388
MDC 7	234	0.024	0.081	0	.512474
MDC 8	234	0.069	0.163	0	.875
MDC 9	234	0.045	0.101	0	.9
MDC 10	234	0.018	0.033	0	.2077509
MDC 11	234	0.058	0.154	0	.8186077
MDC 12	234	0.009	0.037	0	.25
MDC 13	234	0.015	0.078	0	.8
MDC 14	234	0.045	0.190	0	1
MDC 16	234	0.014	0.042	0	.4868421
MDC 18	234	0.026	0.040	0	.5062241
MDC 19	234	0.002	0.005	0	.0448896
MDC 20	234	0.001	0.003	0	.0275344
MDC 21	234	0.033	0.071	0	1
MDC 22	234	0.000	0.000	0	.0037523
MDC 23	234	0.040	0.128	0	1
Proportion theatre	234	0.195	0.248	0	1

Notes: N is number of doctors. Summary statistics for male doctors. The MDC is identified as the MDC that most of their patients are treated in. Proportion theatre variable is the proportion of doctor's patients that have a surgical procedure.

Table 1.32: Female doctor characteristics

	N	Mean	Sd	Min	Max
European doc	150	0.793	0.406	0	1
Asian doc	150	0.073	0.262	0	1
Indian doc	150	0.040	0.197	0	1
MDC 1	150	0.028	0.088	0	.972973
MDC 3	150	0.021	0.104	0	1
MDC 4	150	0.038	0.092	0	.8169935
MDC 5	150	0.079	0.186	0	.9606937
MDC 6	150	0.040	0.085	0	.5754717
MDC 7	150	0.006	0.023	0	.2075472
MDC 8	150	0.040	0.143	0	.9502763
MDC 9	150	0.013	0.024	0	.1666667
MDC 10	150	0.007	0.014	0	.0747331
MDC 11	150	0.013	0.042	0	.3898959
MDC 12	150	0.001	0.005	0	.0630252
MDC 13	150	0.027	0.094	0	.5880952
MDC 14	150	0.516	0.484	0	1
MDC 16	150	0.021	0.093	0	.7712464
MDC 18	150	0.016	0.038	0	.2733119
MDC 19	150	0.001	0.003	0	.0192308
MDC 20	150	0.001	0.002	0	.0240385
MDC 21	150	0.011	0.029	0	.2727273
MDC 22	150	0.000	0.000	0	.0024038
MDC 23	150	0.060	0.196	0	.9963167
Proportion theatre	150	0.114	0.187	0	1

Notes: N is number of doctors. Summary statistics for female doctors. The MDC is identified as the MDC that most of their patients are treated in. Proportion theatre variable is the proportion of doctor's patients that have a surgical procedure.

1.I MDC and DRG information for hospital sample

Table 1.33: hospital admission medical conditions

MDC	*% MDC*	*DRG desc. (trimmed)*	*% DRG*
1	0.061	Stroke W/O Catastrophic or Severe CC	0.064
1	0.061	Cranial and Peripheral Nerve Disorders W	0.089
1	0.061	Seizure W/O Catastrophic or Severe CC	0.058
1	0.061	Headache	0.092
1	0.061	Other Disorders of the Nervous System W/	0.054
2	0.042	Retinal Procedures	0.074
2	0.042	Acute and Major Eye Infections Age <55 W	0.054
2	0.042	Neurological and Vascular Disorders of t	0.060
2	0.042	Hyphema and Medically Managed Trauma to	0.178
2	0.042	Other Disorders of the Eye W/O CC	0.441
3	0.026	Tonsillectomy and/or Adenoidectomy	0.080
3	0.026	Dysequilibrium	0.129
3	0.026	Epistaxis	0.123
3	0.026	Otitis Media and URI W CC	0.067

3	0.026	Otitis Media and URI W/O CC	0.195
3	0.026	Other Ear, Nose; Mouth and Throat Diagno	0.121
4	0.087	Respiratory Infections/Inflammations W S	0.068
4	0.087	Respiratory Infections/Inflammations W/O	0.073
4	0.087	Sleep Apnoea	0.243
4	0.087	Chr Obstruct Airway Disease W Cat/Sev CC	0.073
4	0.087	Chronic Obstructive Airways Disease W/O	0.068
5	0.136	Percutaneous Coronary Intervention W AMI	0.070
5	0.136	Non-Major Arrhythmia and Conduction Diso	0.078
5	0.136	Syncope and Collapse W/O Catastrophic or	0.055
5	0.136	Chest Pain	0.167
6	0.094	Appendicectomy W/O Catastrophic or Sever	0.101
6	0.094	Abdominal Pain/Mesenteric Adenitis no CC	0.190
6	0.094	Oesophagitis, Gastroent + Misc Digestive	0.063
6	0.094	Oesophagitis, Gastroent + Misc Digestive	0.196
7	0.018	Laparoscopic Cholecystectomy W/O Closed	0.073
7	0.018	Malignancy of Hepatobiliary Sys,Panc (Ag	0.053
7	0.018	Disorders of Pancreas Except for Maligna	0.069
7	0.018	Disorders of Liver Excep Malig, Cirrhosi	0.130
7	0.018	Disorders of the Biliary Tract W CC	0.061
7	0.018	Disorders of the Biliary Tract W/O CC	0.145
8	0.068	Humerus, Tibia; Fibula and Ankle Procedu	0.063
8	0.068	Non-surgical Spinal Disorders W/O CC	0.065
8	0.068	Injury to Forearm, Wrist; Hand or Foot A	0.072
8	0.068	Inj to Shoulder,Arm;Leg;etc<65 No CC	0.075
9	0.033	Cellulitis Age >59 W Catastrophic or Sev	0.055
9	0.033	Cellulitis >59 W/O Catast/Sev CC or <60	0.524
9	0.033	Trauma to the Skin, Subcutaneous Tissue	0.072
10	0.013	Diabetic Foot Procedures	0.072
10	0.013	Diabetes W Catastrophic or Severe CC	0.096
10	0.013	Diabetes W/O Catastrophic or Severe CC	0.191
10	0.013	Miscellaneous Metabolic Disorders W Cata	0.057
10	0.013	Miscellaneous Metabolic Disorders Age >7	0.161
10	0.013	Miscellaneous Metabolic Disorders Age <7	0.131
10	0.013	Inborn Errors of Metabolism	0.050
10	0.013	Endocrine Disorders W/O Catastrophic or	0.112
11	0.065	Kidney and Urinary Tract Infections Age	0.064
11	0.065	Kidney and Urinary Tract Infections Age	0.115
11	0.065	Urinary Stones and Obstruction	0.225
11	0.065	Kidney and Urinary Tract Signs and Sympt	0.083
11	0.065	Oth Kidney+Urinary Tr Diag No Cat/Sev CC	0.143
12	0.012	Testes Procedures W/O CC	0.223
12	0.012	Inflammation of the Male Reproductive Sy	0.079
12	0.012	Inflammation of the Male Reproductive Sy	0.326
12	0.012	Other Male Reproductive System Diagnoses	0.163
13	0.018	Other Uterine + Adnexa Procedures for No	0.095
13	0.018	Conisation, Vagina; Cervix and Vulva Pro	0.191
13	0.018	Infections, Female Reproductive System	0.145
13	0.018	Menstrual and Other Female Reproductive	0.070
13	0.018	Menstrual and Other Female Reproductive	0.319
14	0.158	Caesarean Delivery W/O Catastrophic or S	0.094
14	0.158	Vaginal Delivery W Catastrophic or Sever	0.052
14	0.158	Vaginal Delivery W/O Catastrophic or Sev	0.194

14	0.158	Postpartum and Post Abortion W/O O.R. Pr	0.060
14	0.158	Abortion W/O O.R. Procedure	0.057
14	0.158	Antenatal + Other Obstetric Admission	0.106
14	0.158	Antenatal + Other Obstetric Admission, S	0.113
16	0.020	Reticuloendothelial+Immun Dis+Cat/Sev CC	0.089
16	0.020	Reticuloendothelial+Imm Dis No C/S CC	0.138
16	0.020	Reticuloendothelial and Immunity Disorde	0.452
16	0.020	Red Blood Cell Disorders W/O Catastrophi	0.144
16	0.020	Coagulation Disorders	0.074
17	0.035	Acute Leukaemia W Severe CC	0.057
17	0.035	Acute Leukaemia W/O Catastrophic or Seve	0.103
17	0.035	Lymphoma and Non-Acute Leukaemia W/O Cat	0.073
17	0.035	Lymphoma and Non-Acute Leukaemia, Sameda	0.277
17	0.035	Chemotherapy	0.323
18	0.022	O.R. Procedures for Infectious and Paras	0.051
18	0.022	Septicaemia W Catastrophic or Severe CC	0.201
18	0.022	Postoperative + Post-traumatic Infection	0.127
18	0.022	Postop + Post-Traum Infect <55 No C/S CC	0.082
18	0.022	Fever of Unknown Origin W CC	0.070
18	0.022	Viral Illness Age >59 or W CC	0.062
18	0.022	Viral Illness Age <60 W/O CC	0.167
19	0.003	Mental Health Treatment,Sameday;W/O ECT	0.164
19	0.003	Other Affective and Somatoform Disorders	0.055
19	0.003	Anxiety Disorders	0.632
20	0.001	Alcohol Intoxication and Withdrawal W CC	0.179
20	0.001	Alcohol Intoxication and Withdrawal W/O	0.177
20	0.001	Alcohol Intoxication and Withdrawal	0.453
20	0.001	Other Drug Use Disorder and Dependence	0.061
21	0.028	Other Procedures for Other Injuries W Ca	0.058
21	0.028	Other Procedures for Other Injuries W/O	0.087
21	0.028	Injuries Age < 65	0.194
21	0.028	Poisoning/Toxic Effects of Drugs + Other	0.072
21	0.028	Poison/Tox Eff-Drugs,Oth Subs <60 No CC	0.061
21	0.028	Sequelae of Treatment W Catastrophic or	0.065
21	0.028	Sequelae of Treatment W/O Catastrophic o	0.215
22	0.000	Other O.R. Procedures for Other Burns	0.108
22	0.000	Burns,Tran Oth Ac Care Facility< 5 Days	0.216
22	0.000	Severe Burns	0.054
22	0.000	Other Burns Age >64 or W (Cat or Sev CC)	0.216
22	0.000	Oth Burns<65 No C/S CC No Comp Diag/Proc	0.378
23	0.057	Rehabilitation W Catastrophic or Severe	0.503
23	0.057	Rehabilitation W/O Catastrophic or Sever	0.144
23	0.057	F-Up After Completed Treat W/O Endoscopy	0.072
23	0.057	Other Factors Influencing Health Status,	0.171

Notes: Summary of medical conditions in the eligible study sample. For each MDC, the 'prop MDC' column provides the percentage of study sample within that MDC. The top DRGs are listed within each MDC, with the proportion of patients with that DRG in an MDC.

Chapter 2

Emergency Caesarean Procedures and Provider-Patient Ethnic Concordance

2.1 Introduction

Rates of caesarean section procedures have been increasing in many industrialised countries worldwide (Anderson, 2004; Declercq et al., 2011; Knight and Sullivan, 2010). The World Health Organisation estimates the global costs of 'excessive' caesarean procedures at USD 2.32 billion compared to USD 432 million for clinically 'needed' caesarean procedures (Gibbons et al., 2010). Caesarean rates also vary markedly across countries, the highest rates are in Latin American countries, particularly Brazil (45.9%) and Dominican Republic (41.9%) (Gibbons et al., 2010)[1].

In addition to variation across countries in caesarean procedures, several studies have shown significant variation in caesarean rates within countries across ethnicity groupings. In New Zealand, Maori and Pacific peoples have the lowest rates of caesarean procedures (17% and 19.12% respectively), and Asian and European patients have the highest rates of caesarean procedures (27.3% and 27.1% respectively) (Ministry of Health, 2012). A number of studies have also shown that variation across ethnicity groupings exists even after controlling for a variety of confounding factors, like a mother's clinical condition, socioeconomic status, age and so on (von Katterfeld et al., 2011; Vangen et al., 2000; Ibison, 2005). For example, Harris et al. (2007) showed that the difference in caesarean rates for Maori and non-Maori cannot be fully explained by clinical variables, age and socioeconomic factors.

Increasing caesarean section rates and variation in caesarean procedures across ethnicity groupings suggest non-clinical factors, such as preferences or miscommunication, may play a role in the decision to have a caesarean procedure. Non-clinical factors affecting the decision to have a caesarean procedure are of interest to health policy makers concerned with the cost and equity of health care provision. This is because non-clinical factors could potentially be influenced by policy on how health care is delivered.

[1]The World Health Organisation considers national caesarean section rates above 15% 'excessive'.

In this paper I estimate a relationship between doctor-patient ethnic concordance and the likelihood a patient receives an emergency caesarean procedure[2]. This paper is motivated to distinguish between three explanations for ethnic-based variation in caesarean procedures. I identify these three explanations, and suggest that investigating provider-patient ethnic concordance helps to distinguish between the three explanations. I refer to these three explanations as; 'patient ethnic group', 'supply-side', and 'casemanager composition'. I therefore estimate the impact of 'patient ethnic group', 'supply-side', and 'casemanager composition' in explaining ethnic-based variation in caesarean rates. Previous studies have been unable to distinguish between these three explanations, largely because of data limitations.

The first explanation for ethnic-based variation in caesarean procedures, is differences in biological characteristics, health behaviours and cultural preferences across ethnicity groupings. These biological characteristics, health behaviours and cultural preferences are associated with a higher or lower likelihood of a caesarean procedure. Health behaviours and biological characteristics of women which could affect caesarean rates include; weight, height, physical build, smoking rates, child-birth rates, age and so on. For example, pregnant Maori and Pacific women are on average younger, and younger women may be more physically able to undergo a natural labour, and this might therefore explain lower rates of caesarean procedures for Maori and Pacific women (Ministry of Health, 2012). In other words, if some ethnicity groupings are correlated with certain biological characteristics, health behaviours and/or cultural preferences, that are also associated with higher or lower rates of caesarean procedures, this could explain differences in caesarean procedure rates across ethnicity groupings.

These 'patient ethnic group' characteristics are decision factors that are correlated with a patient belonging to an ethnic group, and are relevant decision factors for *all*

[2]A casemanager is defined by the hospital as the primary caregiver for a women who gave birth in hospital, and could include a doctor or nurse. See Section 2.3.3.

casemanagers treating patients from that ethnic group. Often these 'patient ethnic group effects' are not fully observed in administrative data. For example, the weight and height of women, or their stated preferences for a caesarean procedure are not often recorded in hospital administrative data. Therefore, previous studies on the sources of variation in caesarean rates across ethnicity groupings typically do not adequately control for all confounding biological characteristics, health behaviours and cultural preferences associated with a patient's ethnic group.

A second source for ethnic variation in caesarean procedures, are that some ethnicity groupings could be more likely to be treated in regional areas or by casemanagers that use high or low amounts of caesarean procedures on average. In this case, a casemanager has a propensity to advise a caesarean procedure that is *independent of a patient's ethnic group*. For example, Maori women might be more likely to be treated by less specialised casemanagers that use low amounts of caesarean procedures on average. This may accordingly explain low rates of caesarean procedures for Maori patients. This is referred to as a 'supply-side' explanation in this paper. Previous studies that attempt to explain variation in caesarean procedure rates across ethnicity groupings typically do not observe the identity of a women's caregiver, even if regional and hospital-specific factors are controlled for in a regression analysis.

A final source for ethnic variation in caesarean procedures, is an *interaction* between a *patient's ethnic group* and *casemanager characteristics*. Casemanager characteristics could include demographic characteristics and/or beliefs. Demographic characteristics could include a doctor's gender, age, and ethnic group. Casemanager beliefs could include prejudices and stereotypes relating to patient ethnicity groupings. An example of a belief would be a casemanager that is biased against a patient because of their ethnic group. This may result in doctors with that belief undertreating patients from that ethnic group. Casemanager demographic characteristics could also influence caesarean procedure decisions. For example, an older casem-

anager may be more experienced in diagnosing and treating patients from minority ethnicity groupings.

The interaction between a patient's ethnic group and casemanager's demographic characteristics and/or beliefs could explain ethnic-based variation in caesarean procedures if there is a large or small proportion of casemanagers from demographic and belief groups in the population. For example, if the majority of casemanagers are European, and casemanagers from this ethnic group are more likely to believe an emergency caesarean is more clinically appropriate for an Asian patient than an Asian casemanager, this could explain a high rate of caesarean procedures for Asian women. This is referred to as a 'casemanager composition' explanation in this paper, because the composition of casemanager types in the population could be contributing to ethnic disparities in care.

An example of a 'casemanager composition' explanation for ethnic-based variation in medical treatment is Balsa and McGuire's (2003) model for ethnic disparities in medical treatment. In their model, ethnic disparities in medical treatment could be due to doctors' prejudices, stereotypes and/or miscommunication. There are two types of doctors in their model, black and white, and they may alter treatment for black and white patients because of stereotypes, poor communication, or a bias against an ethnic group other than their own. For example, in their miscommunication model doctors and patients that are ethnically discordant have poorer communication, and this can lead to over or under-treatment, relative to a patient's true clinical need. Previous empirical studies often do not have information on the identity of the casemanager and therefore their demographic characteristics. This has limited previous estimates for casemanager-patient ethnic interactions and medical treatment decisions.

In this paper I estimate the effect of a patient and casemanager having the same ethnic group on the likelihood of having an emergency caesarean procedure. I use re-

gression analysis to control for 'supply-side' and 'patient ethnic group' explanations. I then test for the statistical significance of an interaction term for casemanager-patient ethnic group concordance on emergency caesarean outcomes. Based on the above three hypothesised sources for ethnic variation in caesarean procedures, if 'patient ethnic group' (first explanation) and/or 'supply-side' (second explanation) explain all the variation in caesarean rates, I would not expect to find a statistically significant term for casemanager-patient ethnic concordance on emergency caesarean outcomes[3]. This is because I control for each casemanager's average propensity to advise an emergency caesarean across all of their patients in my regression, that is, my 'supply-side' explanation. Secondly, I would not expect a significant interaction term for casemanager-patient ethnic concordance in my regression analysis if all casemanagers use the same 'patient ethnic group' factors when making emergency caesarean decisions. Because 'patient ethnic group' decision factors refer to the clinical need *or* preferences for caesarean procedures (associated with a patient's ethnic group) they should apply equally to *all* doctors treating patients from an ethnic group. Therefore, in my regression analysis, a significant term for casemanager-patient ethnic concordance would suggest that variation in caesarean rates across ethnicity groupings is not solely determined by clinical appropriateness and/or ethnicity-specific preferences.

My empirical strategy uses an interaction model to test for the statistical significance of ethnic concordance. An interaction term is equal to one when a patient and casemanager have the same ethnic group, and is zero otherwise. I estimate a logit model for emergency caesarean outcomes on a cross-section of casemanagers.

I use data on birth events from a large hospital in New Zealand. The three largest casemanager ethnicity groupings in my data are European, Asian and Indian[4]. There

[3]Casemanager-patient is used in this paper to refer to the casemanager and patient pair for an inpatient event (i.e. birth event).

[4]Asian ethnic group includes ethnicities from East and South East Asia, i.e China, Taiwan, Japan, Korea, Burma, Thailand, Laos, Cambodia, Vietnam, Indonesia, Malaysia, Singapore, East Timor, Brunei and the Philippines. I decided to distinguish an Indian ethnic group from Asian ethnicity groupings, even though India is part of the South Asia region

are low numbers of casemanagers from Maori and Pacific ethnicity groupings. I am therefore unable to estimate the impact of Maori and Pacific ethnic concordance on caesarean procedures.

In this paper, I investigate whether a patient from an Asian, European, or Indian ethnic group is more or less likely to receive an emergency caesarean when treated by a casemanager of the same ethnic group relative to a casemanager of a different ethnic group, after controlling for the effect of a patient belonging to a particular ethnic group and a casemanager's propensity to advise caesarean procedures across all of their patients.

My main result is that Asian patients with an Asian casemanager are associated with a statistically significantly lower likelihood of an emergency caesarean procedure relative to treatment by both European and Indian ethnicity groupings. Firstly, an Asian patient with an Asian casemanager is on average 6% less likely to receive an emergency caesarean relative to an Asian patient (or casemanager) with a *European* casemanager (or patient)[5]. Secondly, an Asian patient with an Asian casemanager is on average 3.6% less likely to receive an emergency caesarean relative to an Asian patient (casemanager) with an *Indian* casemanager (patient). After controlling for a number of factors, Indian and European ethnic concordance are not statistically more or less likely to undergo an emergency caesarean procedure, relative to outcomes for ethnically discordant casemanager-patient pairs.

In this paper, a statistically significant interaction term for Asian ethnic concordance only reveals an association, because I am unable to randomly assign patients to casemanagers. That is, some women are able to select their casemanager prior to entering hospital.

[5]Because I use an interaction model, the base category for interpreting ethnic concordance interaction terms includes two types of ethnic discordance. So, Asian patient-Asian casemanager is compared to both Asian patient-European casemanager and Asian casemanager-European patient.

I argue that the unobserved health condition of women in ethnically concordant relative to discordant groups *does not* explain the significant negative relationship for my Asian ethnic group. Because women have entered labour, and have not been selected for a planned caesarean ex-ante, they must have been considered of a sufficient health status to undergo a natural birth. That is, women at risk from labour complications are selected for a *planned* caesarean procedure prior to entering labour. Planned caesarean procedures are identified in my data and are accordingly excluded from my sample so that I only consider women who enter a labour process and either undergo an emergency caesarean or have a vaginal birth. This assumption allows me to argue a significant relationship between ethnic concordance and emergency caesareans is not due to differences in the clinical condition of mothers who are ethnically concordant relative to discordant, because all women who enter labour would have been considered of a sufficient health status to undergo a natural birth.

Because I assume women are of a sufficient health status in ethnically concordant and discordant pairs, and that 'patient ethnic group' and 'supply-side' explanations are controlled for in my regression analysis, a statistically significant Asian-Asian interaction term suggests emergency caesarean outcomes for Asian patients are not always being determined by clinical appropriateness and/or ethnicity-specific preferences. This therefore suggests that health care policy could potentially reduce the high rate of caesarean procedures for Asian women. For example, by encouraging more Asian midwives, or educating casemanagers about the effect of patient ethnicity on their treatment decisions. This is especially relevant to the New Zealand context, because Asian women have the highest rate of caesarean procedures nationally, at 27.3% birth events involving an Asian mother (Ministry of Health, 2012)[6].

[6]Maori and Pacific have the lowest caesarean rate at 17% and 19.12% respectively. An 'Other' ethnic group (including mostly European) have a caesarean rate of 27.1% (Ministry of Health, 2012, pg.26).

I suggest three plausible mechanisms for why Asian women are less likely to receive an emergency caesarean when treated by an Asian casemanager relative to a casemanager from a different ethnic group. The first explanation is a reduction in the clinical need for an emergency caesarean. In this explanation, Asian patients with an Asian casemanager have greater ease, communication, and trust which lowers a women's distress in labour and thereby reduces the clinical need for an emergency caesarean.

A second explanation is a reduction in clinical uncertainty in ethnically concordant pairs. Because emergency caesarean procedures are undertaken in an uncertain environment, an Asian casemanager with an Asian patient may be less likely to over-treat an Asian patient, because they can more accurately interpret their level of distress. The effect of doctor-patient ethnic concordance on a patient's communication of their health condition has been modelled by Balsa and McGuire (2001, 2003). In their model, a patient sends a noisier signal of their condition in ethnically discordant pairings. Poor information from communication could result in over or under-treatment.

A third explanation is that Asian women who prefer an Asian casemanager are also more likely to prefer not to have a caesarean procedure, compared to Asian women that do not have a preference for an Asian caregiver. I argue that this is not a cultural preference because it does not apply to *all* Asian women.

The next section discusses the empirical literature on ethnic variation in caesarean procedures. Later sections outline my data, empirical strategy, results and conclusion.

2.2 Background and previous literature

2.2.1 Ethnic differences in caesarean rates

A number of papers have used regression analysis to investigate explanations for ethnic-based variation in caesarean procedures. These papers test for the significance of patient ethnic group variables, after controlling for a number of explanations. In all of the following reviewed papers, researchers find that ethnic-based variation in caesarean procedures persists even after controlling for various explanations, including; a mother's clinical diagnoses, socioeconomic factors, age, hospital characteristics and so on. In the following, I review a selection of these papers from different countries.

Harris et al. (2007) used regression analysis to investigate the sources for differences in caesarean section rates, both elective and acute, across Maori and non-Maori women in New Zealand. The authors find that both emergency and elective caeserean procedure rates are lower for Maori (13%) than non-Maori (21%), using data from 1997-2001. Maori women were also more likely to have an acute than elective caesarean section procedure. Harris et al. find that differences in caesarean rates remain after controlling for age, deprivation and a number of clinical factors. Clinical factors controlled for include; fetal malpresentation, gestation at delivery (either pre-term delivery at <37 weeks gestation or post-term delivery at >42 weeks gestation), multiple births, maternal hypertension, maternal diabetes, and antepartum haemorrhage. The authors suggest that 'non-clinical factors may be contributing to ethnic differences in CS [Caesarean Section] in New Zealand. While deprivation contributes to this difference it does not fully explain it' (Harris et al., 2007, pg. 125)[7].

[7]Rumball-Smith (2009, p.68) reviews empirical studies of ethnicity-based health care disparities in New Zealand. She concludes that, despite shortcomings in some empirical study designs, 'there is robust evidence for the existence of healthcare disparities for Maori' and that these disparities are most evident in 'obstetric intervention and the incidence of potentially avoidable adverse events'. Lower resource use for Maori and Pacific Peoples has also been indicated in other studies. Tukuitonga and Bindman (2002) compare hospitalization rates for heart failure and cardiac intervention DRGs for Maori, non-Maori, men and women. hospitalization rates for heart failure in Maori are four or more times higher than for non-Maori patients under 65 years (this difference decreases as patients

In the United States, several studies have found higher rates of caesarean procedures in African-American compared to white Americans (Menacker et al., 2010; Braveman et al., 1995; Getahun et al., 2009). Braveman et al. (1995) use data from California, after controlling for various confounding factors, African-Americans were 24% more likely to undergo a caesarean procedure than white Americans. Their control variables include clinical factors, patient characteristics, insurance status, hospital ownership, and hospital teaching status, volume, and region. The authors conclude; 'the findings warrant further research that more directly examines how nonclinical characteristics of patients - particularly race/ethnicity - may inappropriately influence clinical decision making.' (Braveman et al., 1995, pg.630).

A more recent study in the United States used 540,953 births in California during the 1991 to 2008 period to study ethnic variation in caesarean rates (Getahun et al., 2009). The authors also found a higher caesarean section rate for African-American women, after controlling for a number of factors, and conclude that there is 'wide variability in rate of indications for primary cesarean section by race/ethnicity' (Getahun et al., 2009, pg. 423).

In Australia, von Katterfeld et al. (2011) compared caesarean rates for immigrant populations. They find higher caesarean rates for women from Sub-Saharan Africa, Southeast Asia, and Southern and Central Asia compared to Australian-born women. The authors adjusted for differences in perinatal complications and maternal characteristics. The authors conclude that the difference in caesarean rates 'indicates that the increased risk among women from these regions cannot be entirely accounted for by sociodemographic risk factors or identified complications' (von Katterfeld et al., 2011, pg.153).

---

age increases). By contrast, Maori rates for cardiac procedure DRGs are much lower; typically between a third to a half of the rate for non-Maori. This suggests to Tukuitonga and Bindman that Maori are much less likely to receive cardiac interventions, despite having very high rates of hospitalization for heart failure. There are no controls for a patient's medical condition in their study, and it could be the case that Maori are more likely to have less severe heart conditions.

In the United Kingdom, Ibison (2005) find African, West Indian, Bangladeshi, Indian and Pakistani women were at a higher risk of Cesarean procedure compared to 'Caucasian' women. They use data from 1988 to 1997 on 27,667 women in an East London Area.Their control variables include; maternal age, antenatal class attendance, late-booking, year of baby's birth, hospital, intra-uterine growth retardation, baby's birthweight, foetal sex, induction, and augmentation of birth.

In Norway, Vangen et al. (2000) study caesarean rates among immigrant mothers. Of 553,491 live births during the period 1986-1995, 17,891 births were to immigrant mothers. Caesarean section rates ranged from 10.1% for women from in Vietnam, to 25.8% for Filipino-born women. The authors conclude in their study that 'unknown factors come into play in the decision to receive an emergency caesarean' (Vangen et al., 2000, pg.32).

I am unable to find a study that estimates a relationship between provider-patient ethnic concordance and a patient's likelihood of undergoing a caesarean procedure. This is likely due to the lack of widely available data on the ethnic group of casemanagers.

2.3 Data

2.3.1 Data source

Data on birth events was obtained from a large public hospital in New Zealand. This data source has been discussed in Chapter 1 (see Section 1.3).

2.3.2 Patient population

Patients are firstly selected for my study if an inpatient admission has an Australian Refined - Diagnosis Related Group (AR-DRG) corresponding to a birth event in hospital during the 2000 to 2011 time period. Different versions of AR-DRG codes

are used because codes change during the 2000 to 2011 period. All versions of AR-DRGs used to select the patient population are in Appendix 2.D. AR-DRG codes are assigned after an inpatient event. Software is used to assign a DRG based on information from an admission, including; the patient's diagnosis codes, length of stay, and procedures undertaken. AR-DRG codes are used to compensate hospitals for an expected bundle of services for an inpatient event, and are therefore expected to be a reliable indication of a birth event[8].

Secondly, I select only women for the patient population who undergo an emergency caesarean or vaginal birth. Patients that have a planned caesearean are excluded from my study. Patients that enter hospital for a planned caesarean procedure are not at risk of an emergency caesarean during labour. Emergency and planned caesarean procedures are identified with a ICD-10-AM procedure code for *any* of the procedures performed on a patient during their hospital stay. ICD-10-AM codes for emergency caesareans are; 1652001 (Emergency classical caesarean section) and 1652003 (Emergency lower segment caesarean section). ICD-10-AM codes for planned caesareans are; 1652000 (Elective classical caesarean section) and 1652002 (Elective lower segment caesarean section).

2.3.3 Assignment of casemanagers to patients

For each inpatient event, the name of the primary casemanager is recorded in hospital data. The primary casemanager is defined by inpatient data collections as the main casemanager responsible for a patient's hospital stay. I assume that this casemanager plays an important role in the decision for a patient to undergo an emergency caesarean procedure during labour. The decision to have an emergency caesarean would likely involve more than one medical opinion, and an obstetrician would have

[8]This is in contrast to using diagnosis codes to identify a birth event e.g. a 'Singleton birth'. This is because ICD-10 diagnostic codes are subject to a greater risk of error than AR-DRG coding, because AR-DRG coding is the basis of hospital reimbursement.

to perform the procedure. Given a primary casemanager is the person identified as responsible for an inpatient's care, it is a reasonable assumption that there is a relationship between the primary casemanager and patient during labour which allows the casemanager to advise treatment decisions[9].

The assignment of casemanagers to women entering hospital to give birth depends on whether a woman has selected a casemanager prior to entering hospital, or is assigned a casemanager at hospital entry. Women who give birth under the public health care system are mostly assigned a casemanager that is available at the time of hospital entry. It is also possible to pay for private health care and give birth in a public hospital. In this case, a women may have a prearranged casemanager, such as an obstetrician, who will enter hospital to provide medical treatment when a women goes into labour. Importantly, it is possible for some women to select their casemanager, and this may involve selection depending on the ethnicity of the casemanager. The implication of patient self-selection when interpreting my results is discussed in my conclusion. It is not possible in the data to identify women who enter hospital with a pre-arranged casemanager, or are assigned a casemanager from staff at hospital entry.

It is important that emergency caesareans can occur across all casemanagers. That is, that emergency caesareans do not occur only in a small number of 'specialised' casemanagers. I investigate the concentration of caesarean section procedures across casemanagers in the next section (2.4.1). A high concentration of caesarean procedures in a small number of casemanagers could suggest that women change casemanagers if they undergo a caesarean procedure. This kind of endogenous sorting of patients to 'specialised' casemanagers - as a result of a caesarean procedure - would be problematic for estimating the relationship between ethnic concordance and caesarean procedures. I do not find evidence to suggest endogenous sorting of patients

[9]For the rest of the paper I use the term casemanager when referring to primary casemanagers.

to casemanagers depending on caesarean procedures in my hospital data (see Section 2.4.1).

2.3.4 Casemanager ethnic group information

The ethnicity of the casemanager is not provided in inpatient data records. Information on employee ethnic group was obtained from the hospital's HR department. Chapter 1 describes the linking procedure on a casemanager's name between inpatient and HR data to obtain a casemanager's ethnic group.

Casemanagers are associated with one of six main ethnicity groupings; European, Maori, Pacific peoples, Asian, African and Indian. Other ethnicity groupings for casemanagers include Middle Eastern and Latin American. These ethnicity groupings are included in descriptive tables in an 'other' ethnic category, because they comprise a small proportion of casemanagers.

Not all casemanagers complete their ethnic information on an HR form. Summary statistics for casemanagers matched with HR data, for which we have ethnic information on, are in Table 2.1. There are 102 individual casemanagers that have information on ethnicity. Of 102 casemanagers, 75% are European, 10% are Asian and 9% are Indian. These are the three largest ethnicity groupings for casemanagers. The average number of patients per casemanager is 246, and the median number is 64.50[10].

Because not all casemanagers complete their ethnic information, and HR information is only recorded electronically from 2005 onwards, a large number of patients (51,830) in the 2000-2011 period have a casemanager with no ethnic information.

To obtain further information on a casemanager's ethnic group, an ethnic group was manually associated based on the ethnic origin of a casemanager's lastname. This

[10]Birth events are a suitable patient population to study treatment decisions. The population of casemanagers for other clinical conditions, such as Acute Myocardial Infarction (AMI) events, is small and lacks ethnic diversity.

provided a second dataset that is obtained by firstly matching with HR data, and remaining unmatched casemanagers were manually associated an ethnic group. This process of manually associating an ethnic group has been discussed in Chapter 1, Appendix 1.F.

Summary statistics for casemanagers matched with HR data and manually matched are in Table 2.2. There are 389 casemanagers identified with an ethnic group. Of 389 casemanagers, 81% are identified as European, 8.2% are identified as Asian and 4.8% are identified as Indian ethnic origin. There is a high proportion of European casemanagers in Table 2.2, compared to HR matched data (Table 2.1). A possible reason for this is incorrectly associating a European ethnic group. This is suspected to be more likely for Maori and (South) African casemanagers than Asian or Indian ethnicity groupings. For example, there is a greater history of intermarriage and anglicising lastnames for Maori. South African names also tend to share a Dutch heritage which would be associated with a 'European' ethnic group. This raises some concerns about manually associating the ethnic origins of lastnames[11].

The number of patients over the 2000-2011 period using both HR and manual matching that I am unable to match is 4,287 (compared to 51,830 for HR data). All of these patients are not matched because they had a 'consultant' for a primary casemanager. That is, the casemanager is not personally identified.

I use the HR and manually matched dataset for further statistical analysis. The advantage of this dataset is a significantly larger number of inpatient observations, which allows me to more reliably estimate the relationship between ethnic concordance and emergency caesarean procedures.

[11]I am able to check the robustness of my results by using only casemanagers matched to HR data.

Table 2.1: Casemanager characteristics for HR matched

	N	Mean	Sd	Min	Max	P50
European CM	102	0.75	0.44	0.00	1.00	1.00
Asian CM	102	0.10	0.30	0.00	1.00	0.00
Indian CM	102	0.09	0.29	0.00	1.00	0.00
N patients	102	246.59	417.86	1.00	2178.00	64.50
N patients missing CM	51830					

Notes: N=102 is number of casemanagers (CM). Summary statistics for casemanagers that are merged with HR data.

Table 2.2: Casemanager characteristics for manual and HR population

	N	Mean	Sd	Min	Max	P50
European CM	389	0.82	0.38	0.00	1.00	1.00
Asian CM	389	0.08	0.28	0.00	1.00	0.00
Indian CM	389	0.05	0.22	0.00	1.00	0.00
N patients	389	186.88	336.87	1.00	2178.00	47.00
N patients missing CM	4287					

Notes: N=389 is number of casemanagers (CM). Summary statistics for casemanagers that are merged with HR data and manually associated an ethnic group.

2.4 Descriptive statistics

This section summarises features of my dataset that are relevant to the estimation of a relationship between ethnic concordance and emergency caesarean procedures.

2.4.1 Casemanager summary statistics

There is a large number of casemanagers with a small patient history. Fifty three out of 389 casemanagers have one patient during the 2000-2011 period, 26 casemanagers have 2-3 patients, 35 casemanagers have 4-10 patients, 52 casemanagers have 11-30 patients, 34 casemanagers have 31-50 patients, and 45 casemanagers have 51-100 patients. There are 144 casemanagers with greater than 101 patients. Therefore, half of all casemanagers treat 46 or fewer patients.

The variation in caseload size across casemanagers could be problematic if emergency caesarean procedures tend to occur only in casemanagers that treat a large number of cases. For example, if a specialist obstetrician at the hospital treats the

majority of patients that receive an emergency caesarean, this could indicate that patients are transferred from their original casemanager to a obstetrician when they undergo an emergency caesarean. To investigate if this pattern exists in my data, Table 2.3 presents the percentage of emergency caesarean outcomes for patients by the size of their casemanager's total caseload (i.e. number of patients they treat in 2000-2011 period). The first row shows that 26% of patients with a casemanager that treats only one patient have an emergency caesarean. Going down the column for the percentage of emergency caesarean procedures, there is no clear pattern between the number of patients a casemanager treats and emergency caesarean outcomes. There may however be some concern about the high proportion of emergency caesarean procedures (30%) for the small number (4-5) of casemanagers that treat 1500 plus patients. In robustness checks, casemanagers with large and small caseloads are excluded from the sample so that the casemanager population is more homogeneous.

In general, I observe an even spread in the percentage of emergency caesarean outcomes across casemanagers with high and low caseloads. This is a good indication that there is no sorting of patients who undergo an emergency caesarean to a small number of casemanagers. This is relevant to my empirical strategy, because sorting of patients to casemanagers, depending on their likelihood of an emergency caesarean outcome, would be endogenous.

2.4.2 Number of patient and casemanager ethnicity pairs

Casemanagers and patients are associated with one of 6 main ethnicity groupings; African, Asian, European, Indian, Maori, and Pacific peoples. An 'Other' category captures all ethnicity groupings that do not belong to the aforementioned ethnicity groupings. Ethnic concordance occurs when a patient and casemanager have the same ethnic group.

Table 2.3: Percentage of emergency caesarean outcomes by casemanger's caseload size

Number of cm's patients	**Emergency caesarean** %
1 patient (n=53)	26
2-3 patients (n=67)	16
4-10 patients (n=228)	9
11-30 patients (n=987)	10
31-50 patients (n=1,353)	14
51-100 patients (n=3,194)	8
101-200 patients (n=8,279)	12
201-500 patients (n=15,366)	15
501-1000 patients (n=19,893)	21
1001-1500 patients (n=9,184)	21
1501-2001 patients (n=9,771)	32
2001+ patients (n=4,320)	30
Total (n=72,695)	20

Notes: Column (2) is the percentage of emergency caesarean outcomes by casemanager's caseload size. Where n is the number of patients with a casemanager in caseload size group. For example, 53 patients have a casemanager with 1 patient.

The number (percentage) of observations for each casemanager-patient ethnic pair is in Table 2.4 (Table 2.5). The largest ethnic group for *patients* is European, followed by Asian, Pacific, Maori and Indian. For *casemanagers*, 75.8% of all birth admissions have a European casemanager, followed by 9.7% and 9.4% of birth admissions with an Asian or Indian casemanager respectively.

Figures in the center of Table 2.4 and 2.5 provide the number and percentage of observations for each casemanager-patient ethnic pair. European patients have a European casemanager 86.2% of the time. Asian patients have an Asian casemanager 37.4% of the time. Indian patients have an Indian casemanager 12.5% of the time.

Maori and Pacific patients are rarely treated by a casemanager of the same ethnicity. For Pacific patients, 2.5% are treated by a Pacific casemanager. For Maori patients, 3.3% are treated by a Maori casemanager. Therefore, Maori and Pacific ethnic concordance is unlikely to be reliably estimated with such low numbers of observations. There are 306 (183) observations of Pacific (Maori) ethnic concordance.

Table 2.4: Number of observations: casemanager-patient ethnic pairs

	Casemanager ethnicity							
Patient ethnicity	African	Asian	European	Indian	Maori	Other	Pacific	Total
	No.	No.	No.	No.	No.	No.	No.	No.
African	21	70	592	108	7	16	5	819
Asian	213	4712	6241	1151	136	59	72	12584
European	525	869	30021	2439	595	140	221	34810
Indian	99	369	3664	615	69	42	58	4916
Maori	71	242	4074	765	183	36	94	5465
Middle Eastern	21	72	681	133	10	11	20	948
Other	28	40	711	103	9	2	4	897
Pacific	201	681	9090	1542	324	110	306	12254
Total	1179	7055	55074	6856	1333	416	780	72693

Notes: Tabulates the number of observations for each casemanager-patient ethnic pair .

Table 2.5: Percentage of observations: casemanager-patient ethnic pairs

	Casemanager ethnicity							
Patient ethnicity	African	Asian	European	Indian	Maori	Other	Pacific	Total
	%	%	%	%	%	%	%	%
African	2.6	8.5	72.3	13.2	0.9	2.0	0.6	100.0
Asian	1.7	37.4	49.6	9.1	1.1	0.5	0.6	100.0
European	1.5	2.5	86.2	7.0	1.7	0.4	0.6	100.0
Indian	2.0	7.5	74.5	12.5	1.4	0.9	1.2	100.0
Maori	1.3	4.4	74.5	14.0	3.3	0.7	1.7	100.0
Middle Eastern	2.2	7.6	71.8	14.0	1.1	1.2	2.1	100.0
Other	3.1	4.5	79.3	11.5	1.0	0.2	0.4	100.0
Pacific	1.6	5.6	74.2	12.6	2.6	0.9	2.5	100.0
Total	1.6	9.7	75.8	9.4	1.8	0.6	1.1	100.0

Notes: Tabulates the percentage of observations for each casemanager-patient ethnic pair.

2.4.3 Emergency caesarean outcomes by patient and casemanager ethnicity pairs

The proportion (total number) of emergency caesarean procedures for each casemanager-patient ethnic pair is in Table 2.6 (Table 2.7). Out of a total 72,693 birth events, 14,414 patients have an emergency caesarean (excluding planned caesarean procedures.) Maori and Pacific patients have the lowest percentage of emergency caesareans, with 17% and 16% observations respectively.

In my data, Asian casemanagers have a low percentage of emergency caesarean procedures compared to national statistics, (at 19% in my data compared to 27.3%

Table 2.6: Proportion of emergency caesarean outcomes: casemanager-patient ethnic pairs

	Casemanager ethnicity							
Patient ethnicity	African	Asian	European	Indian	Maori	Other	Pacific	Total
African (n=819)	0.24	0.31	0.22	0.30	0.43	0.00	0.40	0.23
Asian (n=12,584)	0.41	0.07	0.24	0.36	0.15	0.49	0.18	0.19
European (n=34,810)	0.35	0.28	0.20	0.32	0.17	0.42	0.19	0.21
Indian (n=4,916)	0.34	0.22	0.23	0.33	0.12	0.43	0.16	0.24
Maori (n=5,465)	0.28	0.18	0.16	0.21	0.09	0.28	0.10	0.17
Middle Eastern (n=948)	0.19	0.17	0.16	0.26	0.00	0.18	0.10	0.17
Other (n=897)	0.46	0.25	0.24	0.40	0.11	0.00	0.00	0.26
Pacific (n=12,254)	0.20	0.17	0.15	0.21	0.11	0.22	0.09	0.16
Total (n=72,693)	0.33	0.12	0.19	0.29	0.14	0.34	0.13	0.20

Notes: Tabulates the proportion of emergency cesarean outcomes for each casemanager-patient ethnic pair.

Table 2.7: Number of emergency caesarean outcomes: casemanager-patient ethnic pairs

	Casemanager ethnicity							
Patient ethnicity	African	Asian	European	Indian	Maori	Other	Pacific	Total
African (n=819)	5	22	128	32	3	0	2	192
Asian (n=12,584)	88	317	1520	413	21	29	13	2401
European (n=34,810)	185	240	5924	791	102	59	41	7342
Indian (n=4,916)	34	82	825	205	8	18	9	1181
Maori (n=5,465)	20	44	672	160	17	10	9	932
Middle Eastern (n=948)	4	12	110	35	0	2	2	165
Other (n=897)	13	10	170	41	1	0	0	235
Pacific (n=12,254)	41	117	1388	331	37	24	28	1966
Total (n=72,693)	390	844	10737	2008	189	142	104	14414

Notes: Tabulates the number of emergency cesarean outcomes for each casemanager-patient ethnic pair.

nationally (Ministry of Health, 2012)). But inspection of Table 2.6 indicates a lower overall percentage can be explained by Asian ethnic concordance. Foreshadowing my main result, I observe the lowest proportion of emergency caesarean outcomes for Asian ethnic concordance (7%), compared to when an Asian patient is treated by a European (24%) or Indian (36%) casemanager.

2.5 Empirical strategy

In this section I firstly select my sample of birth events. I then discuss my econometric model, and how this model distinguishes between three explanations for ethnic-based variation in caesarean procedures. I lastly discuss alternative estimation approaches, and specify the explanatory variables in my model.

2.5.1 Study sample

In this paper I am interested in estimating the impact of ethnic concordance on emergency caesarean procedures for each ethnic group separately. This is because caesarean procedure rates vary by patient ethnic group, it is therefore not clear ethnic concordance would have the same effect (increase or decrease) on emergency caesarean procedures for *all* ethnicity groupings.

To investigate ethnic concordance for each patient ethnic group, I need sufficient numbers of ethnic concordance to estimate the relationship between ethnic concordance and caesarean procedures. The previous section showed that there are low numbers of observations for Maori and Pacific ethnic concordance. It is unlikely I can reliably estimate the effect of ethnic concordance for this ethnic group in my data.

In addition to sufficient numbers of ethnic concordance, there needs to be sufficient variation in casemanager ethnicity groupings treating a patient ethnic group. For example, there needs to be sufficient numbers of non-Asian casemanagers for Asian patients. Ideally, patient ethnicity groupings should be treated by casemanagers from at least two different ethnicity groupings, so that an 'ethnic concordance effect' constitutes something more than a difference between two ethnicity groupings.

In this paper I compare how European, Asian and Indian patients are treated across European, Asian and Indian casemanagers. The number of observations for each casemanager-patient ethnic pair is in Table 2.8. The number of observations for

Table 2.8: Number of observations: casemanager by patient ethnic group

	Casemanager ethnicity			
Patient ethnicity	Asian	European	Indian	Total
Asian	4,712	6,241	1,151	12,104
European	869	3,0021	2439	33,329
Indian	369	3664	615	4,648
Total	5,950	39,926	4,205	50,081

Notes: Tabulates the number of emergency cesarean outcomes for casemanager-patient ethnic pair in study sample.

Table 2.9: Proportion of emergency caesarean outcome: casemanager by patient ethnic group

	Casemanager ethnicity			
Patient ethnicity	Asian	European	Indian	Total
Asian	0.07	0.24	0.36	0.19
European	0.28	0.20	0.32	0.21
Indian	0.22	0.23	0.33	0.24
Total	0.11	0.21	0.34	0.21

Notes: Tabulates the proportion of emergency cesarean outcomes for casemanager-patient ethnic pair in study sample.

each ethnic pair are: Asian patient-Asian casemanager is 4,712, Asian patient-Indian casemanager is 1,151, and Asian patient-European casemanager is 6,241. The number of observations for Indian patient-Indian casemanager is 615, Indian patient-Asian casemanager is 369, and Indian patient-European casemanager is 3,664. The number of observations for European patient-European casemanager is 30,021, European patient-Asian casemanager is 869, European patient-Indian casemanager is 2,439. The number of observations for each of the above ethnic pairs is considered sufficient to identify the relationship between ethnic concordance and emergency caesarean.

I therefore exclude patients and casemanagers from all other ethnicity groupings in my study. Table 2.9 is a reduced table of the proportion of emergency caesarean procedures for each type of casemanager-patient ethnic pair.

2.5.2 Econometric model

The baseline econometric model uses an interaction term to estimate the relationship between ethnic concordance and emergency caesarean outcomes. This interaction

term is composed of two dummy variables for patient and casemanager ethnic group. For example, Asian ethnic concordance is estimated with an interaction term that is equal to one when both the Asian patient and Asian casemanager variables are equal to one, and is zero otherwise. I estimate the following interaction models:

$$\begin{aligned} emergces_{ij} =& \alpha + \beta X_{ij} + \beta_1 asian\_pat_i + \beta_2 asian\_cm_j + \\ & \beta_3 asian\_pa * asian\_cm_{ij} + \beta_4 prop\_ces_{ij} + \epsilon_{ij} \end{aligned} \tag{2.1}$$

$$\begin{aligned} emergces_{ij} =& \alpha + \beta X_{ij} + \beta_1 european\_pat_i + \beta_2 european\_cm_j + \\ & \beta_3 european\_pat * european\_cm_{ij} + \beta_4 prop\_ces_{ij} + \epsilon_{ij} \end{aligned} \tag{2.2}$$

$$\begin{aligned} emergces_{ij} =& \alpha + \beta X_{ij} + \beta_1 indian\_pat_i + \beta_2 indian\_cm_j + \\ & \beta_3 indian\_pat * indian\_cm_{ij} + \beta_4 prop\_ces_{ij} + \epsilon_{ij} \end{aligned} \tag{2.3}$$

$emergces_{ij}$ is a binary outcome that is equal to one if a patient received an emergency caesarean procedure. X_{ij} is a vector of inpatient event control variables, detailed in Section 2.5.3.

I estimate the effect of ethnic concordance relative to different ethnicity groupings in the base category. For example, equation 2.1 is estimated with a base category of (1) European patients and casemanagers and (2) Indian patients and casemanagers. I therefore compare two ethnicity groupings to estimate the relationship between ethnic concordance and emergency caesarean procedures.

I am particularly interested in a result where ethnic concordance has the same sign (positive or negative) relative to *different* (base category) ethnicity groupings. This would suggest that ethnic concordance for a patient ethnic group is consistently

increasing or decreasing the likelihood of an emergency caesarean procedure. For example, if Asian ethnic concordance had a significant positive or negative relationship to caesareans - relative to both Indian and European ethnicity groupings - this suggests Asian ethnic concordance has an expected impact on caesarean procedures. It would therefore suggest (the effect of) Asian ethnic concordance is independent from which ethnic group Asian concordance is compared to[12].

Patient and casemanager ethnic group dummy variables, and their interaction, are interpreted relative to the ethnic group in the base category. For example, when European patients and casemanagers are in the base category, β_1 in equation 2.1 estimates the effect of an Asian patient-European casemanager relative to a European patient-European casemanager on the likelihood of an emergency caesarean. That is, β_1 estimates the difference between Asian and European patients, when the casemanager is European. The coefficient on the $asian\_pa * asian\_cm_{ij}$ term, β_3, would therefore compare caesarean procedures for Asian ethnic concordance relative to Asian patient-European casemanager *and* Asian casemanager-European patient pairs.

In the baseline econometric models, $prop\_ces_{ij}$ is the proportion of emergency caesarean procedures that casemanager j has advised previously up to patient i, but not including patient $i's$ emergency caesarean outcome. This is the sum of emergency caesarean procedures in $j's$ previous case history, divided by the number of patients j has treated previously. $prop\_ces_{ij}$ therefore measures the propensity of j to advise a caesarean procedure. This measure increases if casemanagers have treated a small number of patients, but have had a high number of caesarean procedures. This measure decreases if a casemanagers had a high number of cases and a low number

[12]I also estimate a model with all patients in the sample population, so that the base category for equations 2.1 to 2.3 includes all other ethnicity groupings. For example, Asian ethnic concordance is estimated relative to all non-Asian ethnicity groupings. This is a more crude method to estimate ethnic concordance effects, because it ignores heterogeneity in caesarean procedures across all different casemanager-patient ethnic pairs in the base category.

of caesarean procedures, thereby reflecting a casemanager that had a low propensity to advise a caesarean procedure. This measure uses all information up to patient i to avoid an endogenous relationship between $i's$ outcome and $prop\_ces_{ij}$.

$prop\_ces_{ij}$ is intended to control for 'supply-side' explanations for ethnic-based variation in caesarean procedures. Because I am unable to control for previous emergency caesarean outcomes for casemanagers that have only one observation in the data, I exclude these casemanagers from the study sample. Robustness checks in my result section probe my method to control for 'supply-side' explanations.

A second model includes dummy variables for Asian *and* Indian patients, casemanagers and interaction terms. European patients and casemanagers are the base category:

$$\begin{aligned} emergces_{ij} =& \alpha + \beta X_{ij} + \beta_1 asian\_pat_i + \beta_2 asian\_cm_j + \beta_3 asian\_pat * asian\_cm_{ij} \\ &+ \beta_4 asian\_pat * indian\_cm_{ij} + \beta_5 indian\_pat_i + \beta_6 indian\_cm_j \\ &+ \beta_7 indian\_pat * indian\_cm_{ij} + \beta_8 indian\_pat * asian\_cm_{ij} + \epsilon_{ij} \end{aligned}$$

This model has a larger number of inpatient observations, because all European, Indian and Asian patients and casemanagers are included in the sample population. Coefficient estimates are therefore more efficient. But the interpretation of ethnic dummy variables is only relative to the European ethnic group.

A logit model estimates the relationship between explanatory variables and the binary outcome. Linear models allow interpretation when other variables are held constant, in non-linear models interpretation is undertaken after first setting a reference point for the values of other variables. I estimate marginal effects of variables when other covariates are held at their mean values. Standard errors are clustered on the casemanager. I also estimate a linear regression in robustness checks.

Interpreting results and theoretical framework

My theoretical framework aims to distinguish between three explanations for ethnic-based variation in caesarean rates. These are; 'patient ethnic group', 'supply-side', and 'casemanager composition' explanations.

The 'supply-side' explanation refers to the possibility that some ethnicity groupings are more likely to be treated by casemanagers that have a high or low propensity to advise an emergency caesarean procedure. A casemanager's propensity applies equally to all patients, and is therefore independent of a patient's ethnic group. To control for a 'supply-side' explanation, I include the proportion of a casemanager's previous patients up to i that recieved an emergency caesarean; $prop\_ces_{ij}$. This is intended to control for $j's$ propensity to advise an emergency caesarean. A casemanager may be more or less likely to advise a caesarean procedure for a number of different reasons, including; experience, training and skill level. My measure for the propensity to advise a caesarean procedure aims to control for *all* of these possible explanations by directly measuring the likelihood of a caesarean procedure, based on case history. A dummy variable for casemanager ethnicity also controls for differences across casemanager ethnicity groupings in their average emergency caesarean procedures[13].

The 'patient ethnic group' explanation refers to biological characteristics, health behaviours and cultural preferences relating to both patient ethnic group and caesarean procedures. For example, the clinical need for a caesarean procedure is influenced by a mother's weight, height, and build. Distributions of these characteristics vary across patient ethnicity groupings. Ethnic-based variation in caesarean procedures could therefore be due to differences in characteristics relating to the clinical need or preferences of caesarean procedures across ethnicity groupings. These biolog-

[13]I discuss the possibility of using a casemanager fixed-effect to control for my 'supply-side explanation in the next section.

ical characteristics, health behaviours and cultural preferences are relevant decision factors for *all* doctors treating patients from an ethnic group. I therefore do not expect 'patient ethnic group' decision factors (biological characteristics, health behaviours and cultural preferences) to depend on the ethnic group of the casemanager.

In baseline models (equations 2.1 to 2.3) a statistically significant interaction term for ethnic concordance would therefore suggest that 'patient ethnic group' and 'supply-side' explanations cannot explain *all* ethnic-based variation in caesarean procedures. This is because I would not expect a patient ethnic group to have a different likelihood of a caesarean procedure depending on the casemanager, because all casemanagers should react equally to the same 'patient ethnic group' decision factors.

Accordingly, 'casemanager composition' can explain some ethnic-based variation in caesarean procedures. This could be useful for policy, for example, some demographic groups could be encouraged to train as a casemanager, or assigning casemanagers to patients could take into account ethnicity.

Lastly, I assume the unobserved health of women in ethnically concordant vs. discordant pairs does not systematically differ. An emergency caesarean carries considerable risk to health for both mother and child, and is therefore only undertaken when the risk from complications due to a natural birth are sufficiently high. Because women in the study sample were not selected for a planned caesarean procedure exante, they must have been considered of a sufficient health status to undergo a natural birth. This assumption allows me to argue a significant relationship between ethnic concordance and emergency caesareans is not due to differences in the clinical condition of mothers who are ethnically concordant relative to discordant, because all women in the sample population would have been considered of a sufficient health status to undergo a natural birth.

Method discussion

I discuss two alternative approaches to estimating the relationship between ethnic concordance and emergency caesarean procedures.

The first alternative is to use interaction terms that separate each type of casemanager-patient ethnic pair into dummy variables. In this model, Asian patient-European casemanager observations would be in the *base category* and a separate variable for Asian patient-Asian casemanager would estimate the effect of ethnic concordance relative to Asian patient-European casemanager:

$$\begin{aligned} emergces_{ij} =& \alpha + \beta X_{ij} + \beta_1 european\_pat\_european\_cm_i \\ &+ \beta_2 asian\_cm\_european\_pat_j + \beta_3 asian\_pat\_asian\_cm_{ij} \\ &+ \beta_4 prop\_ces_j + \epsilon_{ij} \end{aligned} \tag{2.4}$$

A similar pairwise comparison would compare the effect of Asian patient-Asian casemanager relative to the base category of an Asian patient-Indian casemanager.

This model however has shortcomings compared to the full interaction model, in equations 2.1 to 2.3. This is because the coefficient on the Asian patient-Asian casemanager interaction term in equation 2.1 is estimated after controlling for both Asian patient-European casemanager *and* Asian casemanager-European patient. In equation 2.4, the base category includes only Asian patients-European casemanagers. Because I want to control for differences across casemanager ethnic group in their propensity to advise an emergency caesarean, the ethnic interaction term in equation 2.1 estimates ethnic concordance relative to both Asian patient-European casemanager *and* Asian casemanager-European patient.

To explain this point further, consider the comparison of Indian and European ethnicity groupings. Using Table 2.9, an Indian patient-European casemanager is

much less likely to receive an emergency caesarean compared to Indian patient-Indian casemanager. However, this comparison does not consider the effect of Indian casemanagers-European patients, for which there is a high likelihood of an emergency caesarean outcome. I may then conclude that Indian patient-Indian casemanager has a positive effect on emergency caesareans, relative to Indian patient-European casemanager. In a full interaction model Indian patient-Indian casemanager may not be significant, because it would account for the (positive) effect of Indian casemanager-European patient.

The second alternative estimation approach is to use a fixed-effect on casemanagers. A fixed-effect would be especially beneficial to controlling for supply-side explanations, because only variation within casemanagers is used to estimate ethnic concordance variables. In my data, half of all casemanagers have 46 or fewer patients. The low number of patients for some casemanagers may present problems for a casemanager fixed-effect analysis. This is because the number of observations within casemanager groups is small. Small group sizes reduce variation in the data to estimate the effect of ethnic concordance on emergency caesarean outcomes.

Fixed-effect analysis is also problematic because the ethnic diversity in casemanagers that treat a large number of patients is small, which makes it difficult to estimate the effect of specific types of ethnic concordance (e.g. Asian-Asian) on an emergency caesarean outcome with fixed-effects.

Table 2.3 shows that there does not appear to be a systematic pattern between caseload size and emergency caesarean outcomes in the data. This indicates that emergency caesarean procedures are not occurring only in a handful of casemanagers that treat a large number of patients in the hospital. This feature of the data mitigates some of the concern about heterogeneity across casemanagers in their capacity to advise a caesarean procedure.

Chapter 1 used a fixed-effect method on casemanagers to estimate the relationship between ethnic concordance and diagnostic resources. I estimated an *average* effect of demographic concordance across different gender and ethnicity groupings. There is greater ethnic variation in caesarean rates than diagnostic resources, and the focus of this paper is on reasons for variation in caesarean procedures for different ethnicity groupings. For example, Asian patients have the highest rate of caesarean procedures nationally, and the causes of a high likelihood are of special interest to policy makers concerned with healthcare budgets. In addition, a fixed-effect method was more appropriate in Chapter 1 because I considered all hospital admissions, and there was therefore greater variation in medical conditions in the study sample. Casemanager fixed-effects controlled to an extent for this variation in medical condition. I am considering only birth events so there is less heterogeneity in the medical condition of patients in the sample.

2.5.3 Other explanatory variables

Summary statistics for explanatory variables are in Table 2.14, Appendix 2.A. Variables for patient characteristics are; age, deprivation scale, dummy variable for whether a patient was transferred from another hospital, if a patient's District Health Board (public provider of health care) is that of the hospital.

Dummy variables for timing of a patient admission include; the hour of admission to hospital, day of week of admission to hospital, and year of admittance[14].

There are a number of different clinical variables used in economic studies for caesarean outcomes. Currie and Gruber (1997) use four clinical variables; previous caesarean, fetal distress, breech presentation and maternal distress. Currie and Gruber used a 'not elsewhere classified' (n.e.c) maternal distress indicator, this excludes other forms of maternal distress that do not fall into a 'n.e.c.' category. Grant (2009)

[14]I do not include explanatory variables for casemanager's characteristics because they are not available for casemanagers not matched with HR data.

includes variables for mothers who have diagnosis codes for the following conditions; diabetes, hypertension, preeclampsia, placenta previa, herpes, and a 'not elsewhere classified' maternal distress category.

In my data, maternal distress could refer to a number of different clinical codes. For example, prolonged pregnancy (O48), long labour (O639), maternal distress during labour and delivery (O750), shock during or following labour and delivery (O751) and so on. Caution is therefore needed when using clinical codes for a health condition which may also be coded in another clinical category. For example, 'obstructed labour due to breech presentation' (O641) and 'maternal care for breech presentation' (O321) are similar health conditions but come under different clinical codes. An advantage of my study is that I use a broad number of ICD codes that refer to the same or similar health conditions. This prevents not indicating a women who is in distress, but may be coded in a different clinical category.

I used the full list of up to 60 diagnosis codes associated with a hospital inpatient event to identify if a patient had a clinical condition. I use dummy variables to control for; large fetus, singleton birth, premature gestation, very premature gestation, antepartum hemorrhage, diabetes, preeclampsia, infection, liver complications, circulatory system complications, respiratory system complications, polyhydramnios, ogliohydramnios, maternal hypotension, abnormal presentation, and fetal stress. All ICD-10-AM codes to identify if a patient had one of these clinical conditions are in Appendix 2.E.

2.6 Results

In this section I interpret results for each ethnic group separately. This allows me to highlight the statistical significance of Asian ethnic concordance on emergency

caesareans, relative to the statistical insignificance of European and Indian ethnic concordance.

Main results are in Tables 2.10, 2.11, and 2.12. Table 2.10 has results for Asian patients and casemanagers (equation 2.1), with a base category of European ethnic group (Column (1)), or Indian ethnic group (Column (2)). Table 2.11 has results for Indian patients and casemanagers (equation 2.3) with a base category of European ethnic group (Column (1)), or Asian ethnic group (Column (2)). Table 2.12 has results for European patients and casemanagers (equation 2.2), with a base category of Asian (Column (1)) and Indian (Column (2)) ethnicity groupings. All other regression results are in Appendix 2.B, including a table of all coefficient values for the emergency caesarean regression (Table 2.15).

Table 2.10: Emergency caesarean for Asian patients

	(1) Euro base	(2) Indian base
Asian pat	0.015*	−0.001
	(2.32)	(−0.09)
Asian doc	0.023	−0.016
	(1.93)	(−1.20)
Asian concordance	−0.062***	−0.036*
	(−4.75)	(−2.21)
Observations	41388	6733

Notes: Outcome: emergency caesarean procedure. Model: logit regression, Column (1) has European base category, Column (2) has Indian base category. Marginal effects for coefficients. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

2.6.1 Asian ethnic group

I interpret the coefficients for Asian patients, casemanagers, and their interaction term in the following.

Table 2.11: Emergency cesarean for Indian patients

	(1) Euro base	(2) Asian base
Indian pat	0.010 (1.24)	0.037** (2.88)
Indian doc	−0.006 (−0.26)	0.053** (3.15)
Indian concordance	0.029 (1.89)	−0.036* (−2.21)
Observations	36451	6733

Notes: Outcome: emergency caesarean procedure. Model: logit regression, Column (1) has European base category, Column (2) has Asian base category. Marginal effects for coefficients. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 2.12: Emergency cesarean for European patients

	(1) Indian base	(2) Asian base
European pat	−0.040** (−3.13)	0.047*** (4.49)
European doc	−0.024 (−1.32)	0.039** (2.96)
European concordance	0.029 (1.89)	−0.062*** (−4.75)
Observations	36451	41388

Notes: Outcome: emergency caesarean procedure. Model: logit regression, Column (1) has Indian base category, Column (2) has Asian base category. Marginal effects for coefficients. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Asian patient

An Asian patient is statistically significantly more likely to have an emergency caesarean compared to a European patient, when the casemanager is European (Column (1), Table 2.10, see also full interaction model in Table 2.15). The marginal effect at mean values of the covariates is estimated at .03, or a 3% increase in the likelihood of having an emergency caesarean[15].

[15]For the odds ratio interpretation, the impact of an Asian patient recieving an emergency caesarean is 1.2 times greater than the odds of a European patient receiving a caesarean, when the casemanager is European (Table 2.15).

An Asian patient is not statistically significantly related to the likelihood of an emergency caesarean procedure relative to an Indian patient, when the casemanager is Indian, (Column (2), Table 2.10).

Asian casemanager

An Asian casemanager is statistically insignificantly related to an emergency caesarean procedure compared to a European casemanager, when the patient is European (Table 2.10).

An Asian casemanager is statistically significantly less likely to advise an emergency caesarean compared to an Indian casemanager, when the patient is Indian. The estimated marginal effect is a 4.1% reduction when other variables are held at their mean values (Table 2.10).

Note that dummy variables for casemanager ethnic group will reflect differences across casemanager ethnicity groupings in their average employee skill level. For example, Indian casemanagers may occupy more specialised roles on average and this may explain a higher likelihood for Indian casemanagers to advise a caesarean procedure relative to other casemanager ethnicity groupings. I am therefore not able to interpret casemanager ethnic group variables as a cultural propensity to advise a caesarean procedure.

Asian casemanager and patient

The interaction term for an Asian patient-Asian casemanager is negative and statistically significant at the 1% level relative to *both* the European and Indian ethnic base category. An Asian patient with an Asian casemanager is on average 6% less likely to receive an emergency caesarean relative to an Asian patient (or casemanager) with a *European* casemanager (or patient). An Asian patient with an Asian casemanager

is on average 3.6% less likely to receive an emergency caesarean relative to an Asian patient (or casemanager) with an *Indian* casemanager (or patient).

Table 2.16 in Appendix 2.B presents estimates for the coefficient on Asian ethnic concordance when *all* other ethnicity groupings are included in the base category. The marginal effect is estimated at a 7.6% reduction in the likelihood of an emergency caesarean for Asian ethnic concordance. This is an estimate of the impact of Asian ethnic concordance compared to all other ethnicity groupings in the patient population.

Robustness tests for Asian ethnic concordance results

I investigate the robustness of my statistically significant negative relationship between Asian ethnic concordance and emergency caesarean outcomes. Regression tables for robustness tests are in Appendix 2.C.

I firstly investigate the robustness of Asian ethnic concordance to different methods to control for 'supply-side' explanations. I firstly exclude $prop\_ces_j$, a control variable for the proportion of emergency caesarean procedures (Column (3), Table 2.17 and Table 2.19). The coefficient for Asian ethnic concordance increases and is estimated at a 15.9% (9%) reduction in the likelihood of an emergency caesarean for European (Indian) base category. This suggests that 'supply-side' explanations are important to control for in my cross-sectional analysis of casemanagers.

I investigate different methods to compute $prop\_ces_j$. Firstly, I use a casemanager's entire patient record to compute $prop\_ces_j$. I use the total number of caesarean procedures divided by the total number of patients for each casemanager (Column (2), Table 2.17 and Table 2.19). This variable is constant within casemanagers, as opposed to changing within casemanagers, based on their history up to patient i. This model also includes casemanagers that treat only one patient, thereby increasing the

number of observations. The coefficient on Asian ethnic concordance increases to 9% (5.5%) with European (Indian) ethnic base category.

I exclude some casemanagers from the sample population to generate a more homogeneous sample of casemanagers. I exclude casemanagers (and their patients) that have less than 20 patients, and casemanagers that have greater than 1500 patients (Column (1), Table 2.17 and Table 2.19). I restrict the sample of casemanagers to avoid comparing caesarean procedures across one-off casemanagers, and casemanagers with extensive experience, thereby reducing heterogeneity in casemanager roles and experience. Asian ethnic concordance is negative and statistically significant for the European base category, but is not statistically significant for the Indian base category. A reason for statistical insignificance of Asian ethnic concordance could be a reduction in sample size (to N=5,722) when casemanagers are excluded from the Indian comparison group.

I also estimate ethnic concordance on the population of patients matched only with HR data (Column (1), Table 2.18 and 2.20). HR data uses only casemanagers that self-identify with an ethnic group, and therefore avoids incorrect association of casemanagers to an ethnic group during manual matching. With HR data, the estimate on Asian ethnic concordance *increases* to 8.1% for the European ethnic comparison group. Asian ethnic concordance is negative but not statistically significant for the Indian ethnic group. The sample size is small (N=4,421) and this likely contributes to the lack of statistical significance.

To test for robustness of Asian ethnic concordance to model specification, I estimate a linear regression model (Column (2), Table 2.18 and 2.20). The coefficient for Asian ethnic concordance is negative but not statistically significant for both European and Indian base categories. When a linear model is estimated with *all* non-Asian ethnicity groupings in the base category, the Asian interaction term is statistically *significant*. Therefore, the lack of statistical significance, when using a linear model,

is likely due to estimating a binary outcome with a linear model and small sample size. Linear models are a poor fit for binary outcomes compared to logit models and this is especially the case when the sample size is small. Asian ethnic concordance is statistically significant in the linear model when the sample size is increased.

My last robustness test estimates the Asian interaction model with *African, Pacific and Maori* ethnicity groupings in the base category (Table 2.21). These results come with a caveat that observation sizes for ethnic pairings (e.g. Asian patient-Maori casemanager) are small. The sign on the Asian interaction term is negative in all models, but lacks statistical significance in the model with Pacific peoples in the base category.

Overall, results from robustness tests support the strength of a negative relationship between Asian ethnic concordance and emergency caesarean outcomes. However, the statistical strength of the relationship is weakened when sample sizes are reduced.

The magnitude of the coefficient changes depending on the method used to control for 'supply-side' factors. A casemanager fixed-effect would remove variation on average caesarean procedures across casemanagers from my data. My small sample size prevents me from estimating a fixed-effect model, because ethnic diversity of casemanagers with large numbers of patients is small.

2.6.2 Indian ethnic group

Indian patient

Indian patients are not statistically significantly related to an emergency caesarean procedure, relative to a European patient, when the casemanager is European (Column (2), Table 2.11).

Indian patients are statistically significantly more likely to receive an emergency caesarean compared to an Asian patient, when the casemanager is Asian. The

marginal effect is estimated at a 3.7% reduction in the likelihood of an emergency caesarean.

Indian casemanager

The coefficient for Indian casemanager is statistically insignificant compared to a European casemanager, when the patient is European (Column (1), Table 2.11).

An Indian casemanager is more likely to advise an emergency caesarean than an Asian casemanager, when the patient is Asian (Column (2), Table 2.11). The marginal effect is estimated at a 5.3% increase in the likelihood of receiving an emergency caesarean.

Indian patient and casemanager

An Indian patient with an Indian casemanager does not have a statistically significantly different likelihood of receiving an emergency caesarean, relative to Indian patient-European casemanagers and Indian casemanager-European patient pairs.

An Indian casemanager-Indian patient is significantly less likely to receive an emergency caesarean, relative to Indian casemanager-Asian patient and Indian patient-Asian casemanager.

The cause of the negative sign on Indian ethnic concordance can be determined by considering each type of casemanager-patient ethnic pair on emergency caesarean outcomes. Firstly, Asian patient-Indian casemanager have a slightly lower likelihood of receiving an emergency caesarean *relative* to an Indian-Indian pair (variable 'Asian pat' in Column (2), Table 2.10). Secondly, Asian casemanager-Indian patient have a lower likelihood of receiving an emergency caesarean *relative* to an Indian-Indian pair (variable 'Asian CM' in Table 2.10). Thirdly, an Indian patient-Asian casemanager is more likely to have a caesarean *relative* to an Asian-Asian pair (Column (2),

Table 2.13: Patient-casemanager combination and relationship with emergency caesarean

Patient-casemanager pair	Comparison group	Effect
Asian patient:		
Asian pat-Ind. CM	Ind. pat-Ind. CM	insig.
Asian pat-Euro. CM	Euro. pat-Euro. CM	+
Asian pat-Asian CM	Asian pat-Ind. CM and Ind. pat-Asian CM	-
Asian pat-Asian CM	Asian pat-Euro. CM and Euro. pat-Asian CM	-
European patient:		
Euro. pat-Ind. CM	Ind. pat-Ind. cm	-
Euro. pat-Asian CM	Asian pat-Asian cm	+
Euro. pat-Euro. CM	Euro. pat-Indian CM and Indian pat-Euro. CM	insig.
Euro. pat-Euro. CM	Euro. pat-Asian CM and Asian pat-Euro. CM	-
Indian patient:		
Ind. pat-Euro. CM	Euro. pat-Euro. CM	insig.
Ind. pat-Asian CM	Asian pat-Asian CM	+
Ind. pat-Ind. CM	Euro. pat-Ind. CM and Ind. pat-Euro. CM	insig.
Ind. pat-Ind. CM	Ind. pat-Asian CM and Asian pat-Ind. CM	-

Notes: The effect column is the coefficient sign (for emergency caesarean outcomes) for each type of patient-casemanager ethnic pair comparison.

Table 2.11). Fourthly, an Indian casemanager-Asian patient is more likely to advise a caesarean *relative* to an Asian-Asian pair (Column (2), Table 2.11).

The significant negative relationship for Indian patient-Indian casemanager pairs is therefore due to Asian-Indian pairs having a high likelihood of emergency caesarean outcomes relative to Asian-Asian pairs (Column (2), Table 2.11), when we consider only Indian and Asian ethnicity groupings. This is because an Asian casemanager-Indian patient relative to Indian-Indian pairs (variable 'Asian doc' in Table 2.10) is *negative*, suggesting that there is no significant negative relationship between Indian ethnic concordance and emergency caesarean procedures. Indian casemanagers are therefore much more likely to advise a caesarean procedure to Asian patients - and by comparison to this - Indian-Indian pairs are negatively related to emergency caesarean outcomes.

2.6.3 European ethnic group

European patient

European patients are statistically significantly less likely to receive an emergency caesarean, compared to Indian patients, when the casemanager is Indian. The estimated marginal effect is a 4% reduction in the likelihood of an emergency caeasarean.

European patients are statistically significantly more likely to receive an emergency caesarean, compared to Asian patients, when the casemanager is Asian. The estimated marginal effect is a 4.7% increase in the likelihood of having an emergency caesarean.

European casemanager

A European casemanager relative to an Indian casemanager, when the patient is Indian, is insignificantly related to emergency caesarean procedures.

A European casemanager is 3.9% more likely to advise an emergency caesarean compared to an Asian casemanager, when the patient is Asian. This effect is statistically significant at the 5% level.

European patient and casemanager

European ethnic concordance is statistically insignificant relative to the Indian base category.

European ethnic concordance is significantly less likely to have an emergency caesarean compared to a European-Asian pair. The reason for a negative relationship is the same as for Indian-Indian pairs, relative to Indian-Asian pairs (discussed in section 2.6.2). Specifically, European and Indian casemanagers are much more likely to advise a caesarean procedure to Asian patients than an Asian casemanager would to an Asian patient. As a result, European-Asian pairs have a higher likelihood

of emergency caesareans, relative to European-European pairs, thus resulting in a negative coefficient for European-European pairs.

2.7 Conclusion

This paper estimated the relationship between ethnic concordance and the likelihood of having an emergency caesarean procedure when a women is in labour. I use the three largest casemanager ethnicity groupings in data collected from a large hospital. Ethnicity groupings are; Asian, Indian and European.

My main result is that Asian patients with an Asian casemanager are associated with a statistically significantly lower likelihood of an emergency caesarean procedure relative to treatment by both European and Indian ethnicity groupings. This result is also clearly visible in descriptive statistics for emergency caesarean outcomes by casemanager-patient ethnic pair (Table 2.6). Asian ethnic concordance is estimated at an average 6.2% reduction in the likelihood of an emergency caesarean, after controlling for a large number of factors including the propensity of the casemanager to advise a caesarean procedure.

In my data, Indian and European ethnic concordance is not statistically significantly related to the likelihood of having an emergency caesarean outcome.

The negative relationship between Asian ethnic concordance on emergency caesarean procedures is robust to estimation with; (1) different measures of a casemanager's propensity to advise caesareans, (2) a restricted sample of casemanagers, (3) a linear model and (4) patients only matched with HR data. However, reducing sample sizes reduces the statistical strength of the relationship.

In addition to the novelty of estimating a relationship between ethnic concordance and emergency caesareans, my result can also be interpreted within a theoretical framework for explaining ethnic-based variation in caesarean procedures. I

distinguish between 'patient ethnic group', 'supply-side' and 'casemanager composition' explanations. I find evidence for a 'casemanager composition' explanation, because Asian patients have a significantly lower likelihood of emergency caesarean procedures when treated by an Asian casemanager relative to a casemanager from a different ethnic group.

I assume women with an ethnically concordant vs. discordant casemanager-patient pair do not have systematic differences in their health condition. This is because women have not been selected for a planned caesarean, and have therefore been considered of a sufficient health status to give birth naturally. Pregnant women are under the care of a midwife, who will advise a planned caesarean procedure if they think a women is unable to successfully undergo a vaginal birth.

I therefore suggest three plausible mechanisms for why Asian women are less likely to receive an emergency caesarean when ethnically concordant. The first explanation is a reduction in clinical need for an emergency caesarean. In this explanation, Asian patients with an Asian casemanager have greater ease, communication, and trust which lowers a women's distress in labour and thereby reduces the clinical need for an emergency caesarean.

A second explanation is a reduction in clinical uncertainty in ethnically concordant pairs. Because emergency caesarean procedures are undertaken in an uncertain environment, an Asian casemanager with an Asian patient may be less likely to over-treat an Asian patient, because they can more accurately interpret their level of distress. The effect of doctor-patient ethnic concordance on a patient's communication of their health condition has been modelled by Balsa and McGuire (2001, 2003). In their model, a patient sends a noisier signal of their condition in ethnically discordant pairings which may then result in over or under-treatment.

Some women are able to select a midwife (especially if they pay privately for a specialist obstetrician), and it is plausible that some women have a preference for

a casemanager of the same ethnic group. There is a large number of ethnically concordant pairs for Asian patients (37.4%), which could indicate a preference some Asian women have for a casemanager of the same ethnic group. A third explanation is therefore that Asian women who prefer an Asian casemanager are also more likely to prefer not to have a caesarean procedure, compared to Asian women who do not have a preference for an Asian casemanager.

A shortcoming of my study is a small sample size. This limits my ability to implement a casemanager fixed-effect model. This could further control for the 'supply-side' explanation, by removing the average propensity to advise a caesarean procedure across casemanagers.

Another limitation is low numbers of ethnic pairs for Maori and Pacific peoples. Maori and Pacific peoples have low rates of caesarean procedures in New Zealand, and are therefore of interest when studying the causes of ethnic-based variation in caesarean rates. Other limitations of my study include a lack of data on some aspects of a mother's health condition, particularly whether a mother received a caesarean procedure previously.

Future work could look at the health outcomes of women and children born with a casemanager of the same ethnic group. My sample size is small, and mortality is a rare occurrence. This limits regression analysis with mortality outcomes. In addition, an Australian study by von Katterfeld et al. (2011) finds higher caesarean rates for Southeast, Southern and Central Asian women compared to Australian-born women. It could be interesting to see if a similar ethnic concordance relationship, for Asian women, also occurs in Australia.

The 2010 maternity figures for New Zealand show that Asian women have the highest rate of caesarean procedures, at 27.3%. Maori and Pacific have the lowest, at 17% and 19.12% respectively. An 'Other' ethnic group (including mostly European) have a caesarean rate of 27.1% (Ministry of Health, 2012, pg.26). Overall, almost a

quarter (23.6%) of women gave birth by caesarean section. Of these women, just over a half (54.9%) had an emergency caesarean section procedure (Ministry of Health, 2012). Because I find a statistically significant negative relationship between emergency caesarean procedures and Asian ethnic concordance, this suggests there might be efficiency gains from ethnically matching Asian women[16]. Efficiency gains would arise because fewer medical resources are used to provide appropriate care[17].

The results from this study support further research into how the doctor-patient relationship affects medical treatment decision-making. This research could encourage doctors to consider how a patient's ethnic group may be entering in to their decision-making, and to accordingly adjust for any inappropriate effects a patient's ethnic group might have on medical treatment decision-making.

[16]The cost per caesarean section procedure is approximately NZD 6000 (Ministry of Health, 2012, pg.26)

[17]I assume care is appropriate because casemanagers are not biased (i.e. prejudiced) against patients from their own ethnic group, for example by providing inadequate care or care against a patient's preferences.

2.A Explanatory variables

Table 2.14: Obstetric population variables

	Mean	*Sd*	*Min*	*Max*
Outcome:				
Emergency caes.	0.21	0.40	0.00	1.00
Patient variables:				
Age	31.41	5.08	14.00	50.00
Deprivation scale	5.18	2.81	0.00	10.00
After 5pm entry	0.29	0.45	0.00	1.00
After 12pm entry	0.17	0.38	0.00	1.00
Overnight admiss.	0.20	0.40	0.00	1.00
Transfer	0.02	0.14	0.00	1.00
AED entry	0.00	0.03	0.00	1.00
Acute admiss.	0.25	0.43	0.00	1.00
Prev admiss with 60days	0.14	0.35	0.00	1.00
Auckland DHB	0.68	0.47	0.00	1.00
Diagnosis count	4.52	2.68	2.00	31.00
Proportion ces proc.	0.09	0.07	0.00	1.00
Admit Monday	0.15	0.35	0.00	1.00
Admit Tuesday	0.14	0.35	0.00	1.00
Admit Wednesday	0.15	0.35	0.00	1.00
Admit Thursday	0.15	0.36	0.00	1.00
Admit Friday	0.15	0.36	0.00	1.00
Admit Saturday	0.13	0.34	0.00	1.00
Admit 2001	0.09	0.29	0.00	1.00
Admit 2002	0.09	0.29	0.00	1.00
Admit 2003	0.09	0.29	0.00	1.00
Admit 2004	0.09	0.29	0.00	1.00
Admit 2005	0.08	0.27	0.00	1.00
Admit 2006	0.08	0.27	0.00	1.00
Admit 2007	0.08	0.27	0.00	1.00
Admit 2008	0.08	0.28	0.00	1.00
Admit 2009	0.08	0.27	0.00	1.00
Admit 2010	0.09	0.28	0.00	1.00
Admit 2011	0.06	0.23	0.00	1.00
Clinical variables:				
Single birth	0.98	0.14	0.00	1.00
Premature gestation	0.03	0.17	0.00	1.00
Very premature gestation	0.01	0.10	0.00	1.00
Antepartum hemorrhage	0.01	0.10	0.00	1.00
Diabetes	0.05	0.21	0.00	1.00
Preeclampsia	0.03	0.16	0.00	1.00
Infection	0.01	0.07	0.00	1.00
Liver complications	0.00	0.05	0.00	1.00
Polyhydramnios	0.01	0.08	0.00	1.00
Oligohydramnios	0.05	0.21	0.00	1.00
Maternal hypotension	0.00	0.06	0.00	1.00
Abnormal presentation	0.06	0.24	0.00	1.00
Fetal stress	0.26	0.44	0.00	1.00

Circ. dis. complications	0.01	0.08	0.00	1.00
Resp. dis. complications	0.01	0.10	0.00	1.00
Large fetus	0.00	0.04	0.00	1.00
N	50081			

Notes: Summary of explanatory variables for the study sample.

2.B Results

Table 2.15: Emergency caesareans for European, Asian and Indian ethnicity groupings

	(1) MEM	(2) AME
Asian pat	0.033***	0.032***
	(4.52)	(4.53)
Indian pat	0.029***	0.028***
	(3.66)	(3.64)
Asian doc	0.038	0.036
	(1.62)	(1.62)
Indian doc	0.043	0.041
	(1.50)	(1.51)
Asian concordance	−0.108***	−0.103***
	(−5.48)	(−5.54)
Asian pat+Indian doc	0.022	0.021
	(1.26)	(1.25)
Asian doc+Indian pat	−0.038	−0.036
	(−1.87)	(−1.87)
Indian concordance	0.022	0.021
	(1.27)	(1.27)
Age	0.003***	0.003***
	(6.50)	(6.97)
Deprivation scale	0.001	0.001
	(0.70)	(0.69)
After 5pm entry	−0.003	−0.003
	(−0.60)	(−0.60)
After 12pm entry	0.011*	0.011*
	(2.16)	(2.16)
Overnight admiss.	−0.012*	−0.011*
	(−2.03)	(−2.04)
Transfer	0.027*	0.026*
	(2.26)	(2.24)
AED entry	−0.106	−0.101

	(−1.96)	(−1.96)
Acute admission	0.010	0.009
	(1.46)	(1.45)
Prev admiss with 60days	−0.013*	−0.012*
	(−2.01)	(−2.03)
Auckland DHB	0.016**	0.015**
	(3.04)	(3.03)
Diagnosis count	0.035***	0.034***
	(18.33)	(24.60)
Prop. cesareans	1.288***	1.229***
	(10.82)	(11.31)
Admit Monday	0.017**	0.016**
	(2.69)	(2.72)
Admit Tuesday	0.014	0.013
	(1.82)	(1.82)
Admit Wednesday	0.009	0.009
	(1.28)	(1.29)
Admit Thursday	0.012	0.012
	(1.57)	(1.57)
Admit Friday	0.014*	0.013*
	(2.14)	(2.16)
Admit Saturday	0.009	0.009
	(1.49)	(1.49)
Admit 2001	0.020*	0.019*
	(2.39)	(2.40)
Admit 2002	0.035***	0.034***
	(3.58)	(3.62)
Admit 2003	0.040***	0.038***
	(3.51)	(3.55)
Admit 2004	0.032**	0.030**
	(2.82)	(2.86)
Admit 2005	0.046***	0.044***
	(3.48)	(3.54)
Admit 2006	0.038*	0.036*
	(1.98)	(2.00)
Admit 2007	0.024	0.023
	(1.57)	(1.58)
Admit 2008	0.014	0.013
	(0.82)	(0.83)
Admit 2009	0.061**	0.058***
	(3.26)	(3.35)
Admit 2010	0.057***	0.055***

	(3.43)	(3.54)
Admit 2011	0.057***	0.054***
	(3.41)	(3.49)
Single birth	0.084***	0.080***
	(5.21)	(5.09)
Premature gestation	0.013	0.013
	(1.24)	(1.23)
Very premature gestation	−0.322***	−0.307***
	(−10.39)	(−11.50)
Antepartum hemorrhage	−0.020	−0.019
	(−1.25)	(−1.25)
Diabetes	−0.081***	−0.077***
	(−6.85)	(−7.28)
Preeclampsia	0.066***	0.063***
	(6.22)	(6.40)
Infection	−0.065**	−0.062**
	(−2.66)	(−2.66)
Liver complications	−0.025	−0.024
	(−0.93)	(−0.93)
Polyhydramnios	0.010	0.009
	(0.53)	(0.53)
Oligohydramnios	−0.004	−0.004
	(−0.56)	(−0.56)
Maternal hypotension	−0.035	−0.033
	(−1.21)	(−1.21)
Abnormal presentation	0.214***	0.204***
	(18.91)	(26.27)
Fetal stress	0.082***	0.078***
	(13.13)	(13.52)
Circ. dis. complications	−0.106***	−0.101***
	(−5.12)	(−5.07)
Resp. dis. complications	−0.045**	−0.043**
	(−2.68)	(−2.69)
Large fetus	0.327***	0.312***
	(9.15)	(9.41)
Observations	49798	49798

Notes: Outcome: emergency caesarean. Model: logit regression, Column (1) has marginal effects at mean values of other variables, Column (2) has the average marginal effects at different values of variables. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 2.16: Emergency caesarean outcome

	(1) European	(2) Asian	(3) Indian
European pat	0.011*		
	(2.41)		
European cm	−0.009		
	(−1.04)		
European concordance	0.003		
	(0.38)		
Asian pat		0.032***	
		(6.37)	
Asian cm		0.021**	
		(2.62)	
Asian concordance		−0.076***	
		(−6.47)	
Indian pat			0.028***
			(5.06)
Indian cm			−0.005
			(−0.31)
Indian concordance			0.024*
			(2.11)
Observations	76167	76167	76167

Notes: Outcome: emergency caesarean. Sample: study sample. Model: logit regression, Column (1) has variables for European ethnic group, base category is all other ethnicity groupings, Column (2) has Asian ethnic variables, Column (3) has Indian ethnic variables. Marginal effects calculated at mean values. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

2.C Robustness tests for Asian ethnic concordance

Table 2.17: Robustness for Asian ethnicity variables with European base

	(1) restrict cm	(2) fixed prop	(3) no prop
Asian pat	0.018***	0.030***	0.028***
	(3.39)	(4.32)	(3.92)
Asian doc	0.028**	0.032	0.029
	(2.93)	(1.51)	(1.33)
Asian concordance	−0.059***	−0.091***	−0.159***
	(−5.44)	(−5.00)	(−7.78)
Observations	33436	41609	41609

Notes: Outcome: emergency caesarean. Sample: Asian and European patient and casemanager ethnicity groupings. Model: logit regression, Column (1) has restricted casemanagers, Column (2) has fixed proportion of caesarean procedure for explanatory variable, Column (3) has no control for proportion of caesarean procedures. Marginal effects calculated at mean values. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 2.18: Robustness for Asian ethnicity variables with European base

	(1) HR only	(2) linear	(3) all linear
Asian pat	0.022*	0.016*	0.037***
	(2.15)	(2.23)	(5.97)
Asian doc	0.032*	0.017	0.010
	(2.09)	(1.10)	(1.23)
Asian concordance	−0.086***	−0.031	−0.043***
	(−3.49)	(−1.85)	(−4.10)
Observations	18908	41388	76167

Notes: Outcome: emergency caesarean. Sample: Asian and European patient and casemanager ethnicity groupings. Model: Column (1) is the HR only population, Column (2) is linear regression model, Column (3) is linear regression with all patient and casemanager ethnicity groupings in the base category. Marginal effects calculated at mean values for logit model. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 2.19: Robustness for Asian ethnicity variables with Indian base

	(1) restrict cm	(2) fixed prop	(3) no prop
Asian pat	−0.007 (−0.47)	−0.002 (−0.15)	0.010 (0.67)
Asian doc	−0.012 (−1.01)	−0.041** (−2.81)	−0.045** (−3.25)
Asian concordance	−0.019 (−1.17)	−0.055** (−3.20)	−0.090*** (−5.67)
Observations	5722	6758	6758

Notes: Outcome: emergency caesarean. Sample: Asian and Indian patient and casemanager ethnicity groupings. Model: logit regression, Column (1) has restricted casemanagers, Column (2) has fixed proportion of caesarean procedure for explanatory variable, Column (3) has no control for proportion of caesarean procedures. Marginal effects calculated at mean values. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 2.20: Robustness for Asian ethnicity variables with Indian base

	(1) HR only	(2) linear
Asian pat	−0.004 (−0.28)	0.001 (0.03)
Asian doc	−0.014 (−0.64)	−0.053** (−2.78)
Asian concordance	−0.032 (−1.30)	−0.012 (−0.56)
Observations R^2	4421	6744 0.333

Notes: Outcome: emergency caesarean. Sample: Asian and Indian patient and casemanager ethnicity groupings. Model: Column (1) is the HR only population, Column (2) is a linear regression model. Marginal effects calculated at mean values for logit model. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 2.21: Emergency caesarean outcome with Asian ethnicity variables

	(1) African base	(2) Pacific base	(3) Maori base
Asian pat	0.054***	0.015	0.030*
	(7.90)	(1.82)	(2.25)
Asian doc	0.039***	0.011	0.012
	(3.52)	(0.76)	(0.62)
Asian concordance	−0.094***	−0.025	−0.040*
	(−10.23)	(−1.58)	(−2.20)
Observations	4923	5677	5183

Notes: Outcome: emergency caesarean. Model: logit regression, Column (1) has African ethnic group as base category, Column (2) has Pacific ethnic group as base category, Column (3) has Maori ethnic group as base category. Marginal effects calculated at mean values. Standard errors clustered on casemanager. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

2.D Obstetric AR-DRG codes

Table 2.22: Obstetric Australian Refined Diagnosis Related Groups (AR-DRG) Version 3 to 5

DRG Code	DRG name
Version 3.1:	
670	Caesarean delivery, w/o complicating diagnosis
671	Caesarean delivery w moderate complicating diagnosis
672	Caesaeran delivery w severe complicating diagnosis
674	Vaginal delivery no complicating diagnosis
675	Vaginal delivery w moderate complicating diagnosis
676	Vaginal delivery w severe complicating diagnosis
677	Vaginal delivery w complicating OR procedures
687	Caesarean delivery w mult complicating diagnoses, at least one severe
688	Vaginal delivery w mult complicating diagnoses, at least one severe
Version 4.1:	
O01A	Caesarean delivery w mult complicating diagnoses, at least one severe
O01B	Caesarean delivery w severe complicating diagnosis
O01C	Caesarean delivery w moderate complicating diagnosis
O01D	Caesarean delivery w/o complicating diagnosis
O02Z	Vaginal delivery w complicating OR procedure
O60A	Vaginal delivery w multiple complicating diagnosis,at least one severe
O60B	Vaginal delivery w severe complicating diagnosis
O60C	Vaginal delivery w moderate complicating diagnosis
O60D	Vaginal delivery w/o complicating diagnosis
Version 5:	
O01A	Caesarean Delivery W Catastrophic CC
O01B	Caesarean Delivery W Severe CC
O01C	Caesarean Delivery W/O Catastrophic or Severe CC
O02A	Vaginal Delivery W O.R. Procedure W Catastrophic or Severe CC
O02B	Vaginal Delivery W O.R. Procedure W/O Catastrophic or Severe CC
O60A	Vaginal Delivery W Catastrophic or Severe CC
O60B	Vaginal Delivery W/O Catastrophic or Severe CC

Notes: AR-DRG codes used to select the study sample.

2.E Obstetric comorbidity codes

Table 2.23: Obstetric comorbidities

ICD-10-AM Code	Diagnosis description
Birth weight:	
O662	Large fetus
Very Premature Gestation:	
O090	Duration of pregnancy <5 completed weeks
O091	Duration of pregnancy 5-13 completed weeks
O092	Duration of pregnancy 14-19 completed weeks
O093	Duration of pregnancy 20-25 completed weeks
Premature Gestation:	
O094	Duration of pregnancy 26-33 completed weeks
O095	Duration of pregnancy 34-36 completed weeks
Singleton pregnancy:	
Z370	Singleton birth
Antepartum haemorrhage:	
O460	Antepartum haemorrhage with coagulation defect
O468	Postcoital antepartum haemorrhage
O469	Antepartum haemorrhage, unspecified
Abnormal presentation:	
O320	Maternal care for unstable lie
O321	Maternal care for breech presentation
O322	Maternal care for transverse and oblique lie
O323	Maternal care for face, brow and chin presentation
O324	Maternal care for high head at term
O325	Maternal care for multiple gest w malpresentation
O326	Maternal care for compound presentation
O328	Maternal care for other malpresentation
O329	Maternal care for malpresent of fetus, unspecified
Pre-eclampsia:	
O140	Moderate pre-eclampsia
O141	Severe pre-eclampsia
O149	Pre-eclampsia, unspecified
Diabetes:	
O244	Diabetes arising in pregnancy
O2441	Diabetes arising in pregnancy, non-insulin
O2442	Diabetes arising in pregnancy, insulin
O2443	Diabetes arising during pregnancy, therapy
O2444	Diabetes arising during pregnancy, diet
O2449	Diabetes arising during pregnancy, unspecified
O240	Pre-existing diabetes, Type 1, in pregnancy
O241	Pre-existing diabetes, Type 2, in pregnancy
O2411	Pre-exist DM Type 2 in pregnoninslntrt
O2412	Pre-exist DM Type 2 in preg insulin trt
O2413	Pre-existing diabetes, Type 2, in pregnancy, the

O2414	Pre-existing diabetes, Type 2, in pregnancy, other
O2419	Pre-existing diabetes, Type 2, in pregnancy, unsp
O242	Pre-existing malnutrition-related diabetes mellitus
O2421	Pre-exist DM oth spec preg
O2422	Pre-exist DM oth spec in preg
O2423	Pre-existing diabetes mellitus, other specified type
O243	Pre-existing diabetes mellitus, unspecified
O2431	Pre-exist DM unspec
Infection:	
O230	Infections of kidney in pregnancy
O231	Infections of bladder in pregnancy
O232	Infections of urethra in pregnancy
O233	Infections of other parts of urinary tract in preg
O234	Unspecified infection of urinary tract in preg
O235	Infections of the genital tract in preg
O239	Other genitourinary tract infection in preg
Maternal hypotension:	
O265	Maternal hypotension
Liver complications:	
O265	Diseases liver complicating preg
Polyhydramnios:	
O40	Polyhydramnios
Oligohydramnios:	
O410	Oligohydramnios
Circulation comp.:	
O994	Diseases circulatory system complicating preg
Respiratory comp.:	
O995	Diseases respiratory system complicating preg
Fetal stress:	
O680	Labour complicated by fetal heart rate anomaly
O681	Labour complicated by meconium in amniotic fluid
O682	Labour complicated by fetal heart rate anomaly with meconium in amniotic fluid
O683	Labour complicated by biochemical fetal stress
O688	Labour complicated by other evidence of fetal stress
O689	Labour complicated by fetal stress, unspecified
O998	Other specified diseases complicating preg

Notes: ICD-10-AM codes used to code comorbid health conditions.

Chapter 3

Ward-level Nursing Hours, Patient Demand and Patient Health Outcomes

3.1 Introduction

The relationship between nursing staff levels and patient health outcomes is of interest to health-policy makers. Nursing staff are a significant component of inpatient health care costs. Nurses also play an important role in the quality of hospital health care, by meeting the needs of ill inpatients. Because of this, legislation in the State of California has mandated the maximum number of patients per nurse in hospital units from 2004 onwards. In response to this legislation, several authors have discussed a lack of consensus from empirical research on the effect of nursing ratios on patient health outcomes (Spetz et al., 2008; Sochalski et al., 2008).

Consensus from empirical research has been limited in part because nursing staff datasets are often aggregated to the hospital and year level. In addition, many studies have not used empirical methods to establish causal inference. In response to the lack of detailed micro-data studies on nursing inputs and patient health outcomes, I estimate a relationship between nursing and patient hours on a *ward* and a patient's health outcome. Health outcomes are length of hospital stay and mortality. A detailed dataset on daily values of nursing and patient hours for 20 wards in a large hospital is used.

My paper makes three main contributions to the empirical literature on hospital nursing staff and patient health outcomes. Firstly, this paper uses a detailed nursing staff dataset. This allows me to calculate patient-level exposure to nursing inputs and patient demand on a ward, and to control for ward-level fixed-effects. Nursing staff datasets in the majority of previous studies are annual and/or cross-sectional across hospitals. The shortcomings of aggregated data sources for studying the nursing relationship have been widely acknowledged in the literature (see, Cook et al. (2010); Evans and Kim (2006); Gruber and Kleiner (2010)). Annual and/or cross-sectional nursing data does not measure differences in nurse ratios across patients within a hospital, and at a high frequency. Cross-sectional studies are also subject to Omitted

Variable Bias (OVB) problems: unobserved hospital-specific factors could influence both the number of nursing staff *and* patient health outcomes. There is therefore a need in the literature to link an inpatient's unique exposure to nurse inputs and their health outcome.

Secondly, I extend the study of nursing inputs on patient health by considering decreasing returns to nursing inputs (on patient health), and proposing an instrumental variable strategy. The relationship between nursing hours on a ward and mortality is initially positive. This is counter-intuitive, because we might expect higher nurse hours to be associated with a lower risk of mortality, through more intensive nursing care. I investigate the possibility that decreasing returns to nursing inputs and/or an endogenous nursing variable explains a positive relationship between nursing inputs and patient health.

Nursing hours on a ward could be endogenous if patients with a severe illness increase nursing staff levels. Because a patient's severity of illness is mostly unobserved in my data, this endogenous relationship could bias the coefficient estimate on nursing inputs upwards. My instrumental variable for nursing hours is the hours of nurse sick and bereavement leave on a ward. This affects the supply of nursing staff, and is argued to be exogenous to the regression model of mortality outcomes. After instrumentation, I find a *negative* but statistically insignificant relationship between nursing hours and 60-day mortality.

A second hypothesis for a positive relationship between nursing hours and mortality is decreasing returns to nursing hours on mortality. Literature on the economics of teamwork have identified potential costs to production from increasing team size. For example, moral hazard may occur when each nurse assumes other nurses will be completing patient care tasks, this results in tasks not completed by anyone. There could also be greater distraction and risks of gaps in coordination when the number of team members increases (Holmstrom, 1982; Hamilton et al., 2003; Delarue et al.,

2008). Team-work by nurses on a ward has also been acknowledged elsewhere as an important factor in the production of quality health care (Ratto et al., 2002). These studies have used measures of nurse team-work developed from surveys of nurses.

In this paper, I use a squared nursing term to indicate if there are decreasing returns to nursing inputs. Studying non-linearities between nursing inputs and health outcomes extends previous studies on nursing inputs and patient health outcomes. Most previous studies measure the quantity of nursing staff, without consideration of how team-production could affect health care. This paper does not find a statistically significant non-linear relationship between nursing inputs and patient health outcomes, though the sign on nursing hours changes from positive to negative after inclusion of a squared nursing term.

A third contribution of this paper is my consideration of the separate effect of nursing and patient hours on a patient's health outcome. Previous studies focused on the relationship between nurse-to-patient *ratios* and health outcomes. Separating nurse and patient variables reveal whether changes in health outcomes are through nursing and/or patient variables. The relative significance of nursing and patient variables have different implications for how the delivery of health care could be improved.

My first main result is a statistically *insignificant* effect of ward nursing hours on either 60-day mortality or ward length of stay, after checking for the robustness of nursing hours to an endogenous and/or non-linear relationship.

My second main result is a *positive* relationship between patients on a ward and the length of ward stay. Patient hours is computed for each inpatient event i as the number of hours of *other* patients (j_{-1}) in $i's$ ward during $i's$ hospital stay[1]. I therefore estimate the relationship between other patients' demand for health care (in

[1]I use the term 'patient hours' as shorthand for my measure of the total hours of other patients in patient $i's$ ward, j. That is, the total ward hours pf patients excluding patient $i's$ hours in that ward. Each inpatient event is associated a measure of the number of hours of other patients (on their ward).

a ward) on a patient's length of stay. I find a negative relationship between patient hours on a ward and the hazard rate of discharge home for 16 of the 20 wards in my study. I estimate a competing risk survival model, in order to model the number of days in a ward before a patient is discharged home.

Increasing demand from other patients is hypothesised to affect a patient's length of stay through a scarcity of fixed hospital resources, such as doctors' and nurses' time, and equipment. A plausible explanation for a positive relationship between patient hours and ward length of stay is that increasing patient hours increases demand on nurse and doctor time and medical equipment, which means patients have to stay in hospital longer to receive the health care they need. This result suggests that lowering length of stay could be achieved through improving patient flow through a hospital, for example by monitoring patients that may be unnecessarily in hospital and who are potentially restricting access to health care for other patients. This could be an example of a negative spillover, because more patients admitted in a ward mean that other patients have a longer hospital stay.

Section 3.2 outlines the empirical literature on the relationship between nursing staff levels and patient health outcomes. Section 3.3 discusses the data used in this paper. The empirical strategy for mortality and length of ward stay are discussed in separate sections 3.4 and 3.5. Results for both health outcomes are discussed in Section 3.6 and 3.7, followed by my conclusion in Section 1.7.

3.2 Literature

3.2.1 Empirical literature on nurse staffing levels and patient outcomes

There is a large empirical literature on the relationship between hospital nursing staff levels and patient health outcomes, particularly mortality (for reviews of empirical

studies see; Lang et al. (2004); Kane et al. (2007); Spetz et al. (2008)). Two prominent studies by Aiken et al. (2002) and Needleman et al. (2002) have found that higher nursing staff levels are associated with fewer adverse patient events and mortality. These papers used hospital-level cross-sectional data on nursing staff. Studies using cross-sectional data may be subject to Omitted Variable Bias (OVB) because unobserved hospital factors affect both staff levels and patient health outcomes. For example, a hospital may use newer medical technology that requires higher nursing staff levels to operate, and medical technology may also affect patient outcomes.

Aiken et al. (2002) uses survey data from 168 hospitals in the State of Pennsylvania. Nurses were mailed a questionnaire to fill in how many patients they were responsible for on their last work shift. This was regardless of whether it was a day, evening, or night shift. Using this data, the authors calculated an average patient load per nurse for each hospital. The authors use the average self-reported patient load, as well as control variables for hospital characteristics (e.g. size, teaching status and technology), to estimate the relationship between nursing staff levels and patient outcomes. The authors find an increase in the likelihood of mortality when nurses care for an extra patient. An advantage of Aiken et al.'s data is that it measures how many patients a nurse is responsible for. I only observe the ward-level number of nursing hours, and not the number of patients a nurse is responsible for on a shift.

Needleman et al. (2002) uses data on nursing staff costs from hospital financial reports for 799 hospitals in the United States. There was a wide variety of formats for reporting nursing staff across states. Nursing staff levels were calculated from the number of reported Full Time Equivalent (FTE) Nursing staff. The definition of a FTE differed across states. For example, in Virginia, FTEs were calculated from total hours of nursing care, and in Missouri and South Carolina FTEs were calculated based on the total number of full- and part-time nursing employees in the hospital. Needleman et al. decided to compute a measure of nursing staff as the

number of nursing hours per inpatient day. To calculate this from FTE reports, the author's assumed nursing hours of 2,080 (52 weeks at 40 hours per week) for each nurse reported working at a hospital. They then multiplied the number of FTEs with nursing hours and divided this by the total number of patient days at that hospital. Using regression analysis at the hospital level, Needleman et al. found higher ratios of nursing hours per patient day were associated with a shorter length of stay and lower rates of some adverse outcomes (such as urinary infection, and pneumonia). They did not find a significant relationship between nursing hours and mortality rates.

Several papers have used econometric techniques to address Omitted Variable Bias (OVB) problems associated with cross-sectional analysis of hospital data. These papers by Cook et al. (2010), Tong (2011), Sochalski et al. (2008), Evans and Kim (2006), and Gruber and Kleiner (2010) are reviewed next.

Sochalski et al. (2008) compare results from a hospital cross-section and fixed-effects regression. They use data from annual reports on nursing staff levels from 343 acute care hospitals in California between 1993-2001. The measure of nursing staff is computed from yearly data on the number of productive nurse hours at the medical/surgical unit level. Productive hours are defined as time worked by permanent and temporary nurses. The authors divide the total productive hours by the total patient days in each medical/surgical unit level. This gives a measure of nursing hours per patient day that is similar to Needleman et al. (2002). They specify a fixed-effect method to use only variation in nursing staff levels over time, within a hospital. After controlling for hospital fixed-effects, Sochalski et al. find *no significant* effect of nursing hours per patient day on patient health outcomes. Patient health outcomes are 30-day AMI mortality and surgical Failure-to-Rescue rates[2].

[2]Failure-to-Rescue patient outcomes are defined by the Agency for Healthcare Research and Quality in the United States. These are defined as complications that arise from a surgical procedure, possibly resulting in mortality.

Cook et al. (2010) use the 2004 California law change mandating maximum patient-to-nurse ratios. The authors argue that the mandate provided a source for an exogenous change in nursing staff levels. This therefore allows the estimation of a causal relationship between nursing staff levels and health outcomes. They examine the effect of the law change for previously under-staffed Medical units. The authors use data on the total hours of Registered Nurses (RNs) and aides/orderlies at the hospital unit level (i.e. medical, surgical units). They calculate (nurse) productive hours per patient day as the ratio of '(nurse) productive hours worked' to 'patient census days'. Patient census days are the total hours a patient spends in a unit. Patients are matched to a unit based on their DRG. Using 294 hospital-unit observations they find *no significant* relationship between nursing ratios and patient health outcomes. Patient health outcomes include Debucitus Ulcers and Failure-to-Rescue rates.

Similarly, Tong (2011) used the 2004 California law change to estimate the impact of nursing staff in Skilled Nursing Facilities (i.e. nursing homes) on patient outcomes. They find a statistically significant decrease in patient mortality in Skilled Nursing Facilities after the nursing law came into effect.

Evans and Kim (2006) use changes in hospital admissions ('shocks') to estimate the effect of staffing shortages on patient outcomes. They use a measure of unexpected admissions on the Friday and Saturday following a patient's admittance to hospital on a Thursday. The authors argue 'admission shocks' only affect patient outcomes through a strain on nursing resources. The authors compare outcomes for patients only admitted on a Thursday. A moving average of Friday and Saturday admissions using an 8 week window is calculated. An 'admission shock' is defined as movement from this average. They find *no significant* effect of admission shocks on mortality, and only a small effect for length of stay and readmission probabilities. A disadvantage of Evans and Kim's study is the inability to observe how staffing variables respond to increases in admissions. For example, hospitals may call in extra

nursing staff if there is an increase in admissions. In addition, the authors are only able to consider hospital-level admissions, and therefore strain on nursing resources may vary across wards depending on the number of patients in a ward.

Gruber and Kleiner (2010) investigate the relationship between nurse strikes and patient outcomes in the State of New York. Their data covers 50 strikes in 43 hospitals in New York State between 1983-2005. Patient outcomes are; in-hospital mortality, readmission within 30 days, LOS, and the number of procedures performed. They aggregate data to the hospital-daily level, so the mortality outcome is the average daily mortality in hospital h on day t. Their regression analysis uses a hospital fixed-effect. The independent variable of interest is whether a strike occurred in a time period or not. Gruber and Kleiner find no significant change in LOS and number of procedures during a strike period. They further split their sample into emergency and non-emergency patients, both of these populations show an increase in mortality during strike periods. Lastly, the authors investigate whether patients with diagnoses that require greater nursing care are more sensitive to nurse strikes. They use nursing intensity weights (NIWs) that were developed by a panel of registered nurses for the New York State Department of Health. They find a larger effect of nursing strikes on mortality for conditions that require greater nursing care.

Kane et al. (2007) completed a systematic review and meta-analysis of 101 empirical studies on the impact of nursing ratios on patient outcomes. These studies are all drawn from health and medical research, and therefore provide a good overview of research in areas outside of economics. Kane et al. (2007) separates studies into those conducted at the hospital and patient level. Papers using data at the patient level do not, however, use nursing data that varies at the ward/unit and/or daily level. That is, each patient is associated with a nursing staff measure using information on nurses that is recorded at the hospital level and/or non-daily (e.g monthly or yearly reporting of staffing levels).

The number of papers that use detailed nursing staff data is small. A key study is by Needleman et al. (2011). They use nurse data on wards in a hospital in the United States. Needleman et al. uses the difference between targeted and actual nurse hours on a ward, to estimate the relationship between under-staffed wards and patient mortality. They measure under-staffed wards by comparing the *targeted* staff for each ward and shift with *actual* staffing numbers. Their independent variable of interest is a time-varying measure of each patient's exposure to shifts where staffing levels were 8 hours or more *below* target. A Cox proportional hazard model is used with in-hospital mortality as the outcome. The authors find a higher mortality hazard ratio for patients that are exposed to under-staffed wards. This means patients are at a greater risk of mortality early in their stay if they are exposed to higher numbers of under-staffed shifts.

Cumulative exposure to under-staffed wards is also likely to be endogenous in Needleman et al.'s model: patients that are more likely to die have higher nursing requirements and therefore will also be more likely to not meet the target rate. Patients that are unlikely to die are less likely to require high amounts of targeted nursing staff, and therefore it is easier to meet patient nursing needs. To deal with this source of confounding, the authors use control variables for targeted staffing rates and number of patients in the ward to control for the average illness severity of the ward.

Another study by Tarnow-Mordi et al. (2000) estimates patient exposure to nursing intensity in an ICU unit in a hospital in the United Kingdom. The authors count the number of nurses reported to work in the unit at four different time periods; January 1st 1992, November 1st 1992, November 1st 1993, and April 1st 1995. Nurses are assumed to work 12 hour shifts. Nursing ratios therefore vary mostly by the number of other patients in the 6-bed ICU unit, because there are only four observations of variation in the number of nurses. Patient outcomes are indexed at the individual

level. They found a higher odds ratio for mortality when patients were exposed to nurses with a higher ICU workload.

There is therefore mixed evidence for the effect of nursing staff levels on hospital patient health outcomes. There is a small literature that uses micro-data on nursing. A key paper (Needleman et al. (2011)) finds a significant positive relationship between under-staffed wards and mortality. However, in the majority of economic studies no statistically significant relationship between nursing levels and patient health outcomes has been found. This may be due to the limitations of using data that has been collected at an aggregate level. This paper contributes to the empirical literature by using both a detailed nursing staff dataset, and a unique instrumental variable strategy.

3.3 Data

3.3.1 Source of data

I obtain data on hours worked by nurses from a hospital's payroll department. The payroll dataset has the total hours worked for each employee on a ward and calendar day. I ensure that only nurse hours are extracted from payroll databases by restricting employees to have 'expense codes' corresponding to nursing staff and health care assistants. Payroll systems also provide information for employees working in a ward that are not nurses, for example, ward clerks and medical staff. I am therefore confident that my staffing measure is only for nurses working on a ward.

Payroll data is also more accurate than rosters maintained by a ward's Charge Nurse. Rosters that are manually managed by Charge Nurses could contain errors, because staff can swap and change shifts informally without changing hard copy rosters. Payroll determines nurse pay and records if nurses leave work early, as well as all other types of leave payments. That is, payment for time not spent working on

a ward (e.g. annual leave, sick leave, education leave) is separated from time paid for work on a ward.

The second source of data is from inpatient data collections. I use a dataset that records all wards a patient stays in, and the date and time they arrived and left each ward. This dataset also records if a patient was transferred to an Operating Room for an operation, or if a patient was transferred to a transition lounge before leaving the hospital. This provides accurate information on the number of patients in each ward for each calendar day. Another hospital inpatient dataset contains information on a patient's demographic characteristics, medical diagnoses and treatments that are used as control variables in my study.

3.3.2 Wards in the study and descriptive statistics

I have payroll and ward stay data for 26 hospital wards. The number of patients staying in each ward, for any amount of time, is tabulated against 60-day mortality in Table 3.1. Sixty-day mortality is calculated as mortality within 60 days of hospital admission. This table shows that there is considerable variation in the proportion of mortality outcomes across wards. The highest proportion are in the Cancer and Hematology and General Medicine wards. The lowest proportion of mortality outcomes occurs in the Maternity ward.

From 26 wards, I exclude 6 wards from my study. Twenty wards are therefore used to estimate the nursing and patient health outcome relationship.

I firstly exclude Maternity and Gynaecology wards from my study. These wards are open-planned in the hospital. It is unclear whether nurses work exclusively on one ward, or on both wards, because there is no physical wall separating these wards. The incidence of mortality in these wards is also amongst the lowest in the hospital, and therefore excluding these wards does not reduce observations of mortality required to identify my regression model.

Table 3.1: Ward population

	60 day mortality occurred			Total	
	Numb.	Row %	Col %	No.	Col %
Cancer & Hematology(Blood) a	728	14.3	4.7	5,094	1.8
Cancer & Hematology(Blood) b	2,956	28.3	19.1	10,442	3.7
Cardiology a	414	2.8	2.7	14,946	5.2
Cardiothoracic	387	2.7	2.5	14,111	4.9
Coronary b	463	4.8	3.0	9603	3.4
General Medicine a	1,143	10.6	7.4	10,831	3.8
General Medicine b	1,026	11.4	6.6	9,003	3.1
General Medicine c	1,134	10.9	7.3	10,365	3.6
General Medicine d	1,138	10.8	7.4	10,571	3.7
General Surgery a	429	5.6	2.8	7,684	2.7
General Surgery b	572	6.0	3.7	9,586	3.4
Gynae	101	0.7	0.7	13,597	4.8
Head & Neck, Ear, Nose & Throat	291	2.2	1.9	13,324	4.7
Kidney & Liver	744	7.0	4.8	10,565	3.7
Long Term a	93	10.2	0.6	910	0.3
Long Term b	94	9.9	0.6	953	0.3
Long Term c	256	9.7	1.7	2,642	0.9
Long Term d	321	0.8	2.1	39,149	13.7
Maternity	10	0.0	0.1	2,9521	10.3
Neuro a	265	3.4	1.7	7,709	2.7
Neuro b	398	6.5	2.6	6,077	2.1
Orthopaedics a	217	2.5	1.4	8,533	3.0
Orthopaedics b	301	4.1	1.9	7,320	2.6
Respiratory	1,152	12.5	7.5	9,189	3.2
Urology	422	2.8	2.7	15,136	5.3
Vascular	407	4.5	2.6	9,036	3.2
Total	15,462	5.4	100.0	285,897	100.0

Notes: Summarises 60-day mortality outcomes by hospital wards. The first three columns give the number, row and column percentage of patients that died within 60 days of Hospital admission. The last two columns give the total number of percentage of patients in each ward.

Secondly, I exclude Long Term Stay wards (a, b, c, d) from my study. Long Term a and b were phased out of use during the study period. Patients in Long Term wards are also generally admitted for an extended physical and/or mental rehabilitation. In addition, they are part of an older hospital building that is separate to the main hospital's acute wards. I focus on a 'standard' set of hospital acute care wards to estimate my nursing input and patient health outcome relationship.

3.3.3 Nurse and patient hours by ward

From payroll and inpatient event data, I calculated the nursing and patient hours for each ward and 24-hour calendar day. Payroll information does not provide the exact timing of a shift (e.g. morning or evening) and therefore ward nursing hours on a date are for a (24 hour) calendar day. The measure of nursing hours on a ward-calender

day is the sum of hours worked by all nurses on that date. A 24 hour period was defined from 7am on a calendar day to 7am the following calendar day.

The measure of patient hours on a ward is calculated from ward stay data. This dataset has the entry and exit times for all wards a patient stays in. For a given ward and calendar day, the amount of time spent by all patients in that ward during a 24-hour period from 7am to 7am the following day is added up. This includes if a patient stayed the full 24 hours, left the ward, arrived at the ward, or passed through that ward during the 24 hour period. For example, if a patient remains in a ward for a whole calendar day, 24 hours is added to the total patient hours for that ward and calendar day. Or, if a patient stayed in that ward for 2 hours between 7am and 7am the following day, 2 hours is added to the total patient hours on a ward.

Figures 3.1 and 3.2 plots the kernel density distribution for daily patient and nurse hours for each ward. The shortened version of ward names is used henceforth in this paper. From these figures, patient hours (Figure 3.1) is slightly more concentrated than nursing hours (Figure 3.2). A reasonable amount of variation in nursing hours is beneficial to identifying the relationship between nursing hours on patient health outcomes.

Summary statistics for nursing and patient hours for each ward are in Table 3.2. The average number of daily nursing hours varies across wards from 128 to 212. This indicates a ward level fixed-effect is important in a regression analysis to account for heterogeneity in average ward staffing levels.

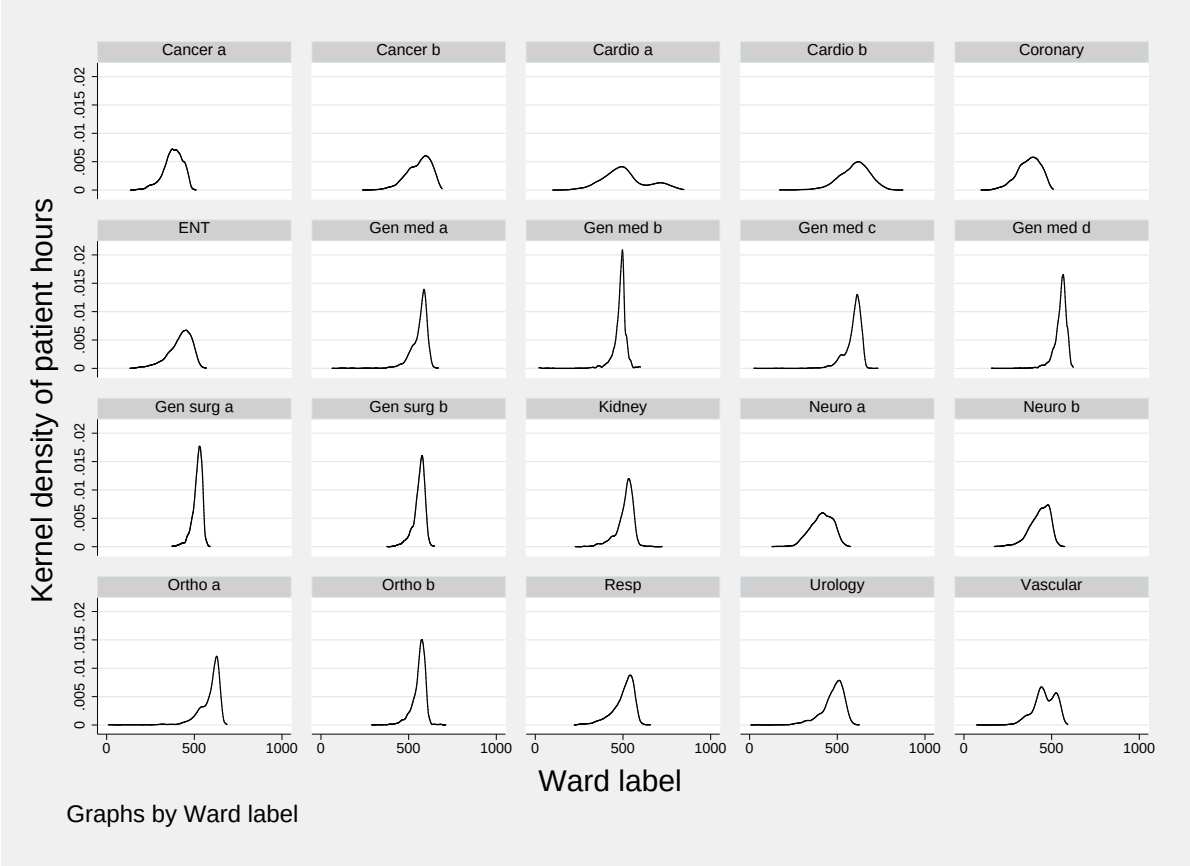

Figure 3.1: Density plots of ward patient hours

Table 3.2: Daily nurse and patient hours for each ward and day

Ward label	Mean		Sd		Min		Max	
	Patient	Nurse	Patient	Nurse	Patient	Nurse	Patient	Nurse
Cancer a	378.8	142.9	56.0	22.6	135.9	83.5	510.3	221.0
Cancer b	557.6	174.8	70.3	20.9	237.1	108.0	691.2	263.5
Cardio a	508.6	128.6	126.7	27.3	98.5	48.0	847.7	244.5
Cardio b	597.3	206.8	89.0	23.5	169.9	95.5	874.8	280.0
Coronary	370.0	212.3	66.1	25.2	97.2	132.0	510.1	357.5
ENT	418.2	130.2	68.4	28.0	132.8	57.8	569.5	216.0
Gen med a	559.8	147.2	56.6	19.0	63.0	16.0	670.5	215.0
Gen med b	482.5	140.3	45.4	18.9	20.4	7.5	600.0	200.0
Gen med c	590.7	161.7	51.2	16.9	24.2	76.0	732.2	227.0
Gen med d	551.5	163.4	37.5	15.7	156.5	84.0	624.0	224.5
Gen surg a	518.1	158.1	28.7	21.9	372.9	68.0	589.9	241.5
Gen surg b	560.7	153.7	34.1	18.5	376.3	80.0	646.9	228.0
Kidney	512.7	167.8	49.9	20.8	227.7	96.0	723.5	237.5
Neuro a	411.0	175.8	64.1	26.9	127.7	99.8	575.4	279.5
Neuro b	432.3	173.0	58.2	21.0	173.1	112.0	574.8	255.5
Ortho a	584.8	172.1	74.0	19.5	9.8	86.8	686.8	239.5
Ortho b	559.5	156.0	39.9	18.3	288.7	88.0	712.4	216.0
Resp	502.8	138.6	63.1	33.1	221.8	48.0	657.4	248.5
Urology	473.5	128.8	68.8	20.8	6.4	48.0	626.0	204.5
Vascular	459.7	156.4	65.0	25.0	72.4	79.5	592.5	244.0
Total	501.4	159.5	94.4	31.8	6.4	7.5	874.8	357.5

Notes: Summarises nurse and patient hours for each ward in the study.

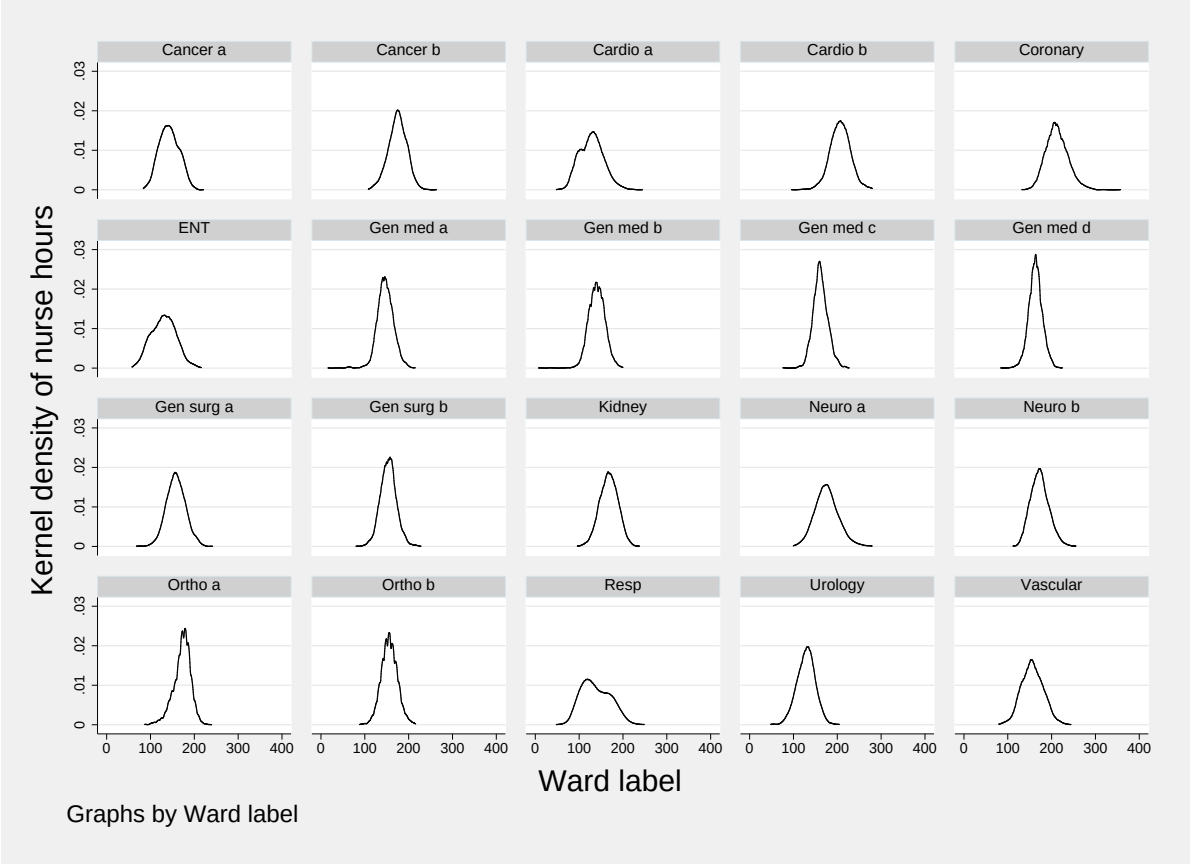

Figure 3.2: Density plots of ward nurse hours

3.4 Empirical strategy: Mortality

I firstly discuss my mortality measure and study sample before presenting my baseline econometric model. I then discuss explanatory variables in my model, including the the nurse and patient variables.

3.4.1 Outcome

I use mortality within 60 days of hospital admission as the patient health outcome. hospital record collections contain the date of death for mortality both inside and outside of hospital. A 60-day time frame is used because it increases the number of mortality observations compared to in-hospital mortality, and also includes patients who may not die in hospital, but shortly after a hospital stay.

Mortality is a widely used indicator for quality of care in hospital empirical studies. It is argued that a lower mortality rate, controlling for a patient's health condition, indicates better quality health care. A shortcoming of using mortality outcomes is that it can be hard to identify patients at a real risk of mortality from the universe of hospital admissions. Combined with the relatively rare occurrence of mortality, it can be hard to identify the effect of explanatory variables on the risk of mortality, particularly in small sample sizes[3].

This paper also includes all inpatient events before a 60-day mortality occurrence. That is, if a patient visited hospital multiple times before mortality within 60-days of their first hospital admission, all hospital events before mortality will be included as separate observations in the regression. It is difficult to identify which ward event

[3]Emergency readmission and adverse hospital events (such as falls in hospital, pneumonias etc.) were also considered for patient health outcomes in this paper. Emergency readmissions could be a problematic measure because they are not observed if a patient turned up at another hospital. For adverse events, patient event data does not record when an adverse event occurs. Patients with an adverse event will likely stay in hospital longer, it is therefore difficult to identify the effect of nursing inputs leading up to an adverse event, compared to nursing inputs after an adverse event. In addition, diagnosis codes are not separated by Present On Admission (POA), or medical conditions that develop in hospital. An adverse event could include patients that enter hospital with a condition (e.g. Decubitus or 'pressure' ulcer) rather than developing this condition in hospital.

led to a mortality outcome, so all inpatient events are included. For example, a patient may return to hospital just before dying, and their health condition could be contributed to by nursing care in a previous ward stay.

A tabulation of 60-day mortality by hospital wards is in Table 3.1. This provides the proportion of patients in each ward where 60-day mortality is equal to one (occurs) and zero (does not occur). There is heterogeneity across wards in patient health outcomes; Cancer and Hematology Ward B has the highest proportion of 60-day mortality (28.1%), the Ear, Nose and Throat ward has the smallest proportion of mortality outcomes (2.2%).

3.4.2 Study sample

I use three restrictions on the population of inpatients for my study. These are; (1) patients must spend at least 80% of their hospital stay in their last ward of stay (2) patients must stay greater than one night and (3) patients must have a DRG code ending in an 'A'. Each of these are discussed next.

This paper excludes patients that stay across multiple wards. I use only patients that stay at least 80% of their whole hospital stay in their last ward. If a patient spends considerable time in more than one ward, it is difficult to identify the relationship between ward nursing and patient hours on health outcomes. Firstly, the average nursing level varies across wards. If patients have nursing hours from different wards, this will add noise to the event-level nursing measure. Secondly, patients cannot be associated with one ward, and I therefore cannot include a ward-level fixed-effect in a regression. The last ward of stay is also the ward stay leading to the discharge decision or mortality outcome and is therefore identified with providing the nursing care required for a patient's recovery.

Many patients enter hospital with minor illnesses and are not at risk of mortality. For example, patients with arm or leg fractures are typically not at risk of mortality.

This paper uses two restrictions on the patient population to select patients with more complex health conditions, and are therefore more likely to be at risk of mortality.

Firstly, patients must stay *greater* than one night. A large number of hospital admissions are for overnight admissions. If a patient was considered at risk for mortality, then it is likely they would stay more than one night. This patient restriction also results in at least three days of ward stay data to calculate average nursing and patient hours. This has the advantage of avoiding fluctuations in nursing measures that are due to a short hospital stay.

Secondly, only patients with a DRG code ending in 'A' are selected. This category refers to the most complex and resource intensive hospital stays. DRG codes are assigned by software that use information from inpatient events. In general, for a given health condition, patients with the highest number of complicating conditions and/or invasive medical treatments are associated with a DRG category of 'A'.

Table 3.3 tabulates the mortality outcome against population restrictions for DRG and ward nights. This illustrates how many observations I exclude from my study sample because they do not meet my inclusion conditions.

Table 3.3 shows that patients in the study sample are at a greater risk of mortality. 12.7% of patients with a DRG code ending in 'A' have a mortality outcome, this is compared to a 4.3% mortality outcome for patients that do not have a DRG code ending in 'A'. In addition, after controlling for all three patient restrictions, 14.6% of patients have a mortality outcome compared to 5.8% of patients that do not meet *all* three patient restrictions.

In the result section, I also estimate the mortality outcome model after relaxing each patient restriction. When estimated on the whole population of hospital admissions standard errors increase. An increase in standard errors is likely due to comparing nurse hours for patients that are not at a real risk of mortality, with those who are. When complex patients are only included in the sample, unobserved het-

Table 3.3: Study population

	60-day mortality occurred		Total
	Number	Row %	Number
DRG group is 'A'			
No	5,467	4.3	127,304
Yes	9,120	12.7	71,821
Ward nights >1			
No	2,358	4.6	50,758
Yes	12,229	8.2	148,367
All three restrictions			
No	9,467	5.8	163,961
Yes	5,120	14.6	35,164
Total	14,587	7.3	199,125

Notes: Summarises 60-day mortality outcome for each patient restriction used to select the study sample from the inpatient population. First two columns give the number and row percentage of patients that died within 60 days of hospital admission. The last column is the total number of patients in each category. For example, 71,821 patients are in DRG group 'A'.

erogeneity in severity of illness is reduced, thereby reducing error in the prediction of nursing variables on mortality. On the whole, patient restrictions are designed to compare 'like with like' because there is a large degree of heterogeneity in hospital admissions.

3.4.3 Econometric model

The baseline model for the mortality outcome is indexed for inpatient event i that stayed in ward j:

$$Mortality_{ij} = \alpha + \beta X_i + \lambda nurse\_hours_i + \theta patient\_hours_i + \gamma_j + \epsilon_{ij} \tag{3.1}$$

This paper specifies a fixed-effect on wards because nurse and patient hours vary across wards, depending on ward size and medical specialty. In addition, unobserved ward-level factors, such as team culture and nurse scheduling practices, could influence the mortality outcome. Wards also tend to specialise in medical condition and a fixed-effect therefore controls for variation in medical condition across patients in the sample.

The method used to calculate $nurse\_hours_i$ and $patient\_hours_i$ is discussed in the next section (3.4.4).

I model my Length of Stay health outcome in a survival framework. Survival models calculate the time to an event of interest, in this case discharge home. Survival models are also used to model mortality, particularly in medical studies. In this paper, the number of in-hospital deaths is small, and it is therefore difficult to identify the effect of nursing hours on in-hospital mortality. Furthermore, Schoenfeld (2005, pg.104) argues that studies of mortality, particularly in Intensive Care Unit stays, should use binary outcome models, rather than competing risk models: 'the problem with these estimators [competing risk] is that they focus on when patients die in the hospital rather than whether they die. The quality of a patient's life in the ICU is very poor. Thus we should avoid any analysis that can confuse longer survival with better mortality.' That is, the outcome of survival modelling is a conditional probability that a person will die per unit of time, rather than whether a patient will die at all.

On the other hand, Wolkewitz et al. (2009) argue that competing risk models have advantages for modelling ICU mortality compared to logistic models. Primarily because they also take into account the effect of time-dependent risk factors, such as the incidence of infection, on the risk of mortality. In addition, the authors argue that the risk of mortality associated with staying in hospital longer is addressed in survival models rather than 'snap-shot' regression methods.

The main advantage of my 'snap-shot' (i.e. event-level) regression model is that I can specify a ward fixed-effect, and implement an instrumental variable strategy for nursing hours. These regression techniques have not been established for survival models.

Endogenous nursing hours and instrumental variable

Nursing hours are potentially endogenous to unobserved patient severity of illness. For example, the arrival of a more severe or complex patient could motivate a ward's Charge Nurse to roster on more nursing care. Because more severe or complex patients are at a greater risk of mortality, this could bias the estimate of the relationship between nursing hours and mortality upwards. As a result of an endogenous relationship $cov(nurse\_hours_{ij}, \epsilon_{ij}) \neq 0$ and the estimate of λ, the coefficient on nursing hours, is inconsistently estimated.

I propose an instrumental variable (Z_{ij}) for nursing hours to consistently estimate the effect of nursing hours on mortality. This is constructed from the hours of sick and bereavement leave of nurses on a ward. Sick leave is defined as nurses with an illness who are not physically able to work. Bereavement leave is defined as leave for a death or serious illness of a nurses' family member.

There are two conditions an instrumental variable must satisfy to be valid for instrumental variable analysis. Firstly, the instrument must be uncorrelated with unobservable factors in the mortality model: i.e. $cov(Z_{ij}, \epsilon_{ij}) = 0$. The second requirement is that the instrument must have an effect on the endogenous variable, even after netting out the effects of all other explanatory variables in the mortality model (Wooldridge, 2002, pg.89-90). The second requirement can be tested. The first requirement must be argued as a reasonable assumption, because it involves all possible *unobservable* factors in the mortality model.

To meet the first requirement, this paper argues that sick and bereavement leave is caused by processes independent of unobserved factors contained in (ϵ_{ij}), after controlling for observable variables. Bereavement leave is easier to satisfy the exogeneity condition, because death of a family member is expected to be unrelated to a ward environment and hospital events. Ward environment could however affect a patient's health outcome and nurse's *sick leave*. I argue that any plausible relationship be-

tween nurse sick leave and patient's likelihood of mortality is controlled for in the main equation (3.1).

The first possible relationship between sick leave and mortality could be caused by the number of patients on a ward. Patient load could induce stress on nurses and this could also affect a patient's risk of mortality, if not controlled for. My patient hours variable controls for the effect of patient demand on mortality. Secondly, nurses' sick leave could be caused by a flu virus, which could also be contracted by a patient and increase their risk of mortality. A dummy variable for month of hospital admission is included to control for seasonal factors, such as the flu season. I also include a dummy variable that is equal to one if a patient has a diagnosis of flu or influenza[4]. Nurses now routinely get flu vaccinations, which lowers the chance of this relationship occurring.

The largest unobserved component of ϵ_{ij} is a patient's unobserved severity and complexity of illness. A patient's illness has unique biological paths that is separate to the biology of nurses and other patients in a hospital. This paper controls for ward-level fixed-effects, patient demand on a ward, and infectious diseases which might affect both nurses' sick leave and patient's likelihood of mortality. This set of control variables is argued to be sufficient for the exogeneity of sick and bereavement leave in equation (3.1).

Other instrumental variable options illustrate the difficulty in meeting the requirement of exogeneity. For example, weekends have lower numbers of nursing staff, which could generate an exogneous change to a patient's nurse hours. Weekends however also have fewer medical staff, which could have an independent effect on a patient's likelihood of mortality.

For the second requirement - instrumental variable is related to the endogenous variable - I argue that sick and bereavement leave affects the supply of nurses able

[4]ICD-10-AM codes are J09-J11x in any of the 60 diagnosis codes available for an inpatient event.

to work, which then affects the number of nurses working. This type of leave is also unplanned, so Charge Nurses are unlikely to make extra staffing provisions. I did not find a significant negative relationship between nurse hours worked and other types of leave, such as annual and educational course leave. An explanation for this could be that other types of leave are often planned in advance, and Charge Nurses are able to cover shortfalls by rostering on replacement staff. Figure 3.3 plots a fitted line for (worked) nurse hours on a ward, against sick and bereavement hours for that day. For most wards, there is a downward slope suggesting that increasing sick-leave hours is associated with lower nursing hours worked on a ward. In the results section, an F-statistic for the explanatory power of the instrument is 202.56, which is above the suggested 'rule of thumb' of 10.

The instrumental variable is measured as the average daily hours of sick and bereavement leave during a patient's hospital stay. This is the sum of sick and bereavement leave, during a patient's stay, divided by the number of ward days (for patient i).

To instrument for nursing hours, with a fixed-effect on wards, this paper uses the two-stage least-squares within estimator implemented in the 'xtivreg2' package for *Stata* statistical software. This package transforms the data by de-meaning variables and then applies the conventional two-stage least-squares instrumental variable method. This implements Ordinary Least Squares to estimate regression coefficients in the model. Linear regression models can provide poor estimates of variable effects when the mean of the dependent variable is very low or high (Bhattacharya et al., 2006). Our mean value is low because mortality is an infrequent occurrence. Non-linear models, such as logit and probit, are designed to model binary outcomes. But there are also drawbacks to these models. They suffer from bias when relevant regressors are omitted from the regression, even if they are uncorrelated with regressors (Wooldridge, 2002, pg. 470). In addition, the inclusion of an endogenous

variable means that all coefficients are inconsistently estimated in a non-linear model. OLS estimates unbiased and consistent coefficients for all variables, except those that are endogenous. An estimation routine for a non-linear binary outcome model with fixed-effects and instrumental variables is also not readily available in *Stata* 11. In my results, I do not find a statistically significant effect of nursing hours on mortality, so concerns about the precision of the coefficient estimate is perhaps not of primary importance. Any robust finding would require that coefficient estimates are statistically significant in *both* linear and non-linear models.

The two-stage least-squares estimation framework can be set up with a first stage and a reduced form equation. A consistent estimate of λ, the effect of nursing hours on mortality, can be obtained from estimation of these two equations. The first stage equation is estimated by regressing the endogenous variable (nursing hours) on covariate vector X from equation 3.1 and the instrumental variable. The first stage equation, with a fixed-effect for within-ward estimation, is:

$$Nurse\_hours_{ij} = \alpha_{FS} + \beta_{FS} X_i + \theta_{FS} patient\_hours_{ij} + \delta_{FS} sick\_and\_bereav_{ij} + \gamma_j + \mu_{ij} \tag{3.2}$$

The reduced form equation specifies the mortality outcome as a function of the instrumental variable:

$$Mortality_{ij} = \alpha_{RF} + \beta_{RF} X_i + \theta_{RF} patient\_hours_{ij} + delta_{RF} sick\_bereav_{ij} + \rho_j + \upsilon_{ij} \tag{3.3}$$

An estimate for λ in equation 3.1 is obtained from the sample estimate of δ_{RF} and δ_{FS} : $\lambda = \delta_{RF}/\delta_{FS}$. Consistent estimation of λ requires that sick and bereavement leave affects a patient's likelihood of mortality *solely* through changes in nurse hours.

Table 3.4: Daily sick and bereavement (Bere.) hours for each ward and day

Ward label	Mean		Sd		Median		Max	
	Sick	Bere.	Sick	Bere.	Sick	Bere.	Sick	Bere.
Cancer a	4.8	0.5	6.7	2.2	0.0	0.0	48.0	24.0
Cancer b	6.4	0.4	8.1	1.9	0.0	0.0	52.0	16.0
Cardio a	5.4	0.2	7.7	1.3	0.0	0.0	44.0	12.0
Cardio b	8.0	0.5	9.3	2.3	8.0	0.0	56.0	12.0
Coronary	7.8	0.4	9.7	2.0	1.3	0.0	60.0	24.0
ENT	5.0	0.3	6.8	1.7	0.0	0.0	44.0	20.0
Gen med a	5.0	0.5	6.9	2.3	0.0	0.0	51.0	20.0
Gen med b	5.2	0.3	6.9	1.6	0.0	0.0	45.0	16.0
Gen med c	6.0	0.5	7.3	2.0	6.0	0.0	42.0	16.0
Gen med d	6.6	0.4	7.7	2.0	8.0	0.0	48.0	16.0
Gen surg a	5.6	0.6	7.4	2.3	0.0	0.0	52.0	23.5
Gen surg b	6.0	0.6	7.7	2.5	0.0	0.0	44.0	24.0
Kidney	6.6	0.4	8.3	2.0	0.0	0.0	48.0	20.0
Neuro a	6.9	0.4	7.9	1.8	8.0	0.0	52.0	13.5
Neuro b	6.0	0.5	7.5	2.1	0.0	0.0	60.0	20.0
Ortho a	6.5	0.5	7.4	1.9	8.0	0.0	40.0	16.0
Ortho b	4.8	0.4	6.1	1.8	0.0	0.0	34.0	16.0
Resp	5.4	0.4	7.1	2.0	0.0	0.0	44.0	16.0
Urology	5.4	0.5	7.4	2.3	0.0	0.0	40.0	24.0
Vascular	6.1	0.7	7.8	2.7	0.0	0.0	44.0	24.0
Total	6.0	0.4	7.7	2.1	0.0	0.0	60.0	24.0

Notes: Summary statistics for sick and bereavement leave hours for each ward in the study.

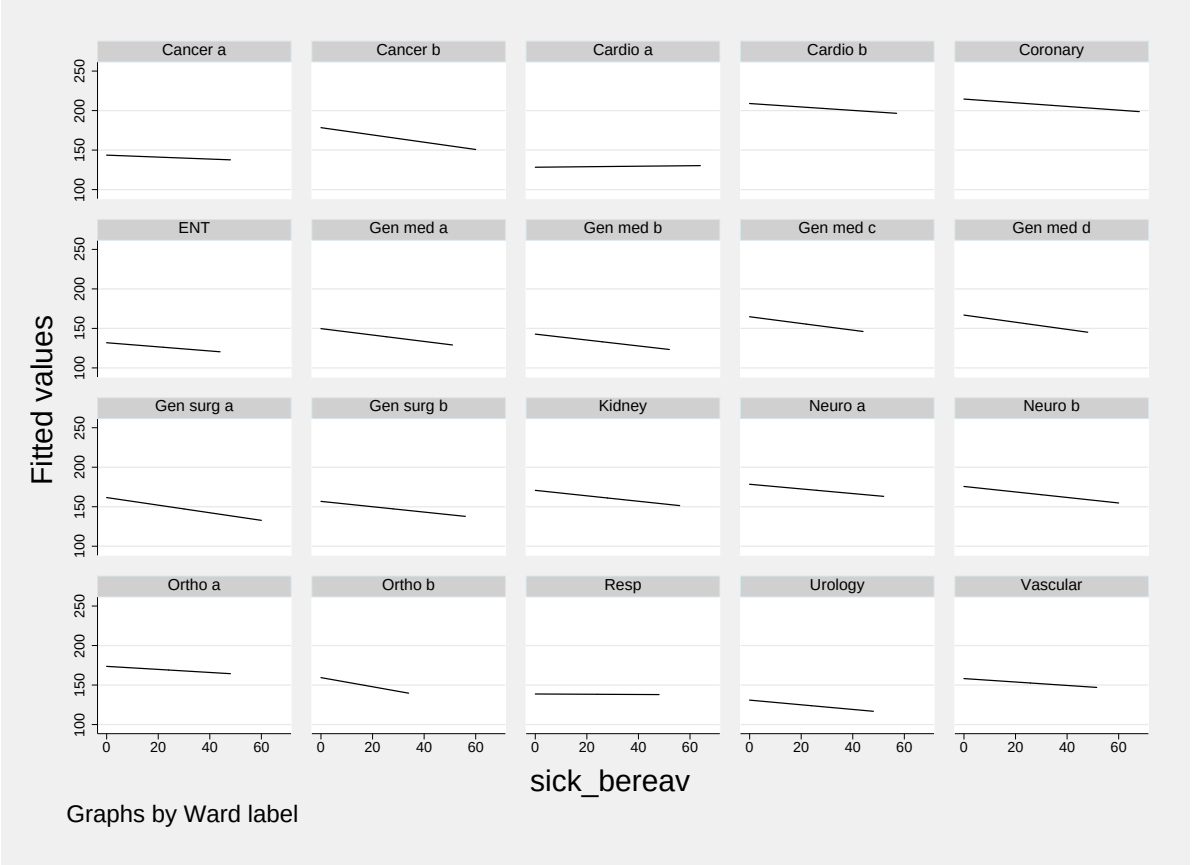

Figure 3.3: Fitted nursing wards by hours of sick and bereavement leave on a ward

3.4.4 Nursing and patient variables

The nurse hours variable, $nurse\_hours_i$, is calculated as the average daily nurse hours on a ward during patient $i's$ ward stay. This is computed by summing the nurse hours ($nurse\_hours_t$) on each calendar day (t) that patient i stayed in their ward, j. I use an index $n = 1, 2, 3...LOS_i$ in the following formula to sum nurse hours ($nurse\_hours_{n,j}$) during each day of $i's$ stay in ward j, starting from $day = n = 1$ to $n = LOS_i$. For each day, $n = 1, 2, 3...LOS_i$, a patient stays in hospital, nurse hours is obtained from payroll data for the calendar day. LOS_i is the total number of days a patient stayed in their ward, j.

$$nurse\_hours_i = \frac{\sum_{n=1}^{LOS_i} nurse\_hours_{n,j}}{LOS_i} \tag{3.4}$$

$nurse\_hours_{n,j}$ is the sum of hours worked by *all* nurses in $i's$ ward j, on day n during $i's$ inpatient stay.

Since I require that patients stay more than one night in hospital, at least three days of patient observations (the arrival day, second day and third day) for nursing and patient hours is used to calculate nursing and patient variables. Using three days reduces fluctuations in nursing and patient measures. Fluctuations are caused by patients staying in hospital for one or two days where extreme values of nursing or patient hours may have occurred. Because exposure to high or low staffed wards is small, given LOS is small, the impact on health outcomes will be limited. This could confound the estimate of the relationship between nurse and patient variables and mortality.

Similarly, the measure of patient hours is averaged over a patient's ward stay. Importantly, a patient's own contribution to ward patient hours is removed. This controls for a patient's own movement in and out of a ward on their measure of

patient hours. For example, on the *last* day of stay, a patient may have lower patient hours, because their own movement from their ward lowered patient hours.

$$patient\_hours_i = \frac{\sum_{n=t_i}^{LOS_i}(patient\_hours_{n,j} - patient\_hours_{i,n,j})}{LOS_i} \tag{3.5}$$

Where $patient\_hours_{n,j}$ is the sum of patient hours on ward j at day n during $i's$ stay. $patient\_hours_{i,n,j}$ is patient $i's$ hours in ward j on day n.

Nurse and patient hours are added to each model separately to test for the independent effect of these variables on patient outcomes. The effect of a nurse-to-patient *ratio* on health outcomes is also estimated. The nurse-to-patient ratio is the total number of nursing hours divided by the total number of patient hours: $nurse\_ratio_{i,n,j} = nurse\_hours_{i,n,j}/patient\_hours_{i,n,j}$. For an event-level measure, the sum of the daily nurse-to-patient ratio is divided by a patient's length of stay to find the average daily nursing ratio.

An alternative to my nursing and patient variables is to use the *total* amount of nursing and patient hours during a hospital stay. That is, the sum of patient and nurse hours for a patient's entire hospital stay. It could be argued that total nursing and patient hours (as opposed to average daily measures) are inputs into a production process determining a patient's health outcome. In this case, when a patient enters hospital they will receive an amount of nursing care which will be a function of their length of hospital stay. A problem with using total nursing and patient hours is that a patient may have a higher level of nursing hours because they stayed in hospital longer, rather than exposure to higher-staffed wards. This will confound attempts to investigate the effect of high or low staffed wards on patient outcomes.

3.4.5 Other explanatory variables

Table 1.13 contains summary statistics of all control variables for the study sample. Most control variables have been discussed in previous chapters, so are not discussed in detail here.

Control variables in dummy variable form are; male patient, patient ethnic group (Asian, Pacific, Indian, Maori, and base category of New Zealand European), if a patient was transferred from another health-care facility, acute admission and arranged admission (base category is wait-list admission), if a surgical theatre event occurred, previous admission to hospital within 60 days, entry to hospital through Accident and Emergency Department (AED), accident as cause of patient's illness, if the patient's DHB is that of the hospital, and Major Diagnostic Categories. Lastly, a set of dummy variables for the day, month and year of admission control for the impact of admission timing on the probability of mortality.

Charlson co-morbidity dummy variables are also included to control for complicating diagnoses. All ICD-10-AM diagnosis codes for a patient were obtained using Sundararajan et al.'s (2004) crosswalk from ICD-9 to ICD-10. A dummy variable is equal to one if a patient has one of Charlson's co-morbid conditions.

Remaining control variables for the mortality outcome model are; the length of hospital stay in days, age of patient, age of patient squared, deprivation scale and number of diagnoses coded. An extra variable - nursing hours squared - is also included in the mortality model to test for a non-linear relationship between nursing inputs and mortality.

3.5 Empirical strategy: length of stay

3.5.1 Outcome

The outcome of interest for the length of stay model is the time to discharge home. Discharge home occurs when a patient has been considered well enough to not require nursing care. The alternative patient discharge options are discharge to another health-care facility and mortality in hospital.

Table 3.5 presents the proportion of patients with each type of discharge. Discharge to another health-care facility includes discharges to a public hospital, private hospital or a rest-home. A key observation from this table are that wards vary in their types of discharges, for example Orthopaedics and Neurology wards show high percentages of total discharges to other health-care facilities, at around 25% of total ward stays.

Table 3.5 suggests that analysis should take into account the different types of discharge when modelling a patient's length of stay. This is because discharge home is a different health outcome to mortality, or discharge to another health-care facility. Discharge to another health-care facility indicates more nursing care is required compared to discharge home. Picone et al. (2003) also accounts for different types of discharge (i.e. competing risks) when modelling hospital length of stay.

A patient's length of ward stay is therefore modelled in a competing risk survival model. Survival analysis models the time to an event of interest. I use a competing risk model because there is more than one possible outcome. In this case, there are different types of hospital discharge; home, to another health care facility and mortality.

I am interested in the time to discharge home because it indicates the quality of health care. Discharge home indicates a patient is well enough to forgo hospital

Table 3.5: Ward by discharge type

Ward label	Home		Facility		Mortality		Total
	No.	Row %	No.	Row %	No.	Row %	No.
Cancer & Haematology(Blood) a	4,508	89.1	381	7.5	169	3.3	5,058
Cancer & Haematology(Blood) b	8,842	84.9	996	9.6	577	5.5	10,415
Cardiology a	13,622	91.6	1,178	7.9	78	0.5	14,878
Cardiothoracic	12,192	87.2	1,756	12.6	39	0.3	13,987
Coronary b	7684	80.7	1,698	17.8	138	1.4	9,520
General Medicine a	9002	83.5	1,387	12.9	394	3.7	10,783
General Medicine b	7297	81.3	1,322	14.7	357	4.0	8,976
General Medicine c	8,476	81.9	1,472	14.2	398	3.8	10,346
General Medicine d	8,725	82.8	1,356	12.9	452	4.3	10,533
General Surgery a	6768	88.5	735	9.6	147	1.9	7,650
General Surgery b	8518	89.2	895	9.4	138	1.4	9,551
Head & Neck, Ear, Nose & Throat	12,984	97.6	266	2.0	54	0.4	13,304
Kidney & Liver	8,992	85.4	1,312	12.5	225	2.1	10,529
Neuro a	5,685	73.9	1,925	25.0	86	1.1	7,696
Neuro b	4,279	71.0	1,650	27.4	96	1.6	6,025
Orthopaedics a	6,529	76.6	1,941	22.8	48	0.6	8,518
Orthopaedics b	5,178	70.9	2,050	28.1	75	1.0	7,303
Respiratory	8,260	90.2	583	6.4	318	3.5	9,161
Urology	14,451	95.6	613	4.1	52	0.3	15,116
Vascular	7,740	86.2	1,141	12.7	99	1.1	8,980
Total	169,732	85.6	24,657	12.4	3,940	2.0	198,329

Notes: Summarises the number of percentage of patients for each discharge type by wards in the study. Facility includes private, public and rest home facilities. Mortality is in-hospital. The last column gives the total number of patients in each ward.

health care. Reducing length of stay therefore suggests an improvement in the quality of health care, as patients recover to sufficient health status in a shorter time period.

3.5.2 Study sample

I use two restrictions when selecting the patient population for the LOS model. These are; (1) patients must spend at least 80% of their hospital stay in one ward and (2) patients must stay greater than one night. These are the same as the mortality model, except I do not use the third restriction that requires patients to have a DRG code ending in 'A'. This is used in the mortality model to select patients at a high risk of mortality.

I use the *last* ward of stay because the discharge decision reveals information about the health of the patient, specifically that they are well enough to forgo further ward nursing care. In addition, the last ward of stay selects the ward stay after an

operation. Patients waiting in hospital for an operation are not at risk of discharge. Because operating room events are separate ward stays in the data, the last ward stay cannot be before an operating room event. Only when a patient enters their last ward of stay do we consider them at risk of discharge. Lastly, to identify the effect of nursing and patient variables in a ward on a patient's health outcome, a patient must have sufficient exposure (at least 80% of stay) to nursing and patient hours on that ward.

3.5.3 Econometric method

In survival models, the time to an event of interest is modelled with a hazard rate. The hazard rate is the probability of 'failure' (i.e. outcome occurs) at a point in time. In this paper, the hazard rate for discharge home is calculated for each day of a ward stay, $t = 2, 3, 4, 5...T$. Variables either increase or decrease the hazard rate of being discharged home. If a variable increases the hazard rate, patients are at a higher risk of discharge home early in a hospital stay, and vice versa for a negative relationship.

I use the Fine and Gray (1999) model for competing risks. Their model is based on the proportional hazards model, which is discussed in Appendix 3.E. Because data on nursing hours is for a 24 hour period, I use a discrete-time set-up with time indexed from $day = t = 0, 1, 2, 3...T$. Appendix 3.E provides details for the Fine and Gray model, and discusses this model in the context of alternative competing risk models.

Fine and Gray's method for modelling competing risks has some advantages. Specifically, it is widely used in medical studies of hospital length of stay, and it is the only estimation routine for competing risks in *Stata* 11. It also requires fewer modelling assumptions, particularly functional form decisions for unobserved heterogeneity terms that are a part of economic approaches. Econometric approaches for

competing risks typically modify a multinomial model to estimate a patient's risk of discharge as a function of time.

The Fine and Gray method estimates model coefficients by comparing the values of variables for individuals who *fail* at time t with individuals who have *not failed* by time t. This is estimated in a partial likelihood function, which is the product of a ratio for each individual in the study, i up to k individuals. The numerator for i's ratio is composed of individuals who fail at LOS_i (i.e. individual i's time of failure) divided by all individuals who remain in the risk set at LOS_i:

$$L(\beta) = \Pi_{i=1}^{k} \left[\frac{\exp\{Z_i^T(LOS_i)\beta\}}{\sum_{j \in R_i} \exp\{Z_j^T(LOS_i)\beta\}} \right]^{I(\varepsilon_i=1)} \tag{3.6}$$

Where ε_i is the cause of failure, in this case discharge home.

$Z_i^T(LOS_i)$ is a vector of possibly time-varying covariates for individual i at time LOS_i. $Z_j^T(LOS_i)$ is a vector of possibly time-varying covariates for individuals j in the risk set at time LOS_i. The subjects that are in the risk set at i's failure time, LOS_i, are all individuals who have not failed by LOS_i, *or* have failed from a competing cause. This partial likelihood can be solved for β, the estimated effect of variables on the hazard rate of discharge home. This is a partial likelihood because it is over the risk set of individuals who have not failed from discharge home, rather than the whole patient population.

Fine and Gray (1999) specify a *proportional hazard* model, which is defined for event m (discharge home) as: $\lambda_m(t;X) = \lambda_{mo}(t)\exp\{Z^T(t)\beta_0\}$. The exponential function remains in the likelihood function 3.6, but the baseline hazard function ($\lambda_{mo}(t)$) and constant term (β_0) cancel out in the ratio of hazard functions, for each individual i at time LOS_i. The likelihood function (3.6) is therefore a product of the

ratio of hazard functions for event m (discharge home) at each individual's failure time.

Z^T includes $nurse\_hours_{iT}$ and $patient\_hours_{iT}$ as well as other explanatory variables discussed in the next section.

3.5.4 Nursing and patient variables

Because the risk of being discharged home is estimated for each day of a patient's hospital stay, it is possible to use time-varying covariates in the competing risk survival model. I use time-varying measures of nurse and patient hours for each day during a patient's hospital stay. I use a cumulative *average* of the nurse and patient hours during a patient's ward stay.

The nurse hours variable is the sum of nurse hours on $i's$ ward j up to day t of patient $i's$ stay, divided by t. For example, nurse hours on day 2 is the average of nurse hours on day 1 and 2 of a patient's hospital stay. The patient hours measure is the same, except patient $i's$ individual contribution to ward patient hours on day t is removed from the ward level measure[5]:

$$nurse\_hours_{it} = \frac{\sum_{n=1}^{t} nurse\_hours_{n,j}}{t} \tag{3.7}$$

$$patient\_hours_{it} = \frac{\sum_{n=1}^{t} (patient\_hours_{n,j} - patient\_hours_{i,n,j})}{t} \tag{3.8}$$

[5]I use the cumulative *average* of nursing and patient hours up to day t, as opposed to the total sum of nurse and patient hours up to time t. In the Fine and Gray competing risk model inpatients who are discharged to a health care facility will remain in the 'risk set' . These patients are not exposed to more nursing hours in the data and therefore their measure of nursing hours will remain constant, while other subjects who are still at risk for the main outcome (discharge home) will increase their nurse measure as they stay in hospital longer. If the total sum of nursing and patient hours up to time t is used, this will affect the comparison of nursing and patient variables at time t for those who fail with those who do not fail. The Fine and Gray model and the survival analysis framework is discussed in detail in Appendix 3.E.

3.5.5 Other explanatory variables

The length of stay model has the same control variables as the mortality model, except for time-varying variables; cumulative average nursing and patient hours on a ward, and dummy variables for day of the week. A dummy variable for day of the week controls for variation in discharges throughout the week. For example, there is a greater risk of discharge on a Friday than during the weekend. These variables are not included in Table 1.13 of study variables, because they are time-varying during each patient's hospital stay.

Lastly, length of stay also has a control variable for average length of stay for a patient's DRG. The average length of stay for a patient's DRG is obtained from WIESNZ11 cost weight tables provided by the Ministry of Health, New Zealand (Ministry of Health, 2011). These can be obtained electronically from the website included in the bibliography. The Ministry of Health calculates the average length of stay for each DRG from nation-wide hospital admissions data. This controls for variation in the expected length of stay for a patient's health condition. Needleman et al. (2002) includes the predicted probability of mortality for a patient's DRG in their survival model for in-hospital mortality. Summary statistics for study variables are in Table 1.13.

3.6 Results: Mortality

Results from the baseline model for mortality are in Table 3.6. Table 3.7 has results from instrumental variable analysis, and tests for a non-linear relationship between nursing hours and 60-day mortality.

A full regression table of coefficients is in Appendix 3.B and robustness tests are in Appendix 3.C.

3.6.1 Baseline mortality model

Table 3.6: Mortality baseline model

	(1) base	(2) age < 75
Patient hours	−0.00005	−0.00008*
	(−1.24)	(−2.15)
Nurse hours	0.00026	0.00042**
	(1.96)	(3.42)
Observations	34727	21795
R^2	0.132	0.128

Notes: Outcome: 60-day mortality. Sample: Column (1) is eligible study sample, Column (2) is eligible study sample under 75 years of age. Model: ward fixed-effect linear regression. Standard errors clustered on wards. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Nursing hours is positively related to mortality outcomes with *all* eligible inpatient events (Column (1)). When I restrict patients to under 75 years of age, the nursing variable is positive and statistically significant at the 5% level.

The size of the coefficient for nursing hours is however small. Increasing nursing hours on a ward is associated with an on average .026% increase in the likelihood of 60-day mortality for patients in the study sample.

Table 3.11 presents coefficient estimates for mortality outcomes that are; (1) within 30 days of hospital admission and (2) in-hospital. Both models exhibit a positive, but statistically insignificant, relationship between nursing hours and mortality.

When I relax my restrictions on the study sample, nursing hours is also positive but statistically insignificant. There are three population restrictions in the mortality model: patients spend at least 80% of their hospital stay in their last ward of stay, patients stay greater than one night and patients have a DRG code ending in 'A'.

Column (1) in Table 3.12 estimates the baseline model on the population of ward admissions that meet the first two conditions, but includes patients with all DRG codes. The nursing hours variable is positive but statistically insignificant. A plausible explanation is that when patients who are not at a risk of mortality are included in the

population, this confounds estimates for nursing inputs coefficient. This explanation is also consistent with the increase in standard errors for the nursing hours variable.

Column (2) in Table 3.12 estimates the relationship on patients that meet the latter two conditions but stay at least 60% of their total hospital stay in the last ward. The coefficient on nursing hours is positive but statistically insignificant.

Column (3) in Table 3.12 only requires that patients spend at least 60% of their total hospital stay in one ward. This allows the largest sample of eligible inpatients. Nurse hours is positively related to 60-day mortality, but is also statistically insignificant.

Table 3.13 presents results for robustness to model specification. Column (2) has no control for ward-level fixed-effects. The relationship for nursing hours is now negative, but is statistically insignificant. A negative relationship could be caused by differences in average nursing staff levels, and mortality outcomes, across wards. This result suggests that ward-level heterogeneity is important to control for when accurately estimating the nursing and health outcome relationship with micro-data.

Column (3) in Table 3.13 uses dummy variables for wards, instead of a fixed-effect transformation. Using dummy variables allows me to test for the Variance Inflation Factors (VIF) for nursing and patient variables. VIFs indicate whether there is a high degree of collinearity between variables in the model. Collinearity between nursing and patient hours could be present if they are correlated. For example, if increasing patient hours also increases the number of nurses scheduled to work. A consequence of multicollinearity is a lack of robustness in results. This is because the particular pattern of collinearity in the data will differ from sample to sample generating different estimates for the coefficients. Variance inflation factors were calculated and are included in the table under standard errors. VIF values for nursing and patient hours are 3.45 and 3.09 respectively. These are within the suggested rule of thumb of 10.

3.6.2 Instrumental variable and non-linear results

Table 3.7: Mortality non-linearities and IV

	(1) nurse sq.	(2) iv
Patient hours	−0.00005	0.00005
	(−1.16)	(0.29)
Nurse hours	−0.00088	−0.00079
	(−1.15)	(−0.49)
Nurse squared	0.00000	
	(1.51)	
Observations	34727	34727
R^2	0.132	0.130
F Statistic		206.845

Notes: Outcome: 60-day mortality. Sample: all eligible study sample. Model: ward fixed-effect linear regression, Column (1) has an additional squared nursing term, Column (2) has instrumental variable results. Standard errors clustered on wards. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

This section investigates two possible explanations for a positive association between a patient's nursing hours and 60-day mortality. Table 3.7 has results from the instrumental variable method (Column (1)) and including a squared nursing term (Column (2)).

Table 3.7 presents results for a linear fixed-effect instrumental variable model. The instrument is average daily hours of sick and bereavement leave of nurses on a ward. After instrumentation, nursing hours is negative but statistically insignificant. The Cragg-Donald Wald F statistic is 202. The suggested rule of thumb is an F-statistic above 10. The Stock and Yogo guideline for instrument validity (at the 10% level) requires an F-statistic of at least 16.38. The F-statistic supports the validity of average sick and bereavement leave as an instrumental variable for nursing hours on a ward.

Standard errors after instrumenting nurse hours are large. It is consequently difficult to test whether the nurse hours variables is endogenous, because endogenous variable tests rely on significant differences between IV and OLS estimates. If IV estimates are not precise, due to large standard errors, this could lead to rejection

of an endogenous variable, even though nursing hours may be endogenous. The p-statistic for the null hypothesis of exogeneity of nursing hours is $p = .74$. Because the sign on nursing hours changes, this suggests nursing hours could be endogenous in the baseline equation.

The second explanation for a positive relationship between nursing hours and mortality are decreasing returns to nursing inputs. Decreasing returns could arise from costs to patient health care production associated with increasing team size. For example, if increasing the number of nurses generates moral hazard in work responsibilities, or nurses are more likely to be distracted on the job. This can result in a positive relationship between nursing hours and mortality, if decreasing returns to nursing inputs outweigh the gains to increased nursing staff on health care provision.

The inclusion of a squared nursing term tests for non-linearities in the relationship between nursing inputs and mortality. Nursing hours squared is statistically insignificant (Column (1), Table 3.7). After including a squared nursing term, the coefficient on nursing hours is negative but statistically insignificant.

The positive association between nursing hours and mortality (for patients with complex health conditions) is not robust to these two hypotheses, though I am unable to conclude whether a non-linear or endogenous relationship explains the positive association in the baseline model.

3.7 Results: Length of stay

This section discusses results for the competing risk survival model. Recall that the outcome is the hazard rate of discharge home, and that nurse and patient hours are time-varying during a patient's stay. I estimate the competing risk survival model

for each ward in my study, because fixed-effect analysis is not possible. Results are in Tables 3.14 to 3.20, Appendix 3.D[6].

My main result is a negative relationship between patient hours and the hazard rate of discharge home in 16 of 20 wards in the study. Of these 16 wards, 7 wards show a statistically significant negative coefficient on patient hours. A negative relationship suggests that increasing patient hours reduces the hazard rate of discharge home, or that patients have a lower risk of discharge early in their stay. In other words, an increase in patient hours is associated with a longer ward stay.

This is somewhat counter-intuitive, because we might expect increasing patient demand to lead to an earlier discharge of patients, in order to make hospital beds available. A plausible explanation is that demand by other patients on fixed hospital resources affects the delivery of health care. For example, doctors may not be able to deliver health care treatments when they are managing many patients, so patients remain in hospital longer.

Wards that do not have a negative relationship between patient hours and the hazard rate of discharge home are wards that treat particular types of patient admissions. Of these four wards, two are cancer wards (Cancer a, b), an Ear, Nose and Throat (ENT) ward and a Urology ward. These wards treat a large proportion of short-stay patients with pre-arranged admissions. Cancer wards have a large proportion of patients entering hospital for radiology treatment. Ear, Nose and Throat and Urology have a high proportion of wait-listed surgical patients. The Urology and ENT wards also has one of the lowest rates of discharges to other health care facilities in the data at 4.1% and 2.0% respectively (Table 3.5). This suggests patients might be entering hospital for planned day-stay operations, and returning home after a procedure. The rate of Mortality is also lowest in ENT and Urology wards. Low

[6]Because nursing hours are potentially endogenous, and instrumentation of nursing hours is not possible in survival models, this section focuses on the interpretation of the *patient hours* variable on the risk of being discharged home.

mortality rates for ENT and Urology suggests that patients in these wards might be disproportionately short-stay and non-intensive patients, and therefore have a lower need for nursing and doctor care that would be sensitive to the fluctuating demands of other patients. If patients arrive at hospital with their treatment plan pre-arranged there is less clinical uncertainty, and therefore a lower burden on doctors to diagnose, treat and monitor a patient's health condition.

3.8 Conclusion

This paper has estimated a relationship between nurse and patient hours on a ward and a patient's health outcome. I have used health outcomes of 60-day mortality and ward length of stay.

My first main result is a positive relationship between nurse hours on a ward and 60-day mortality. To explain this positive relationship, I firstly investigate whether nurse hours is endogenous. I propose an instrumental variable based on the sick and bereavement leave of nurses on a ward. I find some evidence to suggest an endogenous relationship, though standard errors are large after instrumentation and the statistical test for an endogenous variable is therefore inconclusive.

I also considered the possibility that nursing inputs have a non-linear relationship with patient health outcomes. This could arise if increasing team size is associated with decreasing returns from nursing care on patient health. For example, moral hazard may be more likely to occur in large teams.

Overall, I find that nurse hours is statistically *insignificant* after estimation with an instrumental variable and inclusion of a non-linear nursing term. This result is consistent with previous studies that do not find a statistically significant relationship between nursing inputs and patient health outcomes, particularly when Omitted

Variable Bias across hospitals is accounted for (Gruber and Kleiner, 2010; Evans and Kim, 2006; Sochalski et al., 2008).

My second main result is that patient hours on a ward is related to a longer length of stay in 16 of 20 wards in my study. The relationship between patient admissions and length of stay has, to my knowledge, been relatively under-studied in previous literature. Evans and Kim's (2006) estimate the impact of a surge in hospital admissions on patient health outcomes. The authors assume that any effect on patient health outcomes is through a strain on nurses' time. Evans and Kim finds that increasing admissions is associated with a reduction in length of stay, though coefficients are small.

There is a large empirical literature on the relationship between nursing staff levels and patient health outcomes. This paper is a novel contribution to this literature. Firstly, there are few papers that use detailed data, and my instrumental variable - sick and bereavement leave of nurses on wards - is novel. I also investigate the possibility of a non-linear relationship between nursing inputs and patient health outcomes. This extends the study of nursing inputs on patient health, specifically to consider the role of increasing team *size* on the provision of health care.

The ability to robustly investigate these issues is however limited, because my sample size is small. Restrictions on the eligible patient population reduced the number of observations to identify an empirical relationship. This meant that results from tests for an endogenous or non-linear relationship were inconclusive.

Further work could explore the relationship between other patients' demand on health care and a patient's own health outcome. This relates to literature on negative spillovers, because admissions by *other* patients in a hospital could affect a patient's own health outcome. Because patients are sharing medical resources, the impact of admissions on a patient's access to health care is of interest to administrators. It could indicate whether administrators need to increase capacity to avoid negative

spillovers. Further work could also focus on the role of team-work by nurses on a ward and patient health outcomes. This extends the study from the quantity of nursing staff to the quality of nursing staff on wards.

3.A Explanatory variables

Table 3.8: Health outcomes

	Mean	*Sd*	*Min*	*Max*	*P*50	*P*90
60-day mortality	0.09	0.28	0.00	1.00	0.00	0.00
Ward days	7.06	7.22	2.00	221.00	5.00	14.00
N	93382					

Notes: Summary statistics for mortality and length of ward stay. Summarised for the inpatients that stay in hospital greater than one night, and spend at least 80% of total hospital stay in the last ward of stay.

Table 3.9: Study variables

	Mean	Sd	Min	Max	P50	P90
Main variables:						
Patient hours	513.65	81.37	83.20	838.77	521.50	609.84
Nurse hours	158.57	26.37	52.57	275.25	158.14	190.50
Control variables:						
Age	61.03	19.86	0.00	105.00	63.00	85.00
Age sq.	4119.40	2298.30	0.00	11025.00	3969.00	7225.00
Male	0.53	0.50	0.00	1.00	1.00	1.00
Asian pat	0.05	0.23	0.00	1.00	0.00	0.00
Pacific pat	0.12	0.33	0.00	1.00	0.00	1.00
Indian pat	0.04	0.19	0.00	1.00	0.00	0.00
Maori pat	0.09	0.29	0.00	1.00	0.00	0.00
NZ Euro. pat	0.56	0.50	0.00	1.00	1.00	1.00
Deprivation scale	5.90	2.91	0.00	10.00	6.00	10.00
Transfer	0.12	0.33	0.00	1.00	0.00	1.00
Acute admiss.	0.73	0.45	0.00	1.00	1.00	1.00
Arranged admiss.	0.10	0.30	0.00	1.00	0.00	1.00
Admit 2006	0.16	0.36	0.00	1.00	0.00	1.00
Admit 2007	0.16	0.36	0.00	1.00	0.00	1.00
Admit 2008	0.16	0.37	0.00	1.00	0.00	1.00
Admit 2009	0.16	0.37	0.00	1.00	0.00	1.00
Admit 2010	0.17	0.37	0.00	1.00	0.00	1.00
Admit 2011	0.11	0.31	0.00	1.00	0.00	1.00
Myocardial Infarct.	0.06	0.24	0.00	1.00	0.00	0.00
Congestive Heart F	0.08	0.28	0.00	1.00	0.00	0.00
Periphral Vascular s	0.04	0.20	0.00	1.00	0.00	0.00
Cerebrovascular Dis	0.05	0.22	0.00	1.00	0.00	0.00
Dementia	0.03	0.18	0.00	1.00	0.00	0.00
Chronic Pulmonary D	0.06	0.24	0.00	1.00	0.00	0.00
Rheumatic Disease	0.01	0.09	0.00	1.00	0.00	0.00
Peptic Ulcer Disease	0.01	0.09	0.00	1.00	0.00	0.00
Mild Liver Disease	0.02	0.15	0.00	1.00	0.00	0.00
Diabetes w/o compl..	0.07	0.25	0.00	1.00	0.00	0.00
Diabetes w complic.	0.13	0.34	0.00	1.00	0.00	1.00

Paraplegia + Hemip a	0.03	0.18	0.00	1.00	0.00	0.00
Renal Disease	0.12	0.33	0.00	1.00	0.00	1.00
Cancer	0.19	0.39	0.00	1.00	0.00	1.00
Liver Disease	0.01	0.09	0.00	1.00	0.00	0.00
Metastatic Carcinoma	0.09	0.29	0.00	1.00	0.00	0.00
AIDS/HIV	0.00	0.06	0.00	1.00	0.00	0.00
Theatre event	0.33	0.47	0.00	1.00	0.00	1.00
Diagnosis count	6.82	4.54	1.00	65.00	6.00	13.00
Prev admiss 60days	0.30	0.46	0.00	1.00	0.00	1.00
AED entry	0.38	0.49	0.00	1.00	0.00	1.00
Accident	0.11	0.31	0.00	1.00	0.00	1.00
Auckland DHB	0.63	0.48	0.00	1.00	1.00	1.00
MDC 2	0.11	0.31	0.00	1.00	0.00	1.00
MDC 3	0.00	0.03	0.00	1.00	0.00	0.00
MDC 4	0.03	0.17	0.00	1.00	0.00	0.00
MDC 5	0.11	0.31	0.00	1.00	0.00	1.00
MDC 6	0.18	0.38	0.00	1.00	0.00	1.00
MDC 7	0.11	0.31	0.00	1.00	0.00	1.00
MDC 8	0.04	0.19	0.00	1.00	0.00	0.00
MDC 9	0.13	0.34	0.00	1.00	0.00	1.00
MDC 10	0.05	0.21	0.00	1.00	0.00	0.00
MDC 11	0.03	0.17	0.00	1.00	0.00	0.00
MDC 12	0.08	0.27	0.00	1.00	0.00	0.00
MDC 13	0.02	0.14	0.00	1.00	0.00	0.00
MDC 14	0.01	0.08	0.00	1.00	0.00	0.00
MDC 15	0.00	0.03	0.00	1.00	0.00	0.00
MDC 16	0.00	0.00	0.00	0.00	0.00	0.00
MDC 17	0.01	0.11	0.00	1.00	0.00	0.00
MDC 18	0.02	0.15	0.00	1.00	0.00	0.00
MDC 19	0.04	0.18	0.00	1.00	0.00	0.00
MDC 20	0.00	0.04	0.00	1.00	0.00	0.00
MDC 21	0.00	0.03	0.00	1.00	0.00	0.00
MDC 22	0.03	0.16	0.00	1.00	0.00	0.00
MDC 23	0.01	0.11	0.00	1.00	0.00	0.00
February	0.08	0.27	0.00	1.00	0.00	0.00
March	0.08	0.28	0.00	1.00	0.00	0.00
April	0.08	0.27	0.00	1.00	0.00	0.00
May	0.09	0.28	0.00	1.00	0.00	0.00
June	0.09	0.29	0.00	1.00	0.00	0.00
July	0.10	0.29	0.00	1.00	0.00	0.00
August	0.09	0.29	0.00	1.00	0.00	0.00
September	0.08	0.27	0.00	1.00	0.00	0.00
October	0.08	0.27	0.00	1.00	0.00	0.00
November	0.08	0.27	0.00	1.00	0.00	0.00
December	0.08	0.27	0.00	1.00	0.00	0.00
Monday admit	0.17	0.38	0.00	1.00	0.00	1.00
Tuesday admit	0.16	0.37	0.00	1.00	0.00	1.00
Wednesday admit	0.17	0.38	0.00	1.00	0.00	1.00
Thursday admit	0.17	0.37	0.00	1.00	0.00	1.00
Friday admit	0.14	0.35	0.00	1.00	0.00	1.00
Saturday admit	0.08	0.28	0.00	1.00	0.00	0.00
Flu	0.00	0.03	0.00	1.00	0.00	0.00

Extra LOS control						
Avg. LOS	5.74	4.40	1.00	43.40	4.45	11.68
Ward days	7.06	7.22	2.00	221.00	5.00	14.00
N	93382					

Notes: Summary statistics for variables in the mortality and length of stay models. Summarised for patients that stay in hospital greater than one night, and spend at least 80% of total hospital stay in the last ward of stay.

3.B Results: Mortality

Table 3.10: Mortality baseline model

	(1) base	(2) age<75
Patient hours	−0.00005	−0.00008*
	(−1.24)	(−2.15)
Nurse hours	0.00026	0.00042**
	(1.96)	(3.42)
Ward days	−0.00027	−0.00032
	(−0.53)	(−0.71)
Age	−0.00344***	−0.00106
	(−3.92)	(−0.80)
Age sq.	0.00004***	0.00002
	(7.03)	(1.56)
Male	0.01142**	0.00271
	(3.03)	(0.87)
Asian pat	0.00342	−0.00293
	(0.64)	(−0.28)
Pacific pat	−0.00944	−0.01309
	(−2.01)	(−1.96)
Indian pat	−0.01267	−0.02713**
	(−1.54)	(−3.09)
Maori pat	−0.00583	−0.01000
	(−0.86)	(−1.28)
NZ Euro. pat	−0.00649	−0.00810
	(−1.10)	(−0.98)
Deprivation scale	−0.00053	−0.00122
	(−0.79)	(−1.48)
Transfer	0.02066*	0.01641*
	(2.59)	(2.42)

Acute admiss.	0.05891**	0.04908**
	(3.79)	(3.66)
Arranged admiss.	−0.00083	−0.00839
	(−0.04)	(−0.38)
Admit 2006	0.00137	0.00551
	(0.15)	(0.74)
Admit 2007	−0.00872	0.00070
	(−1.14)	(0.10)
Admit 2008	−0.01200	−0.00572
	(−1.64)	(−0.60)
Admit 2009	−0.00652	−0.00241
	(−0.78)	(−0.29)
Admit 2010	−0.01107	−0.00851
	(−1.32)	(−1.21)
Admit 2011	−0.02063	−0.01267
	(−1.77)	(−1.15)
Myocardial Infarct.	0.03462*	0.02186
	(2.48)	(1.49)
Congestive Heart F	0.06160***	0.04516***
	(6.73)	(5.20)
Periphral Vascular Dis	0.02154*	0.01707
	(2.27)	(1.49)
Cerebrovascular Dis	0.02648	0.00398
	(1.40)	(0.19)
Dementia	0.07813***	0.07417**
	(8.32)	(2.91)
Chronic Pulmonary D	−0.01621	−0.02372
	(−1.30)	(−1.35)
Rheumatic Disease	−0.01390	−0.01685
	(−0.89)	(−1.29)
Peptic Ulcer Disease	−0.00149	0.02390
	(−0.07)	(1.36)
Mild Liver Disease	−0.02453	−0.02260
	(−1.46)	(−1.22)
Diabetes w/o complic.	−0.01892	−0.02166
	(−1.61)	(−1.82)
Diabetes w complic.	−0.04367***	−0.03820**
	(−5.87)	(−3.56)
Paraplegia + Hemiplegia	0.04883**	0.05384**
	(3.01)	(3.62)
Renal Disease	0.03982***	0.01878*

	(6.39)	(2.52)
Cancer	0.09491***	0.07409**
	(5.83)	(3.70)
Liver Disease	0.18543**	0.16714**
	(3.49)	(3.17)
Metastatic Carcinoma	0.21707***	0.20681***
	(8.62)	(10.28)
AIDS/HIV	0.00747	0.00169
	(0.27)	(0.06)
Theatre event	−0.05293***	−0.04039***
	(−6.03)	(−5.29)
Diagnosis count	0.00906***	0.00885***
	(8.58)	(5.47)
Prev admiss 60days	0.03322**	0.03985**
	(3.14)	(3.34)
AED entry	−0.00036	0.01203
	(−0.06)	(1.61)
Accident	−0.02633*	−0.02187*
	(−2.53)	(−2.16)
Auckland DHB	0.00806	0.00811
	(1.26)	(1.27)
MDC 2	−0.08266	−0.04574
	(−1.51)	(−0.45)
MDC 3	−0.04550	−0.05302
	(−1.26)	(−1.07)
MDC 4	0.00709	−0.01127
	(0.35)	(−0.56)
MDC 5	−0.05748	−0.04761
	(−1.87)	(−1.15)
MDC 6	−0.01970	−0.03068
	(−0.80)	(−0.93)
MDC 7	0.02208	0.01457
	(0.78)	(0.49)
MDC 8	−0.05253	−0.04974
	(−1.91)	(−1.21)
MDC 9	−0.05951*	−0.05694
	(−2.60)	(−1.83)
MDC 10	−0.03783	−0.05444
	(−0.99)	(−1.31)
MDC 11	−0.02396	−0.04353
	(−0.67)	(−0.92)

MDC 12	−0.05728	−0.07556
	(−1.07)	(−0.94)
MDC 13	−0.03573	−0.06870
	(−0.47)	(−0.91)
MDC 14	−0.03218	−0.04641
	(−1.14)	(−1.23)
MDC 16	−0.16438	−0.18211
	(−1.82)	(−1.76)
MDC 17	−0.04863	−0.07351
	(−1.56)	(−1.88)
MDC 18	−0.06279	−0.09189
	(−1.21)	(−1.49)
MDC 19	−0.16083**	−0.16891**
	(−3.27)	(−2.86)
MDC 20	−0.09326*	−0.13416**
	(−2.49)	(−2.92)
MDC 21	−0.07055*	−0.05950
	(−2.44)	(−1.43)
MDC 22	−0.09904*	−0.12357
	(−2.15)	(−1.90)
MDC 23	−0.02600	−0.02526
	(−0.87)	(−0.72)
February	0.01654*	0.01446
	(2.25)	(1.61)
March	0.00815	0.00851
	(1.16)	(0.93)
April	0.01397	0.00741
	(2.07)	(0.87)
May	0.00243	0.00346
	(0.31)	(0.41)
June	0.00848	0.01287
	(0.98)	(1.43)
July	0.00114	0.00082
	(0.10)	(0.07)
August	0.00267	0.00518
	(0.26)	(0.51)
September	0.00258	0.00828
	(0.46)	(1.49)
October	−0.00111	−0.00314
	(−0.15)	(−0.58)
November	−0.01480	−0.01591

	(−1.10)	(−1.14)
December	−0.00097	0.00095
	(−0.08)	(0.11)
Flu	0.00303	−0.01323
	(0.08)	(−0.36)
Observations	34727	21795
R^2	0.132	0.128

Notes: Outcome: 60-day mortality. Sample: Column (1) is all eligible study sample, Column (2) is inpatients under 75 years of age. Model: fixed-effect linear regression. Standard errors clustered on wards. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

3.C Results: Robustness of mortality

Table 3.11: Mortality robustness checks

	(1) <=30 mortality	(2) In-Hosp. mortality
Patient hours	−0.00004	−0.00006*
	(−1.12)	(−2.31)
Nurse hours	0.00026	0.00024
	(1.87)	(1.62)
Observations	34727	34727
R^2	0.107	0.057

Notes: Outcome: Column (1) has mortality within 30 days, Column (2) has in-hospital mortality. Sample: eligible study sample. Model: fixed-effect linear regression. Standard errors clustered on wards. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 3.12: Mortality robustness checks

	(1) all drg	(2) all los	(3) all ward stays
Patient hours	−0.00002	−0.00002	−0.00001
	(−0.67)	(−0.70)	(−0.57)
Nurse hours	0.00011	0.00019	0.00006
	(1.65)	(1.44)	(1.25)
Observations	91297	43395	137692
R^2	0.127	0.130	0.122

Notes: Outcome: 60-day mortality. Sample: eligible study sample. Model: linear regression, Column (1) has all DRGs, Column (2) has all Length of Stay, Column (3) has all patients that stayed in a ward for at least 60% of total hospital stay. Standard errors clustered on wards. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 3.13: Mortality robustness checks

	(1) no ward fe	(2) dummy ward
Patient hours	−0.00001	−0.00005
	(−0.23)	(−1.24)
	[1.32]	[3.45]
Nurse hours	−0.00014	0.00026
	(−0.91)	(1.96)
	[1.21]	[3.09]
Observations	34727	34727
R^2	0.181	0.185

Notes: Outcome: 60-day mortality. Sample: eligible study sample. Model: linear regression, Column (1) has no ward fixed-effect, Column (2) has dummy variables for wards and allows estimation of Variance Inflation Factors. Standard errors clustered on wards. Robust t-statistics in parentheses. VIF in brackets. ***p<.01; **p<.05; *p<.10.

3.D Results: LOS by ward

Table 3.14: LOS by hospital ward

	(1) Cancer a	(2) Cancer b	(3) Cardio a
Cumul. patient	0.00155*	0.000121	−0.000430**
	(2.41)	(0.32)	(−2.60)
Cumul. nurse	−0.00730***	−0.0000989	0.000458
	(−3.54)	(−0.07)	(0.49)
Male	−0.00740	−0.0576	0.0309
	(−0.17)	(−1.74)	(1.19)
Age	0.00749	0.0142*	0.00984**
	(1.07)	(2.43)	(2.79)

Age sq.	−0.000208**	−0.000212***	−0.000149***
	(−2.78)	(−3.88)	(−4.53)
Asian pat	−0.107	0.00888	−0.0122
	(−1.01)	(0.14)	(−0.20)
Pacific pat	−0.101	0.0379	−0.0270
	(−1.03)	(0.61)	(−0.49)
Indian pat	0.195	0.0397	−0.104
	(1.83)	(0.43)	(−1.47)
Maori pat	−0.222*	−0.0972	−0.0850
	(−2.03)	(−1.53)	(−1.67)
NZ Euro. pat	−0.0476	−0.0412	−0.0128
	(−0.63)	(−0.91)	(−0.34)
Deprivation scale	−0.00671	0.00522	−0.00405
	(−0.84)	(0.94)	(−0.84)
Transfer	−0.287**	−0.355***	−0.509***
	(−3.27)	(−6.71)	(−9.55)
Acute admiss.	−0.593	−0.720***	−0.546***
	(−1.45)	(−8.42)	(−11.99)
Arranged admiss.	−0.275	−0.321***	−0.257***
	(−0.67)	(−4.07)	(−5.87)
Theatre event	0.181	−0.116*	0.0289
	(1.84)	(−2.29)	(0.75)
Diagnosis count	−0.0708***	−0.0851***	−0.0707***
	(−12.65)	(−17.55)	(−14.36)
Avg. LOS	−0.0261***	−0.0622***	−0.0165***
	(−8.37)	(−9.39)	(−4.48)
Prev ad 60days	0.00768	−0.167***	−0.0421
	(0.14)	(−5.41)	(−1.33)
AED entry	−0.253*	−0.426***	−0.0317
	(−2.39)	(−6.75)	(−0.91)
Accident	−0.894*	−0.414	0.00265
	(−2.08)	(−0.98)	(0.04)
Auckland DHB	0.145**	−0.00994	0.130***
	(3.10)	(−0.29)	(3.83)
Admit 2006	0.166	−0.0969	−0.0569
	(1.73)	(−1.48)	(−0.93)
Admit 2007	0.241*	−0.0421	0.0376
	(2.48)	(−0.59)	(0.55)
Admit 2008	0.142	0.0276	−0.0179
	(1.46)	(0.39)	(−0.24)
Admit 2009	−0.0635	−0.0444	0.0594

	(−0.66)	(−0.58)	(0.97)
Admit 2010	0.0667	−0.0433	−0.0673
	(0.60)	(−0.56)	(−0.96)
Admit 2011	0.233*	0.0597	−0.00177
	(2.12)	(0.73)	(−0.02)
Monday	0.129	0.0467	0.199***
	(1.49)	(0.75)	(3.38)
Tuesday	−0.00903	0.241***	0.355***
	(−0.10)	(4.17)	(6.07)
Wednesday	0.125	0.125*	0.179**
	(1.39)	(2.02)	(2.80)
Thursday	0.321***	0.209***	0.319***
	(3.69)	(3.39)	(5.21)
Friday	−0.214*	−0.338***	0.253***
	(−2.25)	(−4.73)	(3.99)
Saturday	−0.281**	−0.549***	−0.730***
	(−3.01)	(−7.54)	(−9.05)
AMI.	−0.112	0.0102	0.138***
	(−0.25)	(0.04)	(3.55)
Congestive Heart	−0.450*	0.0477	−0.155***
	(−2.55)	(0.30)	(−4.87)
Periphral Vascular	0.0883	−0.103	0.0127
	(0.59)	(−0.62)	(0.25)
Cerebrovascular	−0.704*	−0.326	−0.222
	(−2.51)	(−1.43)	(−1.62)
Dementia	0.586***	−0.482	−0.862**
	(3.43)	(−1.13)	(−3.15)
Chronic Pulmonar	−0.185	0.109	−0.107
	(−1.01)	(1.05)	(−1.62)
Rheumatic Dis	0.280	0.502	−0.0538
	(0.89)	(0.72)	(−0.26)
Peptic Ulcer	−0.301	−0.246	−0.00653
	(−0.52)	(−0.97)	(−0.04)
Mild Liver Dis	0.212	−0.113	−0.0609
	(1.69)	(−0.55)	(−0.41)
Diabetes w/o com.	−0.00305	0.0401	0.0716
	(−0.03)	(0.59)	(1.35)
Diabetes w com	0.281*	0.392***	0.235***
	(2.27)	(6.15)	(5.66)
Paraplegia	−0.679	−0.703***	−0.460
	(−1.69)	(−5.03)	(−1.79)

Renal Dis	−0.296*	0.0779	−0.145**
	(−2.15)	(0.85)	(−3.24)
Cancer	−0.270***	−0.109	−0.106
	(−3.46)	(−1.92)	(−0.56)
Liver Dis	−0.696	−2.330***	−0.670
	(−1.55)	(−4.70)	(−1.25)
Metastatic Carc	−0.0918	−0.337***	−0.105
	(−0.49)	(−9.71)	(−0.41)
AIDS/HIV	−0.235	0.952**	−0.563***
	(−0.85)	(2.85)	(−6.63)
Observations	26878	42516	23822

Notes: Outcome: hazard rate for the time-to-discharge home. Sample: eligible study sample in ward Cancer a (Column (1)) and Cancer b (Column (2)) and Cardio a (Column (3)). Model: competing risk survival model. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 3.15: LOS by hospital ward

	(1) Gen med c	(2) Gen med d	(3) Gen surg a
Cumul. patient	−0.00157***	−0.0000446	−0.000921
	(−3.55)	(−0.07)	(−0.92)
Cumul. nurse	−0.00113	−0.00000644	−0.000763
	(−0.78)	(−0.00)	(−0.59)
Observations	43113	39136	32641

Notes: Outcome: hazard rate for the time-to-discharge home. Sample: eligible study sample in ward Gen med c (Column (1)) and Gen med d (Column (2)) and Gen surg a (Column (3)). Model: competing risk survival model. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 3.16: LOS by hospital ward

	(1) Gen surg b	(2) ENT	(3) Kidney
Cumul. patient	−0.00151*	0.0000317	−0.000141
	(−2.22)	(0.11)	(−0.33)
Cumul. nurse	−0.00125	−0.00107	−0.00142
	(−0.91)	(−1.26)	(−1.37)
Observations	36548	25666	34302

Notes: Outcome: hazard rate for the time-to-discharge home. Sample: eligible study sample in ward Gen surg b (Column (1)) and ENT (Column (2)) and Kidney (Column (3)). Model: competing risk survival model. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 3.17: LOS by hospital ward

	(1) Neuro a	(2) Neuro b	(3) Ortho a
Cumul. patient	−0.000963*	−0.0000445	−0.000703
	(−2.29)	(−0.09)	(−1.82)
Cumul. nurse	−0.00225	−0.00542***	0.00277
	(−1.93)	(−3.52)	(1.95)
Observations	27448	30280	45826

Notes: Outcome: hazard rate for the time-to-discharge home. Sample: eligible study sample in ward Neuro a (Column (1)) and Neuro b (Column (2)) and Ortho a (Column (3)). Model: competing risk survival model. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 3.18: LOS by hospital ward

	(1) Ortho b	(2) Resp	(3) Urology
Cumul. patient	−0.00168**	−0.000927*	0.000490*
	(−2.59)	(−2.38)	(1.99)
Cumul. nurse	0.00110	−0.000567	−0.00191*
	(0.65)	(−0.43)	(−2.01)
Observations	41039	29904	32325

Notes: Outcome: hazard rate for the time-to-discharge home. Sample: eligible study sample in ward Ortho b (Column (1)) and Resp (Column (2)) and Urology (Column (3)). Model: competing risk survival model. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 3.19: LOS by hospital ward

	(1) Vascular	(2) Cardio b	(3) Coronary
Cumul. patient	−0.00123*	−0.000340	−0.00161***
	(−2.37)	(−0.85)	(−5.17)
Cumul. nurse	0.00541**	0.00290	−0.000390
	(3.21)	(1.69)	(−0.35)
Observations	24905	12429	19760

Notes: Outcome: hazard rate for the time-to-discharge home. Sample: eligible study sample in ward Vascular (Column (1)) and Cardio b (Column (2)) and Coronary (Column (3)). Model: competing risk survival model. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

Table 3.20: LOS by hospital ward

	(1) Gen med a	(2) Gen med b
Cumul. patient	−0.000404	−0.000945
	(−0.88)	(−1.37)
Cumul. nurse	−0.00129	0.00362*
	(−0.95)	(2.49)
Observations	39059	34036

Notes: Outcome: hazard rate for the time-to-discharge home. Sample: eligible study sample in ward Gen med a (Column (1)) and Gen med b (Column (2)). Model: competing risk survival model. Robust t-statistics in parentheses. ***p<.01; **p<.05; *p<.10.

3.E Survival modelling

Survival analysis

This section provides information on survival models and explains the model used in my paper.

Survival or duration analysis models the time to an event of interest. Or, the time from when a person becomes at risk of event k up until event k occurs. A hazard function models the duration of a subject (e.g. firm or individual). The hazard function estimates the probability an individual will 'fail' in a given time interval $t + \triangle t$, conditional on the individual surviving up to time t. As the interval is reduced, the hazard function gives the instantaneous probability of failure at any given time t. The corresponding survival function is the probability an individual has not failed by time t.

Survival analysis is commonly used for single-outcome events, where a subject is at risk of one event e.g. death, employment, exit from a market etc. My modelling situation has competing risks of different discharge types. For competing risks, there is more than one possible outcome for a subject. A patient's length of hospital stay could terminate for a variety of reasons, including; discharge home, death or transfer to another health-care facility. Discharge home is the main outcome of interest. It is problematic to consider discharge home and transfer as the same outcome, because a

transfer indicates more nursing or health-care is needed. The discharge home decision is interpreted as a patient being well enough to return home. In competing risk models, distributions of failure time (T) for each risk have to be modelled separately or jointly.

Competing risk models

Methods for estimating competing risk models differ in Biostatistics and Econometric literatures. Both approaches are discussed, with a focus on time-varying covariates. Time-varying covariates are nursing and patient hours during a patient's stay in hospital.

Biostatistics

In Biostatistics, competing risks are modelled via the Cumulative Incidence Function (CIF) of a cause or event k. The CIF is derived from cause-specific hazards. The cause-specific hazard, denoted $\lambda_k(t)$, is the hazard function corresponding to each of the competing events (Putter et al., 2007, pg.2398). This is defined as:

$$\lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr ob(t \leq T \prec t + \Delta t, D = k | T \geq t)}{\Delta t} \tag{3.9}$$

Where D is the cause of failure and T is the failure time (Putter et al., 2007, pg.2398). The survival function is defined as the probability a subject survives (i.e not fail) up to time t. In single event data, the survivor function is the integral of the hazard function (from $t = 0$ to t). Following this, the cause-specific survivor function $S_k(t)$ would correspond to $S_k(t) = \exp(-\Lambda_k(t))$ where $\Lambda_k(t) = \int_0^t \lambda_k(s)ds$.

The cause-specific survivor function does not account for the occurrence of competing events. A subject is not always at risk of failing to event k because they could fail to a competing event. That is, $S_k(t)$ ignores individuals that have experienced a

competing event, these individuals are still considered at risk of event k, as a result we over-estimate the probability of failing from cause k, since we think more individuals are at risk from event k than there actually are. This consequently overestimates the probability of failure and underestimates survival from cause k, because individuals may survive from cause k but experience a competing event.

Instead of the survival function, competing risk models specify a cumulative incidence function. The cumulative incidence function (CIF) for event k is defined as the probability of failure from event k before time t, given there are competing risks. This is a function the cause-specific hazard for event k and the probability of survival from all competing events ($S(s)$):

$$I_k(t) = \int_0^t \lambda_k(s) S(s) ds \tag{3.10}$$

Where $\lambda_k(s)$ is the cause-specific hazard and $S(s) = \exp(-\sum_{k=1}^{K} \Lambda_k(t))$ is the summation of the cause-specific survivor functions. The CIF is an alternative measure of the survival function that takes into account the effect of competing risks on the probability of survival from an event as a function of time.

There are two approaches to modelling the CIF in competing risk models in the Biostatistics literature. The first approach is to model each cause-specific hazard separately and then compute an estimate of the cumulative incidence function based on formula 3.10. The other approach is to model the cumulative incidence function directly, i.e. without modelling cause-specific hazard functions individually. The cumulative incidence function is also referred to as the sub-distribution or 'crude' survival function in the literature.

In the first approach, different functional forms can be used to model each cause-specific hazard separately. Examples of functional forms for hazard models are Cox proportional hazard models and additive hazard models. In the additive model, the

hazard function is specified as the sum of a baseline hazard function and another function of covariates and parameters. In Cox's proportional hazard model, the hazard function is the multiplication of a baseline hazard and an exponential function of covariates and parameters. Additive models have the advantage of not specifying proportional effects of covariates on the hazard. Buckley (1984) discusses the advantages and disadvantages of multiplicative and additive models. Andersen (1993) combines Cox's proportional hazard model and Aalen's additive hazards model to estimate each cause-specific hazard function before computing a measure of the CIF (this is also applied in Scheike and Zhang (2003). Shen and Cheng (1999) use the additive risk model as an alternative to the more commonly used Cox's proportional hazards model.

A disadvantage of modelling each cause-specific hazard separately is that the effect of the covariates may differ on the cause-specific hazard compared to the CIF. Because the cause-specific hazard is modelled first, the covariate effects may not follow through when aggregating to the CIF (Fine and Gray, 1999). For example, a covariate may be significantly related to the cause-specific hazard of event k but have little effect on the cumulative incidence function. In the CIF, any effect of covariates is on event k, and competing events. This is because the survivor function in the CIF is an aggregate of the other events' survival functions.

Based on this critique, Fine and Gray (1999) propose a widely used estimation method for modelling the CIF directly. Fine and Gray model the CIF of a cause k at time t with covariates X. Covariates can be time-varying. To estimate the CIF, they specify a sub-distribution hazard formula. Their sub-distribution hazard for event k specifies that individuals who have failed from a previous event $j \neq k$ are still in the risk set:

$$\lambda_k(t;X) = \lim_{\Delta t \downarrow 0} \frac{\Pr ob\{t \leq T \leq t + \Delta t, D = k | T \geq t \cup (T \leq t \cap D \neq k), X\}}{\Delta t} \quad (3.11)$$

This specifies the probability an individual will fail to event k ($D = k$) in the interval $t \leq T \leq t + \Delta t$, given the risk set. The risk set includes individuals who have not experienced cause k at t, or do experience k at time t ($T \leq t$), and individuals that experienced a competing event ($T \leq t \cap D \neq k$). The sub-distribution hazard can also be formulated as a function of the CIF for event k: $\lambda_k(t;X) = -d \log\{1 - I_k(t)\}/dt$. The CIF is the term used for the survival function in competing risk models, (remembering the survival function is probability of not having failed by time t). The CIF is composed of the cause-specific hazards for all events. Because the CIF enters the sub-distribution hazard for event k, the hazard is not independent of event $j \neq k$ hazards, as they are in modelling cause-specific hazards directly.

Fine and Gray include individuals in the risk set who have failed from a competing cause to slow down the hazard rate. The risk set 'should not be interpreted in the way that an individual who has failed from the competing cause is considered to still be able to fail from the cause of interest. Rather, the additional conditioning on [risk set with subjects who fail from a competing cause] should be understood as a way to inflate the risk set, slowing down the SH [sub-distribution hazard], such that the distribution of the attached failure time equals the CIF on the real line' (Beyersmann and Schumacher, 2008, pg. 771).

(Fine and Gray, 1999) specify the sub-distribution hazard as a Cox proportional hazards model: $\lambda_k(t;X) = \lambda_{ko}(t) \exp\{X^T(t)\beta_0\}$. Accordingly the cumulative incidence function of cause k is: $I_k(t,X) = 1 - \exp[-\int_0^t \lambda_{ko}(s) \exp\{X^T(s)\beta_0 ds]$. The likelihood function estimated to solve for β, the covariate effects, is:

$L(\beta) = \Pi_{i=1}^n \left[\frac{\exp\{Z_i^T(T_i)\beta\}}{\sum_{j \in R_i} \exp\{Z_j^T(T_i)\beta\}} \right]^{I(\varepsilon_i = 1)}$

This likelihood function is partial because it is over the risk set of individuals at a time t, that is, the individuals who have not failed from cause k by time t, rather than the risk set of all individuals in the dataset. The use of partial likelihoods originates in Cox (1975), who proved that partial likelihoods were equivalent asymptotically with maximum likelihood methods.

The key difference of Fine and Gray's method to the Cox proportional hazard model for single-outcomes, is firstly a risk set where individuals who have already failed from a competing cause by time t remain in the risk set. Secondly, Fine and Gray specify weights for observations that are censored. Censored observations occur when the outcome for a subject is not observed. There is no censoring in this paper. The weighting function therefore does not apply to the calculation of the likelihood in this paper.

Time-varying regressors and the Fine and Gray model

In theory, the Fine and Gray model can incorporate time-varying regressors. There is however uncertainty in the empirical literature about how to include and interpret time-varying covariates in the competing risk model. Cortese and Andersen (2010, pg. 138) writes: 'In spite of the frequent use of time-dependent covariates, their role in regression modelling (of competing risks) and prediction is, however, still unclear due to both interpretation and practical problems'. In Beyersmann and Schumacher (2008, pg. 765), 'Mathematically, the model (by Fine and Gray) also allows for including random time-dependent covariates, but practical implementation has remained unclear'.

Time-varying covariates are also referred to as time-dependent covariates in the biostatistics literature, following Prentice et al. (1978). The uncertainty mainly arises from the use of 'internal' time-dependent covariates. The problems arising from using internal time-dependent covariates are not unique to competing risk or specifically

Fine and Gray's model of the CIF. The distinction between internal and external covariates has been put forward by Prentice et al. (1978). External covariates satisfy the following condition, for all u, t such that $0 \prec u \leq t$. u, t are points in time, T is failure time, $X(u)$ are covariate values at time u:

$$P[T \in [u, u + du)|X(u), T \geq u] = P[T \in [u, u + du)|X(t), T \geq u\} \tag{3.12}$$

Or equivalently, condition 3.12 holds for $0 \prec u \leq t$:

$$P[X(t)|X(u), T \geq u] = P[X(t)|X(u), T = u] \tag{3.13}$$

Equation 3.13 says that the probability of an individual failing in an interval is the same when conditioning the covariate process over a period of time greater than u, as conditioning the covariate process only up to u. In other words, future values of the time-dependent covariates do not help to predict failure at time t. Or, adding more information about the future path of time-dependent covariates does not help predict an individual's failure at time t.

Prentice et al. (1978) give two examples of external covariates. The first type, is when a covariate is fixed i.e. $x(t) = x$. A second example is when the covariate path is defined, e.g. a treatment regime that is fixed before the study and does not respond to changes in an individual's condition. A third type is when a covariate path is determined by processes external to the individual. This is referred to as an ancillary process and 'a covariate of this sort is the output of a stochastic process that is external to the individual under study and whose probability laws do not involve the parameters in the failure time model under study' (Prentice et al., 1978, pg. 197).

Covariates that do not meet condition 3.13 are called internal covariates. For example, measures on an individual during a study such as health status, blood pressure, infection in an ICU stay and so on. For example, a future blood pressure values could help predict failure at t. The problem with internal covariates is that the direct relation between the hazard and survival function does not hold. For internal covariates, survival up to t depends on more than the hazard rate up to time t, it also depends on future values of covariates in the model (Beyersmann et al., 2009). The key issue is the ability to condition on the covariate path up to time t in the hazard function (probability of failure) so that the direct relation to the survivor function (probability of survival) is maintained. That is, the stochastic covariate path needs to have a distribution in order to allow full modelling of the survivor function.

Wolkewitz et al. (2009) provide a solution to the problem of interpreting and using time-dependent covariates in the Fine and Gray model. They propose using the last recorded value of the time-dependent covariate at the time of failure of an individual, when an individual remains in a risk pool after failure from a competing risk. In *Stata 11*, the last recorded value of the time-dependent covariate is used for future calculations. For this reason, the cumulative average nursing hours is used so that it reflects the history of nursing inputs up to a time t.

Econometrics

In econometric approaches, the majority of competing risk models use the Mixed Proportional Hazards (MPH) framework. The proportional hazard models are mixed because they have an added heterogeneity term. The Mixed Proportional Hazard Model: $\theta_k(tX = x, V) = \psi(t)\exp(x'\beta)V$. Where θ_k is the hazard rate for event k at time t, given covariates X and unobserved heterogeneity term V. This follows the semi-parametric proportional hazard assumption by leaving $\psi(t)$ unspecified function of t and covariate effects enter through $\exp(x'\beta)$.

Following the inclusion of heterogeneity terms by econometricians, much work in econometric competing risk model is in how to model the distribution of the heterogeneity terms in the likelihood function to enable maximisation. The distribution of the heterogeneity term can also allow dependence between the failure time (V) distributions for each of the competing risks.

Applied competing risk methods

Econometric methods have assumed either independence or dependence of the failure time distributions when modelling competing risks. If independence is assumed, k hazard functions are combined into a likelihood function for maximisation. To maximise a likelihood function composed of MPHs with unobserved heterogeneity terms, an assumption needs to be made about the distribution of the unobserved terms. For example, a parametric distribution (e.g. gamma) or the distribution can be non-parametrically estimated (Heckman and Singer, 1984).

If distributions of the failure times for k competing risks are dependent, then researchers commonly introduce dependence by specifying a relationship between the unobserved heterogeneity terms (of the k competing risks). For example, Picone et al. (2003) uses common factor loading to model the relationship between two unobserved heterogeneity terms that enter into one of four hazard functions for competing risks. Picone et al. (2003) models the length of hospital stay decision with competing risks of; discharge to home, skilled nursing facility, home health agency and mortality. Picone et al. models the hazard rate of discharge home with a log-logistic function. A Weibull function is used for the remaining hazard rates: discharge to home health agency, skilled nursing facility or death in hospital. Picone uses the methods of Heckman and Singer to estimate the distribution of the unobserved heterogeneity term in the likelihood function. Estimation of the distribution involves specifying the number of values the heterogeneity terms can take and estimating weights for each

of these values. This is done in the maximisation of the likelihood function. Lastly, a common factor loading relationship of the heterogeneity terms is specified to model dependence between the hazard functions (i.e. $v(1) = pv(2)$).

Sueyoshi (1992) proposes a discrete-time competing risk model with time-varying covariates. Seuyoshi assumes time-varying covariates are constant within an interval, that is, covariates are recorded in discrete time intervals. This assumption allows the formation of a log-likelihood that is a summation over the number of observations, time periods, and competing risks. Therefore, the framework is similar to a multinomial outcome with categories defined over the k competing events and T failure times. Applications of competing risk models with time-varying covariates follow this discrete time framework. For example, Deng et al. (2000) model mortgage terminations with competing risks of prepayment or default. Their time-varying covariates are the interest rate and property values.

3.E.1 Model selection

There are several advantages of Fine and Gray's competing risk model.

Firstly, it imposes few assumptions, except for the proportionality of variables on the hazard rate over time. Econometric models impose parametric assumptions on the failure time distribution (e.g. Weibull or Gamma etc), and distribution of the heterogeneity term. The main assumption for the proportional hazard model is that the hazards for any two people, with their respective covariate vectors, have the same ratio through all t (Jenkins, 2005). Specifically, the effect of a covariate on the hazard rate is constant over time. There has been criticism of this assumption, for example, it may be the case that the effect of a regressor on the hazard function changes with time, such as medical treatments taking time to materialise in a patient. In addition, econometric approaches also specify a proportional hazard assumption.

Secondly, there is a program in *Stata 11* to estimate the Fine and Gray model. There is no program for estimating competing risk models using econometric approaches.

Jenkins (2005) does however propose steps for estimating a discrete-time competing risk model using the multinomial logit command. To set up the model, data is organised into subject-time form. For example, in a hospital this would be patient-day form so there is a record for each day a patient is in hospital. A dependent variable is coded with 0 for censored (patient remains in hospital), 1 for discharge home, 2 for mortality and so on. So each patient episode will end in one of the destinations, and all previous periods for that patient will have a censored (0 outcome) dependent variable. If the base category is the censored category, then the effect of variables on exit to other destinations are estimated relative to the censoring outcome (staying in hospital). The multinomial logit model is however only valid, without further assumptions, when time is intrinsically discrete. That is, there is no underlying process in continuous time that the discrete time is measuring. In our case, we have interval censored data which arises from having data at an interval, daily, weekly, monthly, on an underlying process in continuous time, because patients are discharged at a particular time of day. For interval censored data, the application of a multinomial logit model is less straightforward because the probability of exiting to each of the states is not separable in the likelihood function without additional assumptions. A multinomial based model can be estimated on interval censored data under the assumption that the hazard is constant within intervals.

Multinomial models were also attempted in my study. These models were computationally intensive. There are also no clear guidelines for how to model discrete-time competing risk models with available packages, such as the 'glamm' package. Haynes (2008) uses individual fixed-effects in a panel data model (individuals over time) using multinomial outcomes for employment transitions to full and part time employment.

A disadvantage of the Cox proportional hazard model is that it is more appropriate for settings with continuous time data. This is because it orders failure times and calculates the ratio of the hazards at each step. When there are a lot of tied failure times, approximation methods are used by *Stata 11* to select the relevant risk set at each step. Specifically, *Stata 11* uses Breslow's method for tied failure times in Fine and Gray's model. In this method, the risk set at each day in the data will include all individuals who failed on that day, rather than ordering failures by time of day so that the risk pool decreases throughout a day. This results in a large risk set in which individuals who fail are compared with those who are still at risk. There is no further option for specifying how *Stata 11* should deal with tied data methods in the competing risk model.

Bibliography

Aiken, L., Clarke, S., Sloane, D., Sochalski, J., and Silber, J. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA: the journal of the American Medical Association*, 288(16):1987–1993.

Andersen, P. (1993). *Statistical models based on counting processes*. Springer Verlag.

Anderson, G. (2004). Making sense of rising caesarean section rates. *BMJ*, 329(7468):696–697.

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *The American economic review*, 53(5):941–973.

Balsa, A. and McGuire, T. (2001). Statistical discrimination in health care. *Journal of Health Economics*, 20(6):881–907.

Balsa, A. and McGuire, T. (2003). Prejudice, clinical uncertainty and stereotyping as sources of health disparities. *Journal of Health Economics*, 22(1):89–116.

Balsa, A., McGuire, T., and Meredith, L. (2005). Testing for statistical discrimination in health care. *Health Services Research*, 40(1):227–252.

Berwick, D. and Hackbarth, A. (2012). Eliminating waste in us health care. *JAMA: The Journal of the American Medical Association*, 307(14):1513–1516.

Beyersmann, J., Latouche, A., Buchholz, A., and Schumacher, M. (2009). Simulating competing risks data in survival analysis. *Statistics in Medicine*, 28(6):956–971.

Beyersmann, J. and Schumacher, M. (2008). Time-dependent covariates in the proportional subdistribution hazards model for competing risks. *Biostatistics*, 9(4):765–776.

Bhattacharya, J., Goldman, D., and McCaffrey, D. (2006). Estimating probit models with self-selected treatments. *Statistics in medicine*, 25(3):389–413.

Braveman, P., Egerter, S., Edmonston, F., and Verdon, M. (1995). Racial/ethnic differences in the likelihood of cesarean delivery, california. *American journal of public health*, 85(5):625–630.

Buckley, J. (1984). Additive and multiplicative models for relative survival rates. *Biometrics*, 40(1):51.

Cassel, C. K. and Guest, J. A. (2012). Choosing wisely helping physicians and patients make smart decisions about their care. *JAMA: The Journal of the American Medical Association*, 307(17):1801–1802.

Cassidy, K., Kelly, M., et al. (1999). Inferring gender from name phonology. *Journal of Experimental Psychology: General*, 128(3):362.

Chandra, A. and Staiger, D. (2010). Identifying provider prejudice in healthcare. Technical report, National Bureau of Economic Research.

Chen, J., Rathore, S., Radford, M., Wang, Y., and Krumholz, H. (2001). Racial differences in the use of cardiac catheterization after acute myocardial infarction. *New England Journal of Medicine*, 344(19):1443–1449.

Cook, A., Gaynor, M., Stephens Jr, M., and Taylor, L. (2010). The effect of hospital nurse staffing on patient health outcomes: Evidence from california's minimum staffing regulation. Technical report, National Bureau of Economic Research.

Cooper-Patrick, L., Gallo, J., Gonzales, J., Vu, H., Powe, N., Nelson, C., and Ford, D. (1999). Race, gender, and partnership in the patient-physician relationship. *JAMA: the journal of the American Medical Association*, 282(6):583–589.

Cortese, G. and Andersen, P. (2010). Competing risks and time-dependent covariates. *Biometrical Journal*, 52(1):138–158.

Cox, D. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Currie, J. and Gruber, J. (1997). The technology of birth: Health insurance, medical interventions, and infant health. Technical report, National Bureau of Economic Research.

Declercq, E., Young, R., Cabral, H., and Ecker, J. (2011). Is a rising cesarean delivery rate inevitable? trends in industrialized countries, 1987 to 2007. *Birth*, 38(2):99–104.

Delarue, A., Van Hootegem, G., Procter, S., and Burridge, M. (2008). Teamworking and organizational performance: A review of survey-based research. *International Journal of Management Reviews*, 10(2):127–148.

Deng, Y., Quigley, J., and Order, R. (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica*, 68(2):275–307.

Evans, W. and Kim, B. (2006). Patient outcomes when hospitals experience a surge in admissions. *Journal of Health Economics*, 25(2):365–388.

Fine, J. and Gray, R. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, pages 496–509.

Fiscella, K. and Fremont, A. M. (2006). Use of geocoding and surname analysis to estimate race and ethnicity. *Health services research*, 41(4p1):1482–1500.

Galvao, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, 164(1):142–157.

Getahun, D., Strickland, D., Lawrence, J., Fassett, M., Koebnick, C., and Jacobsen, S. (2009). Racial and ethnic disparities in the trends in primary cesarean delivery based on indications. *American journal of obstetrics and gynecology*, 201(4):422–e1.

Gibbons, L., Belizán, J., Lauer, J., Betrán, A., Merialdi, M., and Althabe, F. (2010). The global numbers and costs of additionally needed and unnecessary caesarean sections performed per year: overuse as a barrier to universal coverage. *World health report*.

Godager, G. (2012). Birds of a feather flock together: a study of doctor-patient matching. *Journal of health economics*, 31(1):296.

Gordon, H., Street Jr, R., Sharf, B., Kelly, P., and Souchek, J. (2006). Racial differences in trust and lung cancer patients' perceptions of physician communication. *Journal of Clinical Oncology*, 24(6):904–909.

Grant, D. (2009). Physician financial incentives and cesarean delivery: new conclusions from the healthcare cost and utilization project. *Journal of health economics*, 28(1):244–250.

Gruber, J. and Kleiner, S. (2010). Do strikes kill? evidence from new york state. Technical report, National Bureau of Economic Research.

Grytten, J., Skau, I., and Sørensen, R. (2011). Do expert patients get better treatment than others? agency discrimination and statistical discrimination in obstetrics. *Journal of Health Economics*, 30(1):163–180.

Hamilton, B., Nickerson, J., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3):465–497.

Harris, R., Robson, B., Curtis, E., Purdie, G., Cormack, D., and Reid, P. (2007). Māori and non-māori differences in caesarean section rates: a national review. *Journal of the New Zealand Medical Association*, 120(1250).

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, pages 271–320.

Hill, S. and Miller, G. (2010). Health expenditure estimation and functional form: applications of the generalized gamma and extended estimating equations models. *Health economics*, 19(5):608–627.

Holmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, 13(2):324–340.

Ibison, J. (2005). Ethnicity and mode of delivery in low-riskfirst-time mothers, east london, 1988–1997. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 118(2):199–205.

Jenkins, S. (2005). Survival analysis. Technical report, Institute for Social and Economic Research, University of Essex, Colchester, UK. Downloadable from http://www.iser. essex. ac. uk/teaching/degree/stephenj/ec968/pdfs/ec968lnotesv6.pdf.

Jones, A. (2011). *Oxford Handbook of Economic Forecasting*, chapter Models for health care, pages 625–654. Oxford, Oxford University Press.

Kane, R., Shamliyan, T., Mueller, C., Duval, S., and Wilt, T. (2007). The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. *Medical care*, 45(12):1195.

Kaplan, S., Gandek, B., Greenfield, S., Rogers, W., and Ware, J. (1995). Patient and visit characteristics related to physicians' participatory decision-making style. results from the medical outcomes study. *Medical care*, 33(12):1176.

King, W., Wong, M., Shapiro, M., Landon, B., and Cunningham, W. (2004). Does racial concordance between hiv-positive patients and their physicians affect the time to receipt of protease inhibitors? *Journal of General Internal Medicine*, 19(11):1146–1153.

Knight, M. and Sullivan, E. (2010). Variation in caesarean delivery rates. *BMJ (Clinical research ed.)*, 341:c5255.

Korenstein, D., Falk, R., Howell, E., Bishop, T., and Keyhani, S. (2012). Overuse of health care services in the united states: An understudied problem. *Archives of Internal Medicine*, 172(2):171.

Lang, T., Hodge, M., Olson, V., Romano, P., and Kravitz, R. (2004). Nurse-patient ratios: a systematic review on the effects of nurse staffing on patient, nurse employee, and hospital outcomes. *Journal of Nursing Administration*, 34(7-8):326.

LaVeist, T. and Nuru-Jeter, A. (2002). Is doctor-patient race concordance associated with greater satisfaction with care? *Journal of Health and Social Behavior*, pages 296–306.

Menacker, F., Hamilton, B., and for Health Statistics (US), N. C. (2010). *Recent trends in cesarean delivery in the United States*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Ministry of Health, N. Z. (2011). *WIESNZ11 Cost Weights*. Retrieved from http://www.health.govt.nz/nz-health-statistics/data-references/weighted-inlier-equivalent-separations/wiesnz11-cost-weights.

Ministry of Health, N. Z. (2012). *Report on Maternity, 2010.* Retrieved from http://www.health.govt.nz/publication/report-maternity-2010.

Mooney, G. and Ryan, M. (1993). Agency in health care: getting beyond first principles. *Journal of Health Economics*, 12(2):125–135.

Needleman, J., Buerhaus, P., Mattke, S., Stewart, M., and Zelevinsky, K. (2002). Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine*, 346(22):1715–1722.

Needleman, J., Buerhaus, P., Pankratz, V., Leibson, C., Stevens, S., and Harris, M. (2011). Nurse staffing and inpatient hospital mortality. *New England Journal of Medicine*, 364(11):1037–1045.

Picone, G., Sloan, F., Chou, S., and Taylor Jr, D. (2003). Does higher hospital cost imply higher quality of care? *Review of Economics and Statistics*, 85(1):51–62.

Prentice, R., Kalbfleisch, J., Peterson Jr, A., Flournoy, N., Farewell, V., and Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4):541.

Putter, H., Fiocco, M., and Geskus, R. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430.

Ratto, M., Propper, C., and Burgess, S. (2002). Using financial incentives to promote teamwork in health care. *Journal of health services research & policy*, 7(2):69–70.

Rumball-Smith, J. (2009). Not in my hospital? ethnic disparities in quality of hospital care in new zealand. a narrative review of the evidence. *Journal of the New Zealand Medical Association. NZMJ*, 122(1297).

Sandhu, H., Adams, A., Singleton, L., Clark-Carter, D., and Kidd, J. (2009). The impact of gender dyads on doctor–patient communication: A systematic review. *Patient Education and Counseling*, 76(3):348–355.

Scheike, T. and Zhang, M. (2003). Extensions and applications of the cox-aalen survival model. *Biometrics*, 59(4):1036–1045.

Schoenfeld, D. (2005). Survival methods, including those using competing risk analysis, are not appropriate for intensive care unit outcome studies. *Critical Care*, 10(1):103.

Scott, A. and Vick, S. (1999). Patients, doctors and contracts: An application of principal-agent theory to the doctor-patient relationship. *Scottish Journal of Political Economy*, 46(2):111–134.

Shen, Y. and Cheng, S. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*, 55(4):1093–1100.

Sochalski, J., Konetzka, R., Zhu, J., and Volpp, K. (2008). Will mandated minimum nurse staffing ratios lead to better patient outcomes? *Medical care*, 46(6):606.

Spetz, J., Donaldson, N., Aydin, C., and Brown, D. (2008). How many nurses per patient? measurements of nurse staffing in health services research. *Health services research*, 43(5p1):1674–1692.

Stepanikova, I., Mollborn, S., Cook, K., Thom, D., and Kramer, R. (2006). Patients' race, ethnicity, language, and trust in a physician. *Journal of health and social behavior*, 47(4):390–405.

Stock, J. and Watson, M. (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*, 76(1):155–174.

Street Jr, R., Gordon, H., and Haidet, P. (2007). Physicians' communication and perceptions of patients: Is it how they look, how they talk, or is it just the doctor? *Social science & medicine*, 65(3):586–598.

Street Jr, R., OMalley, K., Cooper, L., and Haidet, P. (2008). Understanding concordance in patient-physician relationships: personal and ethnic dimensions of shared identity. *The Annals of Family Medicine*, 6(3):198–205.

Sueyoshi, G. (1992). Semiparametric proportional hazards estimation of competing risks models with time-varying covariates. *Journal of econometrics*, 51(1-2):25–58.

Sundararajan, V., Henderson, T., Perry, C., Muggivan, A., Quan, H., and Ghali, W. (2004). New icd-10 version of the charlson comorbidity index predicted in-hospital mortality. *Journal of clinical epidemiology*, 57(12):1288–1294.

Sutton, M. (2011). Ignorance is bliss? *Archives of internal medicine*, 171(17):1600.

Tarnow-Mordi, W., Hau, C., Warden, A., and Shearer, A. (2000). Hospital mortality in relation to staff workload: a 4-year study in an adult intensive-care unit. *The Lancet*, 356(9225):185–189.

Thompson, D. (2011). Overuse of diagnostic exams can harm patient health. USA Today. Retrieved from http://www.usatoday.com/news/health/healthcare/health/healthcare/studies/story/2011-12-30/Overuse-of-diagnostic-exams-can-harm-patient-health/52284494/1.

Tong, P. (2011). The effects of california minimum nurse staffing laws on nurse labor and patient mortality in skilled nursing facilities. *Health Economics*, 20(7):802–816.

Van Ryn, M. and Burke, J. (2000). The effect of patient race and socio-economic status on physicians' perceptions of patients. *Social science & medicine*, 50(6):813–828.

Vangen, S., Stoltenberg, C., Skrondal, A., Magnus, P., and Stray-Pedersen, B. (2000). Cesarean section among immigrants in norway. *Acta obstetricia et gynecologica Scandinavica*, 79(7):553–558.

Volk, M. and Ubel, P. (2011). Better off not knowing: Improving clinical care by limiting physician access to unsolicited diagnostic information. *Archives of internal medicine*, 171(6):487.

von Katterfeld, B., Li, J., McNamara, B., and Langridge, A. (2011). Perinatal complications and cesarean delivery among foreign-born and australian-born women in western australia, 1998–2006. *International Journal of Gynecology & Obstetrics*.

Wolkewitz, M., Beyersmann, J., Gastmeier, P., and Schumacher, M. (2009). Modeling the effect of time-dependent exposure on intensive care unit mortality. *Intensive care medicine*, 35(5):826–832.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA [etc.]: MIT Press.