Decision-Making and Choices in Healthcare

Empirical Essays on the Economics of Healthcare Performance

Ieva Sriubaite

BSc. in Statistics
MSc. in Economics

Supervised by:

Anthony Harris and Andrew M. Jones

April 2, 2021

MONASH
University

To my grandfather
and
my younger sibling

Copyright Notice

Acknowledgements

While adding the finishing touches to my thesis over the last half year, I only wished to begin writing the acknowledgement section. It came to my mind that this section is the most valuable (to me personally) and deserves to be written when the thesis is completed. So here I am just days before the submission. There is an endless list of people who have played a significant part in my PhD journey and to whom I am deeply indebted for their unconditional support during the time I have spent completing this thesis. My only wish at this moment is to say "thank you" to all of them.

First of all, I would like to acknowledge the person who has most greatly inspired me to begin this journey. *Professor Ansgar Wübker*, nearly a decade ago (in 2013), gave me the opportunity to assist his research work at the Ruhr University Bochum when I had just begun studying my master's program. It was my very first contact with Health Economics. As my interest has grown to the point where I have written my master thesis under his outstanding supervision, I have decided to pursue a graduate degree and learn about the peculiarities of this distinct field of economics.

Although my PhD journey has been so extraordinary that took the length of two continents of the world, I would like to express my sincere gratitude to everyone who has welcomed me on my first step towards a graduate degree at *the Health Economic Research Centre CINCH* at the University of Duisburg Essen in Germany. It has been an excellent start that has given me a limitless bag of invaluable knowledge to continue my journey at Monash University in Australia.

Most of all, I would like to extend my most sincere gratitude to my supervisors *Professor Anthony Harris* and *Professor Andrew M. Jones*, who have welcomed me and made the transition to the next part of my journey effortless. Thank you, Anthony and Andrew, for letting me be your student and for your trust and invaluable support when I stood at the edge of the cliff of my PhD. I would venture to say that there could not have been a better decision than to continue my graduate studies under your supervision. I value the time you gave me to discuss the particularities of my research, even though a large part of it was taken by the numerous administrative matters related to my transition to a new continent.

Perhaps most of my research time was spent gazing at a million lines of data and codes. I would not have found the way to solve those data puzzles without the outstanding guidance given by *Corinna Hentschker*, who not only helped me in understanding the largest dataset I have ever seen, but also kept me company while we worked in one of the most depressing rooms accessing sensitive and highly safeguarded German data. I would also like to express my gratitude to *Professor Belinda Gabbe*, who showed me that there are easier ways to protect datasets while working in a virtual secure environment and how to access them from any room I wished. I am also very thankful to Belinda for giving me the opportunity to be a part of her research team at *the School of Public Health and Preventive Medicine* and to discover fascinating public health research. During the time of my graduate studies, I have also been associated with several other excellent research initiatives and institutions that have helped me in building a research network on which I still rely greatly. I would like to thank everyone at *the RWI - Leibniz Institute for Economic Research* in Germany for sharing with me your productive research environment and providing me with a great working space when I travelled back to Europe. I am also indebted to the research initiative *Leibniz Science Campus Ruhr*, that provided the financial support to conduct the interesting research project about hospital closures in Germany. Lastly, I would like to thank everyone at *the Centre of Health Economics* at Monash University who were my companions at the last stop of my journey. I am very grateful for your warm welcome even though the circumstances were so extraordinary. I hope you all know how lost I felt when I came to Melbourne and I am thankful for you accepting me despite some of my direct views that here seem to be less favoured. Most importantly, I would like to express my gratitude to *Monash Business School*, because I could not have met everyone at the centre without the *2019 Monash Business School Graduate Research Scholarship*, which was my financial support in the last years of my PhD.

During this journey across the continents, I had the opportunity to stop at many charming cities in Europe. At various well-organized research events I presented my research and received insightful feedback. I greatly enjoyed this travelling during my PhD, but, most importantly, I have met plenty of interesting researchers and listened to presentations about even more interesting research projects. It was my great inspiration to continue my work in the research environment, especially after listening to the most prominent and inspiring economists such as Janet Currie, Alan Krueger and Jonathan Skinner. I thank all participants of these events for discussing my research in most rigorous ways, for raising your hands when the audience seemed a little drowsy before a cup of coffee and to everyone who engaged in long research discussions during the coffee breaks and pleasant dinners.

I also want to thank all persons who I met in the research environment and who became my partners in crime on this journey. I could not have had a better office mate than my Belarusian friend *Maryna* as we decorated the Christmas tree in our office room and studied long hours with a small, but cosy, light shining in the darkest winter evenings. I am also very grateful to my great travelling companion, *Johann*, at various research events. Your sense of humour, that, honestly, I have not always followed, always put me in a good mood – something one cannot get enough while being a PhD Student. I want to also say thank you to *Daniel*, who kept his office door open for whenever I had any questions, and to *Martin*, who gave me valuable research advice and travelled across Germany just to say goodbye before my journey to Australia. It is noteworthy to mention my four musketeers, *Adam, Alex, Liam and Paul*, who were my "grammar police" and enriched my English language greatly which became so valuable while writing this piece of work. Even though I have found myself in another part of the world, you are my dear friends and I always have you in my mind.

Last, and certainly the most important, I want to thank the members of my family and people in my closest circle: *my mother Renata*, who has supported me my whole life, even if some decisions I made were more or less promising; *my father Linas*, who has also supported me greatly and has given me a view of economics from the business perspective; *Daniel*, who made me engage in his endless discussions of academic and political chatter, but also taught me the most about research and made me rethink what really matters; *Farzaneh*, my dearest and beloved friend, who stood beside me during all my battles, both personal or professional; and, finally, *my grandfather Rimantas* and *my younger sibling Povilas* – to whom I dedicate this thesis. My grandfather is the person who has been the most interested in my research despite having less acquaintance with the research world. He has introduced me to my muse of knowledge and education, and it is my sincere hope that with this piece I will inspire my brother to find his.

Ačiū.

Su meile,
I.

ABSTRACT

Decision-makers in the healthcare market, be they a healthcare consumer, provider, insurer or regulator, constantly interact and exchange information with each other. However, one party often holds greater knowledge than the other. Hospitals expect that physicians will choose the most efficient treatment, patients have to rely on medical decisions made by their physician, and insurers must trust the information provided by their client about their medical condition. To equalise the information available in the healthcare industry by reducing information gaps between agents, has long been seen as a way to improve market efficiency. The aim of this collection of empirical studies is to address such information gaps and to understand the incentives and decision-making behaviour of each agent and its consequences in the healthcare market.

This collection consists of four self-contained essays, each focusing on a different decision-maker in the healthcare system. The first essay studies the role of the demand-side and the behaviour of healthcare consumers. In the context of maternal care in Germany, the study analyses how a prospective mother responds to both objective and subjective quality information and quantifies this response as a distance-performance trade-off. The second essay shifts focus to the behaviour of healthcare providers. In the context of the choice between using bare metal stents or drug-eluting stents among interventional cardiologists in Sweden, the study analyses the formation of physician practice styles after a physician has relocated to a new hospital. It decomposes the change in the physician practice style to provider-specific and a peer group-specific components and evaluates the effects of these decisions on the patient outcomes. The third essay concentrates on the trade-off between the quality and the quantity of care from the perspective of the regulator. In the context of hospital closures in Germany, the study analyses the consequences of hospital closures on healthcare access and patient outcomes. Finally, the fourth essay draws attention to the main intermediary in the healthcare system – health insurers. In the context of road traffic injuries in Victoria, the study builds a predictive model using supervised machine learning methods to identify the main risk triggers and risk groups for high costs and poor patient outcomes.

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes 1 original papers published in peer reviewed journal, 1 published in a Working Paper Series and 2 unpublished manuscripts. The core theme of the thesis is health economics. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Centre for Health Economics under the supervision of Anthony Harris and Andrew M. Jones.

Table 1 presents my contribution to paper published in peer reviewed journal.

Student Ieva Sriubaite
26 November 2020

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor Anthony Harris
26 November 2020

Table 1: CRediT author statement for contribution to published essays

Essay 1. Subjective and Objective Quality and Choice of Hospital	
Status	Published at the *Journal of Health Economics 68 (2019)*
Nature and %	*of contribution:*
Student's	70%. Conceptualisation, Formal analysis, Visualisation, Writing - Original draft
Co-author's	(1) D. Avdic, 10%. Conceptualization, Writing - Review & Editing
	(2) G. Moscelli, 10%. Methodology, Writing - Review & Editing
	(3) Adam Pilny, 10%. Writing - Review & Editing
	(all co-authors are not Monash Students)

Contents

Introduction

From a consumer perspective healthcare is a commodity unlike any other. It is fundamental for human well-being, yet individual autonomy to make decisions is limited and may put consumers at a health risk and leave them with extremely expensive choices. Information on how to alter health stakes through the use of medical services is costly and incoherent, making the third party agents such as healthcare providers major decision makers. As the Nobel prize laureate *Kenneth Arrow* argues,[1] the uncertainty in healthcare arising from such information asymmetry contributes to various decision errors and misinformation and is a major reason for market imperfections. As the private healthcare market is unable to cover risks to health using fair pricing, public regulation and funding is needed to ensure efficient delivery of healthcare and manage the uncertainty.

The healthcare market in practice is an interaction between consumers, providers, insurers and regulators who constantly interact and exchange information with each other; however, available information is incomplete and one party often holds greater knowledge than others. Patients are not trained in medicine and may have to rely on choices made by their physicians with respect to their treatments. The relationship between patient and physician could be based on trust and the belief that the physician will act as a perfect agent for the patient. This information asymmetry may create situations in which perfect conditions for physicians are faced with monetary incentives to under-treat or over-treat their patients depending on the reimbursement system in place. Similarly, insurers have limited information about their client's health condition, creating incentives for the latter to provide false information on actual health problems leading to variations in the client's anticipated behaviour, or moral hazard.

To reduce the uncertainty in healthcare markets by reducing the information gap between agents has long been seen as a way to improve market efficiency. For example, the U.S. officials released information about hospital quality (Luft *et al.*, 1990), that showed substantial geographic variation in physician practice styles without supporting clinical evidence. Exces-

[1] Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. American Economic Review, 53(5), 941-973.

sive use and misuse of medical treatments led to rising healthcare expenditures, laying the basis for large inefficiencies in the U.S. healthcare market (Steinwachs and Hughes, 2008). The release of the quality information has actively drawn patient choice into healthcare systems reducing the dominance of healthcare professionals over decisions made in medical treatment. Patient choice refers to patient's opportunity to choose the preferred hospital or healthcare professional according to the accessible information, financial independence and preferences. The main determinants of a choice have been extensively studied and vary from the proximity to a hospital to the quality of care.[2,3] A large body of empirical evidence[4] suggests that patients generally respond to reported quality-related measures and their choice of better performing healthcare provider leads to general improvements in the quality of care in the market.[5]

Regardless of increasing patient autonomy in choosing a healthcare provider, providers still have a great deal of power in making decisions during the course of treatment. One of the most substantial critiques of market-based healthcare systems appealed to the ignorance of healthcare provider behaviour. As *Victor Fuchs* argues,[6] the dominance in the decision-making in the healthcare market requires better understanding of supply-side incentives and is a crucial step in moving towards greater efficiency. The system itself may create stimulus for physicians and hospitals to act in one way, or the other, and based on empirical evidence they do in fact respond to it.[7] Although incentives play an important role, they may not be the only factor affecting decision-making. Some decisions may also be influenced by preferences or beliefs, that are difficult to quantify. The physician's set of beliefs

[2]Several studies highlighted the importance of the distance when choosing a healthcare provider, see e.g., Lee and Cohen (1985); Mcguirk and Porell (1984); Porell and Adams (1995); Sivey (2012).

[3]One prominent strain of literature has examined the impact of quality reporting on the choice of healthcare providers. For an overview, see Brekke *et al.* (2014). Other relevant studies found positive association between quality measure and hospital demand, see e.g., Bundorf *et al.* (2009); Gaynor (2006); Gaynor *et al.* (2016); Luft *et al.* (1990); Varkevisser *et al.* (2012).

[4]There are other ways to deepen the knowledge about patient's preferences and choices for healthcare such as discrete choice experiments (Ryan, 2004). While recognised for their benefits and contributions, this piece of work solely relies on econometric techniques, thus the experiment designs will not be further discussed.

[5]Much of the literature finds a positive response to the public hospital quality reporting and observes improvements in the overall quality, see e.g., Chou *et al.* (2014); Cutler *et al.* (2004); Dranove *et al.* (2003); Gutacker *et al.* (2016); Wang *et al.* (2011), whereas several empirical works find it less effective, particularly in the markets with strong pre-existing beliefs about quality of the provider, see e.g., Dranove and Sfekas (2008); Epstein (2010).

[6]Victor R. Fuchs (1974). *Who Shall Live? Health, Economics, and Social Choice*. New York: Basic Books, Inc., 1974.

[7]Studies like Egdahl and Taft (1986); Hillman *et al.* (1989); Manning *et al.* (1987); McGuire and Pauly (1991); Relman (1985, 1988) laid the groundwork for studying the physician's response to financial incentives. For further and more recent evidence, see e.g., Grant (2009); Gruber *et al.* (1999); Mitchell *et al.* (2000); Shafrin (2009); Shen *et al.* (2004); Yip (1998). Hospital incentives were generally introduced through the performance incentives model, *Pay-for-Performance*, that is recognized as a substantial improvement in quality and efficiency both of hospitals and physicians (Chingos, 2002).

may be formed by a multitude of factors such as their medical training and the influence of their teachers and leaders, their speciality and respective set of skills, the scientific evidence they are exposed to, or even the environment they work in (Stano, 1993). All these factors, some more, some less, influence the physician's behaviour and choices made in the course of medical treatment and are key determinants of variation in healthcare utilisation and spending.[8]

The nature of the healthcare market, information asymmetry and the behaviour of agents makes healthcare delivery one of the most difficult conundrums for economists. If we are to design a policy that delivers equitable and efficient healthcare, we need to understand incentives and the decision-making behaviour of each stakeholder under these constraints in order to equalise the information available in the healthcare industry and bring the utmost benefit to public health.

This dissertation consists of four self-contained essays, each focusing on a different decision-maker in the healthcare system. The thesis contributes to the literature by providing evidence about the behaviour of each participating party, their response to system incentives and the consequences different decisions lead to. The first essay studies the role of the demand side and the behaviour of healthcare consumers. It examines choices made by patients as a response to information. In the context of maternal care, the study analyses how a prospective mother chooses a hospital. It provides evidence on how patients respond to reported quality and quantifies this response as a distance-performance trade-off. The second essay shifts focus to the behaviour of healthcare providers. To build further evidence on how variation in clinical decision-making generates substantial geographical variation in healthcare spending, this essay analyses the malleability of physician practice styles. Particularly, it investigates the formation of physician's choices of how to treat a patient and further evaluates the effects of these decisions on patient outcomes. As healthcare performance and quality is inseparable from healthcare access, the third essay analyses the trade-off between the quality and quantity of care from the perspective of the decision-making by regulators and local healthcare organisers. A high density of hospitals provides patients with rapid access and short travel times to healthcare when needed. However, having many small hospitals may impose risks to patient's health as small hospitals might not provide the same variety of services and quality of care and are expensive to maintain. This essay studies the implications of hospital closures on the quality and costs of healthcare in the German healthcare system. Lastly, the fourth essay draws attention to the main intermediary in the healthcare system - health insurers. Using statistical machine learning methods, this study builds a predictive model aimed

[8]Physician practice style as a significant factor for variations in health has been first recognised by Wennberg and co-workers. See e.g., Mcpherson *et al.* (1982); Wennberg and Gittelsohn (1973, 1982); Wennberg (1984, 1985).

at assisting healthcare payers in the contracting with healthcare providers. The model predicts healthcare costs and patient outcomes that inform the payer about the resource use and the quality of healthcare providers. Such predictions are important for the sustainability of the healthcare system as it allows the insurer both to detect systematic overuse of resources and to provide clients with efficient and evidence-based management strategies.

The empirical evidence in this thesis aims to provide a step forward in the understanding of how choices are made, the mechanisms leading to these decisions and, most importantly, if and to what extent these decisions have consequences. It is my sincere hope that the essays included in this monograph will provide valuable knowledge for future policy makers about choices and decisions made in healthcare markets. As for my research community, I hope that the variety of empirical methods used in this thesis will beneficially aid future research. The evidence comes from three different healthcare systems – the German, the Swedish and the Australian – and provides diversity in this thesis and contributes to a better understanding of the decision-making and its consequences in the healthcare market.

Subjective and Objective Quality and Choice of Hospital: Evidence from Maternal Care Services in Germany

The focal point of *the first essay of the thesis (Essay 1)* is the behaviour of healthcare consumers. Most of the modern healthcare systems are designed to provide opportunities for patients to choose their healthcare provider. To some extent it ensures the patient's freedom to meet their preferences. However, in some markets particularly characterised by fixed prices of healthcare, it becomes a powerful tool to achieve desired policy goals such as greater efficiency or improved quality of care (Brekke *et al.*, 2014; Gaynor *et al.*, 2016). When providers cannot compete for prospective patients by offering more favourable fees, quality becomes the main factor in this bargaining game. Yet this is only feasible if demand is in fact responsive to quality.

This study analyses if and to what extent prospective healthcare consumers respond to quality when choosing a healthcare provider. In the context of the German maternity care market, which is highly competitive due to an extensive number of clinics, this essay focuses on the expectant mother's choice of a maternity clinic. Universal healthcare coverage and the absence of a formal gate-keeping system stimulate a concerted effort by consumers to scrutinize their options and make an informed decision over the course of their pregnancy. Using a rich patient-level dataset of hospital discharge records, the study relates the choice of maternal care provider to the „objective" (clinical indicators) as well as the „subjective" (patient satisfaction scores) performance measures, that jointly provide a broad spectrum of quality of care in the hospital. To quantify patient's responsiveness and better understand the trade-off between the distance to the hospital and the quality, this essay builds a conditional logit model and computes marginal utilities to provide direct estimates of the patient's willingness to travel to a provider with higher reported quality levels.

In line with the literature, this study provides empirical evidence that patients are responsive to quality.[9] In fact, they respond to both objective and subjective quality measures, suggesting that patient satisfaction scores may constitute important complements to clinical indicators when choosing a provider and describe a different dimension of quality such as personal comfort and staff friendliness. This finding demonstrates that the different quality measures may not necessarily be substitute to each other as both of these appear to be valued by consumers. The comparison of these two qualitatively different dimensions of quality is the main contribution of this essay to the literature on the competition/quality/choice nexus, that has relied heavily on clinical quality measures so far. To date, only three studies[10] have described patient choice by subjective quality indicators, yet most of these rely on measures such as the patient's self-reported health status rather than satisfaction across hospital organisational domains. The findings in this essay suggest that patients are willing to travel 0.1-2.7 additional kilometres for a one standard deviation increase in quality, providing compelling evidence for both a strong healthcare consumer response to the quality of care and a large variation in the magnitude of this response depending on the quality indicator.

Providers, Peers and Patients. How does the Physician Environment Affect Patient Outcomes?

The second essay of the thesis (Essay 2) shifts focus to the behaviour of healthcare providers. Substantial geographic variation in healthcare spending has been a cornerstone of the health economic literature for decades, yet traditional demand factors, such as preferences or patient health status, have been found to explain only a small part of this variation.[11] For this reason, literature began to emerge relating to the understanding of supply side factors and so called unwarranted variation in healthcare delivery, referring to treatment patterns of physicians and medical practitioners. These patterns often cannot be explained by the patient's medical needs and, in the absence of clear clinical guidelines, the physician's decision on how to treat a patient could result in both different clinical outcomes and costs. As it

[9] See e.g., for literature on the patient's response to quality Gaynor *et al.* (2016); Moscelli *et al.* (2016); Moscone *et al.* (2012); Pope (2009); Santos *et al.* (2016); Varkevisser *et al.* (2012); and the trade off between the quality and the distance Beckert *et al.* (2012); Gutacker *et al.* (2016); Jung *et al.* (2011); Moscelli *et al.* (2016); Pilny and Mennicken (2014); Santos *et al.* (2016); Tay (2003).

[10] Gutacker *et al.* (2016); Moscone *et al.* (2012); Pilny and Mennicken (2014a).

[11] See e.g., Chandra *et al.* (2012); Finkelstein *et al.* (2016); Skinner *et al.* (2011); Wennberg and Gittelsohn (1973).

is unclear whether high-spending regions perform better than low-spending regions (Skinner, 2011), this could potentially lead to resource waste and, as a result, such choices made by physicians may serve as a source of this variation.

This essay provides evidence on the determinants of physicians practice styles and it's effect on the quality of care. Using rich administrative data on the universe of coronary stenting procedures in Sweden between 2005 and 2013, the study relates the physician's treatment decision to the choice of stent, bare metal stent (BMS) or drug-eluting stent (DES). Whilst the medical procedures of inserting a stent are identical irrespective of the type of stent, the treatment decision may result in different patient outcomes and, to this end, clinical guidelines to assist on such decisions are uncertain. This essay further extends the approach taken by Molitor (2018) and analyses how these decisions are determined by the change of the physician's work environment. Specifically, the empirical setting first identifies physicians who migrate between hospitals and relates the variation in the rate of use of DES between the physician's origin and the destination hospitals to changes in the physician's own use of DES over time in a difference-in-differences empirical design. To better understand potential environmental channels mediating practice style, the study then decomposes the work environment into provider and peer group-specific factors. In contrast to most previous literature,[12] this study links the choice of the stent to relevant patient outcomes and directly measures the effect of physician treatment behaviour.

The empirical findings in this study contribute to the small literature on the adaptability of physician practice style with respect to the work environment. They shed light on underlying mechanisms shaping the decision-making process by studying the extent to which the effect relates to the organisational structure of the hospital and the physician's social group factors. Distinguishing these mechanisms is important as provider-specific factors may be less informative about the physician's preferences if the hospital follows local clinical practice guidelines and is restricted by technological resources. In contrast to the most of previous peer effects literature,[13] the findings of this study suggest that studying both peer- as well as provider-specific factors is crucial. The decomposition results reveal that physicians strongly respond to a change in the work environment and half of this response is driven by the physician's peers. Unlike previous studies, this empirical setting relates these changes in practice style to patient outcomes and further study the physician's decision errors, measured by the application of a random forest machine learning algorithm. The evidence reveals that the physician's choices, though affected by the change in their environ-

[12] See e.g., Currie *et al.* (2016); Epstein and Nicholson (2009); Grytten and Sørensen (2003); Molitor (2018).

[13] See e.g., Burke *et al.* (2003); Epstein and Nicholson (2009); Huesch (2011); Nair *et al.* (2010); Yang *et al.* (2014).

ment, are in line with the prevailing clinical guidelines and do not have any adverse effect on the quality of care.

Altogether, these results have broad implications for healthcare system efficiency and may inform policy makers about potential mechanisms explaining the large geographical variation in healthcare. As the quality of care appears to be insensitive, the empirical evidence suggests that the unwarranted practice variation can be mitigated through care coordination and profound clinical guidelines and may be vastly resource-saving during a time of rapidly rising costs and expense of healthcare (Wennberg, 2010).

Hospital Closures, Patient Outcomes and Local Politics. Evidence from Germany

The third essay of the thesis (Essay 3) empirically analyses the effects of current consolidation patterns in the healthcare industry and switches attention to the decisions made by local healthcare organisers. Modern healthcare systems continue to evolve and shift with vigorous focus on both quality and patient safety and, most importantly, incentivised by the government, the efficient delivery of services. Accelerated by the patient's ability to choose, as discussed in the first essay, and the substantial geographical variation in healthcare spending due to supply side factors, as analysed in the second essay, competition among healthcare providers has become ever stronger. As a result, some hospitals are not able to remain financially viable, compelling regulators to close the hospital and raising public concerns with respect to both the access and the quality of healthcare provision in the remaining market.

The main focus of this essay is to understand the consequences of hospital closures as a result of consolidation in healthcare. In the context of emergency health services for individuals with an acute myocardial infarction (AMI) or a hemorrhagic stroke, this essay estimates the effects of hospital closures on geographical access to care and clinical outcomes as well as some efficiency indicators in hospitals. Studying cardiovascular diseases, for which AMI and stroke are the two most common manifestations, provides close to an ideal empirical setting to analyse hospital closures. It is the leading cause of death globally (World Health Organization, 2011) and is the number one reason for all medical emergencies (Linden, 2006), for which healthcare centralisation plays a crucial role in determining chances for survival, as time is of the essence for patients with these medical conditions (American Heart Association, 2003).

Regardless of the compelling context this study is built on, one of the major empirical challenges in studying the effects of hospital closures is the endogeneity between hospital quality and market structure. While a hospital may close for numerous reasons, one of the most common is the poor quality of care. It is often the final verdict for the future of a hospital, particularly in the healthcare market characterized by the free choice of the provider, as discussed in the first essay. Most of the existing evidence relies on the policy-induced variation in distance as an indirect predictor of a hospital closure.[14] While this measure is relevant for more concentrated markets such as the U.S. or Sweden, it is less informative in a market with high hospital density such as the one in Germany. To tackle this, I employ an instrumental variable methodology and build a strong and relevant instrument for hospital closure in the context of municipal politics. Closing a hospital becomes a very unpopular political act among local politicians, who often fear that substantive policies such as hospital closure are often "punished" by a lower share of votes in the next election. This study follows a similar strategy to Bloom *et al.* (2015) and constructs a highly relevant instrument for hospital closure based on political pressure in the local governmental area.

This essay contributes to the literature on consolidation policies. Building on the political mechanism, the instrument used in this study provides compelling evidence that local politics play a substantial role in shaping a hospital's future. It is an important channel that could mediate potentially adverse effects on social welfare and address public concerns with healthcare consolidation trends. Applying this instrument gives empirical evidence for alternative empirical methodologies to study less concentrated markets. Additionally employed official reports on hospital closures and public hospital quality reports support the evidence and eliminate any concerns arising from potentially unobserved heterogeneous effects among healthcare providers. As the results suggest, even during one of the strongest periods in healthcare consolidation in Germany, this phenomenon did not result in any adverse clinical outcomes. The increase in travel time to hospital due to closure does not result in a higher mortality rate following an AMI or a stroke. On the contrary, patients living in closure-affected urban areas have a significantly shorter length of stay, that is not accompanied by any adverse clinical outcomes. As for access to care, policy-makers should only be concerned by hospitals closing in less densely populated areas.

Economic Consequences of Road Traffic Injuries. Application of the Super Learner algorithm

The fourth essay of the thesis (Essay 4) focuses on the role of healthcare payer in the decision-making process and provides evidence on how such decisions could be made using data-driven insights. In markets characterised by supply-side incentives healthcare providers often have information advantage in decisions related to healthcare and resource use. Payers can only imperfectly observe the cost and the quality of healthcare delivery, which

[14] See e.g., Avdic (2016); Blondel *et al.* (2011); Buchmueller *et al.* (2006); Ravelli *et al.* (2011).

may lead to the increased risk for over- or under-use of resources and reduced quality of care (Eggleston, 2000). To share adequate information on financial risk in order to bridge this information gap payers ought to develop statistical models aimed at predicting patient's healthcare costs and informing about provider's quality of care (Ellis *et al.*, 2018).

A great deal of research has focused on developing statistical models to predict healthcare costs and outcomes. Researchers have utilised various parametric and semi-parametric models,[15] but, as healthcare data have become increasingly detailed, the focus has shifted towards new data science approaches, such as supervised machine learning.[16] Supervised machine learning algorithms offer the ability to uncover complex data structures not known in advance and, due to their functional flexibility, have demonstrated a strong potential in healthcare research (Mullainathan and Spiess, 2017). While still in its infancy, one such method is the Super Learner proposed by van der Laan *et al.* (2007).[17] The Super Learner is an ensemble machine learning algorithm based on a weighted combination of various parametric and non-parametric statistical models. This study employs the Super Learner in the context of road traffic injuries and thereby contributes to the literature on predicting healthcare costs and patient outcomes in the environment where a single payer contracts with multiple providers. The algorithm utilises the insurance claims dataset provided by the statutory insurance company Transport Accident Commission and patient- and treatment-related information of all major traumas in the state of Victoria from the Victoria State Trauma Register. The study employs advanced statistical methods to improve the predictive power of complex patterns of healthcare data and adds to the rapidly emerging field of the application of the Super Learner in the context of the economics of healthcare. The study also contributes to existing research by, in addition to healthcare costs, predicting several patient outcomes that are relevant for risk-sharing between payers and providers. Following the concept of net benefits by Stinnett and Mullahy (1998), the study computes net benefits gained from the major trauma treatment and estimates the monetary value of patient lifetime health impacts. Because the participation in paid employment is an important factor for patient's well-being, the study also predicts return to work.

Altogether, the results reported in this essay provide further evidence on the benefits of advanced data science methods in health economics. The Super Learner is a powerful tool in predicting the economic consequences of

[15] See e.g., Duan (1983); Jones (2011); Jones *et al.* (2014); Manning *et al.* (2005); Mullahy (2009).

[16] See e.g., Arandjelović (2015); Bertsimas *et al.* (2008); Einav *et al.* (2016); Lahiri (2014).

[17] The Super Learner algorithm has been used in the economic context to study health insurance markets (Park and Basu, 2018; Rose, 2016; Rose *et al.*, 2017; Shrestha *et al.*, 2018).

road traffic injuries, including healthcare spending, net benefits and return to work. Its strong performance in predicting over half of the variation of considered outcomes shows that it may be used by policy-makers to both bridge the information gap and share the financial risk between payers and providers.

References

American Heart Association (2003). *Heart and Stroke Facts*. Tech. rep., American Heart Association.

Arandjelović, O. (2015). Prediction of health outcomes using big (health) data. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2543–2546.

Avdic, D. (2016). Improving efficiency or impairing access? Health care consolidation and quality of care: Evidence from emergency hospital closures in Sweden. *Journal of Health Economics*, **59**, 44–60.

Beckert, W., Christensen, M. and Collyer, K. (2012). Choice of NHS-funded Hospital Services in England. *The Economic Journal*, **122** (560), 400–417.

Bertsimas, D., Bjarnadóttir, M., Kane, M., Kryder, J., Pandey, R., Vempala, S. and Wang, G. (2008). Algorithmic prediction of health care costs and discovery of medical knowledge. *Operations Research*, **56** (6), 1382–1392.

Blondel, B., Drewniak, N., Pilkington, H. and Zeitlin, J. (2011). Out-of-hospital births and the supply of maternity units in France. *Health & Place*, **17** (5), 1170–3.

Bloom, N., Propper, C., Seiler, S. and van Reenen, J. (2015). The Impact of Competition on Management Quality: Evidence from Public Hospitals. *Review of Economic Studies*, **82**, 457–489.

Brekke, K. R., Gravelle, H., Siciliani, L. and Straume, O. R. (2014). Patient choice, mobility and competition among health care providers. *Developments in health economics and public policy*, **12**, 1–26.

Buchmueller, T. C., Jacobson, M. and Wold, C. (2006). How far to the hospital? The effect of hospital closures on access to care. *Journal of Health Economics*, **25**, 740–761.

Bundorf, M. K., Chun, N., Goda, G. S. and Kessler, D. P. (2009). Do markets respond to quality information? The case of fertility clinics. *Journal of Health Economics*, **28** (3), 718–727.

Burke, M. A., Fournier, G. and Prasad, K. (2003). *Physician social networks and geographical variation in medical care*. Center on Social and Economic Dynamics Washington, DC.

Chandra, A., Cutler, D. and Song, Z. (2012). Who ordered that? The economics of treatment choices in medical care. In T. M. Mark, V. Pauly and P. P. Barros (eds.), *Handbook of Health Economics*, vol. 2, *Chapter 6*, Elsevier, pp. 397–432.

Chingos, P. T. (2002). *Paying for Performance : A Guide to Compensation Management*. New York: John Wiley & Sons, Inc.

Chou, S.-Y., Deily, M. E., Li, S. and Lu, Y. (2014). Competition and the impact of online hospital report cards. *Journal of Health Economics*, **34**, 42–58.

Currie, J., MacLeod, W. B. and Van Parys, J. (2016). Provider practice style and patient health outcomes: The case of heart attacks. *Journal of Health Economics*, **47**, 64–80.

Cutler, D. M., Huckman, R. S. and Landrum, M. B. (2004). The Role of Information in Medical Market: An Analysis of Publicly Reported Outcomes in Cardiac Surgery. *American Economic Review*, **94** (2), 342–346.

Dranove, D., Kessler, D., McClellan, M. and Satterthwaite, M. (2003). Is more information better? the effects of health care quality report cards. *Journal of Political Economy*, **111**, 555–588.

— and Sfekas, A. (2008). Start spreading the news: A structural estimate of the effects of New York hospital report cards. *Journal of Health Economics*, **27** (5), 1201–1207.

Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, **78**, 605—-610.

Egdahl, R. H. and Taft, C. H. (1986). Financial incentives to physicians. *The New England Journal of Medicine*, **315** (1), 59–61.

Eggleston, K. (2000). Risk selection and optimal health insurance-provider payment systems. *The Journal of Risk and Insurance*, **67** (2), 173–196.

Einav, L., Finkelstein, A., Kluender, R. and Schrimpf, P. (2016). Beyond Statistics: The Economic Content of Risk Scores. *American Economic Journal: Applied Economics*, **8** (2), 195–224.

Ellis, R. P., Martins, B. and Rose, S. (2018). Chapter 3 - risk adjustment for health plan payment. In T. G. McGuire and R. C. van Kleef (eds.), *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets*, Academic Press, pp. 55–104.

Epstein, A. J. (2010). Effects of report cards on referral patterns to cardiac surgeons. *Journal of Health Economics*, **29** (5), 718–731.

— and Nicholson, S. (2009). The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics*, **28** (6), 1126–1140.

Finkelstein, A., Gentzkow, M. and Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics*, **131** (4), 1681–1726.

Gaynor, M. (2006). What Do We Know about Competition and Quality in Health Care Markets? *Foundations and Trends in Microeconomics*, **2** (6).

—, Propper, C. and Seiler, S. (2016). Free to choose? Reform, Choice, and Consideration Sets in the English National Health Service. *American Economic Review*, **106** (11), 3521–3557.

Grant, D. (2009). Physician financial incentives and cesarean delivery: New conclusions from the healthcare cost and utilization project. *Journal of Health Economics*, **28**, 244–250.

Gruber, J., Kim, J. and Mayzlin, D. (1999). Physician fees and procedure intensity: the case of cesarean delivery. *Journal of Health Economics*, **18** (4), 473–490.

Grytten, J. and Sørensen, R. (2003). Practice variation and physician-specific effects. *Journal of Health Economics*, **22** (3), 403–418.

Gutacker, N., Siciliani, L., Moscelli, G. and Gravelle, H. (2016). Choice of hospital: Which type of quality matters? *Journal of Health Economics*, **50**, 230–246.

Hillman, A. L., Pauly, M. V. and Kerstein, J. J. (1989). How do financial incentives affect physicians' clinical decisions and the financial performance of health maintenance organizations? *The New England Journal of Medicine*, **321** (2), 86–92.

Huesch, M. D. (2011). Is blood thicker than water? Peer effects in stent utilization among Floridian cardiologists. *Social Science & Medicine*, **73** (12), 1756–1765.

Jones, A. M. (2011). Models for health care. In M. P. Clements and D. F. Hendry (eds.), *Oxford Handbook of Economic Forecasting*, Oxford: Oxford University Press, pp. 625–654.

—, Lomas, J. and Rice, N. (2014). Applying beta-type size distributions to healthcare cost regressions. *Journal of Applied Econometrics*, **29**, 649–670.

Jung, K., Feldman, R. and Scanlon, D. (2011). Where would you go for your next hospitalization? *Journal of Health Economics*, **30** (4), 832–841.

Lahiri, B. (2014). *Predicting Healthcare Expenditure Increase for an Individual from Medicare Data*. Retrieved from: https://www.academia.edu/7836580/Predicting_Healthcare_Expenditure_Increase_for_an_Individual_from_Medicare_Data [accessed 21.04.2020].

Lee, H. L. and Cohen, M. A. (1985). A Multinomial Logit Model for the Spatial Distribution of Hospital Utilization. *Journal of Business and Economic Statistics*, **3** (2), 159–168.

Linden, A. (2006). What Will It Take for Disease Management to Demonstrate a Return on Investment? New Perspectives on an Old Theme. *The American Journal of Managed Care*, **12**, 217–222.

Luft, H. S., Garnick, D. W., Mark, D. H., Peltzman, D. J., Phibbs, C. S., Lichtenberg, E. and Mcphee, S. J. (1990). Does Quality Influence Choice of Hospital? *JAMA : the journal of the American Medical Association*, **263** (21), 2899–2906.

Manning, W. G., Basu, A. and Mullahy, J. (2005). Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*, **24** (3), 465–488.

—, Newhouse, J. P., Duan, N., Keeler, E. B. and Leibowitz, A. (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *The American Economic Review*, **77** (3), 251–277.

McGuire, T. G. and Pauly, M. V. (1991). Physician response to fee changes with multiple payers. *Journal of Health Economics*, **10** (4), 385–410.

Mcguirk, M. A. and Porell, F. W. (1984). Spatial patterns of hospital utilization: the impact of distance and time. *Inquiry : A Journal of Medical Care Organization, Provision and Financing*, **21** (1), 84.

Mcpherson, K., Wennberg, J. E., Hovind, O. B. and Clifford, P. (1982). Small-Area Variations in the Use of Common Surgical Procedures: An International Comparison of New England, England, and Norway. *The New England Journal of Medicine*, **307** (21), 1310–1314.

Mitchell, J. M., Hadley, J. and Gaskin, D. J. (2000). Physicians' responses to medicare fee schedule reductions. *Medical Care*, **38** (10), 1029–1039.

Molitor, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy*, **10** (1), 326–356.

Moscelli, G., Siciliani, L., Gutacker, N. and Gravelle, H. (2016). Location, quality and choice of hospital: Evidence from England 2002-2013. *Regional Science and Urban Economics*, **60**, 112–124.

Moscone, F., Tosetti, E. and Vittadini, G. (2012). Social Interaction in Patients' Hospital Choice: Evidences from Italy. *Journal of the Royal Statistical Society Series A*, **175** (2), 453–472.

Mullahy, J. (2009). Econometric modeling of health care costs and expenditures: A survey of analytical issues and related policy considerations. *Medical Care*, **47** (7), S104–S108.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, **31**, 87–106.

Nair, H. S., Manchanda, P. and Bhatia, T. (2010). Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research*, **47** (5), 883–895.

Park, S. and Basu, A. (2018). Alternative evaluation metrics for risk adjustment methods. *Health Economics*, **27** (6), 984–1010.

Pilny, A. and Mennicken, R. (2014). *Does Hospital Reputation Influence the Choice of Hospital?* Ruhr Economic Papers No. 516, RWI-Leibniz-Institut für Wirtschaftsforschung, Ruhr-University Bochum, TU Dortmund University, University of Duisburg-Essen.

— and — (2014a). *Does Hospital Reputation Influence the Choice of Hospital?* Ruhr Economic Papers No. 516, RWI-Leibniz-Institut für Wirtschaftsforschung, Ruhr-University Bochum, TU Dortmund University, University of Duisburg-Essen.

Pope, D. G. (2009). Reacting to rankings: Evidence from "America's Best Hospitals". *Journal of Health Economics*, **28** (6), 1154–1165.

Porell, F. W. and Adams, E. K. (1995). Hospital Choice Models: A Review and Assessment of their Utility for Policy Impact Analysis. *Medical Care Research and Review*, **52** (2), 158–195.

Ravelli, A., Jager, K., de Groot, M., Erwich, J., Rijninks-van Driel, G., Tromp, M., Eskes, M., Abu-Hanna, A. and Mol, B. (2011). Travel time from home to hospital and adverse perinatal outcomes in women at term in the Netherlands. *BJOG: An International Journal of Obstetrics & Gynaecology*, **118** (4), 457–465.

Relman, A. S. (1985). Dealing with conflicts of interest. *The New England Journal of Medicine*, **313** (12), 749–751.

— (1988). Salaried physicians and economic incentives.

Rose, S. (2016). A machine learning framework for plan payment risk adjustment. *Health Services Research*, **51**, 2358–2374.

—, Bergquist, S. L. and Layton, T. J. (2017). Computational health economics for identification of unprofitable health care enrollees. *Biostatistics*, **18** (4), 682–694.

Ryan, M. (2004). Discrete choice experiments in health care. *BMJ*, **328** (7436), 360–361.

Santos, R., Gravelle, H. and Propper, C. (2016). Does Quality Affect Patients' Choice of Doctor? Evidence from England. *The Economic Journal*, **127** (600), 445–494.

Shafrin, J. (2009). Operating on commission: Analyzing how physician financial incentives affect surgery rates. *Health economics*, **19**, 562–580.

Shen, J., Andersen, R., Brook, R., Kominski, G., Albert, P. and Wenger, N. (2004). The effects of payment method on clinical decision-making. *Medical care*, **42**, 297–302.

Shrestha, A., Bergquist, S., Montz, E. and Rose, S. (2018). Mental health risk adjustment with clinical categories and machine learning. *Health Services Research*, **53** (S1), 3189–3206.

Sivey, P. (2012). The effect of waiting time and distance on hospital choice for English cataract patients. *Health Economics*, **21** (4), 444–456.

Skinner, J. (2011). Causes and consequences of regional variations in health care. In T. M. Mark, V. Pauly and P. P. Barros (eds.), *Handbook of Health Economics*, vol. 2, *Chapter 2*, Elsevier, pp. 45–93.

—, Gottlieb, D. J. and Carmichael, D. (2011). A new series of Medicare expenditure measures by hospital referral region: 2003-2008. *Dartmouth Atlas Project*.

Stano, M. (1993). Evaluating the policy role of the small area variations and physician practice style hypotheses. *Health policy (Amsterdam)*, **24** (1), 9–17.

Steinwachs, D. and Hughes, R. (2008). *Health Services Research: Scope and Significance*.

Stinnett, A. A. and Mullahy, J. (1998). Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analaysis. *Medical Decision Making*, **18** (2), S68–S80.

Tay, A. (2003). Assessing Competition in Hospital Care Markets: The Importance of Accounting for Quality Differentiation. *The RAND Journal of Economics*, **34** (4), 786–814.

van der Laan, M. J., Polley, E. C. and Hubbard, A. E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, **6**.

Varkevisser, M., van der Geest, S. A. and Schut, F. T. (2012). Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands. *Journal of Health Economics*, **31** (2), 371–378.

Wang, J., Hockenberry, J., Chou, S.-Y. and Yang, M. (2011). Do bad report cards have consequences? Impacts of publicly reported provider quality information on the CABG market in Pennsylvania. *Journal of Health Economics*, **30** (2), 392–407.

Wennberg, J. and Gittelsohn, A. (1973). Small area variations in health care delivery: a population-based health information system can guide planning and regulatory decision-making. *Science*, **182** (4117), 1102–1108.

— and — (1982). Variations in medical care among small areas. *Scientific American*, **246**, 120–134.

Wennberg, J. E. (1984). Dealing with medical practice variations: a proposal for action. *Health affairs (Project Hope)*, **3** (2), 6.

— (1985). On patient need, equity, supplier-induced demand, and the need to assess the outcome of common medical practices. *Medical care*, **23** (5), 512.

— (2010). *Tracking medicine: a researcher's quest to understand health care*. Oxford University Press.

World Health Organization (2011). Tech. Report. In *Global Health and Aging*, World Health Organization, Geneva.

Yang, M., Lien, H.-M. and CHOU, S.-Y. (2014). Is there a physician peer effect? Evidence from new drug prescriptions. *Economic Inquiry*, **52** (1), 116–137.

Yip, W. C. (1998). Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft (CABG) surgeries in the Medicare and private sectors. *Journal of Health Economics*, **17** (6), 675–699.

Essay 1: Subjective and Objective Quality and Choice of Hospital: Evidence from Maternal Care Services in Germany[*][†]

1 Introduction

The role played by quality as a factor explaining patients' choices of healthcare provider is a key component of the quality-competition theory, according to which providers have incentives to compete on quality when prices are fixed (Brekke *et al.*, 2014; Gaynor, 2006). However, hospital competition on quality is possible only if demand for healthcare is not inelastic with respect to quality. As such, flexible patient choice of provider has been introduced in many healthcare systems across the world as a way to make healthcare demand more responsive to quality (Propper, 2018). Over the last decade, a number of studies have evaluated the association between quality and choice for elective care, finding that patient choice is to some extent responsive to quality (see, e.g., Gaynor *et al.*, 2016; Moscelli *et al.*, 2016; Moscone *et al.*, 2012; Pope, 2009; Santos *et al.*, 2016; Varkevisser *et al.*, 2012).[1] However, most studies to date have only considered clinical quality indicators.

In this paper we empirically investigate to what extent healthcare consumers vary in their choices of provider depending on the nature of the re-

[*]Published at the *Journal of Health Economics* 68 (2019) with co-authors: Daniel Avdic, Giuseppe Moscelli and Adam Pilny.

[†]The authors thank the editor, two anonymous referees, Matthias Bäuml, Tuğba Büyükdurmuş and seminar participants at the RGS doctoral conference in Dortmund, SMYE conference in Halle an der Saale, LAGV conference in Aix-en-Provence, ESPE conference in Glasgow, the Annual Lithuanian Economics Meeting 2017 in Vilnius, and EEA conference in Lisbon for valuable comments and suggestions. We are also grateful to Tautvydas Januškevičius and Joseph Montalbo for excellent SQL programming and research assistance. Financial support from the Federal Ministry of Education and Research (BMBF) is gratefully acknowledged.

[1]This result also holds in other settings such as the choice of health insurance plans with higher reported ratings. See, e.g., Beaulieu (2002); Bünnings *et al.* (2019); Chernew *et al.* (2008); Jin and Sorensen (2006); Scanlon *et al.* (2002); Wedig and Tai-Seale (2002).

ported quality information. Specifically, we relate the choices of maternity clinics of expectant mothers to "objective" (clinical indicators) and "subjective" (patient satisfaction scores) performance measures using rich German data from administrative hospital discharge records linked to publicly available information about provider quality.[2] We focus on maternal care in Germany for several reasons: first, healthcare consumers in Germany are entirely free to choose hospital due to the universal health insurance system (which covers treatment in all hospitals) and the absence of a formal gatekeeping system (which regulates access into specialized care).[3] Furthermore, the market for hospital childbirths is highly competitive with many producers and consumers of the service.[4] Finally, consumers in this market are likely to exert effort to make substantiated choices because they value any information that allows them to scrutinize their options over the course of their pregnancy.[5] Thus, the context of German maternal care suggests a close to optimal market setting where high-stakes patients are able to make informed choices between competing providers.

Our empirical analysis entails the use of three merged datasets from Germany. First, we use a ten percent nationally representative sample of all German hospital births for years 2009–2012 from a rich patient-level dataset of hospital discharge records, including a wide range of patient characteristics, services, clinical outcomes, and geographical locations of both hospitals and patients. We link this information to a set of objective quality indicators taken from standardized public report cards that all hospitals are required to provide. These report cards disclose relevant information for prospective patients, such as availability of medical services, clinical patient outcomes, capacity and competency of the medical staff. Finally, we complement the objective quality indicators with subjective quality information in the form of patient satisfaction scores from a nationwide survey administrated by one of Germany's largest public health insurance providers. The survey includes information on patients' satisfaction with their medical treatment, staffing, communication, organization, and accommodation in the hospital. Linking the hospital discharge records to the quality indicators in this manner allows us to directly use information that prospective patients have access to when choosing provider in contrast to relying on indirect information derived from, for example, hospital episodes.

We first estimate a simple linear probability model for the dichotomous choice between the closest hospital and any other hospital in a pre-defined

[2] We define an objective, relative to a subjective, quality indicator as a performance measure which is not based on patients' own experiences or perceptions (e.g., rates of obstetric trauma).

[3] See, e.g., Busse and Blümel (2014) for a review of healthcare provision in Germany.

[4] Germany has the highest density of hospital beds in Europe. See https://www.destatis.de/Europa/EN/Topic/PopulationLabourSocial/Health/HospitalBeds.html.

[5] Giving birth is an activity frequently involving a substantial amount of anxiety for the patient. For example, pregnancy-related anxiety (PrA) is a disorder affecting 14 percent of all childbearing women (see, e.g., Alder *et al.*, 2007; Blackmore *et al.*, 2016).

choice set as a function of the former hospital's distance and quality. We subsequently model patient choice structurally using a conditional logit model from which we are able to compute marginal utilities to provide a direct estimate of a patient's willingness to travel (WTT) to a provider with higher reported quality. To this end, we use information on the distance between an individual's home and the chosen hospital to construct a measure of a representative patient's WTT for a given improvement in reported quality.

Our results indicate that patients generally respond to hospital quality, but also that responses vary substantially across quality indicators. Patients appear to respond to subjective quality also after conditioning on objective quality, suggesting that patient satisfaction scores provide a complementary source of information about the performance of a hospital that goes beyond established clinical indicators. We estimate that, depending on the specific quality indicator, an expectant mother is on average willing to travel an additional 0.1–2.7 kilometres (0.2–4.5 minutes by car) to give birth in a hospital with a one standard deviation higher reported quality. This corresponds to a WTT of up to one-third of the average distance to the closest hospital for individuals in our sample. Our findings are largely robust to a set of sensitivity checks with respect to model specification, choice set and variable definitions. Despite somewhat attenuated coefficients, most of our results still hold when we incorporate hospital fixed effects in our model to account for time-invariant unobserved heterogeneity in hospital quality.

The literature on the competition-quality-choice nexus, upon which many of today's healthcare systems are built, is relatively scarce but growing[6] and only a few papers have explicitly considered the trade-off between distance and quality.[7] We contribute to this important area of research mainly in two ways: First, we compare how patients respond to two qualitatively different dimensions of quality in their choice of healthcare provider for maternal delivery services. To our knowledge, while some studies have found that distance to the hospital has a significant effect on patients' choice (Porell and Adams, 1995; Sivey, 2012), only three studies (Gutacker *et al.*, 2016; Moscone *et al.*, 2012; Pilny and Mennicken, 2014) analyzed the influence of social interactions and subjective quality on patient's choice of hospital. In this work, however, the contrast of the effects of objective and subjective quality is in many ways different from the one analysed by Gutacker *et al.* (2016). The subjective quality that we use captures patients' satisfaction across various hospital organizational domains, whereas in Gutacker *et al.* (2016) it is a measure of change in the patient's health status captured by routinely collected Patient Reported Outcomes Measures (PROM) like the Oxford Hip Score (OHS). In other words, the subjective quality indicators

[6] See, e.g., Baker *et al.* (2003); Bundorf *et al.* (2009); Cutler *et al.* (2004); Dranove and Sfekas (2008); Gaynor *et al.* (2016); Hodgkin (1996); Mukamel and Mushlin (1998); Pope (2009); Santos *et al.* (2016); Varkevisser *et al.* (2012); Werner *et al.* (2012)

[7] See, e.g., Beckert *et al.* (2012); Gutacker *et al.* (2016); Jung *et al.* (2011); Moscelli *et al.* (2016); Pilny and Mennicken (2014); Santos *et al.* (2016); Tay (2003)

in our paper are proxies of *non-clinical* hospital quality. Using non-clinical subjective quality indicators has several advantages. It differs from previous literature, shedding light on the role that the perception of organizational and management quality has in the choice of elective hospital care; a question related and possibly preliminary to the one investigated by Bloom *et al.* (2015). It also represents a general measure of the broad quality of hospitals, thus being a sensible quality measure with respect to the choice of hospital service we analyse, whereas a subjective clinical measure like the OHS would be unsuitable to use, even if available, given that pregnancy is not a disease like osteoarthritis. Moreover, and differently from the OHS and other PROMs, a non-clinical subjective quality measure is less prone to the confounding bias due to differences in the post-treatment health status.[8]

Second, we contribute to the literature on the determinants of women's choice of birth clinic which has previously only rarely been investigated. O'Cathain *et al.* (2002) report evidence for Wales that a large minority of women giving birth did not feel like they exercised an informed choice in their maternity care. They show that evidence based leaflets were not effective in promoting informed choice in women using maternity services from a sample of 13 maternity units in Wales. Moreover, Wagle *et al.* (2004) show that distance to hospital and higher socioeconomic status are the main drivers of choice of place of maternal delivery (i.e., home versus hospital) in Nepal, but the study does not include any quality measure. Differences in healthcare experience or environment at critical times have also been found to affect psychological status of the mothers during pregnancy (Jomeen and Martin, 2008). Our results suggest that expectant women are highly responsive to reported quality when choosing clinic to give birth in, but also that the particular performance indicator appears to be crucial for the magnitude of the response.

The paper is organized as follows. The next section provides a summary of the relevant characteristics of the German healthcare system. Section 3 describes the data we use for our empirical analysis. Section 4 outlines our econometric framework. Section 5 reports results from estimation and Section 6 concludes.

2 Institutional context

The German healthcare system is jointly organized by federal and state level institutions and provides healthcare for all citizens and permanent residents. The German health insurance system is characterized by the coexistence of public statutory health insurance (SHI) and substitute private health

[8]Such bias may arise when the patients' health status change is driven not just by the quality of the treating hospital, but by the quality of other healthcare services (e.g., good or bad rehabilitation units, family doctors, osteopaths, etc.) located in the same catchment area of the treating hospital and playing a significant role in patients' post-surgical recovery.

insurance (PHI). Access to healthcare is ensured by mandatory membership in one of the approximately 110 SHI funds or 50 PHI funds.[9] The SHI covers about 90 percent of the German population.[10] In the SHI family insurance, nonworking spouses and dependent children under 25 years are covered free of charge. Further exemptions from insurance premiums apply for students and unemployed. Insurance under the SHI is mandatory for employees with gross wage earnings below a defined threshold (€59K/$73K annually in 2018). Specific groups of the population may opt out of SHI and buy substitute PHI or remain publicly insured as voluntary members, including high-income earners, self-employed and civil servants. Each SHI fund only offers one standardized health plan, which by law comprises full coverage of healthcare services and free choice of healthcare provider for all types and levels of care. By contrast, PHI providers are allowed to offer highly diverse health plans with varying components, such as co-payment levels and complementary care services. In general, PHI health plans also offer full coverage and include free hospital of treatment choice. However, PHI providers do not have to contract with healthcare providers and do not negotiate tariffs and prices as in the SHI. The maximum fee providers may charge for the treatment of PHI clients is regulated by the German Federal Ministry of Health (Wasem *et al.*, 2004).

A set of regulations have been implemented in order to maintain and improve high levels of care quality delivered by healthcare providers. For example, all providers are obliged to establish a quality management system based on continuous medical education for all physicians as well as a health technology assessment for drugs and medical procedures. Moreover, minimum case volume requirements of complex inpatient procedures force hospitals to adapt to the development of new healthcare technologies to stay competitive. The overall treatment process as well as the outcomes are regularly controlled through a mandatory quality reporting system (Busse, 2008; Busse and Blümel, 2014).

Large parts of German hospital policy are decentralized to the level of the 16 federal state governments (Länder). In particular, the state governments are responsible for hospital planning, meaning that they decide on the extent, location and specialization of hospitals in their respective region. To this end, each state assembles a hospital plan and schedules the allocation of hospital capacities, investment funding and, to some extent, quality requirements for particular departments (Karmann and Roesel, 2017; Pilny, 2017). Hospitals that are included in a state's hospital plan are, since 2006, legally obliged to publish standardized quality report cards.[11] Individuals

[9] Numbers as of January 2019.

[10] Pilny *et al.* (2017) provide a detailed overview about the German SHI and characteristics of its clients.

[11] Hospitals not included in a hospital plan can still contract with the SHI, in which case they are also legally obliged to publish quality report cards. Together these hospitals comprise about 90 percent of all hospitals and 99 percent of all bed capacities in

are free to choose the healthcare provider for their next elective hospitalization among those hospitals included in a hospital plan, or those hospitals that contract with the SHI. The dissemination of hospital quality among the public is a key strategy used by policy makers in the competitive hospital market to stimulate choice among healthcare recipients.

The performance indicators in the standardized quality report cards are analysed by independent and impartial institutes: the Institute for Quality and Patient Safety (BQS), the Institute for Applied Quality Improvement and Research in Healthcare (AQUA), and state-level specialized groups providing various services, such as individual feedback for hospitals, to assure high quality standards in the German healthcare market.[12] However, the quality report cards contain technical terms too complex to understand without significant clinical knowledge. Therefore, with the aim of giving patients the opportunity to form an opinion about hospital quality in a more digestible format, several web-based hospital comparison portals have been launched to provide a comprehensive and easily accessible hospital quality ranking for prospective patients.

3 Data

3.1 Inpatient care data

Our empirical analysis uses patient-level data collected from hospital discharge records. The discharge data is based on diagnosis related group (DRG) reimbursement claims from a nationally representative sample of clients hospitalized between 2009 and 2012 and provided by a large German health insurance company. It includes a wide range of patient characteristics and comprehensive information about medical symptoms and administered treatments during the hospital visit. Clinical procedures performed by hospital physicians are coded according to the German classification of medical operations and procedures, *Operationen- und Prozedurenschlüssel (OPS-12)*. To identify deliveries we use information on the cause for each admission, classified according to the *World Health Organization's International Statistical Classification of Diseases and Related Health Problems (ICD-10)*.[13]

Our population of interest is restricted to expectant mothers, 18 to 51 years of age, who gave birth at a maternity clinic located in a German hospital. We identify and sample patients in this age range with a singleton

the market.

[12] BQS managed the development and implementation of the external quality assurance system in Germany from 2001 to 2009, after which AQUA took over responsibility of this task (Busse *et al.*, 2009).

[13] Specifically, to identify deliveries we rely on ICD-10 codes: O80 (spontaneous delivery), O81 (delivery by forceps and vacuum extractor) O82 (delivery by cesarean section). We do not include multiple births in our analysis as they are considered risky deliveries and therefore subject to additional patient choice restrictions.

hospital delivery (in total around 250,000 deliveries), excluding births occurring outside of specialized departments (6,457 births or around 2% of the sample). Furthermore, we apply the Elixhauser (Elixhauser *et al.*, 1998) index, computed from secondary medical diagnoses coded in the hospital data, to account for patient case-mix variation in terms of baseline health status.[14]

3.2 Quality data

We merge the inpatient data described in the previous section with a set of objective (OQ) and subjective (SQ) hospital quality indicators. These indicators are obtained from publicly available quality report cards, which all hospitals are by law required to provide, and a patient satisfaction survey conducted by Techniker Krankenkasse (TK),[15] a large German SHI fund. In order to use quality information that prospective patients are most likely to use, we attempt to match as far as possible the criteria that the largest provider search platform in Germany, *weisse-liste.de*, bases its hospital ranking on.[16,17] Appendix B provides a brief summary of the search features the website offers. However, it is important to note that the search portal has changed considerably over time and the information currently reported is quite different from the information provided during the time period our analysis covers.

One important aspect of the quality data is that it is reported biannually while we base our empirical analysis on annual information from the hospital discharge data. However, except for reducing empirical variation in our data, this does not constitute an important problem; it simply means that the information prospective patients have access to (which we are primarily interested in) is only updated every second year. Hence, for years where hospital quality was not updated, we retain the previous year's quality measures for each hospital. Below we give a brief description of the different quality indicators we use in our analysis.

---

[14] The Elixhauser Comorbidity Index (ECI) distinguishes 31 different comorbidities and is often used as a risk-adjustment tool to predict hospital resource use and in-hospital mortality. For a list of comorbidities we include in our analysis, see Table A.1 in the Appendix A.

[15] Techniker Krankenkasse, founded in 1884, is one of Germany's largest social health insurance funds with a market share of about 14 percent, or 10 million clients (as of 2018).

[16] *Weisse Liste* is administered and maintained by the independent Bertelsmann Foundation and can be reached at https://www.weisse-liste.de. Pross *et al.* (2017) show that this online platform is frequently used for provider search in Germany. Although our main empirical specification does not exactly correspond to the variable definitions provided on *weisse-liste.de*, we have performed sensitivity analyses where our quality indicators are defined exactly as in the provider search platform, with qualitatively similar results.

[17] The subjective quality data provided by TK closely corresponds to a patient satisfaction survey conducted jointly by two other large German SHI funds, Allgemeine Ortskrankenkasse (AOK) and BARMER. Results from the latter are since 2012 published on *weisse-liste.de* while results from the former are published on a similarly widely used search portal: https://www.tk.de/tk/klinikfuehrer.

Quality report cards

The quality report cards include detailed information on numbers of cases and procedures performed for each hospital department. Furthermore, they also provide an overview of available medical and nursing services, existence of special departments and equipment, and a set of quality indicators measuring the structure, process, and clinical outcomes in the hospital. We employ three OQ indicators that account for the quality of mandatory services in the maternity clinic. For consistency and ease of interpretation, we redefine these quality indicators in our empirical analysis so that a more positive value of the indicator always corresponds to higher quality. Furthermore, we include a set of indicators for available services that a given clinic offers in addition to mandatory maternal services. These are categorized into medical and nursing services and care specialities, respectively. Figure 3.1 presents the hospital distribution of the OQ indicators we include in our analysis. We define and explain the different quality indicators in turn below.

- *Decision-to-delivery interval (DDI):* In some cases an emergency cesarean section (ECS) is necessary in order to avoid irreversible damage to the infant (e.g., due to a lack of oxygen). The time span between the decision made for performing an ECS and the delivery is termed decision-to-delivery interval (DDI). According to current recommendations by the German Association for Gynecology and Obstetrics, the procedure should be performed within 20 minutes from the decision (DGGG, 1995). Hospitals may improve their process structure and organization through a reduction of DDI, for example, by providing stand-by facilities or staff for emergency duties. DDI is a process quality indicator calculated as

$$\text{DDI} = \frac{\text{All ECS deliveries with DDI below 20 minutes}}{\text{All ECS deliveries}}.$$

  The upper left panel of Figure 3.1 shows that almost all hospitals fully comply with a DDI below 20 minutes, i.e., a DDI indicator close to one.

- *Availability of paediatrician:* This process indicator refers to deliveries of premature infants with a gestational age (GA) of less than 37 weeks. In such cases, a paediatrician should attend the delivery and, if needed, provide necessary medical treatment to the newborn. This indicator is calculated as

$$\text{Pediatrician} = \frac{\text{Access to pediatrician for births with GA} < 37 \text{ weeks}}{\text{All live births with GA} < 37 \text{ weeks}}.$$

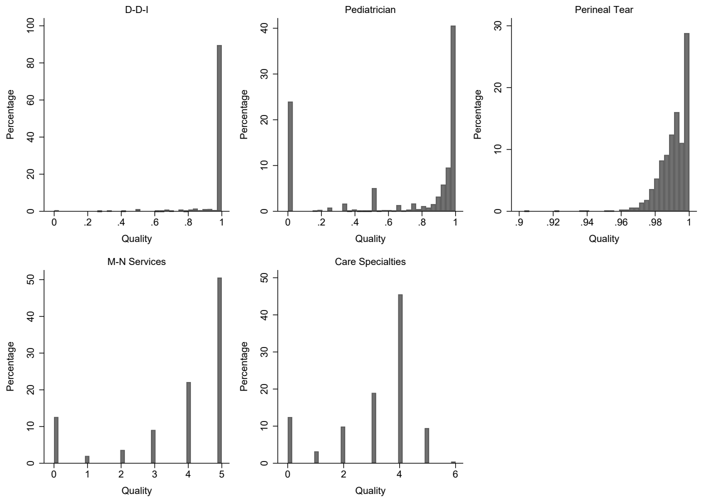
  The distribution of this indicator is depicted in the upper middle panel of Figure 3.1. The figure shows that, while most hospitals have a paediatrician attending the vast majority of premature births, a substantial proportion do not have this option available at all.

- *Perineal tear:* A perineal tear is a category of obstetric trauma which can be either light and curative (degree 1-2), or heavy and potentially chronic (degree 3-4). A heavy perineal tear is considered a preventable condition and, as such, a commonly used patient safety indicator for hospital quality. Since assisted and surgical deliveries are generally more risky births, this indicator is calculated as the ratio of the absence of heavy perineal tears among all spontaneous (i.e., non-assisted) births
$$\text{Perineal tear} = \frac{\text{Absence of heavy perineal tear}}{\text{All spontaneous deliveries}}.$$
The upper right panel of Figure 3.1 indicates that this outcome indicator exercises some variation across hospitals, although trauma rates are very unlikely to exceed 0.05.

- *Medical & Nursing services:* Medical and nursing services (M-N Services) comprise a maximum of five complementary medical services a hospital may offer to expectant mothers: postpartum exercises, prenatal classes, infant care classes, breastfeeding advice, and additional midwife services (such as, e.g., water births). Figure 3.1 shows considerable variation across the maternity clinics with respect to the availability of these services.

- *Care Specialities:* Care specialities comprise a maximum of six complementary medical care specialities a hospital may offer to expectant mothers: prenatal diagnosis, surgery to ease delivery, assistance for high-risk pregnancies, advice for high-risk pregnancies together with a gynaecologist, examination of diseases during the pregnancy, delivery, and the postpartum period, and (outpatient) delivery without a stay at the maternity clinic. Also for this indicator, Figure 3.1 shows substantial heterogeneity across maternity clinics.

FIGURE 3.1: Distribution of objective quality (OQ) indicators



NOTE.— Empirical distributions of the objective quality (OQ) indicators (see Section 3.2). The distributions of D-D-I, Paediatrician and Perineal Tear refer to shares between zero and one while the distributions of M-N Services and Care Specialities refer to discrete values between zero and five and six, respectively.

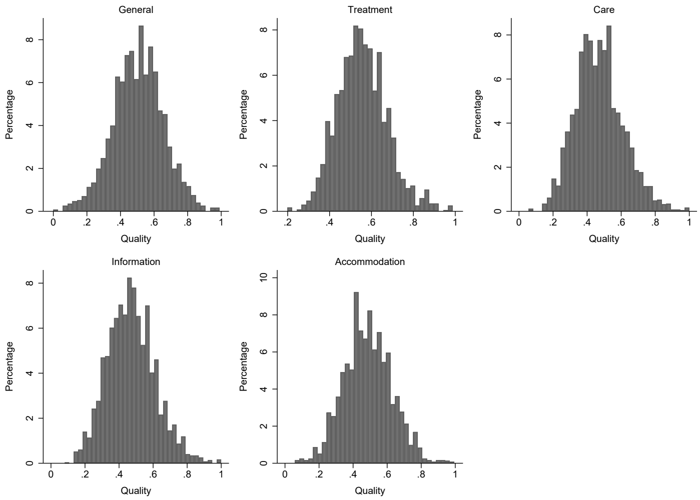
Patient satisfaction

Since 2006, TK has conducted a biannual survey of its clients' experiences with the care they received during their last hospital visit (Techniker Krankenkasse, 2010)[18] The questionnaires are sent to a random sample of clients, with exceptions for individuals older than 80 years or in need of long-term care.[19] The survey consists of 41 questions partitioned into five categories where the participants are asked to rate their general satisfaction with the hospital visit, the results of treatment, the medical and nursing care, the communication with the hospital staff, and the organization and accommodation during the stay. Each question is evaluated by assigning points on a 12 point likert scale where more points indicate higher satisfaction. For ease of interpretation, answers were subsequently aggregated to the category level and rescaled to lie within the unit interval. Figure 3.2 shows the distributions for each satisfaction category.

[18]The hospital ranking and a document explaining the survey method can be found at https://www.tk.de/tk/klinikfuehrer.

[19]For each hospital between 150 and 1,000 patients were asked to participate in the survey. The response rates were quite high. For example, in 2010 more than 61% of surveyed patients responded (Pilny and Mennicken, 2014). However, the results were only reported when at least 60 questionnaires were fully completed. In our sample around 22% of all hospitals were unable to comply with this requirement. We account for missing quality information by including a dummy variable for each hospital where satisfaction data is unavailable.

FIGURE 3.2: Distribution of subjective quality (SQ) indicators



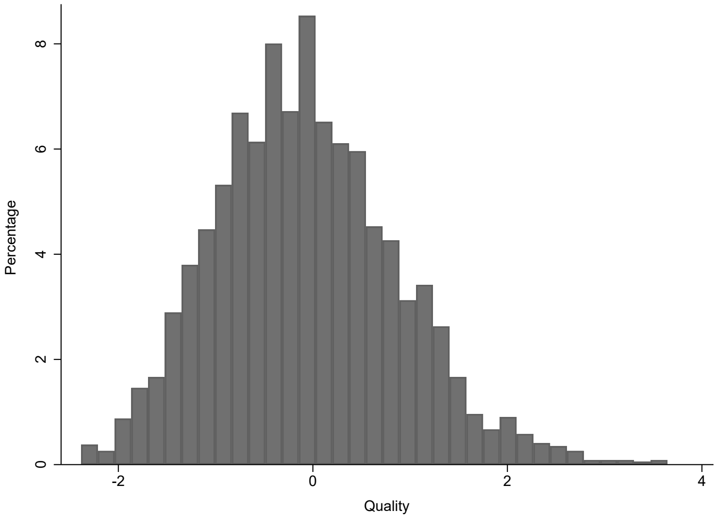
NOTE.— Empirical distributions of the subjective quality (SQ) indicators (see Section 3.2). The distributions refers to (rescaled) shares of patients in the Techniker Krankenkasse (TK) survey who were satisfied with the overall, treatment, care, information and accommodation of their last hospital visit, respectively.

One potential issue with including all the five satisfaction categories of the TK survey simultaneously in our econometric model is that they are likely to be highly internally correlated. For example, a patient who was unsatisfied with the treatment she received is also more likely to respond more negatively with respect to general satisfaction. The first panel of Table 3.1 displays a correlation matrix across the five SQ indicators, confirming our suspicion that correlations across the different satisfaction categories are indeed very high. In comparison, the correlation coefficients across the OQ indicators, reported in the middle panel of the same table, are much smaller in magnitude. Finally, the bottom panel of Table 3.1 reports the correlations between OQ and SQ indicators. Interestingly, coefficients are in general negative, implying that hospitals with high OQ are associated with lower SQ and vice versa.

Due to the high correlations across the SQ indicators, we apply a principal component analysis (PCA) to extract and summarize the information content of the five categories into a single satisfaction score index.[20] Since results from estimation will be interpreted in units of standard deviations from standardized coefficients, the exact scaling of the variable is unimportant. Figure 3.3 illustrates the distribution of the composite subjective quality (CSQ) score.

[20]Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. In our case, the number of principal components turns out to equal exactly one.

Figure 3.3: Distribution of the composite subjective quality (CSQ) indicator



Note.— Empirical distribution of the composite subjective quality (CSQ) indicator (see Section 3.2). The CSQ is constructed by application of principal component analysis (PCA) on the five satisfaction categories of the Techniker Krankenkasse (TK) patient satisfaction survey (see Figure 3.2).

Table 3.1: Correlation coefficients across quality indicators

I. Subjective quality (SQ)	General	Treatment	Care	Information	Accommodation
General	–				
Treatment	0.727***	–			
Care	0.872***	0.772***	–		
Information	0.860***	0.786***	0.937***	–	
Accommodation	0.823***	0.660***	0.822***	0.786***	–
CSQ score	0.894***	0.868***	0.963***	0.957***	0.889***
II. Objective quality (OQ)	D-D-I	Paediatrician	Perineal Tear	M-N Services	Care Specialities
D-D-I	–				
Paediatrician	0.055***	–			
Perineal Tear	0.021	0.192***	–		
M-N Services	0.054***	0.220***	0.100***	–	
Care Specialities	0.013	0.285***	0.143***	0.521***	–
III. OQ/SQ	D-D-I	Paediatrician	Perineal Tear	M-N Services	Care Specialities
General	−0.004	−0.152***	−0.046	0.044	−0.061*
Treatment	−0.064*	−0.275***	−0.099***	0.030	−0.117***
Care	−0.065**	−0.277***	−0.097***	0.034	−0.169***
Information	−0.050*	−0.253***	−0.098***	0.046	−0.134***
Accommodation	−0.044	−0.226***	0.005	0.006	−0.071**
CSQ score	−0.060*	−0.279***	−0.079**	0.032	−0.134***

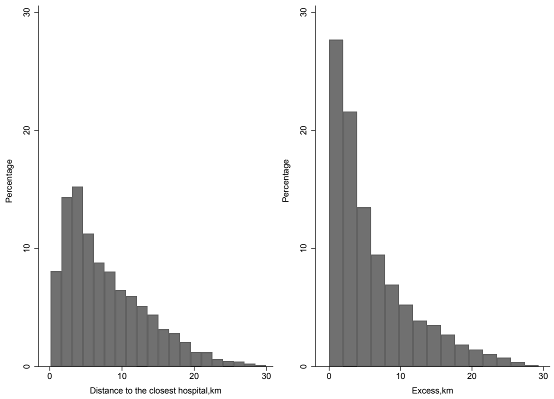
Note.— Estimated Pearson correlation coefficients across subjective quality (SQ) indicators (panel I), objective quality (OQ) indicators (panel II), and cross-correlation coefficients across SQ and OQ indicators (panel III). See Section 3.2 for definitions. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.3 Distance from hospital and choice sets

To measure the geographical distance for a patient to a hospital with maternal care capacity, we use the 5-digit postal code of patient's registered home and the postal address of each hospital, both of which are available in our data.[21] We estimate both the travel distance and the travel time for each patient-hospital combination using geocoding API software from Google® and Open Source Routing Machine (OSRM).[22]

The left panel of Figure 3.4 presents the distance distribution from each patient's home to the closest hospital in our sample. The resulting distribution is highly right-skewed with a range between 0 to 30 kilometres and a mean of 7.8 kilometres. In addition, the right panel of the figure shows the distribution of the *excess* distance patients travel between the closest and the chosen hospital. Although the mean of the excess distribution is only three kilometres, it has a substantial range. For example, more than ten percent of patients travel at least ten kilometres more than necessary to give birth. This suggests that patients value other factors than only geographical distance when choosing hospital.

FIGURE 3.4: Distribution of patient choice of maternity clinic: Distance and excess distance



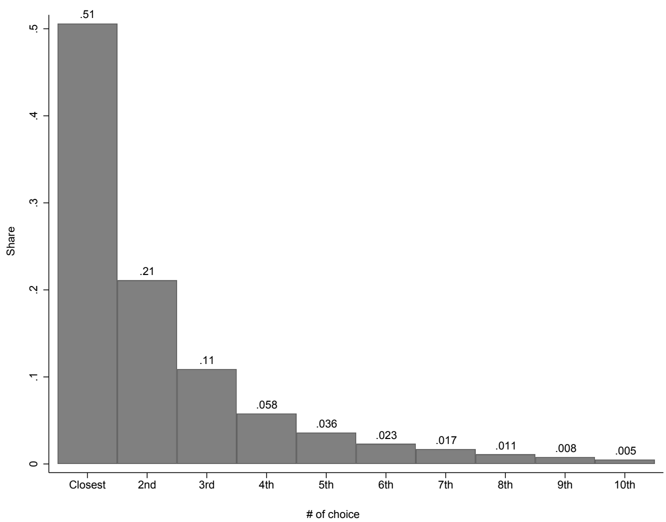
NOTE.— Empirical distribution of the distance to the *closest* hospital (left panel) and the *excess* distance between the closest and the chosen hospital (right panel) in a patient's choice set.

[21]This approach follows, e.g., Hentschker and Mennicken (2015, 2018); Mennicken *et al.* (2014) and implicitly assumes that patients travel from the geographic centroid of each 5-digit postal code area corresponding to its geographic center. There are about 8,200 5-digit postal code in Germany with a median size of 27 km^2 and the vast majority below 100 km^2. When interpreting the results from estimation, it is worth noting that there are no obvious reasons why any measurement errors arising from this simplification would be systematically related to quality indicators of individual hospitals. In fact, if patients and hospitals are randomly located with respect to the centroid of a given postal code, the measurement error would have zero unconditional expectation.

[22]For a documentation of the latter resource, see http://project-osrm.org/ and Huber and Rust (2016). We exclude a few cases where measuring the distance to a hospital was not possible, such as patients living on an island without a road connection to a hospital.

In order to estimate our choice model (described below), we define a choice set (i.e., a local hospital market) for each patient. To this end, we include all hospitals within a radius of 30 kilometres from the individual's place of residence (corresponding to the 90$^{\text{th}}$ percentile of the sample distance distribution to the chosen hospital). Consequentially, since 10% of patients choose a hospital outside of their choice set, our sample is reduced to around 225,000 births. From this definition the maximum number of choices provided to any patient in our sample is 25.[23] Figure 3.5 shows a histogram displaying the share of patients in our sample who gave birth in hospitals ranked by distance from the patient's home. Roughly half of patients chose to give birth in their closest hospital, while the remaining patient share in a gradually declining fashion chose hospitals located farther away from their homes.

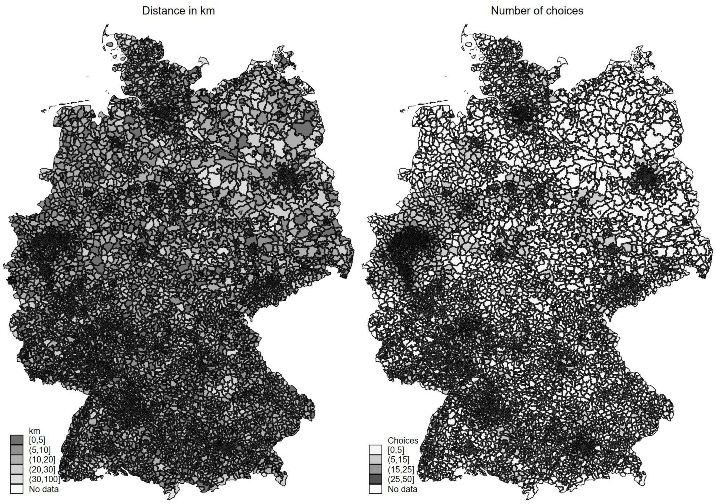
FIGURE 3.5: Distribution of patient choice of maternity clinic: Hospital distance rank



NOTE.— Empirical distribution of patient choice of hospital ranked by distance from patient residence (indicated from left/closer to right/farther).

[23] We have evaluated the robustness of our results with respect to the definition of the choice set by estimating separate models for a maximum of 15 and 20 choices with very similar results.

Figure 3.6: Geographical distribution of maternity clinics in Germany

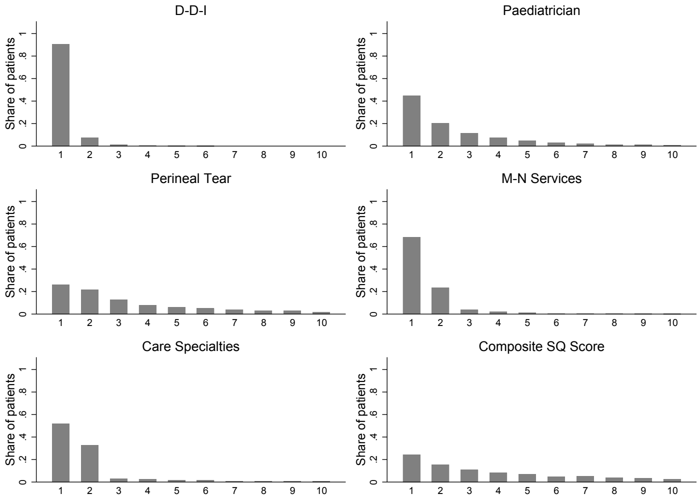


Note.— Distance in km to the closest hospital with a maternity clinic (left panel) and number of clinics within a 30 kilometres radius (right panel) by postal code.

To visualize the geographical variation in access to maternal care in Germany, Figure 3.6 provides two maps of Germany showing the average distance to the closest maternity clinic (left panel) and density of maternity clinics within a radius of 30 km (right panel) by postal code, respectively. While inhabitants of most parts of Germany have less than 30 kilometres to the nearest maternity clinic, the number of choices varies substantially across the country. The metropolitan areas of North Rhine-Westphalia, Hamburg, Berlin, Frankfurt, Stuttgart and Munich often have more than 25 choices while the sparsely populated areas in Eastern Germany often have less than five.

Figure 3.7 presents the unconditional distribution of patients' choice of hospital ranked from best to worst in their choice set by quality indicator. For the purpose of presentation, we show only the ten highest-ranked hospitals as the shares of patients choosing lower ranked hospitals become very small. Although the intensity of the pattern varies across indicators, all measures exhibit a positive relationship between a hospital's reported quality and its relative popularity. There are strong positive associations for the decision-to-delivery interval, paediatrician availability, medical and nursing services and care specialities indicators, while patterns are less clear for the remaining quality indicators.

FIGURE 3.7: Distribution of patient choice of maternity clinic: Hospital quality rank



NOTE.— Shares of patients who chose the hospital with the best (1) to the worst (10) quality rating in their choice set by quality indicator (see Section 3.2).

3.4 Sample summary statistics

Table 3.2 reports sample summary statistics by different levels of data aggregation. From upper-left to lower-right the panels refer to variable means and standard deviations on the patient, choice-set, hospital and closest hospital levels, respectively. Around one-third of the roughly 225,000 births in our sample are classified as emergencies. To account for the fact that emergency patients are unlikely to have full discretion in their choice of hospital, we include an indicator variable for whether the hospital admission was coded as an emergency in our regression models. We also adjust for other factors that may have affected the individual's choice of maternity clinic, such as patient age and case-mix controls for the number of Elixhauser comorbidity indicators, whether the admission occurred on the weekend or during rush hour, whether the patient lived in a rural or an urban postal code and whether the birth was coded as being risky or not. Summary statistics for these variables are reported in the two top panels of Table 3.2, corresponding to the level of the patient (left) and the choice set (right).

Table 3.2 also provides some summary statistics on the distance variables. Specifically, although the average patient had approximately eight kilometres (12 minutes) to the closest hospital, she nevertheless chose a hospital at a distance of 11 kilometres (14 minutes) from her home. Around one-half of the expectant mothers did not choose their closest hospital, but instead went to a hospital located at an additional three kilometres distance, on average. The corresponding figures for the choice sets closely resemble the individual level counterparts, except for a larger share of rural choice sets and increased distances and travel times to the closest and chosen hospitals, respectively. Our econometric approach (explained below) accounts for bias from such heterogeneity by only using within-individual variation across choices to estimate the parameters of the model.[24]

The two bottom panels of Table 3.2 present hospital-level summary statistics of the quality indicators we include in our analysis. The left panel refers to the universe of hospitals while the right panel only considers the closest hospital in each choice set. Around 22 percent of hospitals lacked information about SQ (see footnote 19). To handle this missing data issue and simultaneously keeping the choice sets intact, we impute a value of zero for each observation for which quality information is unavailable and include an binary indicator variable in our econometric model to distinguish these missing values from "true" zeros.

To avoid confounding between our quality indicators and other hospital characteristics, we include a set of control variables related to the (perceived) performance of a hospital in our models, such as ownership type and whether the hospital is a teaching or a university hospital. We also include a set of capacity-related variables, such as the number of midwives and nurses, number of beds and the share of specialized physicians in the hospital. Finally, to account for any time-invariant unobserved heterogeneity in perceived hospital quality, we estimate models with hospital fixed-effects.

[24]To study potential heterogeneous effects between rural and urban choice sets and between emergency and non-emergency admissions, we estimate models where these regressors are interacted with the quality indicators. The results from this analysis are provided in Section 5.2 below.

Table 3.2: Descriptive sample statistics

	Patient		*Choice-set*	
	Mean	SD	Mean	SD
Patient characteristics				
Age in years	31.16	[5.05]	31.26	[2.14]
# Elixhauser conditions	0.17	[0.41]	0.15	[0.17]
If emergency	0.30	[0.46]	0.29	[0.29]
If weekend	0.23	[0.42]	0.23	[0.16]
If rush hour	0.37	[0.48]	0.37	[0.18]
If risky	0.04	[0.21]	0.04	[0.08]
Choice-set characteristics				
If urban postal code	0.76	[0.43]	0.72	[0.45]
If closest hospital chosen	0.51	[0.50]	0.48	[0.35]
Excess distance	2.95	[4.97]	3.29	[3.45]
Distance closest hospital (km)	7.81	[5.66]	9.80	[6.15]
Travel time closest hospital (min)	11.54	[7.11]	13.32	[7.51]
Distance chosen hospital (km)	10.76	[7.36]	13.09	[6.70]
Travel time chosen hospital (min)	14.32	[8.14]	16.33	[7.63]
Observations	225,352		8,666	
	Hospital		*Closest hospital*	
	Mean	SD	Mean	SD
Objective quality indicators (OQ)				
D-D-I	0.99	[0.08]	0.98	[0.08]
Paediatrician	0.32	[0.44]	0.48	[0.44]
Perineal Tear	0.99	[0.01]	0.96	[0.07]
M-N Services	3.78	[1.70]	3.94	[1.56]
Care Specialities	3.04	[1.46]	3.22	[1.39]
Subjective quality indicators (SQ)				
General	0.26	[0.27]	0.33	[0.25]
Treatment	0.29	[0.29]	0.37	[0.26]
Care	0.25	[0.26]	0.31	[0.23]
Information	0.25	[0.26]	0.31	[0.23]
Accommodation	0.25	[0.26]	0.32	[0.23]
Composite SQ Score	-0.03	[0.68]	-0.06	[0.72]
Hospital characteristics				
If public	0.42	[0.49]	0.40	[0.49]
If private	0.17	[0.37]	0.44	[0.50]
If university	0.03	[0.17]	0.03	[0.16]
If teaching	0.40	[0.49]	0.52	[0.50]
Birth-staff ratio	176.72	[136.88]	192.92	[136.14]
Share specialized physicians	0.56	[0.16]	0.41	[0.27]
# hospital beds	389.95	[345.94]	403.85	[345.59]
# hospital midwives	8.03	[7.82]	9.26	[8.59]
# hospital nurses	3.34	[5.84]	3.82	[5.05]
Observations	6,545		8,666	

Note.— Descriptive statistics for different levels of data aggregation. Characteristics on the patient level (top-left), choice-set level (top-right), hospital level (bottom-left) and on closest hospital level (bottom-right). Emergency, rush hour and risky are dummy variables indicating whether a patient was admitted as an emergency case, whether admitted between 6am–10am on weekdays, and whether the patient was diagnosed with an secondary ICD-10 code of Z.35 (supervision of high-risk pregnancy), respectively. Excess distance is defined as the additional distance (in kilometres) between the closest and the chosen hospital. The Composite SQ score is constructed by application of principal component analysis (PCA) based on 5 satisfaction variables (see Section 3.2).

4 Econometric framework

4.1 Theoretical predictions

We assume that a patient i choosing a hospital $k = 1, 2, ..., K$ values both higher quality (Q) and shorter distance (D). However, the patient may restrict her search to providers within the same hospital market $j = 1, 2, ..., J$, yielding $k_j = 1, ..., K_j$ choices for each of the J hospital markets. Formally, patient utility is modelled by the additively separable utility function

$$U_{ijk} = f(Q_{k_j}, D_{k_j}) + \varepsilon_{ijk}, \tag{4.1}$$

where we assume that utility is (weakly) increasing in the first and decreasing in the second argument. We also allow the patient to have heterogeneous preferences for a particular hospital as long as these preferences are unrelated to Q and D.

Assume that patient i is choosing between two hospitals in the same hospital market (j subscript omitted), 1 and 2, with corresponding utility functions

$$\begin{aligned} U_i(1) &= f(Q_1, D_1) + \varepsilon_{i1}, \\ U_i(2) &= f(Q_2, D_2) + \varepsilon_{i2}, \end{aligned} \tag{4.2}$$

The probability that the patient will choose hospital 1 is then

$$\begin{aligned} p_{i1} =& \Pr[U_i(1) > U_i(2)] = \\ & f(Q_1, D_1) - f(Q_2, D_2) + \mathbb{E}[\varepsilon_{i1} - \varepsilon_{i2}] > 0. \end{aligned} \tag{4.3}$$

Given identical preferences for the two hospitals, the last term collapses and the decision rule is to choose hospital 1 whenever $f(Q_1, D_1) > f(Q_2, D_2)$.[25]

In order to define an economically relevant measure for the trade-off the patient is facing, we consider the additional distance an individual is willing to travel in order to obtain a given increase in quality, i.e., $\partial D/\partial Q > 0$. Totally differentiating (4.1) we get (omitting subscripts)

$$dU(Q, D) = (\partial U(Q, D)/\partial Q)\, dQ + (\partial U(Q, D)/\partial D)\, dD, \tag{4.4}$$

which is equivalent to

$$\frac{dU(Q, D)}{dQ} = \partial U(Q, D)/\partial Q + (\partial U(Q, D)/\partial D)\, \frac{dD}{dQ}. \tag{4.5}$$

Since we evaluate the trade-off between quality and distance, implying that

[25] Generalizing to k hospitals the corresponding decision rule is $p_{ik} = \Pr[U_{ik_j} > U_{i-k_j}, \text{all} - k_j \neq k_j]$.

total utility is held constant, the left hand side is equal to zero and so

$$\frac{\partial D}{\partial Q} = -\frac{\partial U(Q,D)/\partial Q}{\partial U(Q,D)/\partial D}, \tag{4.6}$$

where the right hand side term is the marginal rate of substitution of quality for distance, $MRS_{Q,D}$. Given a suitable empirical specification for individual utility, we can estimate the willingness to travel for a patient for a given increase in hospital quality.

Our hypotheses can be summarized as follows: (i) a patient's likelihood of choosing a particular hospital will increase with reported maternal care quality and decrease with the distance to the hospital (i.e., $f_1(Q,D) > 0$ and $f_2(Q,D) < 0$) and; (ii) a patient will trade off additional distance to a maternity clinic in a hospital with a higher reported quality, i.e., $\partial D/\partial Q > 0$.

4.2 Reduced form analysis

We first consider a simple linear probability model (LPM) for choosing the closest hospital in the choice set as a function of hospital quality. Specifically, for individual i in choice set j and year t, the LPM is defined by

$$Closest_{ijt} = \alpha_0 + f(d^c_{jt};\alpha_d) + q^{c\prime}_{jt}\beta_q + X'_{it}\Theta_X + Z^{c\prime}_{jt}\Theta_{Z^c} + \bar{Z}'_{jt}\Theta_{\bar{Z}} + \epsilon_{ijt}, \tag{4.7}$$

where $Closest_{ijt}$ is a binary indicator for whether a patient chose the closest hospital in her choice set.[26] Similarly, d^c_{jt} and q^c_{jt} indicate the distance (scalar) and quality (vector) of the closest hospital in the individual's choice set, where $f(\cdot)$ is a cubic polynomial function of d^c_{jt} with corresponding parameter vector α_d. Furthermore, X_{it}, Z^c_{jt}, and $\bar{Z}_{jt} = N^{-1}\sum_k z_{jkt}$ are vectors of patient, closest hospital, and average choice set specific variables as reported in Table 3.2, respectively. Finally, ϵ_{ijt} is an assumed random regression error term. We cluster standard errors on the level of spatial planning regions (96 clusters) to account for any residual correlation across individuals living in the same region.[27] We are primarily interested in the signs of the $\hat{\beta}_q$ vector, which inform us about whether an improvement in a given quality indicator of the closest hospital increases the likelihood of choosing it relative to other hospitals in the same choice set. Since higher values of all quality correspond to better quality, we expect all coefficients to be positive.[28]

[26] That is, $Closest_{ijt}$ evaluates to one if the chosen hospital k_j satisfies $k_j : d_{jkt} = \min(d_{jt})\ \forall\, k_{jt} \in (j,t)$.

[27] We have also clustered standard errors on the local (kreise) and the state (land) levels, yielding very similar results.

[28] We have also estimated models where we use the relative quality compared to the average quality in the choice set instead of including the absolute quality of the closest

4.3 Structural choice modelling

Inferences derived from estimation of the LPM in (4.7) are generally uninformative about the hospital distance-quality trade-off a patient faces. Therefore, we also consider a structural econometric framework for hospital choice based on estimation of a conditional logistic regression model. The advantage of this approach is that it allows us to derive and estimate the additional distance a patient is willing to travel to a hospital in exchange for an increase in reported quality.

Departing from our theoretical framework in (4.1), the random utility model specifies

$$U_{ikt} = V_{ikt} + \xi_{kt} + \mu_{ikt} \quad \text{for } (i,k) \in j, \tag{4.8}$$

where utility of individual i of choosing hospital k in year t in choice set j is a linear function of observable individual and hospital characteristics V_{ikt} (e.g., reported quality indicators), unobservable hospital characteristics ξ_{kt} (e.g., hospital reputation), and unobserved individual heterogeneity, μ_{ikt} (e.g., patient preferences). Assuming that μ_{ikt} is i.i.d. and type I extreme value distributed, the probability that patient i chooses hospital k can be written on the logistic form as (see, e.g., Cameron and Trivedi, 2005)

$$\begin{aligned} p_{ikt} = &\Pr[y_{it} = k] = \\ &\exp\left(V_{ikt} + \xi_{kt}\right)\left[\sum_{k' \in j} \exp\left(V_{ik't} + \xi_{k't}\right)\right]^{-1}, \quad k = 1, \ldots, K_j, \end{aligned} \tag{4.9}$$

where the dependent variable y_{ikt} is defined as

$$y_{ikt} = \begin{cases} 1 & \text{if } y_{it} = k, \\ 0 & \text{if } y_{it} \neq k. \end{cases} \tag{4.10}$$

Individual utility is assumed to be represented by the linear model

$$\begin{aligned} U_{ikt} = &\sum_{p}^{P} \gamma_{pt}^{q} q_{kt}^{p} + \sum_{s}^{S} \gamma_{st}^{d} d_{ikt}^{s} + \sum_{p}^{P} \sum_{m}^{M} \gamma_{pmt}^{qx} q_{kt}^{p} \tilde{x}_{imt} \\ &+ \sum_{s}^{S} \sum_{m}^{M} \gamma_{mst}^{dx} d_{ikt}^{s} \tilde{x}_{imt} + \sum_{l}^{L} \gamma_{lt}^{z} z_{kt}^{l} + \nu_{ikt}, \end{aligned} \tag{4.11}$$

where q_{kt}^{p} refers to the pth quality indicator and d^s to the sth polynomial order for the (cubic) distance relation. Furthermore, $\tilde{x}_{imt} = x_{imt} - \bar{x}_m$ is

hospital. This alternative specification does not change our results to any important extent.

the mean-centered value of the mth individual characteristic with $\bar{x}_m = N^{-1} \sum_t \sum_i x_{imt}$ and z^l_{kt} is the lth hospital characteristic reported in Table 3.2. The vector **fl** $= (\gamma^q, \gamma^d, \gamma^{qx}, \gamma^{dx}, \gamma^z)$ comprises the set of coefficients to be estimated. Finally, the joint error term $\nu_{ikt} = \xi_{kt} + \epsilon_{ikt}$ is assumed to be i.i.d. conditional on included individual and hospital level control variables.

Endogeneity concerns could arise if, for example, private or teaching hospitals are perceived by individuals as being of different quality than public or non-teaching hospitals, or if lower quality hospitals are exiting the market due to fierce competition. We assume that our included hospital specific variables account for the former concern, and the fact that very few hospitals are closed during the relevant years suggests that the latter is unlikely to be a severe problem here.[29] Furthermore, the conditional logit model uses within-individual variation to estimate the model parameters, implying that any potential biases from choice-invariant factors (e.g., average distance to and quality of hospitals in the choice set) are effectively accounted for in the analysis. Nevertheless, in order to account for potential unobserved heterogeneity in (perceived) hospital quality, we also estimate models with hospital fixed-effects. The inclusion of hospital fixed-effects implies that empirical variation in quality across hospitals within a choice set are purged from the analysis and the model's parameters are exclusively estimated using changes within a hospital's quality indicators across time. Since the source of identifying variation is qualitatively different in the models with and without hospital fixed-effects, and consequently also the interpretation of the estimated parameters, we retain both specifications in the discussion of our results.

Mean-centering the individual patient characteristics allows us to both control for potential confounding factors and interpret the estimated γ^q_{pt} and γ^d_{st} as marginal utilities with respect to quality and distance for a patient with average characteristics in a given year. From the conditional logit model, described by equations (4.8)-(4.11), we can thus produce an estimate of the willingness to travel (WTT) for a representative patient to a hospital with a one standard deviation increase in the pth reported quality measure as (see, e.g., Moscelli *et al.*, 2016)

$$\begin{aligned} WTT(p) &= \sigma_p \frac{\partial d_{ikt}}{\partial q^p_{kt}} = \sigma_p \left(-\frac{\partial U_{ikt} / \partial q^p_{kt}}{\partial U_{ikt} / \partial d_{ikt}} \right) \\ &= \sigma_p \frac{-\gamma^q_{pt}}{\gamma^d_{1t} + 2\gamma^d_{2t}\zeta_d + 3\gamma^d_{3t}\zeta^2_d}, \end{aligned} \tag{4.12}$$

[29]According to the Federal German Statistical Office, there were 67 hospital market exits across Germany between 2009 and 2012, corresponding to around three percent of all German hospitals. Note that these numbers refer to all hospitals and not necessarily to hospitals with maternal health services. In addition, the total number of hospital beds barely changed from 503,341 to 501,475 (-0.4 percent) over the same time period (Destatis, 2018, p.11).

where the second equality is the negative of the marginal rate of substitution (see equation (4.6)) and the third equality is obtained from differentiation of (4.11) with a cubic distance representation. Furthermore, σ_p is the standard deviation of the pth quality measure and ζ_d is the average distance to the chosen provider for all patients over all years. To obtain standard errors for the WTT, we apply the delta method (see, e.g., Cameron and Trivedi, 2005).

5 Results

5.1 Main results

Table 5.3 reports results from estimation of the linear probability model for choosing the closest hospital in the choice set as specified in equation (4.7). In column (1) all OQ indicators are included together with a cubic distance polynomial and the full set of patient and hospital control variables from Table 3.2, while SQ is included through the five satisfaction categories in the TK survey, respectively. In column (2), SQ is instead included through the composite SQ score. Finally, the specification reported in column (3) additionally includes hospital fixed effects to account for time-invariant unobserved heterogeneity in hospital quality.

As expected, choosing the closest provider is negatively associated with distance and positively associated with the OQ indicators throughout the table. The estimated coefficients of the SQ categories in column (1) are only statistically distinguishable from zero for satisfaction with accommodation while the coefficients of the four other categories have negative signs, highlighting the issue of multicollinearity.[30] When we instead include the composite SQ score in columns (2) and (3) we obtain a positive and, in the latter case, also statistically significant point estimate of the SQ indicator.

Table 5.4 reports the estimated coefficients from the conditional logit model, defined by equations (4.8)-(4.11), including the full set of controls from Table 3.2. As before, SQ is first included using the five satisfaction categories, in column (1), and subsequently through the composite score, in columns (2) and (3), respectively. Choice of hospital is again negatively correlated with distance and the higher order terms suggest a diminishing association as distance increases. Furthermore, all OQ indicators have a positive impact on hospital choice across the different specifications. The issue of multicollinearity is again visible from observing the highly variable coefficient values of the individual SQ categories in column (1), but positive and statistically significant in column (2). Effect sizes are substantially

[30] Columns (1)-(5) of Table A.2 reports results, corresponding to the specification in column (3) of Table 5.3, where each satisfaction score category is included separately. All categories are individually positively associated with choosing the closest hospital, but, once all are included simultaneously in column (6), only one coefficient remains significantly distinguishable from zero.

attenuated for D-D-I, paediatrician and the composite SQ score quality indicators when hospital fixed effects are included in column (3), which, together with lower statistical precision, render the impact of these variables statistically indistinguishable from zero.[31]

TABLE 5.3: Linear probability model estimates for choosing the closest hospital: Main results

	(1)	(2)	(3)
Distance	-0.030***	-0.031***	-0.032***
	(-5.51)	(-5.58)	(-6.89)
Distance^2	0.001*	0.001*	0.001**
	(2.20)	(2.38)	(2.91)
Distance^3	-0.000	-0.000	-0.000*
	(-1.36)	(-1.58)	(-2.07)
D-D-I	0.033	0.049	0.150**
	(0.65)	(0.88)	(3.24)
Pediatrician	0.061***	0.070***	0.072***
	(4.10)	(4.59)	(5.43)
Perineal Tear	0.051	0.085	0.032
	(0.52)	(0.81)	(0.41)
M-N Services	0.031***	0.030***	0.026***
	(4.98)	(4.88)	(4.58)
Care Specialities	0.010	0.013*	0.002
	(1.78)	(2.31)	(0.46)
General	-0.021		
	(-0.14)		
Treatment	-0.041		
	(-0.35)		
Care	-0.226		
	(-1.19)		
Information	-0.096		
	(-0.53)		
Accommodation	0.340**		
	(2.75)		
Composite SQ score		-0.003	0.013*
		(-0.61)	(2.34)
Patient characteristics	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes
Hospital fixed effects	No	No	Yes
Observations	225,352	225,352	225,352

NOTE.— Linear probability model (LPM) estimates for whether a patient chose the closest hospital in her choice set. Patient characteristics include age, # of Elixhauser conditions, whether a patient was admitted as an emergency case, on a weekend, between 6am–10am on weekdays, and whether the patient was admitted with an secondary ICD-10 code of Z.35 (supervision of high-risk pregnancy). Hospital characteristics include ownership type, # of beds, if university, if teaching, # of midwives; Birth-Staff ratio (# of cases per doctor); share specialized doctors (# of specialized doctors as a share of all doctors); # of nurses and an indicator for whether subjective quality was missing. Composite SQ score is based on five satisfaction variables: General, Treatment, Care, Communication, Accommodation. Standard errors are clustered on spatial planning regions (96 clusters). t-statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[31] The inclusion of hospital fixed effects in the model implies that empirical variation in the outcome is restricted to within-hospital temporal changes in quality. This may restrict both the total variation in the model and in single regressors and lead to imprecisely estimated parameters. To study the extent of this problem, we analyzed the coefficient of variation (CV) for each of the quality indicators to see how much variation remains after removing the cross-sectional variation. Table A.3 shows that the CV is quite low for several quality indicators, suggesting that they change little over time.

TABLE 5.4: Conditional logit model estimates for choice of hospital: Main results

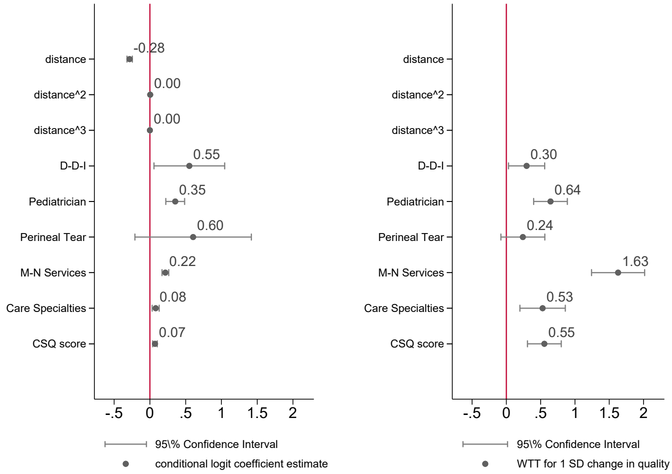
	(1)	(2)	(3)
Distance	-0.289***	-0.284***	-0.299***
	(-13.81)	(-14.35)	(-14.53)
Distance^2	0.004**	0.004**	0.003*
	(3.21)	(3.29)	(2.47)
Distance^3	-0.000	-0.000	-0.000
	(-1.37)	(-1.39)	(-0.30)
D-D-I	0.534**	0.551*	0.150
	(2.60)	(2.19)	(0.76)
Paediatrician	0.305***	0.354***	0.067
	(5.58)	(5.28)	(0.90)
Perineal Tear	0.270	0.604	0.493
	(0.72)	(1.46)	(1.70)
M-N-Services	0.217***	0.216***	0.384***
	(8.69)	(8.99)	(4.52)
Care Specialities	0.070**	0.081**	0.153**
	(3.08)	(3.21)	(2.72)
General	10.21***		
	(7.72)		
Treatment	-0.676		
	(-0.53)		
Care	-6.457*		
	(-2.40)		
Information	0.800		
	(0.34)		
Accommodation	-3.537**		
	(-2.72)		
Composite SQ score		0.071***	0.017
		(4.35)	(0.62)
Patient characteristics	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes
Hospital fixed effects	No	No	Yes
Observations	2,191,422	2,191,422	2,191,422

NOTE.— Conditional logit estimates for patient choice of hospital. Patient characteristics include age, # of Elixhauser conditions, whether a patient was admitted as an emergency case, on a weekend, between 6am–10am on weekdays, and whether the patient was admitted with an secondary ICD-10 code of Z.35 (supervision of high-risk pregnancy). Hospital characteristics include ownership type, # of beds, if university, if teaching, # of midwives; Birth-Staff ratio (# of cases per doctor); share specialised doctors (# of specialized doctors as a share of all doctors); # of nurses and an indicator for whether subjective quality was missing. Composite SQ score is based on five satisfaction variables: General, Treatment, Care, Communication, Accommodation. Standard errors are clustered on spatial planning regions (96 clusters). t-statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Finally, we use the estimated parameters from our conditional logit model to estimate the average WTT for a one standard deviation increase in reported quality. The results are presented graphically in Figure 5.8 (without hospital fixed effects) and Figure 5.9 (with hospital fixed effects), where the left and right panels refer to the point estimates from the conditional logit model and the WTT estimates using equation (4.12), respectively.

From Figure 5.8, a one standard deviation increase in reported quality for the three process quality indicators (D-D-I, perineal tear and paediatrician availability) are associated with increases in the WTT of between 0.2 to 0.6 kilometres, while an equivalent increase in quality for the service categories increases WTT by 0.5 (Care Specialities) and 1.6 (Medical and Nursing Services) kilometres, respectively. Finally, the corresponding figure is 0.6 kilometres for the composite SQ score. Hence, the range of WTTs are highly variable and dependent on the specific quality indicator.

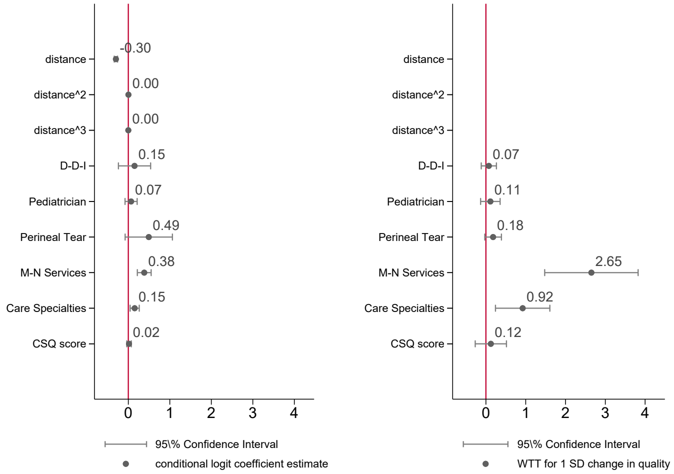
FIGURE 5.8: Willingness-to-travel (WTT) estimates *excluding* hospital fixed effects



NOTE.— Conditional logit estimates coefficients (left panel) and willingness-to-travel (WTT) estimates (right panel) *excluding* hospital fixed effects. WTT estimates are obtained from equation 4.12 using parameter estimates from the conditional logit model as inputs.

Turning to the results with hospital fixed effects in Figure 5.9, WTTs are generally attenuated (0.1-0.2 kilometres), except for the service categories in which WTT estimates for Care Specialities and Medical and Nursing Services increase to 0.9 and 2.7 kilometres, respectively. Given that the mean difference between the closest and the chosen hospital is about three kilometres (see Table 3.2), this upper bound of the WTT does not appear to be an unreasonable estimate. Taken together, depending on the specific indicator, a statistically representative patient is willing to travel between virtually no distance at all to about one-third of the average distance to the closest hospital to reach a hospital with a one standard deviation higher reported quality.

FIGURE 5.9: Willingness-to-travel (WTT) estimates *including* hospital fixed effects



NOTE.— Conditional logit estimates coefficients (left panel) and willingness-to-travel (WTT) estimates (right panel) *including* hospital fixed effects. WTT estimates are obtained from equation 4.12 using parameter estimates from the conditional logit model as inputs.

5.2 Robustness checks

We perform a falsification test to study whether our quality indicators pick up unobserved hospital quality not captured by the control variables we include in the model. Specifically, we use hospital-level information on mortality from heart attack patients, which is arguably irrelevant for the choice of maternity clinic of expectant mothers, to assess the extent to which such factors influence the choice of hospital for our sample. If such information is predictive of hospital choice, we might suspect that there are unobserved factors determining choice that we do not account for in the analysis. Table 5.5 reports the results from our main specification where we have included in-hospital and 30-days post-discharge acute myocardial infarction (AMI) mortality as additional regressors. Reassuringly, the choice of maternity clinic in our sample is not significantly associated with AMI mortality, irrespective of mortality definition.

We have also evaluated the robustness of our results with respect to our definition of choice set and travel distance. The results are shown in Table 5.6. The first four columns show the baseline specification using distance in kilometres with a maximum of 15 choices in each patient's choice set (columns 1 and 2) and with a maximum of 20 choices (columns 3 and 4), respectively. We also estimate models for a travel time-based definition (in minutes) with a maximum of 25 choices (columns 5 and 6). Our main findings do not change to any important extent when these definitions are changed.

To study heterogeneous effects across patient groups, we re-estimate the conditional logit model by additionally including interaction terms between our quality measures and indicators for whether the birth was classified as an emergency and whether the patient was residing in an urban area, respectively. Table 5.7 presents the results from this exercise. Columns (1) and (2) report level and interaction effects with the emergency indicator for each distance and quality indicator, where the former and latter set of parameters are interpreted as the magnitude of response from non-emergency cases and the additional impact from emergency cases, respectively. While the interaction coefficients with distance are estimated with negative signs, there is no evidence that emergency patients differ significantly from non-emergency patients in their preferences for distance. With respect to the set of quality indicators, one general discernible pattern is that indicators for clinical service capacity (D-D-I, Paediatrician and Care Specialities) seem to be more relevant for the choice of emergency patients, while indicators for clinical quality (Perineal Tear and CSQ) are less relevant. This is an intuitive result given that the priority of emergency services should be to direct the patient to a provider with the necessary capacity to assist the delivery of high-risk births.

Heterogeneity with respect to urban and rural choice sets is presented in columns (3) and (4) of Table 5.7. The impact of distance for patients residing in urban choice sets has an effect size that is more than twice that of patients living in rural areas. This suggests that competition in this dimension is much greater in urban areas. The results for quality suggest, with the exception of maternal trauma, that patients living in urban areas do not have systematical different preferences than those living in rural areas.

Finally, we test the robustness of our results by redefining the definition of the quality indicators more in line with the definition provided on *weisse-liste.de*. Specifically, the information provided on the provider search portal defines a quality threshold for each of the treatment relevant indicators for which a hospital has “passed” if the quality is at least as good or better than the required quality threshold according to a specified reference value set by the responsible authority. The first two columns of Table 5.8 report results from our linear probability model where we have redefined all quality indicators in this way. Furthermore, the last two columns provide estimates of two general quality categories: Mandatory Quality Assurance and Treatment-Related Services, defined as objective quality scores related to the quality and availability of services, respectively (see Appendix B for details). The results from this exercise show that the implications from our main analysis are largely unchanged.

TABLE 5.5: Conditional logit model estimates for choice of hospital: Falsification test

	(1)	(2)	(3)	(4)
Distance	-0.287***	-0.324***	-0.286***	-0.316***
	(-13.34)	(-15.25)	(-13.34)	(-14.63)
Distance^2	0.004**	0.005***	0.004**	0.004***
	(3.29)	(3.94)	(3.29)	(3.30)
Distance^3	-0.000	-0.000	-0.000	-0.000
	(-1.60)	(-1.69)	(-1.60)	(-1.08)
In-hospital AMI mortality	-0.119	-0.046		
	(-0.62)	(-0.28)		
30-day AMI mortality			-0.058	0.007
			(-0.30)	(0.04)
Patient characteristics	Yes	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes	Yes
Hospital fixed effects	No	Yes	No	Yes
Observations	2,191,422	2,191,422	2,191,422	2,191,422

NOTE.— Conditional logit estimates for patient choice of hospital including in-hospital (columns 1-2) and 30-days post-discharge (columns 3-4) acute myocardial infarction (AMI) mortality as additional regressors. Patient characteristics include age, # of Elixhauser conditions, whether a patient was admitted as an emergency case, on a weekend, between 6am–10am on weekdays, and whether the patient was admitted with an secondary ICD-10 code of Z.35 (supervision of high-risk pregnancy). Hospital characteristics include ownership type, # of beds, if university, if teaching, # of midwives; Birth-Staff ratio (# of cases per doctor); share specialized doctors (# of specialized doctors as a share of all doctors); # of nurses and an indicator for whether subjective quality was missing. Composite SQ score is based on five satisfaction variables: General, Treatment, Care, Communication, Accommodation. Standard errors are clustered on spatial planning regions (96 clusters). t-statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE 5.6: Conditional logit model estimates for choice of hospital: Heterogeneity analysis I

	Choice set definition				Travel time definition	
	15 choices		20 choices		25 choices	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance	-0.284***	-0.309***	-0.284***	-0.297***	-0.177***	-0.297***
	(-14.30)	(-15.57)	(-14.36)	(-14.18)	(-7.77)	(-8.96)
Distance^2	0.004**	0.004**	0.004**	0.003*	-0.004*	0.003
	(3.24)	(3.13)	(3.23)	(2.46)	(-2.52)	(1.31)
Distance^3	-0.000	-0.000	-0.000	-0.000	0.000**	-0.000
	(-1.25)	(-0.74)	(-1.26)	(-0.25)	(3.19)	(-0.36)
D-D-I	0.555*	0.171	0.552*	0.171	0.555	0.171
	(2.20)	(0.84)	(2.19)	(0.88)	(1.91)	(1.01)
Paediatrician	0.350***	0.074	0.354***	0.068	0.328***	0.0427
	(5.44)	(0.96)	(5.31)	(0.94)	(4.69)	(0.70)
Perineal Tear	0.620	0.508	0.614	0.497	0.449	0.401
	(1.50)	(1.76)	(1.48)	(1.83)	(1.06)	(1.48)
M-N Services	0.217***	0.397***	0.217***	0.369***	0.208***	0.175***
	(9.23)	(4.45)	(9.02)	(4.55)	(8.13)	(4.67)
Care Specialities	0.083**	0.128*	0.081**	0.115*	0.091**	0.099**
	(3.24)	(2.48)	(3.23)	(2.42)	(3.29)	(2.72)
Composite SQ score	0.066***	0.035	0.070***	0.038	0.057**	0.001
	(3.77)	(1.24)	(4.34)	(1.35)	(3.28)	(0.03)
Patient characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Hospital fixed effects	No	Yes	No	Yes	No	Yes
Observations	1,919,790		2,118,517		2,191,422	

NOTE.— Conditional logit estimates for patient choice of hospital varying baseline choice set and distance definitions. Columns (1)-(2) and (3)-(4) report results when choice sets have been restricted to a maximum of 15 and 20 choices, respectively. Columns (5)-(6) report results by replacing distance in kilometres by travel time in minutes by car. Patient characteristics include age, # of Elixhauser conditions, whether a patient was admitted as an emergency case, on a weekend, between 6am–10am on weekdays, and whether the patient was admitted with an secondary ICD-10 code of Z.35 (supervision of high-risk pregnancy). Hospital characteristics include ownership type, # of beds, if university, if teaching, # of midwives; Birth-Staff ratio (# of cases per doctor); share specialized doctors (# of specialized doctors as a share of all doctors); # of nurses and an indicator for whether subjective quality was missing. Composite SQ score is based on five satisfaction variables: General, Treatment, Care, Communication, Accommodation. Standard errors are clustered on spatial planning regions (96 clusters). t-statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.7: Conditional logit model estimates for choice of hospital: Heterogeneity analysis II

	I : Emergency		I : Urban	
	(1)	(2)	(3)	(4)
Distance	-0.312***	-0.356***	-0.098**	-0.192***
	(-10.88)	(-13.59)	(-2.78)	(-7.00)
Distance $\times I$	-0.039	-0.062	-0.245***	-0.198***
	(-1.13)	(-1.62)	(-5.59)	(-5.42)
Distance^2	0.005**	0.007***	-0.004	0.001
	(2.89)	(3.86)	(-1.69)	(0.77)
Distance^2 $\times I$	0.002	0.005	0.011***	0.007**
	(0.84)	(1.77)	(3.54)	(2.78)
Distance^3	-0.000	-0.000	0.000	-0.000
	(-1.21)	(-1.83)	(1.90)	(-0.25)
Distance^3 $\times I$	-0.000	-0.000	-0.000**	-0.000
	(-0.74)	(-1.85)	(-2.67)	(-1.73)
D-D-I	0.452	0.368	0.513	0.208
	(1.79)	(1.70)	(1.88)	(0.86)
D-D-I $\times I$	0.391	0.034	-0.023	0.240
	(1.36)	(0.11)	(-0.11)	(1.39)
Paediatrician	0.318**	-0.025	0.326**	0.058
	(3.28)	(-0.22)	(3.02)	(0.40)
Paediatrician $\times I$	0.053	0.097	0.015	0.010
	(0.29)	(0.42)	(0.13)	(0.08)
Perineal Tear	0.770	0.578	1.129*	0.552
	(1.83)	(1.77)	(2.37)	(1.37)
Perineal Tear $\times I$	-0.525	-0.054	-0.718	0.049
	(-1.74)	(-0.15)	(-1.87)	(0.19)
M-N Services	0.233***	0.388***	0.240***	0.291**
	(6.58)	(4.06)	(6.25)	(3.18)
M-N Services $\times I$	-0.064	-0.053	-0.038	0.085
	(-1.14)	(-0.75)	(-0.83)	(1.40)
Care Specialities	0.041	0.032	0.064	0.018
	(1.12)	(0.55)	(1.50)	(0.32)
Care Specialities $\times I$	0.084	0.107	0.011	0.058
	(1.31)	(1.33)	(0.26)	(1.23)
Composite SQ score	0.099***	0.056	0.060*	0.055
	(3.72)	(1.64)	(2.17)	(1.38)
Composite SQ score $\times I$	-0.067	-0.099	0.020	-0.006
	(-1.32)	(-1.54)	(0.79)	(-0.24)
Patient characteristics	Yes	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes	Yes
Hospital fixed effects	No	Yes	No	Yes
Observations	2,191,422		2,191,422	

Note.— Conditional logit estimates for patient choice of hospital including interactions between emergency cases (columns 1-2) and patients living in urban areas (columns 3-4) as additional regressors, respectively. Patient characteristics include age, # of Elixhauser conditions, whether a patient was admitted as an emergency case, on a weekend, between 6am–10am on weekdays, and whether the patient was admitted with an secondary ICD-10 code of Z.35 (supervision of high-risk pregnancy). Hospital characteristics include ownership type, # of beds, if university, if teaching, # of midwives; Birth-Staff ratio (# of cases per doctor); share specialized doctors (# of specialized doctors as a share of all doctors); # of nurses and an indicator for whether subjective quality was missing. Composite SQ score is based on five satisfaction variables: General, Treatment, Care, Communication, Accommodation. Standard errors are clustered on spatial planning regions (96 clusters). t-statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.8: Linear probability model estimates for choosing the closest hospital: Alternative specification

	(1)	(2)	(3)
Distance	-0.030***	-0.031***	-0.032***
	(-5.51)	(-5.58)	(-6.89)
Distance^2	0.001*	0.001*	0.001**
	(2.20)	(2.38)	(2.91)
Distance^3	-0.000	-0.000	-0.000*
	(-1.36)	(-1.58)	(-2.07)
D-D-I	0.033	0.049	0.150**
	(0.65)	(0.88)	(3.24)
Pediatrician	0.061***	0.070***	0.072***
	(4.10)	(4.59)	(5.43)
Perineal Tear	0.051	0.085	0.032
	(0.52)	(0.81)	(0.41)
M-N Services	0.031***	0.030***	0.026***
	(4.98)	(4.88)	(4.58)
Care Specialities	0.010	0.013*	0.002
	(1.78)	(2.31)	(0.46)
General	-0.021		
	(-0.14)		
Treatment	-0.041		
	(-0.35)		
Care	-0.226		
	(-1.19)		
Information	-0.096		
	(-0.53)		
Accommodation	0.340**		
	(2.75)		
Composite SQ score		-0.003	0.013*
		(-0.61)	(2.34)
Patient characteristics	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes
Hospital fixed effects	No	No	Yes
Observations	225,352	225,352	225,352

Note.— Linear probability model (LPM) estimates for whether a patient chose the closest hospital in her choice set with alternative definition of objective quality indicators. In columns (1)-(2) D-D-I, Paediatrician and Perineal Tear are binary variables indicating whether the hospital passed a required quality threshold according to a specified reference value. In columns (3)-(4), Mandatory Quality Assurance is a summary score of passing the thresholds of D-D-I, Paediatrician and Perineal Tear (out of 3) and Treatment-Relevant Services is a combined score of included M-N Services and Care Specialities (out of 11). Patient characteristics include age, # of Elixhauser conditions, whether a patient was admitted as an emergency case, on a weekend, between 6am–10am on weekdays, and whether the patient was admitted with an secondary ICD-10 code of Z.35 (supervision of high-risk pregnancy). Hospital characteristics include ownership type, # of beds, if university, if teaching, # of midwives; Birth-Staff ratio (# of cases per doctor); share specialized doctors (# of specialized doctors as a share of all doctors); # of nurses and an indicator for whether subjective quality was missing. Composite SQ score is based on five satisfaction variables: General, Treatment, Care, Communication, Accommodation. Standard errors are clustered on spatial planning regions (96 clusters). t-statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

6 Conclusion

We study patient choice of hospital with respect to both objective and subjective information about provider quality in the context of maternal care in Germany. Objective quality indicators are obtained from mandatory hospital quality report cards and subjective indicators are based on patient satisfaction scores from a large, nationwide patient survey. The quality information is linked to hospital discharge records including information on the place of residence of both patients and hospitals. We use the data to estimate econometric models of hospital choice to quantify the additional distance expectant mothers are willing to travel to give birth in a hospital of higher reported quality. Our results indicate that patients are on average willing to travel between 0.1 and 2.7 additional kilometres to obtain a one standard deviation increase in reported quality.

Our findings contribute to the existing literature on the determinants of consumer choice of healthcare provider. In line with previous findings, we obtain empirical evidence that prospective patients are responsive to quality; other papers have estimated a willingness to travel (WTT) of at most 0.9 kilometres (Gutacker *et al.*, 2016) or 0.7 kilometres (Moscelli *et al.*, 2016) for a one standard deviation increase in objective quality measures, related to elective hip replacement surgery. The magnitudes of the average WTTs, estimated with hospital fixed effects, for two of the objective quality measures (the number of Medical and Nursing Services and Care Specialities) in a given hospital, are large (2.7 and 0.9 kilometres, respectively) in comparison when compared to the WTTs estimated for other healthcare services in the literature, suggesting that there is scope for patient choice to respond to hospital quality. One reason for this strong patient response could be the importance that medical and nursing services can have for both the mother's and the child's health and well-being both before, during, and after childbirth.[32]

We also find that patients appear to value not only objective but also subjective quality information. This is an important finding since it highlights that there are dimensions of quality of care that are not subsumed within standard objective quality metrics despite their richness and variety. Subjective quality is in general negatively correlated with the objective quality indicators within a hospital, suggesting that hospitals with high clinical excellence, such as low risks of mortality or complications, perform relatively worse with respect to "softer" dimensions of quality, such as personal comfort and staff friendliness, that might contribute to patient well-being in ways that are not captured by physical health events. Different quality measures may thus not necessarily be substitutes and could even involve

[32] In Germany women tend to keep loyalty to the hospital where they gave birth. Anecdotal evidence is provided in e.g. Süddeutsche Zeitung (2017). Availability of such services may thus be important when making a long-term commitment to a hospital.

conflicting information. Studying the mechanisms through which different provider quality indicators in the healthcare sector interact with each other and how this affects patient choice may be a fruitful area for future research.

References

Alder, J., Fink, N., Bitzer, J., Hösli, I. and Holzgreve, W. (2007). Depression and anxiety during pregnancy: a risk factor for obstetric, fetal and neonatal outcome? A critical review of the literature. *The Journal of Maternal-Fetal & Neonatal Medicine*, **20** (3), 189–209.

Baker, D. W., Einstadter, D., Thomas, C., Husak, S., H.Gordon, N. and Cebul, R. D. (2003). The Effect of Publicly Reporting Hospital Performance on Market Share and Risk-Adjusted Mortality at High-Mortality Hospitals. *Medical Care*, **41** (6), 729–740.

Beaulieu, N. D. (2002). Quality information and consumer health plan choices. *Journal of Health Economics*, **21**, 43–63.

Beckert, W., Christensen, M. and Collyer, K. (2012). Choice of NHS-funded Hospital Services in England. *The Economic Journal*, **122** (560), 400–417.

Blackmore, E. R., Gustafsson, H., Gilchrist, M., Wyman, C. and O'Connor, T. G. (2016). Pregnancy-related anxiety: Evidence of distinct clinical significance from a prospective longitudinal study. *Journal of Affective Disorders*, **197**, 251–258.

Bloom, N., Propper, C., Seiler, S. and Van Reenen, J. (2015). The Impact of Competition on Management Quality: Evidence from Public Hospitals. *The Review of Economic Studies*, **82** (2), 457–489.

Brekke, K. R., Gravelle, H., Siciliani, L. and Straume, O. R. (2014). Patient choice, mobility and competition among health care providers. *Developments in health economics and public policy*, **12**, 1–26.

Bundorf, M. K., Chun, N., Goda, G. S. and Kessler, D. P. (2009). Do markets respond to quality information? The case of fertility clinics. *Journal of Health Economics*, **28** (3), 718–727.

Bünnings, C., Schmitz, H., Tauchmann, H. and Ziebarth, N. R. (2019). The Role of Prices Relative to Supplemental Benefits and Service Quality in Health Plan Choice. *Journal of Risk and Insurance*, **86**, 415–449.

Busse, R. (2008). The Health System in Germany. *Eurohealth - Health system snapshots: perspectives from six countries*, **14** (1), 5–6.

— and Blümel, M. (2014). Germany: Health system review. European Observatory on Health Systems and Policies. *Health Systems in Transition*, **16** (2), 1–296.

—, Nimptsch, U. and Mansky, T. (2009). Measuring, Monitoring, and Managing Quality in Germany's hospitals. *Health Affairs*, **28** (2), w294–w304.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge University Press.

Chernew, M., Gowrisankaran, G. and P.Scanlon, D. (2008). Learning and the value of information: Evidence from health plan report cards. *Journal of Econometrics*, **144** (1), 156–174.

Cutler, D. M., Huckman, R. S. and Landrum, M. B. (2004). The Role of Information in Medical Market: An Analysis of Publicly Reported Outcomes in Cardiac Surgery. *American Economic Review*, **94** (2), 342–346.

Deutschen Gesellschaft für Gynäkologie und Geburtshilfe (DGGG) (1995). Mindestanforderungen an prozessuale, strukturelle und organisatorische Voraussetzungen für geburtshilfliche Abteilungen. *Der Frauenartzt*, **36** (1), 27–28.

Dranove, D. and Sfekas, A. (2008). Start spreading the news: A structural estimate of the effects of New York hospital report cards. *Journal of Health Economics*, **27** (5), 1201–1207.

Elixhauser, A., Steiner, C., Harris, D. R. and Coffey, R. (1998). Comorbidity Measures for Use with Administrative Data. *Medical Care*, **36** (1), 8–27.

Gaynor, M. (2006). What Do We Know about Competition and Quality in Health Care Markets? *Foundations and Trends in Microeconomics*, **2** (6).

—, Propper, C. and Seiler, S. (2016). Free to choose? Reform, Choice, and Consideration Sets in the English National Health Service. *American Economic Review*, **106** (11), 3521–3557.

German Federal Statistical Office (Destatis) (2018). *Gesundheit: Grunddaten der Krankenhäuser, 2017*. Fachserie 12, reihe 6.1.1, German Federal Statistical Office (Statistisches Bundesamt), Wiesbaden.

Gutacker, N., Siciliani, L., Moscelli, G. and Gravelle, H. (2016). Choice of hospital: Which type of quality matters? *Journal of Health Economics*, **50**, 230–246.

Hentschker, C. and Mennicken, R. (2015). The Volume-Outcome Relationship and Minimum Volume Standards – Empirical Evidence for Germany. *Health Economics*, **24** (6), 644–658.

— and — (2018). The Volume-Outcome Relationship Revisited: Practice Indeed Makes Perfect. *Health Services Research*, **53** (1), 15–34.

Hodgkin, D. (1996). Specialized service offerings and patients' choice of hospital: The case of cardiac catheterization. *Journal of Health Economics*, **15** (3), 305–332.

Huber, S. and Rust, C. (2016). Calculate Travel Time and Distance with OpenStreetMap Data using the Open Source Routing Machine (OSRM). *The Stata Journal*, **16** (2), 416–423.

Jin, G. Z. and Sorensen, A. T. (2006). Information and Consumer Choice: The Value of Publicized Health Plan Ratings. *Journal of Health Economics*, **25** (2), 248–275.

Jomeen, J. and Martin, C. R. (2008). The impact of choice of maternity care on psychological health outcomes for women during pregnancy and the postnatal period. *Journal of Evaluation in Clinical Practice*, **14** (3), 391–398.

Jung, K., Feldman, R. and Scanlon, D. (2011). Where would you go for your next hospitalization? *Journal of Health Economics*, **30** (4), 832–841.

Karmann, A. and Roesel, F. (2017). Hospital Policy and Productivity – Evidence from German States. *Health Economics*, **26** (12), 1548–1565.

Mennicken, R., Kolodziej, I. W., Augurzky, B. and Kreienberg, R. (2014). Concentration of gynaecology and obstetrics in Germany: Is comprehensive access at stake? *Health Policy*, **118** (3), 396–406.

Moscelli, G., Siciliani, L., Gutacker, N. and Gravelle, H. (2016). Location, quality and choice of hospital: Evidence from England 2002-2013. *Regional Science and Urban Economics*, **60**, 112–124.

Moscone, F., Tosetti, E. and Vittadini, G. (2012). Social Interaction in Patients' Hospital Choice: Evidences from Italy. *Journal of the Royal Statistical Society Series A*, **175** (2), 453–472.

Mukamel, D. B. and Mushlin, A. I. (1998). Quality of Care Information Makes a Difference: An Analysis of Market Share and Price Changes After Publication of the New York State Cardiac Surgery Mortality Reports. *Medical Care*, **36** (7), 945–954.

O'Cathain, A., Thomas, K., Walters, S. J., Nicholl, J. and Kirkham, M. (2002). Women's perceptions of informed choice in maternity care. *Midwifery*, **18** (2), 136–144.

Pilny, A. (2017). Explaining Differentials in Subsidy Levels Among Hospital Ownership Types in Germany. *Health Economics*, **26** (5), 566–581.

— and Mennicken, R. (2014). *Does Hospital Reputation Influence the Choice of Hospital?* Ruhr Economic Papers No. 516, RWI-Leibniz-Institut für Wirtschaftsforschung, Ruhr-University Bochum, TU Dortmund University, University of Duisburg-Essen.

—, Wübker, A. and Ziebarth, N. R. (2017). Introducing risk adjustment and free health plan choice in employer-based health insurance: Evidence from Germany. *Journal of Health Economics*, **56**, 330–351.

Pope, D. G. (2009). Reacting to rankings: Evidence from "America's Best Hospitals". *Journal of Health Economics*, **28** (6), 1154–1165.

Porell, F. W. and Adams, E. K. (1995). Hospital Choice Models: A Review and Assessment of their Utility for Policy Impact Analysis. *Medical Care Research and Review*, **52** (2), 158–195.

Propper, C. (2018). Competition in health care: lessons from the English experience. *Health Economics, Policy and Law*, **13** (Special Issue 3-4), 492–508.

Pross, C., Averdunk, L.-H., Stjepanovic, J., Busse, R. and Geissler, A. (2017). Health care public reporting utilization–user clusters, web trails, and usage barriers on Germany's public reporting portal Weisse-Liste.de. *BMC Medical Informatics And Decision Making*, **17:48**.

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.-C., Saunders, L. D., Beck, C. A., Feasby, T. E. and Ghali, W. A. (2005). Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care*, **43** (11), 1130–1139.

Santos, R., Gravelle, H. and Propper, C. (2016). Does Quality Affect Patients' Choice of Doctor? Evidence from England. *The Economic Journal*, **127** (600), 445–494.

Scanlon, D. P., Chernew, M., McLaughlin, C. and Solon, G. (2002). The impact of health plan report cards on managed care enrollment. *Journal of Health Economics*, **21** (1), 19–41.

Süddeutsche Zeitung (2017). *Eine schöne Geburt? Ist nicht das Wichtigste*. Retreived from: https://www.sueddeutsche.de/leben/hebammen-eine-schoene-geburt-ist-nicht-das-wichtigste-1.3682528 [accessed 09.04.2019].

Sivey, P. (2012). The effect of waiting time and distance on hospital choice for English cataract patients. *Health Economics*, **21** (4), 444–456.

Tay, A. (2003). Assessing Competition in Hospital Care Markets: The Importance of Accounting for Quality Differentiation. *The RAND Journal of Economics*, **34** (4), 786–814.

Techniker Krankenkasse (2010). *Qualitätstransparenz im Krankenhaus: TK-Krankenhausbefragung*. Tech. rep., Techniker Krankenkasse.

Varkevisser, M., van der Geest, S. A. and Schut, F. T. (2012). Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands. *Journal of Health Economics*, **31** (2), 371–378.

Wagle, R. R., Sabroe, S. and Nielsen, B. B. (2004). Socioeconomic and physical distance to the maternity hospital as predictors for place of delivery: an observation study from Nepal. *BMC Pregnancy and Childbirth*, **4** (8).

Wasem, J., Gress, S. and Okma, K. G. (2004). The role of private health insurance in social health insurance countries. In R. B. Saltman, R. Busse and J. Figueras (eds.), *Social health insurance systems in western Europe*, New York: World Health Organization on behalf of the European Observatory on Health Systems and Policies, pp. 227–247.

Wedig, G. J. and Tai-Seale, M. (2002). The effect of report cards in consumer choice in the health insurance market. *Journal of Health Economics*, **21** (6), 1031–1048.

Werner, R. M., Norton, E. C., Konetzka, R. and Polsky, D. (2012). Do consumers respond to publicly reported quality information? Evidence from nursing homes. *Journal of Health Economics*, **31** (1), 50–61.

Appendix A: Additional tables and figures

TABLE A.1: Classification of Elixhauser Comorbidities

Variable	Comorbidity
el1	Congestive heart failure
el2	Cardiac arrhythmias
el3	Vascular disease
el4	Pulmonary circulation disorders
el5	Peripheral vascular disorders
el6	Hypertension, uncomplicated
el7	Hypertension, complicated
el8	Paralysis
el9	Other neurological disorders
el10	Chronic pulmonary disease
el11	Diabetes, uncomplicated
el12	Diabetes, complicated
el13	Hypothyroidism
el14	Renal failure
el15	Liver disease
el16	Peptic ulcer disease (excluding bleeding)
el17	AIDS/HIV
el18	Lymphoma
el19	Metastatic cancer
el20	Solid tumor without metastasis
el21	Rheumatoid arthritis/collagen vascular diseases
el22	Coagulopathy
el23	Obesity
el24	Weight loss
el25	Fluid and electrolyte disorders
el26	Blood loss anemia
el27	Deficiency anemia
el28	Alcohol abuse
el29	Drug abuse
el30	Psychoses
el31	Depression

NOTE.— Classification of Elixhauser Comorbidities. For detailed construction using ICD-9 and ICD-10 codes, see Quan *et al.* (2005).

TABLE A.2: Linear probability model for choosing the closest hospital: Successive inclusion of satisfaction scores

	(1)	(2)	(3)	(4)	(5)	(6)
Distance	-0.032***	-0.032***	-0.032***	-0.032***	-0.032***	-0.030***
	(-6.83)	(-6.87)	(-6.86)	(-6.86)	(-6.83)	(-5.51)
Distance^2	0.001**	0.001**	0.001**	0.001**	0.001**	0.001*
	(2.87)	(2.87)	(2.87)	(2.87)	(2.84)	(2.20)
Distance^3	-0.000*	-0.000*	-0.000*	-0.000*	-0.000*	-0.000
	(-2.04)	(-2.02)	(-2.03)	(-2.03)	(-1.98)	(-1.36)
D-D-I	0.146**	0.151***	0.152***	0.153***	0.149**	0.033
	(3.26)	(3.34)	(3.38)	(3.38)	(3.25)	(0.65)
Paediatrician	0.073***	0.070***	0.072***	0.070***	0.072***	0.061***
	(5.60)	(5.18)	(5.38)	(5.30)	(5.44)	(4.10)
Perineal Tear	0.025	0.021	0.029	0.031	0.021	0.051
	(0.32)	(0.27)	(0.37)	(0.39)	(0.26)	(0.52)
M-N Services	0.026***	0.026***	0.026***	0.026***	0.026***	0.031***
	(4.68)	(4.57)	(4.59)	(4.57)	(4.69)	(4.98)
Care Specialities	0.002	0.002	0.002	0.002	0.002	0.010
	(0.42)	(0.32)	(0.46)	(0.42)	(0.34)	(1.78)
General	0.237***					-0.021
	(4.07)					(-0.14)
Treatment		0.144*				-0.041
		(2.14)				(-0.35)
Care			0.137*			-0.226
			(2.28)			(-1.19)
Information				0.134*		-0.0957
				(2.13)		(-0.53)
Accommodation					0.159*	0.339**
					(2.14)	(2.75)
Patient characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Hospital fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	225,352	225,352	225,352	225,352	225,352	225,352

NOTE.— Linear probability model (LPM) estimates for whether a patient chose the closest hospital in her choice set. Columns (1)–(5) report results where each of the five SQ indicators are included separately. Column (6) reports results from the joint inclusion of all SQ indicators in the model. Patient characteristics include age, # of Elixhauser conditions, whether a patient was admitted as an emergency case, on a weekend, between 6am–10am on weekdays, and whether the patient was admitted with an secondary ICD-10 code of Z.35 (supervision of high-risk pregnancy). Hospital characteristics include ownership type, # of beds, if university, if teaching, # of midwives; Birth-Staff ratio (# of cases per doctor); share specialized doctors (# of specialized doctors as a share of all doctors); # of nurses and an indicator for whether subjective quality was missing. Composite SQ score is based on five satisfaction variables: General, Treatment, Care, Communication, Accommodation. Standard errors are clustered on spatial planning regions (96 clusters). t-statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE A.3: Coefficients of variation for quality indicators

	Coefficient of variation
Objective quality indicators (OQ)	
D-D-I	0.118
Paediatrician	0.569
Perineal Tear	0.097
M-N Services	0.428
Care Specialities	0.463
Subjective quality indicators (SQ)	
General	0.059
Treatment	0.038
Care	0.049
Information	0.045
Accommodation	0.062
Composite SQ Score	0.265

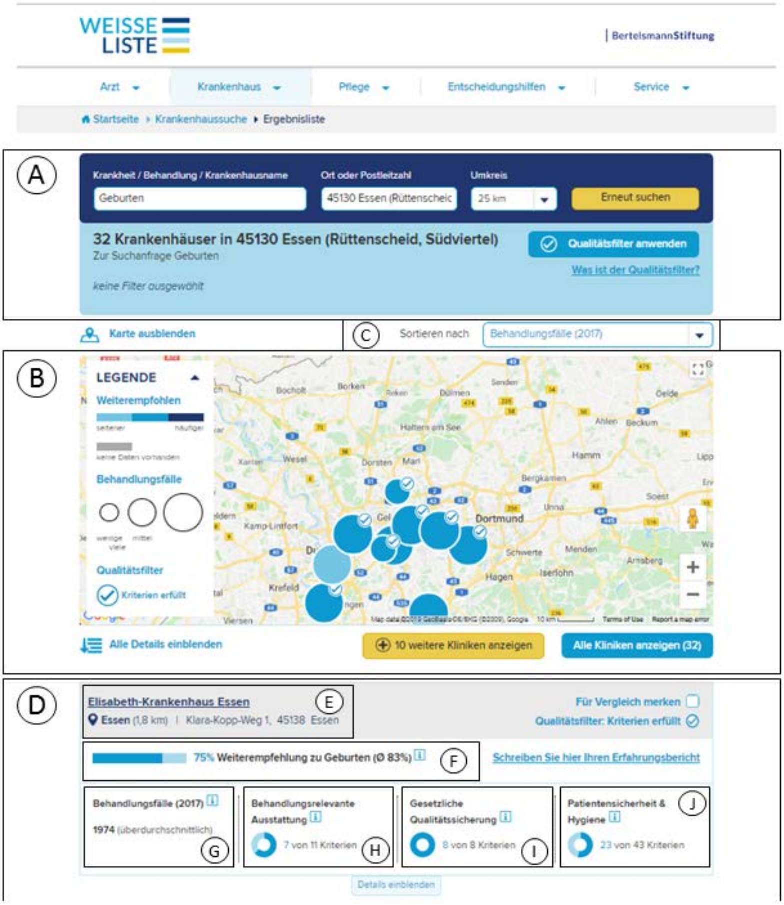
NOTE.— Estimated coefficients of variation (CV) of the included quality indicators. The CV is defined as the ratio of the standard deviation to the mean of a variable. The Composite SQ score is constructed by application of principal component analysis (PCA) based on 5 satisfaction variables (see Section 3.2).

Appendix B: Healthcare provider search

This appendix provides basic information on the use of the *Weisse Liste* hospital search portal (https://www.weisse-liste.de). Figure B.1 presents a screenshot of the website's main window in which relevant parts have been marked with letters A–J for reference. The user first types in search criteria in the three white boxes of panel A; the service (*Krankheit / Behandlung*); the name of a city or postal code (*Ort oder Postleitzahl*); and the desired maximum distance from the midpoint of the specified area the user is considering (*Umkreis*). After clicking the yellow search button (*Krankenhaussuche*), the search engine lists a set of hospitals that meet the specified search criteria together with information about their location and performance indicators.

The website first provides some general information about the selected geographical area, such as the number of hospitals which met the search criteria and a map with their exact locations (B). The hospitals can be ranked (C) in different ways, such as travel distance (*Entfernung*), general satisfaction (*Weiterempfehlung*), treatment-related services (*Behandlungsrelevante Ausstattung*) and quality assurance indicators (*Gesetzliche Qualitätssicherung*). A separate section for each hospital (D) provides relevant information; name and address together with the distance from the selected postal area (E); average patient satisfaction with the hospital (F); number of treatment-specific annual cases (G); number of treatment-related services provided (H); number of quality assurance indicators passed (I); and number of patient safety and hygiene quality measures passed (K). Detailed quality information can be found by expanding the panel (click *Details einblenden*). Note that the screenshot shows the information that the search portal currently provides, which is different from the quality information available during the years studied in the article.

FIGURE B.1: *Weisse Liste* hospital search web portal



NOTE.— Screenshot taken from *www.weisse-liste.de*.

Essay 2: Providers, Peers and Patients. How does Physician Environment Affect Patient Outcomes?[*]

1 Introduction

It is well known that traditional demand factors, such as patient preferences and needs, are largely unable to explain the substantial geographic variations in healthcare utilization observed in many countries (see, e.g., Chandra *et al.*, 2012; Finkelstein *et al.*, 2016; Mafi *et al.*, 2017; Skinner, 2011; Skinner *et al.*, 2011; Wennberg and Gittelsohn, 1973).[1] Furthermore, it is unclear whether areas with higher-than-average healthcare spending per capita perform better than lower-spending areas with respect to quality of care to legitimate such discrepancies (Baicker and Chandra, 2004; McClellan and Newhouse, 1997). These observations serve to fuel long-standing questions on the extent of resource waste and cost-efficiency in healthcare delivery (see, e.g., Doyle *et al.*, 2017, 2015; Fisher *et al.*, 2003a,b; Shrank *et al.*, 2019; Wennberg *et al.*, 2002).

The lack of explanatory power by demand factors in decomposing geographic variations in healthcare utilization has led some researchers to shift focus to the supply side and the behaviour of healthcare providers. A relatively small but growing literature have sought to understand the causes of variation in physician practice styles and their consequences for efficiency

[*]I acknowledge the contribution of Daniel Avdic (who has also supervised this project) and Maryna Ivets to the conceptualization of the project, valuable suggestions for the analysis and writing. Data collection and curation performed by Bo Lagerqvist. The author thanks to Amitabh Chandra, Jonathan Skinner, Sofia Amaral-Garcia and seminar participants in Essen, the 4$^{\text{th}}$ EuHEA student-supervisor conference in Lausanne, 27$^{\text{th}}$ European Workshop on Econometrics and Health Economics in Groningen, 6$^{\text{th}}$ Health Econometrics Workshop in Bergamo and the 30$^{\text{th}}$ Annual EALE Conference in Lyon for insightful comments.

[1]For studies based on non-US data, see Bojke *et al.* (2013); Corallo *et al.* (2014); Kopetsch and Schmitz (2014); Phelps (2000); Prieto and Lago-Peñas (2012); Reich *et al.* (2012).

in healthcare delivery (see, e.g., Chandra and Staiger, 2020; Currie *et al.*, 2016; Currie and MacLeod, 2020; Cutler *et al.*, 2019; Epstein and Nicholson, 2009; Grytten and Sørensen, 2003; Molitor, 2018; Skinner and Staiger, 2015).[2] Although still in its infancy, this research has provided robust evidence that the decision-making process of healthcare providers is complex but nevertheless important to account for when designing policies to reduce resource waste in healthcare.

This paper contributes to the literature on the determinants of provider practice styles in healthcare by studying how physicians' treatment choices are influenced by their physical and social practice environment and the consequences these choices have for their patients' welfare. To this end, we use rich administrative data from the Swedish Coronary Angiography and Angioplasty Register (SCAAR) on all percutaneous coronary interventions (PCI) performed in Sweden between 2004 and 2013 and study how interventional cardiologists' choices of stent type, bare-metal stent (BMS) or drug-eluting stent (DES), are determined by their work environment. Since the medical procedure is executed identically irrespective of stent type, but the choice of stent is subject to different clinical indications associated with potentially life-threatening consequences, our context provides a nearly ideal empirical setting to study how the practice environment shapes physician preferences for treatment. In contrast to most existing studies on physician practice styles, we are also able to gauge and directly measure the impact of changes in physician treatment behaviour on variation in the quality of care received by patients by linking practice style changes to relevant patient outcomes. Relating environmentally induced variation in physician treatment behaviour to changes in patient outcomes is useful for policies seeking to mitigate unwarranted variations in healthcare use by providing evidence on how such variations may arise from a physician's environment (OECD, 2014).

To provide a framework for the identification and consistent estimation of causal effects, we apply and extend the physician migration approach in Molitor (2018). That is, we identify physicians who move (migrate) between hospitals and relate variation in the rate of DES use between physician's origin and destination hospitals to changes in the physician's own DES use over time in a difference-in-differences empirical design. However, while empirical evidence on the extent to which physician practice styles are influenced by their work environment is informative, it does not per se convey much detail on exactly *which* environmental factors are the drivers of such changes. Yet, knowledge of the specific mechanisms mediating practice style changes

[2] Chandra *et al.* (2012) provide an overview of different explanations for why provider treatment decisions may vary across similar patients. Such reasons include (i) "defensive medicine", where providers perform unnecessary procedures to avoid complaints, bad reputation and possible lawsuits from patients; (ii) financial incentives associated with fee-for-service reimbursement models (McClellan, 2011); and (iii) unobserved heterogeneity across providers (Doyle *et al.*, 2010).

could be important. For example, physical, or provider-specific, factors may be less informative about the malleability of physicians' true preferences if the possibility to operate in line with such preferences is restricted by factors beyond the individual physician's control, such as hospital material resources or restrictive top-down management structures. In contrast, social, or peer group-specific, factors are directly related to the adjustment of physician beliefs for which much of the economic literature on physician practice styles lies at the heart of (see, e.g., Epstein and Nicholson, 2009).

To distinguish between behavioural and more mechanical drivers of practice style changes of physicians, we suggest and implement a method to decompose the combined impact of the environment on a physician's treatment style into a provider-specific and a peer group-specific factor by exploiting quasi-random variation on physicians working together on given days in a hospital. Specifically, given sufficient practice style variation among migrating physicians' co-workers (peers) *within* a hospital, the inclusion of hospital fixed effects in the empirical model will effectively purge all time-invariant provider-specific variation in practice styles *across* hospitals from the analysis. Any remaining practice variation will consequently be derived from changes in the migrating physicians' co-worker mix, so that effect estimates with and without hospital fixed effects gauge the relative magnitude of the adjustment in physician practice style arising from provider- and peer group-specific factors, respectively.

Our estimation results show that Swedish cardiologists' use of DES in angioplasty treatments is strongly determined by the prevailing practice style of the hospital they currently work in. Migrating cardiologists rapidly adapt to their prevailing environment after relocation by changing their DES use with on average half a percentage point for each percentage point difference in DES-utilization rates between the origin and destination hospitals. This result is robust to a number of alternative specifications and definitions and close to the corresponding estimate found in Molitor (2018). Specifically, to test the sensitivity of our results to the definition of the counterfactual (i.e., our identifying assumption), we use non-migrating cardiologists to form a synthetic practice environment from which estimates across definitions can be directly compared. Furthermore, when decomposing the overall effect into a provider-specific and a peer group-specific effect, we find that each is responsible for roughly half of the practice style adjustment. Finally, we provide results from a series of split-sample regressions to assess the extent of effect heterogeneity and find that our main results are driven by younger migrants moving to more innovative (in terms of DES use) hospitals.

In contrast, we find no evidence that environmentally induced changes in migrating physicians' practice styles had any important consequences for the quality of care received by patients. Specifically, in addition to studying a set of adverse clinical events related to the medical procedure, we employ a machine-learning algorithm to classify appropriate stent choice based on out-of-sample predictions from teaching hospitals and a range of patient characteristics. Neither of these outcome categories reveal systematic impacts as a result of a change in their physician's environment. This result suggests that environmentally induced changes in physicians' practice behaviour are mainly performed on "gray-zone" cases who run little risk of serious adverse medical events as a consequence of such choices.

Our findings contribute to the scant literature on peer effects and social learning in healthcare. Social learning is broadly defined as the process of information transmission between economic agents when they observe and interact with each other within their social networks (see, e.g., Lin *et al.*, 1981). In line with our findings, Huesch (2011) finds evidence for intra-group spillovers in the use of DES, suggesting that physicians are influenced by their peers. Furthermore, Nair *et al.* (2010) study peer effects in prescribing choices of physicians and find that prescribing behaviour is particularly influenced by research-active peers within physician groups. Heijmans *et al.* (2017) find similar results studying peer effects in cardiovascular risk management in networks with and without opinion leaders. On the other hand, Yang *et al.* (2014) only document small peer effects in prescription behaviour for new drugs among physicians working in the same hospital at the same time. Epstein and Nicholson (2009) find that physician's treatment styles are responsive to changes in treatment styles of other physicians in the same hospital region in the context of Cesarean sections, but the effect dampens when accounting for common shocks at the hospital level. This is in line with our finding that both providers and peers are influential in altering practice styles of physicians. Finally, Burke *et al.* (2003) find that patients are more likely to receive certain procedures if an attending physician is in a group that performs these procedures more frequently and Yuan *et al.* (2020) show that shared beliefs are crucial for successful implementation of new health technology within a peer network. Complementing these findings, we provide results from heterogeneity analyses showing that our main effects are driven by younger cardiologists who move to more DES-intensive practice environments.

We also add contextual depth to the more general economic literature on peer effects. A number of papers have investigated the influence of peers on academic performance, yielding mixed results. While some authors find significant peer effects (Sacerdote, 2001; Zimmerman, 2003), others find no effects at all (Foster, 2006; Lyle, 2007), or effects only for particular subgroups (Stinebrickner and Stinebrickner, 2006). In contrast, there exists strong evidence for positive social spillovers on task-oriented work behaviour and productivity in non-academic settings. Mas and Moretti (2009) study peer effects at the workplace by analysing the productivity of co-workers within the same team. They find evidence of positive productivity spillovers when working with highly productive peers, especially when they interact more frequently. Moreover, in an experimental setting, Falk and Ichino (2006) study individuals working on separate tasks within the sight of one another, finding that the productivity of workers is influenced by the productivity of their peers. These results support our approach to use physicians working on the same days as relevant peers in our analysis. Finally, Bandiera *et al.* (2010) study whether workers' behaviours are affected by the presence of peers that they are socially tied to, with the main finding that a given worker's productivity is positively correlated with the ability of a worker's personal friends. Our results also have broad implications for healthcare system efficiency. The fact that physicians treatment behaviours are influenced not only by their physical but also by their social environment suggests a rationale for why specific practice styles cluster in specific areas. While such clustering may generate positive productivity and learning spillovers as in Chandra and Staiger (2007), it also implies that patients may receive suboptimal care depending on the prevailing practice style at the admitting hospital. In particular in supply-sensitive areas of healthcare, where the frequency of use of a given activity is related to its local capacity, and where the choice of healthcare provider is subject to restrictions, such as place of residence, this may lead to substantial allocation inefficiencies. If the quality of provided care is largely insensitive to such variations, as this paper shows in the context of cardiac catheterizations, a more integrated system where inappropriate practice variation can be mitigated through enhanced care coordination, monitoring, and followup based on evidence-based clinical guidelines could be vastly resource-saving (Wennberg, 2010). The paper proceeds as follows. Section 2 gives an overview of the Swedish healthcare system and the clinical context. Section 3 outlines our empirical framework. Section 4 describes the data, sample and variables we include in our analysis. Section 5 presents our estimation results. Section 6 concludes.

2 Institutional Setting

The empirical analyses in this paper are based on inpatient medical records on all percutaneous coronary interventions performed in Sweden between 2004 and 2013. In this section, we first provide relevant background information on the Swedish healthcare system. This is followed by clinical information on the general treatment of coronary heart disease and on the specific medical procedure studied in this paper.

2.1 Healthcare in Sweden[3]

Healthcare in Sweden is mainly funded by direct income taxes raised by the three different levels of government: central, regional (21 county councils)

[3] www.kliniskastudier.se/english/sweden-research-country/swedish-healthcare-system.html provides a concise summary of the main features of the Swedish healthcare system in English.

and local (290 municipalities). Responsibilities for health and medical care are shared between the governments according to a scheme stipulated in the Swedish Health and Medical Service Act. Within each government tier, principals (elected politicians and bureaucrats) have substantial discretion in designing the system in their area of administration, subject to a few general principles such as that all citizens are entitled to accessible and high-quality healthcare services based on their individual needs. Both county councils and municipality executive boards are political bodies that consist of representatives elected by residents every four years coinciding with the national election.

The main responsibilities of the central government are to set goals for national health policy, coordinate and provide advice to health and medical care providers and to regulate prices and approval of new medical services and products. Municipalities are mainly responsible for organizing long-term care for the elderly in their home or in aged care facilities and to accommodate the needs of residents with physical or psychological disabilities. Lastly, the county councils are the main providers and financiers of healthcare in Sweden by virtue of being responsible for primary and specialized healthcare on both the in- and outpatient basis in their respective geographical area. Since the end of the 1990's, both local and regional healthcare boards are allowed to contract out healthcare services to private providers in purchaser-provider split models. While outsourcing of healthcare to private agents has become commonplace over time within the primary, outpatient and long-term care sectors, virtually all inpatient care is still operated by public providers.

The vast majority of healthcare spending in Sweden is paid for by county and municipal-level direct income taxes raised from area residents. Contributions from the central government are relatively small and mainly refer to provider pay-for-performance incentive schemes and redistribution between regions. Each county sets its own patient fees, although there is a national limit for the amount a patient has to pay out of pocket (approximately $130 per annum as of 2020). Consequently, patient fees only account for around three percent of total spending on healthcare. Both employed and unemployed Swedish citizens are also covered by a statutory national sickness and disability insurance financed through employer social contributions, replacing up to eighty percent of lost earnings which can be topped up for employees covered by collective agreements or complementary private insurance schemes. Hence, virtually all Swedish citizens have strong financial protection from both direct healthcare costs as well as indirect income losses from temporary or permanent work inability.

One feature of the Swedish inpatient healthcare system that is important for our empirical strategy is that recipients of healthcare are constrained in their choices of provider and treating physician. Specifically, each hospital is responsible for providing care to all residents within a geographical catchment area. This means that place of residence largely determines which hospital a patient will be admitted to when seeking care. Furthermore, hospitals are not obliged to accommodate patient requests for a specific treating physician. As a general rule therefore, a patient will be assigned to an onduty physician on the day of admission. This context suggests that patients are quasi-randomly allocated to physicians with the implication that selection bias arising from hypothetical patient-physician sorting should be less concerning.

2.2 Treatment of coronary heart disease

Coronary arteries supply oxygen and blood to the heart. When cholesterol and other fatty plaque build up inside these arteries, the wall of the blood vessel thickens, narrowing the channel within the artery and reduces blood flow to the heart. This process, called atherosclerosis, starves the heart muscle of oxygen and may cause heart tissue damage, known as Myocardial Infarction (MI) or, more commonly, a heart attack. Worldwide, about 15.9 million myocardial infarctions occurred in 2015 (Vos *et al.*, 2016).

Coronary heart disease is generally treated by interventional cardiologists applying a catheter-based treatment method called percutaneous coronary intervention (PCI), or coronary angioplasty.[4] In a PCI, the cardiologist first inserts a catheter through the femoral or radial arteries, which is subsequently transported to the site of the blockage using a guide wire. Once the obstructed area is reached, a tiny balloon attached to the catheter is inflated, compressing the atherosclerotic plaque against the artery wall, thereby restoring the blood flow. To keep the artery open at the site of the blockage after balloon dilation, the cardiologist may also place and leave a stent (an expandable small metal mesh tube) in the artery to reinforce the blood vessel's wall and prevent it from re-occluding.

Prior to invasive treatment, a diagnostic technique, angiography, is used to determine the size, severity and location of the suspected artery blockage(s). To this end, a catheter is guided into one of the major coronary arteries to inject a contrast dye into the blood passing through the heart. The diagnosing physician, the *angiographer*, can then determine the locations with restricted blood flow from a series of images (angiograms) taken by an X-ray machine. Sometimes, when considered suitable by the responsible physician, the angiography is directly followed by a PCI in the same treatment session, a procedure known as *ad-hoc* PCI.

[4] PCI began as percutaneous transluminal coronary angioplasty (PTCA), a term still found in the literature. It now encompasses balloons, stents, and other modifications to the catheter tip, including devices that cut out plaque to open narrowed arteries.

2.3 Bare-Metal and Drug-Eluting Stents

Two main types of stents are used when performing a PCI: Bare-Metal Stents (BMS), commonly referred to as first-generation stents, and the newer Drug-Eluting Stents (DES), first approved in Europe in 2002. The principal difference between the BMS and the DES is that the latter is coated with a drug that reduces the incidence of restenosis, the medical term for the gradual re-narrowing of a coronary artery after a blockage has been treated with angioplasty. Because the process of compressing, or "crushing", the atherosclerotic plaque often causes trauma to the artery wall, the body will attempt to heal itself by repairing the tissue damage caused by the intervention by proliferation of endothelial cells (a layer on the surface of blood vessels). Restenosis occurs from excessive tissue growth as a consequence of such healing processes, which re-occludes the blood vessel at the site of the stent. In contrast to the BMS, the DES was developed to counteract re-occlusion of the artery by being coated with drugs that inhibit cell proliferation, thereby sharply reducing the risk of restenosis.

Although the DES represents a major medical advance for angioplasty over the BMS, it has also been associated with the significantly more severe side-effect of stent thrombosis (ST): the formation of blood clots in the blood vessels caused by the stent itself.[5] As the drugs coated on the DES inhibit the body's natural healing process (i.e., the formation of an endothelial layer), they simultaneously expose the body to an increased risk of thrombus formation (blood clots). Thus, the DES has been linked with an increased risk of ST occurring up to several years after the initial intervention. So-called Dual Anti-Platelet Therapy (DAPT), most commonly involving acetylsalicylic acid (aspirin) and clopidogrel, is considered crucial to reduce the risk of ST. Early cessation of these drugs after angioplasty using DES significantly increases the risk of both ST and MI.

The above discussion suggests that the choice between a BMS and a DES when performing angioplasty is not trivial. Although clearer guidelines exist today as to which stent type should be used for each case, this choice belonged to the "gray zone" of medical decision-making (where guidance from clinical evidence is inadequate in providing clear indications for use) during the time period we study in this paper. In addition, the choice between a BMS and a DES does not involve significant differences in other categories of use, such as prices[6] (e.g., costs of equipment necessary for

[5]While this is true for the first generation of DES (Taxus and Cypher), the second generation DES has been associated with significantly less ST than its predecessor (Chitkara and Gershlick, 2010). However, the latter stent type only began gaining popularity at the end of our analysis period.

[6]See, e.g., Ekman *et al.* (2006) who estimate that the expected one-year cost of a PCI with a Taxus DES in 2004 amounted to SEK 72,000 (USD 7,900) versus SEK 67,000 (USD 7,400) for a BMS. Both direct and indirect (i.e., repeat revascularization) treatment costs are included as Swedish hospitals are paid on a capitation basis. This contrasts, for example, with much larger cost differences in the US (see, e.g., Karaca-Mandic *et al.*, 2017). In addition, we can rule out large incentives for adoption from

the procedure), mode of treatment (e.g., minimally invasive versus highly invasive), or physical attributes of the clinician (e.g., visual acuity or motor skills). This context provides us with a close to ideal setting for studying how physician preferences for treatments vary with their environment, since observed choices are likely to be mainly a function of the physician's personal preferences regarding the relative efficacy of each treatment option.

3 Econometric framework

In this section we describe our empirical approach for quantifying the effect of the environment on physician treatment styles in the context of the choice of stent type in angioplasty treatments. We first define how we measure physician exposure to their practice environment and how the overall environment can be partitioned into a provider-specific and a peer group-specific component. Next, we describe our empirical model from which physician responses to a change in their practice environment can be identified and estimated using empirical variation from cardiologists moving across hospitals.

3.1 Definition of physician practice environment

The practice environment a physician is exposed to is a latent variable, meaning that it exists but is not directly quantifiable. A challenge is therefore to define a variable that captures the relevant features of the practice environment for our purposes. Following the methodology taken in Molitor (2018) and adapted to our setting, we characterize cardiologist $j \in J$'s practice environment in hospital $h \in H$, where patient $i \in N_{ht}$ received a PCI in time period $t \in T$, as the ratio

$$E_{jht} = \frac{\sum_{i \in N_{kht}} \mathbb{1}(DES_i = 1)}{N_{kht}} \quad \forall\, k \neq j \in J, \tag{3.1}$$

where $N_{kht} \subset N_{ht}$ is the subset of patients *not* treated by cardiologist j. Hence, E_{jht} is j's exposure to the practice environment with respect to the rate of DES use among eligible patients in hospital h and time t. Next, we define the *difference* in practice environments between a migrating cardiologist's origin (h_{O_j}) and destination (h_{D_j}) hospital at a given point in time as

$$\Delta_{jt} = E_{jh_{D_j}t} - E_{jh_{O_j}t}. \tag{3.2}$$

In other words, Δ_{jt} is the period-specific difference in DES leave-out shares between the hospital that cardiologist j practiced in before and after relo-

lobbying by the medical device industry as this is much more muted in the Swedish centralized healthcare system compared to more market-oriented systems.

cation, respectively. Note that this setup provides an intuitive framework for defining counterfactual treatment states of migrating physicians that we will use to motivate our empirical approach below.

Equations (3.1) and (3.2) constitute the basic framework for quantifying physicians' exposure to their practice environment over time and across hospitals. We now extend this framework by partitioning the overall practice environment into two separate dimensions: a physical (provider-specific) and a social (peer group-specific) component, respectively. Conceptually, we can think of a physician's practice environment as a combination of physical (e.g., hospital infrastructure, technology, assets and resources) and social (e.g., peers, physician networks and co-workers) factors. The former component may be less relevant from a behavioural point of view, since physician responses to the availability of physical resources are not directly related to his or her preferences for a particular treatment. On the other hand, social interactions may be highly influential in forming and developing physician preferences for treatments and beliefs in their efficacy. Studying the net as well as the relative impact of these components in their capacity to alter physician practice styles is therefore important both theoretically, in terms of understanding the anatomy of physician decision-making, and in practice, to provide a basis for policy to improve the effectiveness of healthcare delivery.

To empirically disentangle provider- and peer group-specific components in physician practice environment, we postulate that cardiologists who are working in the same hospital on the same day form a relevant peer group from which we can draw inference.[7] Formally, let

$$P_{k_j ht} = \frac{\sum_{i \in N_{k_j ht}} \mathbb{1}(DES_i = 1)}{N_{k_j ht}} \quad \forall\, k_j \neq j \in K_j. \tag{3.3}$$

be the average DES share used by cardiologist j's *peers* k_j in hospital h and period t. Cardiologist j's peers are defined as all other K_j cardiologists who performed PCI's on patients in the same hospital and at the same point in time as physician j. We use this within-hospital variation to define and estimate physician j's *peer exposure* in time period t by the relation

$$E^P_{jht} = \sum_{k_j \in K_j} \sum_{d_t \in D_t} \mathbb{1}(d_{t_j} = 1, d_{t_{k_j}} = 1) \times P_{k_j ht}, \tag{3.4}$$

where $d_t \in D_t$ is the specific calendar date *within* period t, and d_{t_j} and $d_{t_{k_j}}$ are indicator variables for whether physicians j and k_j were both treating patients on day d_t. In other words, E^P_{jht} is a weighted average of the overall

[7]While this definition makes intuitive sense, as individuals who work together are able to observe and directly influence each other, it is also supported by the economic literature on peer effects in the workplace (see, e.g., Falk and Ichino, 2006; Mas and Moretti, 2009).

practice environment of hospital h in time period t, with weights defined by the correspondence between cardiologist j and each of his or her peers with respect to the days they both performed PCI's on admitted patients. Note that giving all K_j peers the same weight in Equation (3.3) would return E_{jht} from Equation (3.1).

The difference in peer practice environment between a migrating cardiologist's origin and destination hospitals, Δ^P_{jt}, is correspondingly defined by replacing E with E^P in Equation (3.2). The counterfactual practice environment (i.e., the environment in the hospital cardiologist j is not currently working in) is simply defined as the potential peer exposure derived from all cardiologists who worked in the counterfactual hospital over that period,

$$\Delta^P_{jt} = E^P_{jh_{D_j}t} - E^P_{jh_{O_j}t}. \tag{3.5}$$

The total variation in the hospital's practice environment is equal to the sum of the within- and the between-components, implying that we can decompose physician j's overall practice environment as

$$E_{jht} = E^P_{jht} + E^H_{ht}, \tag{3.6}$$

where E^H_{ht} is equal to the provider-specific component, varying only across hospitals and time, and E^P_{jht} as the peer group-specific component, varying across cardiologists within hospitals over time. It follows that the total change in a migrating physician's practice environment can be decomposed as

$$\Delta_{jt} = \Delta^P_{jt} + \Delta^H_{jt}. \tag{3.7}$$

That is, the total impact of the change in environment of a migrating cardiologist at a given point in time consists of a physician-specific and a hospital-specific effect. Our approach to empirically disentangle these two effects is described in the following subsection.

3.2 Empirical model

The point of departure for our empirical modelling is based on the method in Molitor (2018) who uses longitudinal administrative data on cardiologists moving across hospitals to obtain empirical variation in physicians' practice environment. This variation is used to estimate causal effects of changes in the migrating physicians' practice environment on their own treatment choices in a difference-in-differences (DD) empirical design. The idea is simple yet intuitive: if physicians' practice styles are malleable to the environment they operate in, then we would expect to observe patients managed by migrating physicians to receive treatments more aligned with the practice environment in the destination hospital after, but not prior to, their relocation.

Formally, the patient-level DD model for patient $i \in N$, treated by cardiologist $j \in J$ at time $t \in T$ can be described by the equation

$$y_{ijt} = \alpha Post_t + \beta \Delta_{jt} + \gamma(\Delta_{jt} \times Post_t) + X'_{ijt}\Gamma + \lambda_j + \lambda_t + \epsilon_{ijt}. \qquad (3.8)$$

The outcome y_{ijt} is defined by a dummy indicator variable equal to one if a patient undergoing PCI received a DES, and equal to zero if a BMS was used. Moreover, $Post_t = \mathbb{1}_{t \geq t_0}$ is a dummy variable which equals one for all time periods subsequent to cardiologist j's move to a new hospital at time t_0. The model also includes controls for cardiologist, λ_j, and time, λ_t, cluster-specific effects (i.e., $\sum_z \theta_z \mathbb{1}_{\lambda_{z'=z}}$ for $z = j,t$) and a vector of potentially time-varying observable patient and cardiologist characteristics, X_{ijt}, to adjust for observed and unobserved heterogeneity across patients, physicians and time. Finally, Δ_{jt}, defined in Equation (3.7), is a continuous variable with range $[-1,1]$, characterized as the difference in physician j's practice environment between the *origin* (pre-migration) and *destination* (post-migration) hospitals with respect to the share of DES used in patients undergoing PCI at time t.

The main parameter of interest in Equation (3.8) is γ, which, under standard identifying assumptions of the DD estimator, captures the average physician response in their DES use to the difference in practice environments between the origin and destination hospitals after, relative to before, their relocation. Defining practice environment as the hospital's risk-adjusted share of DES used on patients undergoing PCI, γ can be interpreted as the percentage change in physician j's *own* DES practice style for each percentage point difference in practice style *environment*. We refer to Equation (3.8) as our baseline model in order to provide a link to and compare the results in Molitor (2018) to our decomposition approach described below.

To study the dynamic pattern of the migrating cardiologists' responses to their practice environment and test the common trend assumption, we extend our baseline model in Equation (3.8) by replacing $Post_t$ with a set of period-specific indicators

$$y_{ijt} = \beta \Delta_{jt} + \sum_{s=-T'}^{T'} \mathbb{1}(s = t')\left(\alpha_{t'} + \gamma_{t'} \Delta_{jt'}\right) + X_{ijt}\Gamma + \lambda_j + \lambda_t + \epsilon_{ijt}, \qquad (3.9)$$

where $t' = t - t_0 \in (-T', T')$ is the period-specific index re-centered around the time of the cardiologist's move, t_0. This modification allows us to interpret the average period-specific cardiologist responses by time from their move on a common time index that can be plotted in an event-study fashion.

3.3 Effect decomposition and quality of care

Our approach to identify physician responses to their practice environment relies on empirical variation derived from cardiologists moving across hospitals at different points in time. Whenever this happens, we maintain that they are exposed to a combined shift in practice environment arising from two sources: a provider-specific, Δ^H_{jt}, and a peer group-specific, Δ^P_{jt}, component, as defined in Equation (3.7). To empirically disentangle these two effects, we make use of the fact that the former component is assumed to be constant within a hospital provider. Therefore, the additional inclusion of hospital-specific effects, λ_h, in Equations (3.8) and (3.9) will effectively purge the practice environment of the hospital-specific component and any remaining variation will hence be attributed to the peer effect, Δ^P_{jt}. Thus, we estimate Equations (3.8) and (3.9) with and without hospital fixed effects for our sample of movers and attribute the estimated γ without hospital fixed effects as the net impact of the practice environment. In contrast, the estimated effect with hospital fixed effects will be attributed to the peer group-specific effect component. Finally, the relative difference between these two effects as a share of the net effect is interpreted as the provider-specific effect.

So far our model framework has focused on changes in the practice styles of cardiologists induced by their practice environment. However, we are also interested in knowing whether any environmentally induced behavioural changes of physicians translate into changes in the quality of care received by patients who were treated by the migrating cardiologists. In particular, knowing *how* these behavioural changes affect the appropriateness of the treatment and patient health outcomes would provide useful information on whether and to which extent physician adaptation to their practice environment improved or worsened quality of healthcare delivery. To this end, we consider two additional sets of outcomes within our regression framework: physician decision errors and patient health outcomes. The latter category is based on a composite measure of relevant post-intervention adverse clinical events, including death, myocardial infarction and restenosis requiring a new intervention. The former outcome category is based on defining a measure of stent appropriateness using an auxiliary sample from which we employ a classification exercise based on machine learning techniques. We defer the details of this approach to the next section.

4 Data

We use data from the Swedish Coronary Angiography and Angioplasty Registry (SCAAR) for our empirical analyses.[8] Since 1998, SCAAR registers

[8] SCAAR is maintained by the Uppsala Clinical Research Center (UCR), sponsored by the Swedish Health Authorities and independent of commercial funding. Reporting

cardiac catheterization procedures performed in Swedish hospitals, including detailed clinical information on patient health status and comorbidities (e.g., diabetes mellitus, smoking status and BMI), angiography diagnostic results (e.g., location and severity of blockage by coronary artery segment) and relevant treatment outcomes (e.g., complications and adverse clinical events such as myocardial infarction or death). Importantly, the register also includes information on the treating hospital and responsible physician, performed procedure(s) and the time and dates of intervention, hospital admission and discharge.

4.1 Analysis sample

We initially sample all instances of PCI performed in Swedish hospitals and reported in SCAAR between 2004 and 2013.[9] To clearly identify our main outcome variable, the cardiologist's choice between using a DES and a BMS in the procedure, we drop patients who received multiple stents in the same treatment session from the sample. This restriction leaves us with a total of 51,381 PCI cases performed by 199 cardiologists in 28 hospitals.

The data include daily information on each cardiologist's angioplasty treatments and the hospital the activity takes place in. We use this information to define physician practice episodes by indicating the first and the last date a cardiologist practiced in a particular hospital. This method defines an origin and a destination hospital and a specific time-stamp for when the move took place. As a few cardiologists may occasionally practice in several hospitals, we classify physician practice episodes to hospitals where the cardiologist continuously treated patients over a period of at least six months.[10] In total, we identify 51 migrating cardiologists treating 8,589 patients across 25 hospitals over the analysis period. Remaining cardiologists, who were based at the same hospital throughout the analysis period, are referred to as non-migrating cardiologists.

Columns (1) and (2) of Table 4.1 present means and standard deviations for our analysis sample of migrating cardiologists while columns (3) and (4) present corresponding figures for non-migrating physicians for comparison. The upper, middle and lower panels of the table partition this

in the SCAAR is Internet-based. The data are recorded online through a Web interface in the cardiac catheter laboratory, encrypted and sent to the UCR central server. Each hospital receives a feedback on the processes and quality of care measures. To monitor and maintain quality, a continuous screening process of the registry data is in place, operating by comparing 50 entered variables in 20 randomly selected interventions per hospital-year with the patients' hospital records. The overall correspondence in data during the study period is 95.2%.

[9] We restrict the starting year of our analysis to 2004 as this is the first year all hospital in Sweden that performed PCI contributed to the registry. The endpoint is chosen because the market for stents included additional options from 2013 onward due to the entry of a new second-generation DES and the corresponding sharp decline in the use of the BMS.

[10] We exclude a few cases where a cardiologist systematically practices in two hospitals over a long period of time (e.g., Karolinska hospital in Solna and Huddinge in Stockholm county and Lund and Malmö hospital in Skåne county).

information into hospital-, cardiologist- and patient-specific characteristics, respectively. With respect to hospital characteristics, we observe no major differences across the two groups other than that non-moving cardiologists seem to work in moderately larger hospitals in terms of annual case volume. With respect to the characteristics of the cardiologists themselves, migrants tend to be somewhat younger and more likely to have a specialization in cardiology (in contrast to, e.g., radiology or surgery). Patient case-mix is remarkably similar in all aspects across the groups on average, although migrating cardiologists appear to be somewhat less prone to use DES. However, there are no differences in terms of patient health outcomes between migrants and non-migrants.

TABLE 4.1: Descriptive sample statistics

	Moving cardiologists		Non-moving cardiologists	
	Mean	SD	Mean	SD
Hospital characteristics				
Teaching hospital	0.38	0.49	0.41	0.49
RiksHIA quality index	3.73	1.95	3.84	1.95
Case volume	7,861	7,349	8,912	7,468
Hospitals	25		28	
Cardiologist characteristics				
Male	0.93	0.25	0.90	0.30
Age	46.59	6.45	49.00	7.20
Specialization in cardiology	0.85	0.35	0.70	0.46
Total error rate	0.40	0.05	0.39	0.07
Type I error rate	0.14	0.06	0.15	0.08
Type II error rate	0.26	0.08	0.24	0.10
Cardiologists	51		148	
Patient characteristics				
Risk factors				
Male	0.73	0.45	0.72	0.45
Age	65.81	10.94	66.00	11.11
Smoker	0.79	0.79	0.82	0.79
Diabetes	0.17	0.37	0.17	0.37
Chronic obstructive pulmonary disease	0.01	0.11	0.02	0.12
Peripheral vascular disease	0.00	0.05	0.00	0.07
Hypertension	0.49	0.50	0.50	0.50
Previous infarction	0.20	0.40	0.18	0.39
Previous CABG	0.09	0.28	0.08	0.27
Previous PCI	0.11	0.31	0.10	0.30
Outcomes				
DES treatment	0.36	0.48	0.42	0.49
Death (1 year)	0.04	0.19	0.04	0.19
MI (1 year)	0.07	0.26	0.07	0.26
TLR (1 year)	0.06	0.24	0.06	0.23
Total error rate	0.42	0.49	0.40	0.49
Type I error rate	0.12	0.32	0.15	0.36
Type II error rate	0.30	0.46	0.25	0.43
Cases	8,589		51,381	

NOTE.— SCAAR data for years 2004–2013. Means and standard deviations for samples of moving and non-moving cardiologists. Patient characteristics are missing for a subset of observations: gender (28 cases), smoking (4,893 cases), diabetes (680 cases), hypertension (1,535 cases), previous infarction (1,724 cases), previous CABG (158 cases), previous PCI (168 cases); and cardiologist characteristics: age (739 cases); specialization (692 cases); and hospital characteristics: RiksHIA quality index (693 cases). All observations with missing characteristics are included in the analysis by defining dummy variables for the missing categories.

4.2 Decision errors and patient health outcomes

To study the impact of migrating cardiologists' changes in practice environment on quality of care, we replace our main outcome variable from Equations (3.8) and (3.9) with two sets of outcomes proxying for the appropriateness of the chosen treatment and for any adverse patient health consequence of such choices. We first define a dummy indicator variable for whether the treatment decision was the appropriate choice based on a risk-adjusted measure of treatment suitability and classified using an machine learning method for classification. To this end, we employ the Random Forest (RF) algorithm which has been demonstrated to have improved prediction accuracy in comparison with other supervised learning methods (Breiman, 2001; Svetnik *et al.*, 2003).[11]

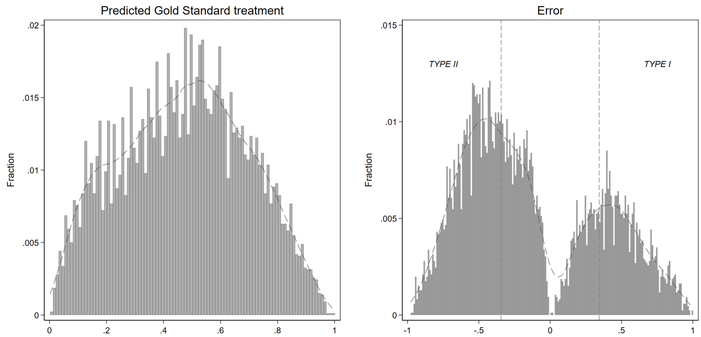
We assess the appropriateness of cardiologists' stent choices by relating actual physician choices to predicted "gold standard" choices derived from the RF algorithm using auxiliary data based on angioplasty procedures performed in Swedish teaching hospitals with no migrating cardiologists in 2011–2012.[12] We predict the appropriate stent choice for each case in our analysis sample and define a dummy variable for overall error, equal to one whenever the observed choice does not match the predicted choice irrespective of the choice of stent. Figure 4.1 shows the distribution of predicted probabilities (left panel) and respective error rates (right panel).

We furthermore decompose the overall decision error into Type I and Type II errors under the null hypothesis that the BMS is the suitable choice. To this end, a Type I error (i.e., a false positive) pertains to incorrectly inserting a DES when a BMS is suitable and a Type II error (i.e., a false negative) is defined by inserting a BMS when a DES was the correct option. This decomposition may provide additional insights into the consequences of inappropriate treatment choices since incorrect use of the DES is subject to more severe adverse events, such as ST. Table 4.2 presents a matrix of the cardiologists' treatment decisions in our sample and corresponding error rates.

[11] RF is a supervised machine learning method for classification based on the construction of decision trees. The computational steps of the RF algorithm are illustrated in Figure A.7 in the Appendix. A decision tree splits the data into a set of subsamples defined by a classification rule represented by a tree branch. Each branch could either lead to another sub-tree or have a leaf/terminal node with an assigned class. The most frequently classified outcome among all individual decision trees performed defines the terminal prediction (class) of the RF. Application of this data splitting method can be further pruned by setting constraints on model parameters to boost the accuracy on the out-of-sample predictions and stability of the tree.

[12] The auxiliary data sample was randomly divided into two parts: a training sample that is used to fit the RF algorithm and a validation sample used to validate the performance. This re-sampling procedure is based on 70:30% split. We grow 500 individual decision trees to improve the performance of the RF and achieve the best prediction accuracy in the validation sample. Each tree's terminal node has at least 15 observations, but the total number of terminal nodes in each tree does not exceed 200 nodes in total. Out of total 190 predictors, we randomly sampled 50 variables at each split. The tuning of all parameters is based on the performance evaluation on the validation sample. Figure A.8 in the Appendix presents the importance of variables used in prediction.

FIGURE 4.1: Distributions of predicted gold standard DES probabilities and cardiologists' decision errors



NOTE.— SCAAR data for years 2004–2013. Left panel presents distribution of predictions of "gold standard" treatment, with respect to use of DES in angioplasty treatments, from estimation of the random forest (RF) machine learning algorithm explained in Section 4.2. Predictions are based on an auxiliary sample of non-moving cardiologists working in university hospitals years 2011–2012. See also Breiman (2001); Svetnik *et al.* (2003). Right panel shows corresponding decision errors by comparing migrating cardiologists' actual choices to gold standard predictions. Vertical lines correspond to thresholds for classification into Type I and Type II errors.

TABLE 4.2: Cardiologist treatment decision matrix

	BMS recommended	DES recommended	Error rate
Treated BMS	3,026	2,603	46%
Treated DES	982	1,972	33%

NOTE.— SCAAR data for years 2004–2013. Recommended treatments are classified according to predictions from estimation of the random forest (RF) machine learning algorithm explained in Section 4.2. Predictions are based on an auxiliary sample of non-moving cardiologists working in university hospitals years 2011–2012. Error rates are defined as the share of chosen non-recommended treatments among all treatments using the specific stent type. See also Figure 4.1.

Finally, we include a set of patient outcomes based on the prevalence of one-year post-intervention adverse clinical events, including patient death, myocardial infarction (MI), and total leison revascularization (TLR) to our regression model. The bottom panel of Table 4.1 shows the rates of these events in our analysis sample.

4.3 Estimation of physician practice environment

Since both the absolute number and the case-mix of patients treated by cardiologists may vary substantially, we modify each cardiologist's use of DES using the Empirical Bayes (EB) method. To this end, we estimate a mixed-effects model with both fixed (risk-adjustment) and random (shrinking imprecise physician DES shares to the population mean) elements to correct for potentially biased estimates of the physicians' practice environment (see, e.g., Rabe-Hesketh and Skrondal, 2008) as well as any existing risk selection between cardiologists and their patients.

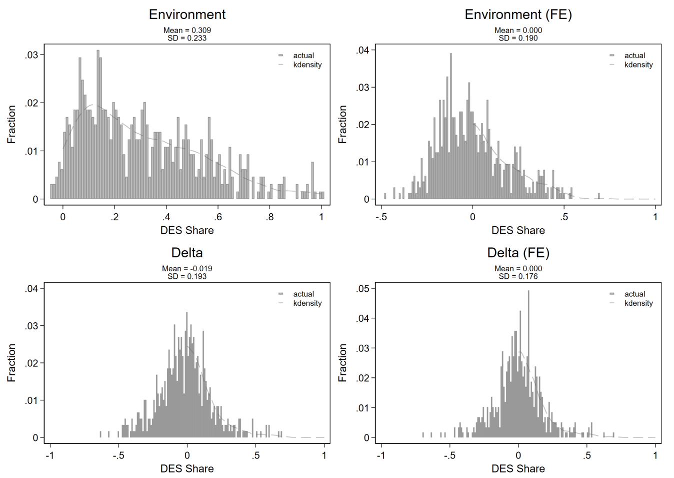
Variation in the EB-adjusted practice environment across all migrating cardiologists and periods in our sample is shown in the upper left panel of Figure 4.2. The variation is large, covering almost the full range of the variable, and slightly skewed to the left with a mean of 0.31. The corresponding distribution after regression adjustment for hospital fixed effects (i.e., the within-hospital variation) is visualized in the upper right panel of the same figure. There is substantial variation remaining even after the hospital-specific component has been eliminated from the environment, suggesting that including provider-specific effects is unlikely to generate problems of model overfitting.[13] The lower left and right panels of Figure 4.2 show corresponding distributions of Δ_{jt} with and without hospital-specific fixed effects, respectively. Interestingly, the change in practice environment among migrating cardiologists in our sample is symmetrically distributed across higher and lower shares of DES. Hence, our empirical approach captures a wide range of treatment effects in both the positive and negative domains of changes in the physicians' practice environment.

Figure 4.3 provides a graphical illustration of the intuition behind the identification approach we use in our empirical analysis. The solid lines indicate the average practice style environment, measured by the average quarterly share of DES used among migrating cardiologists' peers, by time from their relocation. To avoid cancelling out positive and negative changes in the practice environment, physicians moving from more to less DES-intensive environments and from less to more DES-intensive environments are plotted in the left and right panels of the figure, respectively. Moreover, the dashed lines show the corresponding estimated *counterfactual* environment in the hospitals associated with the migrating cardiologists: the destination hospital, prior to the relocation, and the origin hospital, after the relocation took place. At any point in time, the vertical difference between the two lines is computationally equivalent to the average difference in physician practice environments, Δ_{jt}, averaged over all J migrating cardiologists.

The figure shows that there are significant jumps in the practice environment for both groups of migrating cardiologists at the time of relocation when the actual and the counterfactual environments are switched. The quarter of the move has been interpolated in the graph (and omitted from our analysis), since the cardiologist may treat patients in both the origin and destination hospitals during this period. The counterfactual environment can hence be interpreted as an estimate of the hypothetical environment that would have prevailed if the migrating physician would not have relocated. We can use this estimate to derive and evaluate the common trend assumption when estimating our DD model. In particular, if migrants react to the counterfactual environment prior to their move, we would conclude that our empirical approach is invalid. We study this in further detail in the next section.

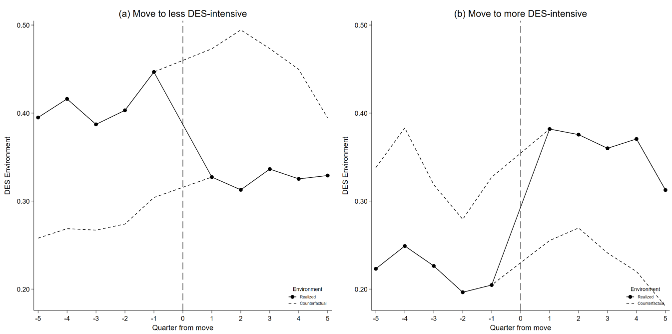
[13]The distribution of the risk-adjusted DES rates across the 21 county councils in Sweden is displayed in Figure A.9 in the Appendix.

FIGURE 4.2: Distributions of migrating cardiologists' practice environments



NOTE.— SCAAR data for years 2004–2013. Upper panels pertain to physicians' practice environment prior to relocation without (left panel) and with (right panel) adjustment for hospital fixed effects. Lower panels show corresponding distributions for the difference in practice environment between migrating cardiologists' origin and destination hospitals, Δ_{jt}

FIGURE 4.3: Average trends in migrating cardiologists' practice environments



NOTE.— SCAAR data for years 2004–2013. Practice environment defined as the share of DES used in angioplasty treatments in realized (solid lines) and counterfactual (dashed lines) hospitals by quarter from the cardiologist's move. Separate plots for cardiologists moving to hospitals with lower and higher intensity of DES use. Vertical dashed line indicates re-centered quarter of physician relocation from the origin to the destination hospital. Quarter of move linearly interpolated.

5 Results

This section reports results from estimation of the econometric models described in 3 using our analysis sample explained in Section 4. We first provide main results obtained from estimation of our DD model on migrating cardiologists' responses to a change in their practice environment with respect to their use of DES when performing PCI. Next, we investigate the extent to which these responses improved or worsened the appropriateness of physicians' treatment choices and whether they were associated with changes in patient health outcomes. Finally, we provide results from a set of robustness checks and heterogeneity analyses to evaluate the stability of our inference with respect to model specification and variable definitions.

5.1 Do physicians adapt to their practice environment?

Columns (1)–(4) of Table 5.3 report results from estimation of different models using our sample of migrating cardiologists. Column (1) provides corresponding coefficient estimates from the model used in Molitor (2018) to estimate the response of migrating cardiologists to changes in their practice environment. Our reported DD estimate of 0.72, interpreted as the average percentage change in the physician's own practice style for each percentage change in the practice environment between the origin and destination hospitals after relocation, is very close to the estimate of 0.67 found in Molitor (2018). Moreover, the coefficient of Δ_{jt}, interpreted as migrating physicians' average response to the destination hospital's practice environment *prior* to the move, is insignificant. This result supports our maintained common trend assumption that migrating cardiologists do not systematically change their own practice style in response to the destination hospital's practice environment before they relocate.

Next, Columns (2) and (3) show estimation results from our baseline DD model, defined in Equation (3.8), by successive inclusion of control variables. While the results from Column (2), in which only the control variables listed in Table 4.1 have been added, suggest a marginally significant response to Δ_{jt} prior to the move, this coefficient is once again insignificant after further adjustment for period-specific and cardiologist-specific effects in Column (3). The DD point estimates for these model specifications suggest a somewhat smaller physician response of between 0.49 and 0.52. In other words, about half of the migrating cardiologists' DES use can be attributed to their overall practice environment in our sample.

Finally, in Column (4) we decompose the overall effect from the change in practice environment by including hospital fixed effects in our regression model. Recall that migrating cardiologists face both a change in the provider-specific and the peer group-specific practice environment when they move across providers. Assuming that the provider-specific component is constant within a hospital, whereas the peer group-specific component varies within hospitals, we can include hospital fixed effects to eliminate the impact of the former from the practice environment variable. This adjustment reduces the DD estimate by another fifty percent to 0.25. We interpret this result as that the peer group-specific effect is responsible for roughly half of the response in physician practice style. This suggests that physicians' reactions to their practice environment embody both the characteristics of the hospital itself, such as infrastructure, management and resources, as well as the social environment, captured by the physicians' workplace peers.

TABLE 5.3: Difference-in-Differences estimates of migrating cardiologists' changes in practice environment: Use of DES

	(1) DES	(2) DES	(3) DES	(4) DES
Post	-0.003	-0.030	0.014	0.003
	(0.022)	(0.034)	(0.020)	(0.023)
Δ_{jt}	-0.131	-0.253**	-0.164	0.013
	(0.085)	(0.126)	(0.105)	(0.087)
$Post \times \Delta_{jt}$	0.719***	0.485**	0.523***	0.247***
	(0.130)	(0.201)	(0.114)	(0.090)
Covariates		✓	✓	✓
Year FE	✓			
Origin hospital FE	✓			
Year-quarter FE			✓	✓
Cardiologist FE			✓	✓
Hospital FE				✓
Cardiologists	51	51	51	51
Observations	8,589	8,589	8,589	8,589

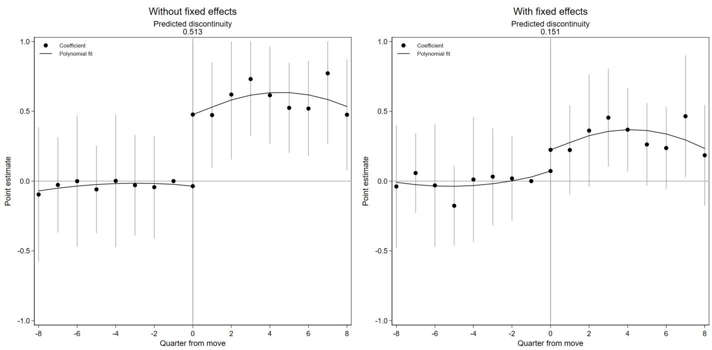
NOTE.— SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (3.8). Dependent variable is an indicator for whether a patient undergoing PCI received a DES. Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The left and right panels of Figure 5.4 display estimation results from the event study model in Equation (3.9) without and with hospital fixed effects, corresponding to the specifications in Columns (3) and (4) of Table 5.3, respectively. Each dot in the figure refers to an estimated $\gamma_{t'}$ parameter and the associated vertical spikes indicate corresponding 95% confidence bands. The solid vertical line in each panel pertains to the specific re-centered year-quarter of cardiologists' move from the origin to the destination hospital. The quarter of relocation is omitted from the analysis and replaced with the predicted value based on a cubic polynomial, indicated by the solid line, and estimated separately for quarters before and after the move. The predicted discontinuity at the quarter of move is reported in the panel header. To ensure sufficient number of leads and lags while simultaneously keeping the panel of migrating cardiologists balanced, we follow the migrating cardiologist for eight quarters before and after the move. As the estimated parameters are only identified up to scale, we use the quarter prior to the move normalized to zero as the baseline.

The estimated parameters prior to the physician's relocation are not significantly distinguishable from zero (i.e., the baseline period), suggesting that migrating physicians did not systematically respond to the counterfactual practice environment prior to their move. Moreover, for the model without hospital fixed effects, there is a visible sharp discontinuity occurring at the time of cardiologist relocation where the estimated $\gamma_{t'}$ coefficients become positive and highly significant. The estimated magnitude of this discontinuity is around 0.51 and close to the one reported in Column (3) of Table 5.3. Interestingly, the cardiologists appear to rapidly and permanently adapt to the prevailing practice style at the destination hospital for the entire duration of the follow-up period.

The corresponding period-specific effect pattern in the right figure panel, where hospital fixed effects have been added to the model, describes a smaller, but still pronounced, change in the moving cardiologist's behaviour at the time of relocation. In this case, we observe a somewhat more gradual adaptation to the destination hospital's practice environment over time and that the initial discontinuity at the time of relocation is somewhat smaller. We conclude from this analysis that cardiologists in our sample are partially malleable to their practice environment in terms of their own practice behaviour, and that they are equally responsive to their social environment as they are to their physical environment.

FIGURE 5.4: Event study estimates of migrating cardiologists' changes in practice environment: Use of DES



NOTE.— SCAAR data for years 2004–2013. Dots correspond to coefficient estimates of $\gamma_{t'}$ from OLS estimation of Equation (3.9). Dependent variable is an indicator for whether a patient undergoing PCI received a DES. Covariates include hospital, cardiologist characteristics and patient risk factors reported in Table 4.1 and fixed effects for year-quarter, cardiologist, and hospital (right panel only). Vertical spikes around coefficient estimates pertain to robust 95 percent confidence intervals clustered by hospital.

5.2 Impact on quality of care

We next study the extent to which the environmentally induced changes in migrating cardiologists' DES use affected the appropriateness of physician treatment choice and their consequences for patients' health outcomes. To this end, we estimate versions of Equation (3.8) and Equation (3.9) by replacing our outcome variable with the three indicators for major adverse cardiac events we consider: patient death, myocardial infarction (MI), and total lesion revascularization (TLR) within a year from the initial intervention. Moreover, we compare changes in physicians' rates of decision errors before and after their relocation using predictions from the RF machine learning algorithm to predict optimal treatment choice.

Decision errors

Table 5.4 reports DD estimation results using decision errors, based on the correspondence between migrating cardiologists' choices and predictions from our RF machine learning algorithm, as outcomes. Columns (1), (2) and (3) show the estimates on the overall propensity to make inappropriate decisions, and for Type I and Type II errors, respectively. Recall that Type I errors (false positives) refer to the application of DES when BMS is the recommended treatment choice, and vice versa for Type II errors (false negatives). This distinction is relevant as it is likely that making errors of the former type may be subject to more severe risks for the patient due to the possibility of stent thrombosis. In contrast, the latter error type may be more associated with higher medical costs in the form of a higher prevalence of restenosis and the need for subsequent interventions.

The results from estimation show that the overall probability of making a treatment error is positive, although not significantly different after, relative to before, cardiologist relocation. Splitting the decision errors into Type I and Type II errors, we find that physicians are somewhat more likely to make Type I errors after their change in practice environment. In contrast, the risk of committing Type II errors is reduced, but not significantly so. Hence, this result suggests that migrating cardiologists are more likely to overuse DES when they move to a hospital with higher use of DES than they are to overuse BMS when moving to a hospital with lower DES use. In the next subsection we explore whether these changes were associated with changes in patients' health outcomes.

TABLE 5.4: Difference-in-Differences estimates of migrating cardiologists' changes in practice environment: Decision errors

	(1) Error	(2) Type I	(3) Type II
Post	0.005	0.026	-0.020
	(0.027)	(0.018)	(0.024)
Δ_{jt}	-0.025	-0.014	-0.014
	(0.068)	(0.053)	(0.069)
Post $\times$ Δ_{jt}	0.096	0.185**	-0.081
	(0.081)	(0.075)	(0.077)
Covariates	✓	✓	✓
Year-quarter FE	✓	✓	✓
Cardiologist FE	✓	✓	✓
Hospital FE	✓	✓	✓
Cardiologists	51	51	51
Observations	8,589	8,589	8,589

NOTE.— SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (3.8). Dependent variables are indicators for whether a patient undergoing PCI received a non-recommended stent type. See Section 4.2 for details. Column (1) reports results for the propensity to commit any error while Column (2) and (3) reports error decomposition results for false positives and false negatives, respectively. Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Patient health outcomes

Columns (2)-(4) of Table 5.5 report results from estimation of Equation (3.8) for the three adverse patient health outcomes we consider: patient death, myocardial infarction (MI), and total lesion revascularization (TLR) within a year from the initial intervention. For comparison, the first column of the table reproduces the results from our preferred specification in Column (4) of Table 5.3. Each column corresponds to a specific outcome for our model with hospital fixed effects, implying that the reported point estimates refer to physician responses to the change in their peer environment. As before, the reported parameter estimates are interpreted as the rate of change in the outcome from a one percentage point change in the physicians' practice environment between the origin and destination hospitals. A negative sign implies that the risk of the event is less likely, whereas a positive coefficient indicates a higher risk.

The reported parameter estimates suggest that rates of changes in patient outcomes are generally small and statistically indistinguishable from zero. The point estimate of 0.04 for MI is greatest in magnitude, but is only one-sixth of the response for the choice of stent. We interpret this finding as indicating that patient health outcomes are not systematically related to migrating physicians' adaptation to their peer practice environment. One possible reason for this result could be that the estimated changes in the cardiologists' use of DES after relocation were mainly based on low-risk patients for which the choice between a BMS and a DES was unlikely to put patients at serious health risks.

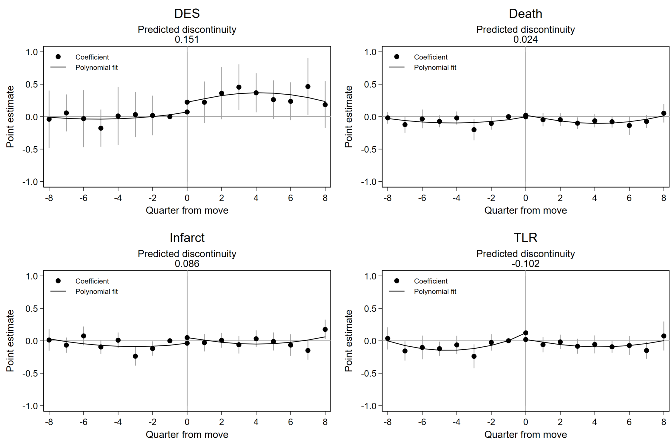
TABLE 5.5: Difference-in-Differences estimates of migrating cardiologists' changes in practice environment: Patient outcomes

	(1) DES	(2) Death	(3) Infarct	(4) TLR
Post	0.003	-0.009	0.001	-0.009
	(0.023)	(0.008)	(0.011)	(0.011)
Δ_{jt}	0.013	-0.047	-0.069*	-0.053
	(0.087)	(0.030)	(0.037)	(0.033)
Post × Δ_{jt}	0.247***	-0.011	0.041	0.028
	(0.090)	(0.027)	(0.042)	(0.033)
Covariates	✓	✓	✓	✓
Year-quarter FE	✓	✓	✓	✓
Cardiologist FE	✓	✓	✓	✓
Hospital FE	✓	✓	✓	✓
Cardiologists	51	51	51	51
Observations	8,589	8,589	8,589	8,589

NOTE.— SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (3.8). Dependent variables from left to right are indicators for whether a patient undergoing PCI received a DES and whether the patient died, suffered a myocardial infarction, or had another angioplasty within one year from the intervention, respectively. See Section 4.2 for details. Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5.5 illustrates the corresponding event study graphs based on Equation (3.9) and the outcomes from Table 5.5. The four panels in the figure, separated by patient outcome, provide a similar pattern as above with no indications of important changes in patient health outcomes at any point over the two years before or after cardiologists' relocation. These results show that the changes in treatment behaviour induced by variation in the migrating cardiologists' peer practice environment did not affect the quality of care in terms of patient outcomes to any important extent.

FIGURE 5.5: Event study estimates of migrating cardiologists' changes in practice environment: Patient outcomes



NOTE.— SCAAR data for years 2004–2013. Dots correspond to coefficient estimates of $\gamma_{t'}$ from OLS estimation of Equation (3.9). Dependent variables from top left to bottom right are indicators for whether a patient undergoing PCI received a DES and whether the patient died, suffered a myocardial infarction, or had another angioplasty within one year from the intervention, respectively. Covariates include hospital, cardiologist characteristics and patient risk factors reported in Table 4.1 and fixed effects for year-quarter, cardiologist, and hospital. Vertical spikes around coefficient estimates pertain to robust 95 percent confidence intervals clustered by hospital.

5.3 Robustness and sensitivity checks

Lastly, we report estimation results from a set of extensions to our main analysis to gauge the sensitivity of our findings to alternative model and sample specifications. We first study effect heterogeneity with respect to physician age and the direction of the change in practice environment of migrants. Next, we analyse the stability of our results with respect to the definition of the practice environment by reestimating our main DD model using a synthetic environment and non-moving cardiologists to predict counterfactual states.

Heterogeneity across physicians and change in practice environment

Table 5.6 reports split-sample results from estimation of our main DD model separately for cardiologists moving to hospitals with higher and lower shares of DES, displayed in Columns (1) and (2), and for younger and older migrants, based on the median age of migrating cardiologists, displayed in Columns (3) and (4), respectively. Again, we focus on the peer environment by including hospital fixed effects in the model. The intuition behind this analysis is to evaluate whether our main results are driven by specific subgroups. We anticipate that relatively younger physicians' practice styles are likely to be more malleable due to their lower practical experience and being in an earlier stage of their careers, consistent with the theory of champions, or opinion leaders, of clinical care (see, e.g., Shortell *et al.*, 2004). Furthermore, it is possible that migrating physicians are more susceptible to adopting treatment styles in more innovative practice environments, here characterized as a higher share of the relatively newer DES, due to the attractiveness of new technology (see, e.g., Hofmann, 2015).

Our predictions align with the empirical evidence reported in Table 5.6 in that the estimated response to the change in practice environment is mainly driven by younger cardiologists who move to a more innovative environment in terms of a higher average use of DES. While the first two columns suggest that the effect is positive for both positive and negative Δ_{jt}'s (albeit the latter coefficient is not statistically significant), the last two columns indicate that older cardiologists do not respond at all to their peer practice environment when relocating. We conclude that heterogeneity in the effect across both physicians and environments appear to be important to understand how clinicians react to their practice environment.

TABLE 5.6: Difference-in-Differences estimates of migrating cardiologists' changes in practice environment: Heterogeneity analyses

	Environment $\pm$		Physician age	
	(1) $\Delta_{jt} > 0$	(2) $\Delta_{jt} < 0$	(3) Below median	(4) Above median
Post	-0.021	-0.002	0.020	-0.059
	(0.051)	(0.043)	(0.025)	(0.038)
Δ_{jt}	-0.077	0.075	0.161	-0.032
	(0.129)	(0.146)	(0.142)	(0.106)
Post $\times$ Δ_{jt}	0.323**	0.184	0.292*	-0.080
	(0.154)	(0.187)	(0.159)	(0.121)
Covariates	✓	✓	✓	✓
Year-quarter FE	✓	✓	✓	✓
Cardiologist FE	✓	✓	✓	✓
Hospital FE	✓	✓	✓	✓
Cardiologists	24	27	23	28
Observations	3,776	4,813	4,429	4,160

NOTE.— SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (3.8). Dependent variable is an indicator for whether a patient undergoing PCI received a DES for different subsamples. Columns (1) and (2) splits the sample into cardiologists moving to more and less DES-intensive hospitals. Columns (3) and (4) splits the sample into younger and older cardiologists with median cardiologist age as threshold. Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Synthetic environment

One empirical issue with the DD approach outlined so far is that migrating cardiologists are unlikely to randomly relocate between hospitals. This leads to two inferential problems with respect to the interpretation of our main findings. The first problem relates to the external validity of our estimated effects. Migrating physicians may constitute a selected group that is unrepresentative for the physician population at large. While Table 4.1 suggests some differences in observable characteristics between moving and non-moving physicians, such as age, the case-mix of patients they treat and the quality of care they provide is indistinguishable from those of non-moving cardiologists. We take this as evidence supporting the notion that the subpopulation of cardiologists moving across hospitals is not widely different from non-moving cardiologists with respect to relevant characteristics.

The second problem relates to the internal validity of our estimates and is potentially more severe as it may invalidate our approach altogether. Specifically, if physicians generally move to hospitals based on their preferences for using DES, the associations we estimate and interpret as caused by changes in practice environment cannot be empirically distinguished from the sorting of physicians to hospitals with practice environments based on their clinical preferences. Although the results from Figure 4.3 and Table 5.3 are reassuring in the sense that the common trend assumption is not rejected, we may still be concerned that the counterfactual practice environment is estimated with bias. To test whether our approach is robust to alternative definitions of practice environments, we propose to extend our analysis by using a synthetic control method derived from a different source of variation to estimate the counterfactual practice environment.

To find a suitable control group that can serve to identify the counterfactual state of migrating cardiologists should they not have moved, we define a synthetic practice style environment from the pool of non-migrating cardiologists (see, e.g., Abadie *et al.*, 2010, 2015; Abadie and Gardeazabal, 2003).[14] For each migrating cardiologist $j \in J$, we define $\tilde{\Delta}_{jt} = \sum_c w_c \Delta_{ct}$ as the counterfactual environment based on non-migrating cardiologists, $c \in C \notin J$. The weights, w_c, are chosen to minimize functions of pre-migration DES share levels ($\sum_{s \in t < t_0} \Delta_{js} - \tilde{\Delta}_{js}$) and slopes ($\sum_{s \in t < t_0} \partial \Delta_{js} / \partial s - \partial \tilde{\Delta}_{js} / \partial s$) based on a constrained quadratic optimization routine. A corresponding approach is applied to estimate the counterfactual environment in the pre-migration period using post-migration DES share levels and slopes. Finally, the re-

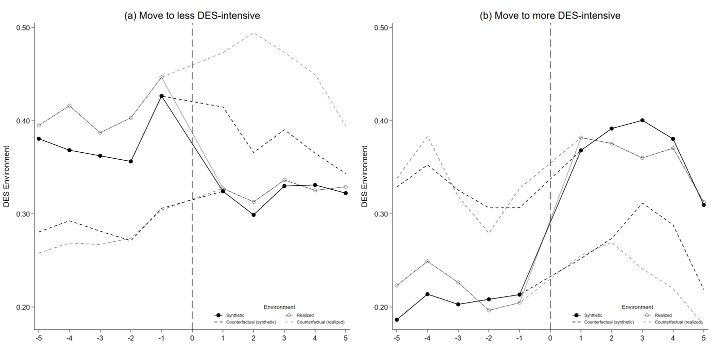
[14] Although the synthetic control method was originally developed for a single treated unit, the framework can easily accommodate estimation with multiple treated units by fitting separate synthetic controls for each of the treated units (see, e.g., Abadie, 2020). While there is no important conceptual difference in the contexts of one versus multiple treated units, practice issues relating to the non-uniqueness of the solution to the minimization problem when selecting weights for the synthetic controls are exacerbated in the latter. To address this issue, Abadie and L'Hour (2019) propose a synthetic control estimator that incorporates a penalty for pairwise matching discrepancies between the treated units and each of the units that contributes to their synthetic controls.

sulting counterfactual estimates are applied to versions of the event study model in Equation (3.9) where the original practice style environment, Δ_{jt}, has been replaced with its synthetic equivalent, $\tilde{\Delta}_{jt}$.

Figure 5.6 illustrates the synthetic environment approach (darker-coloured lines) and how it relates to the previous approach by overlaying the corresponding trends in practice environment from Figure 4.3 (brighter-coloured lines). The two definitions mostly overlap, with the exception of the post-migration counterfactual environment among cardiologists moving to less DES-intensive hospitals that is somewhat lower than the corresponding environment using the original approach. This suggests that, while the two types of counterfactual environments are partially based on the same empirical variation, there are also important differences between them.

Finally, we study whether our main estimation results are sensitive to the definition of practice environment. Table 5.7 reports estimation results from our main DD model where we have replaced Δ_{jt} with $\tilde{\Delta}_{jt}$ in the analysis. Reassuringly, the results are close to our main estimation from Table 5.5: a change in DES use of migrating cardiologists of around 0.31 percentage points for each percentage point change difference in synthetic practice environment between origin and destination hospitals but no corresponding impacts on adverse patient outcomes. We conclude from this analysis that our main results are robust to the definition of practice environment with respect to whether it is derived from the hospital or from the pool of non-migrating cardiologists.

FIGURE 5.6: Average trends in migrating cardiologists' realized and synthetic practice environments



NOTE.— SCAAR data for years 2004–2013. Practice environment defined as the share of DES used in angioplasty treatments in realized (solid lines) and counterfactual (dashed lines) hospitals by quarter from cardiologist move. Brighter lines pertain to estimates of Δ_{jt} while darker lines pertain to the estimated synthetic practice environment, $\tilde{\Delta}_{jt}$. See Section 5.3 for details on the construction of this variable. Separate plots for cardiologists moving to hospitals with higher and lower intensity of DES use. Vertical dashed line indicates re-centered quarter of physician relocation from the origin to the destination hospital. Quarter of move linearly interpolated.

TABLE 5.7: Difference-in-Differences estimates of migrating cardiologists' changes in synthetic practice environment: Patient outcomes

	(1) DES	(2) Death	(3) Infarct	(4) TLR
$Post$	-0.022	-0.009	0.005	-0.011
	(0.023)	(0.008)	(0.012)	(0.011)
$\tilde{\Delta}_{jt}$	0.122	-0.060	-0.019	-0.047
	(0.139)	(0.036)	(0.025)	(0.043)
$Post \times \tilde{\Delta}_{jt}$	0.312**	0.019	0.006	0.056
	(0.128)	(0.028)	(0.038)	(0.053)
Covariates	✓	✓	✓	✓
Year-quarter FE	✓	✓	✓	✓
Cardiologist FE	✓	✓	✓	✓
Hospital FE	✓	✓	✓	✓
Cardiologists	51	51	51	51
Observations	6,729	6,729	6,729	6,729

NOTE.— SCAAR data for years 2004–2013. Coefficient estimates from OLS estimation of Equation (3.8) using the estimated synthetic practice environment, $\tilde{\Delta}_{jt}$ in place of Δ_{jt}. See Section 5.3 for details on the construction of this variable. Dependent variables from left to right are indicators for whether a patient undergoing PCI received a DES and whether the patient died, suffered a myocardial infarction, or had another angioplasty within one year from the intervention, respectively. See Section 4.2 for details. Covariates include all hospital and cardiologist characteristics as well as patient risk factors reported in Table 4.1. Robust standard errors clustered by hospital in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6 Conclusion

This paper empirically analyses how physicians' treatment decisions are influenced by their practice environment and how such decisions affect the quality of care received by patients. We study these questions in the context of the choice between using bare metal stents (BMS) or drug-eluting stents (DES) among interventional cardiologists in Sweden performing percutaneous coronary interventions (PCI) on patients diagnosed with coronary artery disease. To obtain empirical variation in a physician's practice environment, we identify cardiologists who moved between hospitals and relate changes in their own treatment behaviour and subsequent patient outcomes to differences in the hospital's practice environment before and after they relocated. The overall physician response to their environment is then decomposed into a physical (provider-specific) and a social (peer group-specific) component by exploiting quasi-random information on the practice behaviour of migrating physicians' co-workers within a hospital. Finally, we relate the environmentally induced changes in practice environment to variations in physicians' rate of decision errors and patient adverse clinical events to gauge whether the practice style changes led to important changes in quality of care provision.

Similar to the results reported in Molitor (2018), we find that migrating cardiologists rapidly, but not fully, adapt to the prevailing practice environment in their use of DES after relocating. Our estimates suggest that cardiologists change their use of DES with around 0.5 percentage points for each percentage point difference in practice environment between the origin and destination hospitals. Decomposing the overall effect into a provider-specific and a peer group-specific component, we find that around half of the response is driven by the latter effect, suggesting that a physician's peer group is as influential as the physical work environment in altering treatment styles. Furthermore, we find no evidence that neither major adverse cardiac events, such as heart attacks or patient death, or physician decision errors, measured using a Random Forest (RF) machine learning algorithm, were strongly associated with changes in the migrating physicians' treatment styles. This could potentially be explained by that medical decisions were still made within prevailing medical guidelines and did not lead to significantly increased health risks for cardiac patients. Finally, estimation results from a set of split-sample heterogeneity analyses show that our main effects are primarily driven by younger cardiologists who move to more innovative environments (i.e., with higher use of DES), suggesting that both environmental as well as individual characteristics appear to be important for the magnitude of physician response.

In conclusion, the results obtained in this paper have important bearing on current health policy with respect to the causes and consequences of unwarranted regional variations in healthcare use (see, e.g., Corallo *et al.*, 2014). Recent evidence on the extent to which regional variations are driven by providers or individual clinicians have emphasized the role of the latter (see, e.g., Gutacker *et al.*, 2018). That physicians strongly respond and adapt to their prevailing practice environment, and that such conforming arises from both the provider itself and from the workplace peers, suggest a rationale for why physician treatment styles may cluster in specific areas. The absence of an impact on patient outcomes from such adjustments also provides an explanation for the conundrum of a weak observable correlation between regional variations in the costs and the quality of healthcare provision (see, e.g., Fisher *et al.*, 2003a,b). Although concrete policy advice may require more substantiated evidence, which we leave for further work, we believe that our results show that information campaigns aimed at harmonizing treatment choice among healthcare professionals, such as clinical guidelines, may not suffice to significantly reduce unwarranted variations in healthcare use.

References

Abadie, A. (2020). *Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects*. Unpublished manuscript.

—, Diamond, A. and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, **105** (490), 493–505.

—, — and — (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, **59** (2), 495–510.

— and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, **93** (1), 112–132.

— and L'Hour, J. (2019). *A Penalized Synthetic Control Estimator for Disaggregated Data*. Unpublished manuscript.

Baicker, K. and Chandra, A. (2004). Medicare spending, the physician workforce, and beneficiaries' quality of care. *Health Affairs*, **23** (Suppl1), W4–184–W4–197.

Bandiera, O., Barankay, I. and Rasul, I. (2010). Social incentives in the workplace. *The Review of Economic Studies*, **77** (2), 417–458.

Bojke, C., Castelli, A., Street, A., Ward, P. and Laudicella, M. (2013). Regional variation in the productivity of the English National Health Service. *Health Economics*, **22** (2), 194–211.

Breiman, L. (2001). Random forests. *Machine Learning*, **45** (1), 5–32.

Burke, M. A., Fournier, G. and Prasad, K. (2003). *Physician social networks and geographical variation in medical care*. Center on Social and Economic Dynamics Washington, DC.

Chandra, A., Cutler, D. and Song, Z. (2012). Who ordered that? The economics of treatment choices in medical care. In T. M. Mark, V. Pauly and P. P. Barros (eds.), *Handbook of Health Economics*, vol. 2, *Chapter 6*, Elsevier, pp. 397–432.

— and Staiger, D. O. (2007). Productivity spillovers in health care: Evidence from the treatment of heart attacks. *Journal of Political Economy*, **115** (1), 103–140.

— and — (2020). Identifying sources of inefficiency in healthcare. *The Quarterly Journal of Economics*, **135** (2), 785–843.

Chitkara, K. and Gershlick, A. (2010). Second versus first-generation drug-eluting stents. *Journal of Interventional Cardiology*, **5** (1), 23–26.

Corallo, A. N., Croxford, R., Goodman, D. C., Bryan, E. L., Srivastava, D. and Stukel, T. A. (2014). A systematic review of medical practice variation in OECD countries. *Health Policy*, **114** (1), 5–14.

Currie, J., MacLeod, W. B. and Van Parys, J. (2016). Provider practice style and patient health outcomes: The case of heart attacks. *Journal of Health Economics*, **47**, 64–80.

Currie, J. M. and MacLeod, W. B. (2020). Understanding doctor decision making: The case of depression treatment. *Econometrica*, **88** (3), 847–878.

Cutler, D., Skinner, J. S., Stern, A. D. and Wennberg, D. (2019). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *American Economic Journal: Economic Policy*, **11** (1), 192–221.

Doyle, J. J., Ewer, S. M. and Wagner, T. H. (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of Health Economics*, **29** (6), 866–882.

—, Graves, J. A. and Gruber, J. (2017). Uncovering waste in US healthcare: Evidence from ambulance referral patterns. *Journal of Health Economics*, **54**, 25–39.

—, —, — and Kleiner, S. A. (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy*, **123** (1), 170–214.

Ekman, M., Sjögren, I. and James, S. (2006). Cost-effectiveness of the Taxus paclitaxel-eluting stent in the Swedish healthcare system. *Scandinavian Cardiovascular Journal*, **40** (1), 17–24.

Epstein, A. J. and Nicholson, S. (2009). The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics*, **28** (6), 1126–1140.

Falk, A. and Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, **24** (1), 39–57.

Finkelstein, A., Gentzkow, M. and Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics*, **131** (4), 1681–1726.

Fisher, E. S., Wennberg, D. E., Stukel, T. A., Gottlieb, D. J., Lucas, F. L. and Pinder, E. L. (2003a). The implications of regional variations in Medicare spending. Part 1: The content, quality, and accessibility of care. *Annals of Internal Medicine*, **138** (4), 273–287.

—, —, —, —, — and — (2003b). The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care. *Annals of Internal Medicine*, **138** (4), 288–298.

Foster, G. (2006). It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics*, **90** (8-9), 1455–1475.

Grytten, J. and Sørensen, R. (2003). Practice variation and physician-specific effects. *Journal of Health Economics*, **22** (3), 403–418.

Gutacker, N., Bloor, K., Bojke, C. and Walshe, K. (2018). Should interventions to reduce variation in care quality target doctors or hospitals? *Health Policy*, **122** (6), 660–666.

Heijmans, N., Van Lieshout, J. and Wensing, M. (2017). Information exchange networks of health care providers and evidence-based cardiovascular risk management: An observational study. *Implementation Science*, **12** (7), 1–12.

Hofmann, B. M. (2015). Too much technology. *British Medical Journal*, **350** (7996), h705.

Huesch, M. D. (2011). Is blood thicker than water? Peer effects in stent utilization among Floridian cardiologists. *Social Science & Medicine*, **73** (12), 1756–1765.

Karaca-Mandic, P., Town, R. J. and Wilcock, A. (2017). The effect of physician and hospital market structure on medical technology diffusion. *Health Services Research*, **52** (2), 579–598.

Kopetsch, T. and Schmitz, H. (2014). Regional variation in the utilisation of ambulatory services in Germany. *Health Economics*, **23** (12), 1481–1492.

Lin, N., Ensel, W. M. and Vaughn, J. C. (1981). Social resources and strength of ties: Structural factors in occupational status attainment. *American Sociological Review*, **46** (4), 393–405.

Lyle, D. S. (2007). Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point. *The Review of Economics and Statistics*, **89** (2), 289–299.

Mafi, J. N., Russell, K., Bortz, B. A., Dachary, M., Hazel, W. A. and Fendrick, A. M. (2017). Low-cost, high-volume health services contribute the most to unnecessary health spending. *Health Affairs*, **36** (10), 1701–1704.

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, **99** (1), 112–145.

McClellan, M. (2011). Reforming payments to healthcare providers: The key to slowing healthcare cost growth while improving quality? *The Journal of Economic Perspectives*, **25** (2), 69–92.

— and Newhouse, J. P. (1997). The marginal cost-effectiveness of medical technology: A panel instrumental-variables approach. *Journal of Econometrics*, **77** (1), 39–64.

Molitor, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy*, **10** (1), 326–56.

Nair, H. S., Manchanda, P. and Bhatia, T. (2010). Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research*, **47** (5), 883–895.

OECD (2014). Geographic variations in health care: What do we know and what can be done to improve health system performance? OECD Health Policy Studies, OECD Publishing.

Phelps, C. E. (2000). Information diffusion and best practice adoption. In A. Culyer and J. Newhouse (eds.), *Handbook of Health Economics*, vol. 1, *Chapter 5*, Elsevier, pp. 223–264.

Prieto, D. C. and Lago-Peñas, S. (2012). Decomposing the determinants of health care expenditure: The case of Spain. *The European Journal of Health Economics*, **13** (1), 19–27.

Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.

Reich, O., Weins, C., Schusterschitz, C. and Thöni, M. (2012). Exploring the disparities of regional health care expenditures in Switzerland: Some empirical evidence. *The European Journal of Health Economics*, **13** (2), 193–202.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, **116** (2), 681–704.

Shortell, S. M., Marsteller, J. A., Lin, M., Pearson, M. L., Wu, S.-Y., Mendel, P., Cretin, S. and Rosen, M. (2004). The role of perceived team effectiveness in improving chronic illness care. *Medical Care*, **42** (11), 1040–1048.

Shrank, W. H., Rogstad, T. L. and Parekh, N. (2019). Waste in the US health care system: Estimated costs and potential for savings. *Journal of the American Medical Association*, **322** (15), 1501–1509.

Skinner, J. (2011). Causes and consequences of regional variations in health care. In T. M. Mark, V. Pauly and P. P. Barros (eds.), *Handbook of Health Economics*, vol. 2, *Chapter 2*, Elsevier, pp. 45–93.

—, Gottlieb, D. J. and Carmichael, D. (2011). A new series of Medicare expenditure measures by hospital referral region: 2003-2008. *Dartmouth Atlas Project*.

— and Staiger, D. (2015). Technology diffusion and productivity growth in health care. *Review of Economics and Statistics*, **97** (5), 951–964.

Stinebrickner, R. and Stinebrickner, T. R. (2006). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics*, **90** (8-9), 1435–1454.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P. and Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, **43** (6), 1947–1958.

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., Carter, A., Casey, D. C., Charlson, F. J., Chen, A. Z. *et al.* (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, **388** (10053), 1545–1602.

Wennberg, J. and Gittelsohn, A. (1973). Small area variations in health care delivery: A population-based health information system can guide planning and regulatory decision-making. *Science*, **182** (4117), 1102–1108.

Wennberg, J. E. (2010). *Tracking medicine: A researcher's quest to understand health care*. Oxford University Press.

—, Fisher, E. S. and Skinner, J. S. (2002). Geography and the debate over Medicare reform. *Health Affairs*, **21** (Suppl1), W96–W112.

Yang, M., Lien, H.-M. and Chou, S.-Y. (2014). Is there a physician peer effect? Evidence from new drug prescriptions. *Economic Inquiry*, **52** (1), 116–137.

Yuan, C. T., Nembhard, I. M. and Kane, G. C. (2020). The influence of peer beliefs on nurses' use of new health information technology: A social network analysis. *Social Science & Medicine*, **255**, 113002.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, **85** (1), 9–23.

Appendix: Additional tables and figures

FIGURE A.7: Random Forest machine learning algorithm

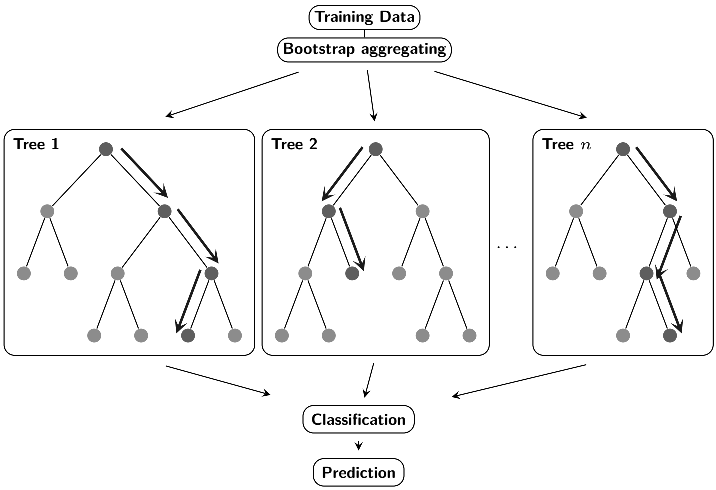
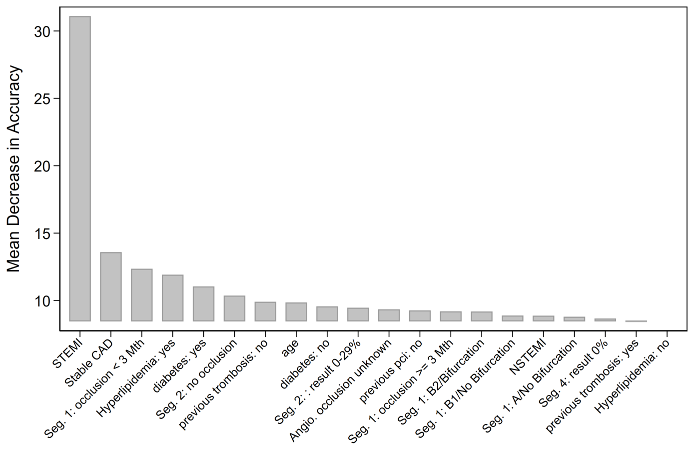
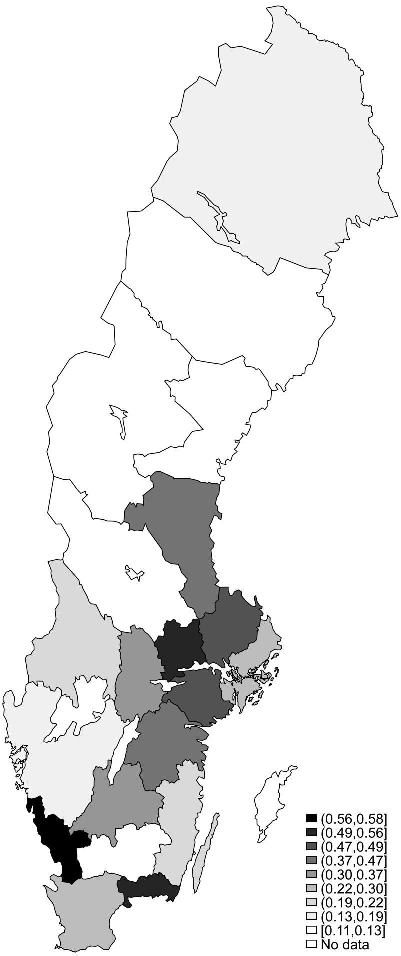


FIGURE A.8: Variable importance weights in Random Forest prediction



NOTE.— SCAAR data for years 2011–2012. Higher values indicate greater importance of variable in predicting outcomes. Included variables: patient's gender; age; reason for hospitalization; diabetes; COPD; peripheral vascular disease; hypertension; hyperlipidemia; previous infarction; previous CABG; previous PCI; previous stroke; patient creatinin clear; hemoglobin test; any occlusion; angiography results by segment including degree of stenosis severity and duration ; left ventricular ejection fraction; location of lesions; 3-vessel and/or LM lesion; number of treated segments; primary diagnosis.

Figure A.9: Distribution of raw DES rates across hospital regions in Sweden, 2004–2013



Note.— SCAAR data for years 2004–2013. Regional administrative map of the 21 county councils in Sweden. Intensity of shaded areas reflect average shares of DES use among patients undergoing angioplasty treatment across all years.

Essay 3: Hospital Closures, Patient Outcomes and Local Politics. Evidence from Germany[*]

1 Introduction

Pledges of greater efficiency, better coordinated healthcare and lower costs have led many healthcare systems to experience strong consolidation tendencies over the last two decades in the form of hospital closures, mergers and acquisitions[1] and privatizations[2] (Harrison, 2007; Town *et al.*, 2006). Driven by rapidly rising healthcare costs such policies resulted in multiple benefits to the efficiency in healthcare, in particular related to the coordination and the management of healthcare (Ciliberto and Lindrooth, 2007; Deily *et al.*, 2000; Lindrooth *et al.*, 2003). Yet public concerns with respect to both the equity of healthcare access and the quality of care still remain at the forefront of the political arena.[3] Competition-driven market concentration and the shift of medical resources from remote to metropolitan areas lead to the establishment of large hospital complexes in urban areas at the

[*]The author thanks to Ansgar Wübker and Adam Pilny for insightful comments about the German healthcare system; Christian Wittrock for helpful insights and information about the German political system. Valuable comments and suggestions by Anthony Harris, Daniel Avdic, Adam Irving, Terence C. Cheng and participants at the Australian Health Economic Doctoral workshop 2019, and the Australian Health Economics Society Conference 2019 are much appreciated. Financial support from the RWI – Leibniz-Institut für Wirtschaftsforschung and the Leibniz Science Campus Ruhr is gratefully acknowledged.

[1]For an overview of the hospital merger literature see: Coenen *et al.* (2012); Cuellar and Gertler (2003); Dor and Friedman (1994); Dranove (1998); Dranove and Lindrooth (2003); Dranove and Shanley (1995); Hansmann *et al.* (2007); Ho and Hamilton (2000); Huckman (2006); Krishnan *et al.* (2004); Pilny (2014); Schmid and Varkevisser (2016); Schmitt (2017); Sloan *et al.* (2003).

[2]A common policy to increase healthcare savings and improve the efficiency in the provision of healthcare is privatization of healthcare providers. See i.e., Cutler and Horwitz (2007); Mark (1999); Shen (2003).

[3]See, i.e., Avdic (2016); Bindman *et al.* (1990); Buchmueller *et al.* (2006); Burkey *et al.* (2017); Capps *et al.* (2010); Countouris *et al.* (2014); Harrison (2007); Hsia *et al.* (2012)

expense of smaller units in rural areas for which populations are often too small to financially support resource-intensive medical centers. This pattern causes geographical imbalances in catering to the demand for and access to emergency healthcare for urban and rural populations. For instance, the German Federal Statistical Office states that only 64% of rural residents are able to reach the closest hospital offering basic healthcare within 15 minutes, while the corresponding figure is over 90% for the urban population.[4] In addition to the deteriorated geographical access, researchers found that the concentration of healthcare markets inhibit provider competition that consequently lead to increases in costs of care (Gaynor, 2011) and further promote the migration of medical professionals away from under-served areas (Benham *et al.*, 1968; Kuhn and Ochsen, 2019; Vogt, 2016; Zuckerman *et al.*, 1990).

This study contributes to the literature on the effects of healthcare consolidation policies. I study a case of hospital market exits in Germany to empirically assess the causal impact of hospital closure on geographical access to healthcare and multiple clinical patient outcomes in the context of acute myocardial infarction (*AMI*) and hemorrhagic stroke. Cardiovascular diseases, for which AMI and stroke are the two most common manifestations, are the leading cause of death globally causing nearly 18 million deaths worldwide each year (World Health Organization, 2011) and are the number one reason for all medical emergencies (Linden, 2006). Timely access to healthcare is essential for patients with these conditions, thus deteriorating access, as i.e. due to healthcare consolidation, might impinge the chances of survival as well as medical complications during the recovery.[5]

A number of studies has investigated the effects of healthcare consolidation trends on the geographical healthcare access. For instance, a study by Burkey *et al.* (2017) analysed the closure of several hospitals in the Southeastern U.S. and did not find any significant impact on healthcare access. Similarly, Hentschker and Mennicken (2014) estimated only a marginal increase in the travelling distance for patients with a hip fracture or abdominal aortic aneurysm after hospital closures in Germany. However, these findings rely on a strong assumption that all hospitals provide universal care and patients could receive similar care in any given hospital. As hospitals often differ in both the services they provide and in the quality of their care, the empirical setting used in these studies potentially biases the true effect of travel distance downwards. To address this limitation, Mennicken *et al.* (2014) studied the centralisation of hospital services in gynaecology and obstetrics and, similar to previous studies, found that patients did not travel

[4]Based on the *Hospital Atlas (Krankenhausatlass)* 2016 published by DESTATIS - the Federal Statistical office of Germany. More statistics about the German healthcare sector can be found at www.destatis.de.

[5]According to the report published by American Heart Association every minute without treatment for a patient with an *AMI* reduces survival chances by 7 to 10 percent (American Heart Association, 2003).

further after hospital closures. However, while aligned with the discussed literature, the study raises concerns about potential patient sorting caused by differences in the quality of care that are particularly important when studying planned procedures such as maternal care (Avdic *et al.*, 2019). To address this empirical drawback, several studies investigated the effect of hospital closures on patient outcomes. If inferior access to healthcare is affected by trending healthcare consolidation policies, it is likely to result in worse health outcomes, in particular for patient who require immediate medical attention. Studying hospital closures in Los Angelos County, Buchmueller *et al.* (2006) found that an increase in travel distance results in higher mortality rates from heart attacks and unintentional injuries. In line with this evidence, Avdic (2016) concluded that after closures of emergency departments in Sweden, patients had lower chances of surviving an acute myocardial infarction. Together, these and other studies[6] suggest that healthcare efficiency gains from consolidating services are likely to be accompanied with a deterioration in patient care for the most sensitive groups. In a similar manner to previously discussed research designs, this study also analyses patients requiring emergency care. However, instead of focusing on policy-induced variation in distance, I study the direct effects of hospital closures using an instrumental variable approach to overcome the empirical challenges arising from the endogeneity between the hospital quality and market structure.

My empirical analysis entails the use of a nationally representative sample of hospital discharge records provided by a large German health insurer. I identify the sample of interest using comprehensive clinical information about the diagnosis assigned at the time of admission. This sample is further augmented by the addition of two auxiliary data sources. First, I obtain information about all hospital closures in Germany that occurred in the years 2006 - 2012 from a report published by the Federal Joint Committee of German physicians (Preusker *et al.*, 2014). Detailed information about each hospital closure provided in this report allows me to identify all particularities of hospital market exits such as details about the process and execution, primary reasons and the exact time of closure that is particularly important in this empirical analysis. Using the information on the geographical location, for each hospital I define a hospital emergency market based on the radius distance and assign to treatment if the hospital closed during the study period. I employ a Propensity Score Matching technique to find a non-treated market similar to treated market in observed characteristics and reduce the bias arising from systematic differences between hospitals. I use a linear regression to estimate the effect of the treatment on patient outcomes while controlling for a number of observed confounding factors such as patient demographics, medical condition, hospital- and

[6]For more research findings that draw similar conclusions, see i.e., Blondel *et al.* (2011); Engjom *et al.* (2014); Grzybowski *et al.* (2011); Ravelli *et al.* (2011).

market-related characteristics. There may be several reasons for a hospital to close; however, in most of the cases, hospitals close when they are not able to cater to the demand due to remoteness, poor quality of care or financial performance. In order to adjust for such unobserved factors that might influence both hospital closures and patient outcomes, I employ an instrumental variable strategy. I collect data on political party composition of the local councils in German municipalities from the German Federal Statistical Offices and estimate the political party's voting shares. This measure represents the political dynamics and the distribution of political powers in each council and serve as an instrument for a decision regarding hospital closure.

Political decisions play a substantial role in shaping the German healthcare market. In the last decade federal policy-makers have adopted several major policies to encourage market efficiency. First, the 1993 Healthcare Structure Act introduced a number of changes in the hospital payment system that substantially limited hospital expenditure. Additional financial pressure from the introduction of a prospective payment system based on the Diagnosis-related groups was introduced in 2004 (Augurzky and Schmitz, 2010; Schulten, 2006). These reforms placed all healthcare providers under significant financial pressure, that, combined with high competition, resulted in a reduction of hospitals and hospital beds over the last two decades. The German healthcare market in 2020 was expected to be approximately 20% smaller than that of the early 1990s (Schulten, 2006). The most significant factors for this phenomenon are the size of a hospital (Augurzky and Schmitz, 2010; Pilny, 2014), the variety of services a hospital provides, the ownership type of a hospital as well as the financial status (Ciliberto and Lindrooth, 2007; Pilny, 2014; Succi *et al.*, 1997; Williams *et al.*, 1992). Even though the financial status is mainly the responsibility of the federal state, maintenance of a hospital falls on the political decision-making bodies in the municipality. Closing hospitals is a very unpopular politically, especially for local politicians, who often worry about their political decisions losing them votes in the next election. When the electoral margin is small, it is more likely that the politicians currently in office will be very cautious about implementing an unpopular hospital closure policy. Thus, hospitals are more likely close when the ruling political party received a significant majority of votes in the previous election. I follow a similar strategy to Bloom *et al.* (2015)[7] and construct a highly relevant instrument for hospital closure based on political pressure in the local governmental area.

The main findings of this paper provide evidence of the benefits of health-

[7]Political pressure has played a role as an instrument in the previous literature. Bloom *et al.* (2015) employed the political winning margin to instrument for the competition in the healthcare market; the political affiliation has also been used to instrument the size of police force in the area (Levitt, 1997) and the employment in the public hospitals (Clark and Milcent, 2011) .

care consolidation policies. Even though I find that a hospital closure causes a significant increase in travel distance of, on average, 4 kilometres (or 3 minutes of travel time) for patients residing in closure-affected areas, the reduced access to emergency services does not lead to worse patient outcomes. The results also suggest that both the likelihood for death in-hospital and within 30 days decrease in closure-affected areas after the closure; however, the effect is not statistically significant. I also explore several other treatment-related outcomes that relate to the efficiency of healthcare provision. I find that, likely due to an increased number of patients attending neighbouring hospitals, hospital closures lead to a more efficient provision, reducing length of stay of emergency admissions by approximately 2.5 days without impacting on readmission rates.

This study contributes to the growing research literature analysing healthcare consolidation in several ways. First, I explore potential mechanisms affecting hospital financing through politics in the German healthcare setting. By exploiting this mechanism, I minimize potential endogeneity bias arising from the hospital quality and market structure, a common manifestation when studying healthcare market exits in empirical settings. I employ electoral turnout data on all German municipalities to construct a strong and highly relevant instrument and provide evidence for geographical access, patient outcomes and the efficiency of healthcare provision. Studying emergency patients addresses concerns related to patient sorting to hospital that was rarely addressed in the previous literature. Information about each hospital closure from official reports provides further evidence for the validity of the results. They are also supported by a comprehensive administrative data with a rich set of explanatory variables to control for variation in outcomes and to reduce further endogeneity issues that arise from unobserved heterogeneity.

This paper is structured as follows. The next section (Section 2) presents a detailed overview of the healthcare system in Germany and the main causes and consequences of market exits. The remainder of this article introduces the data and sampling in Section 3, the econometric framework and the definition of the instrument in Section 4 and Section 5 respectively. Section 6 summarizes the main results and Section 7 presents a number of robustness checks. Finally, Section 8 concludes.

2 Institutional context

Health care in Germany

The German healthcare system has universal health insurance coverage that is based on a multi-payer insurer system. A mandatory membership in either the public statutory health insurance (SHI) and/or the private health insurance (PHI) ensures healthcare for all citizens and permanent residents.

Whether an insuree belongs to SHI or PHI is distinguished by the gross wage earnings[8] and the nature of work itself. For instance, the self-employed and civil servants can voluntarily choose PHI (Bünnings *et al.*, 2019). Currently around 90 percent of the German population is covered by SHI.[9]

Regardless of the type of insurance the healthcare provision is similar; both PHI and SHI offer a full range of healthcare services for all types and levels of care. The major difference between these insurance policies is the selection of health insurance plans as PHI allows for an individual to choose a tailored plan (e.g. cost-sharing, better accommodation at the hospital such as private wards), while SHI offers only one standardised health insurance plan. Due to the additional benefits, PHI insurees may face different tariffs and prices since PHI companies do not have to contract with healthcare providers; however, the maximum fee that providers may charge is regulated by the German Federal Ministry of Health to ensure fair pricing and impartiality among different insurees (Wasem *et al.*, 2004). Despite the type of insurance an insuree has, everyone is entitled to choose their healthcare provider, which fosters competition in the healthcare market (Avdic *et al.*, 2019).

The German healthcare system provides high quality, attentive care and professional services; however, increasing public expenditures on healthcare raise serious concerns and is often debated by policy makers.[10] Higher healthcare expenditures do not necessarily mirror in better quality of care or patient health outcomes (Garber and Skinner, 2008), highlighting both equity and efficiency concerns. In fact, efficiency is often at the forefront of the German political arena. Germany has the highest number of hospital beds per capita in the European Union. In 2017[11] there were approximately 800 hospital beds available per 100,000 inhabitants, significantly above the average of 541 beds in the European Union. Partly, as a consequence, healthcare in Germany accounts for nearly 12 % of GDP in 2018 compared to approximately 8% in the United Kingdom (OECD, 2019). High healthcare costs led to the political pressure to reduce public spending on hospitals, particularly on those identified as less efficient. Hospital consolidation and closures was a policy response and, as a result, the German healthcare market has shrunk by nearly 20 % since the 1990s.[12]

[8]In 2018 employees earning more than 59K €/$73K per year qualified for receiving the PHI.

[9]A detailed overview about German SHI is given in Pilny *et al.* (2017).

[10]Healthcare expenditure as a share of GDP has risen from 9.4% in 1992 to 11.7% in 2018. Despite some stability in recent years, expenditure per inhabitant rose from 3,771 € in 2012 to 4,712 € in 2018 due to demographic changes. (Federal Statistical Office Destatis.)

[11]Statistics provided by Eurostat, accessed on 04.05.20 at https://ec.europa.eu/eurostat/databrowser/view/tps00046/default/table?lang=en.

[12]In 1991 2,411 hospitals operated in Germany an in 2017 this number reduced to 1,942. Statistics accessed on the official website of the Federal Statistical Office (Destatis) at https://www.destatis.de/EN/Themes/Society-Environment/Health/Hospitals/Tables/gd-hospitals-years.html.

Why do hospitals close?

One of the main risk factors for market exit is the financial status of a hospital (Williams *et al.*, 1992). According to the findings of Succi *et al.* (1997) and Pilny (2014), hospitals that operate with higher cash flows are less likely to close or merge with other entities. This is closely linked to the size of the hospital, another significant factor when describing market exits. Hospitals with smaller capacities, especially located in rural and remote areas, suffer from financial distress more often as they offer fewer services. They usually provide only basic healthcare services including emergency care that, as a result, leads to weakly designed inpatient/outpatient care programs without access to sophisticated and high-tech services (Ciliberto and Lindrooth, 2007; Pilny, 2014; Succi *et al.*, 1997; Williams *et al.*, 1992). In addition to the hospital's size, the ownership status plays an important role. Based on evidence from the U.S. healthcare market, for-profit as well as public healthcare providers are more likely to exit the market than non-profit hospitals (Ciliberto and Lindrooth, 2007; Succi *et al.*, 1997; Williams *et al.*, 1992); however in the German healthcare market public hospitals together with non-profit organisations are more likely to experience mergers and acquisitions due to a high protection of the federal state (Pilny, 2014).[13]

Preusker *et al.* (2014) suggest that the most common reasons for both full as well as partial market exits in the German healthcare market are similar to those described in the wider literature. The majority of hospitals closed due to economic insolvency that accounts for approximately 68% of all closures from 2006 – 2012. Failure to meet regulatory quality requirements outlined in the *Hospital Plans* imposed by each federal state resulted in around 12 % of hospital exits. The introduction of prospective funding system on the DRGs (2004) penalized comparatively inefficient hospitals and forced additionally around 3% of hospitals to close. Some public hospitals were in practice only small clinics offering several inpatient care beds and closed due to high competition with their larger competitor. However, these and other similar reasons describe more exceptional cases, which occurred less frequently (Preusker *et al.*, 2014).

What are the consequences?

Hospital efficiency is often discussed in the literature both as a reason for as well as a consequence of the closure. Larger hospital complexes with a higher concentration of services often have better care coordination with access to high-tech services and more resources to support continuous learning for healthcare professionals. Using a measure of hospital relative efficiency calculated via a frontier cost function, Deily *et al.* (2000) suggest that inefficient hospitals are more likely to close regardless of their ownership status.

[13] A detailed overview about the hospital financing in Germany and the role of federal state is provided in Appendix B.

A similar conclusion is drawn by Ciliberto and Lindrooth (2007) who measure the efficiency by hospital fixed effects previously suggested by Skinner (1994). When a less efficient hospital closes, it places pressure on the remaining hospitals in the market. Therefore, this stimulates local healthcare providers and leads to a more efficient delivery of services. Based on the evidence from the urban hospital closures in the U.S., Lindrooth *et al.* (2003) found that the closure of a less efficient hospital leads to lower costs for their competitors due to an increased number of inpatient as well as emergency admissions at neighbouring hospitals. However, overall market efficiency gains does not necessarily reflect on a better quality of care for some groups of patients when the immediate access to care is crucial (Avdic, 2016; Blondel *et al.*, 2011; Buchmueller *et al.*, 2006; Engjom *et al.*, 2014; Grzybowski *et al.*, 2011; Ravelli *et al.*, 2011).

The closure of a hospital is often followed by critical feedback from local residents who are frightened to lose their access to healthcare. To attract attention, local and regional press often emphasise closures and do not discuss other perspectives, which magnifies the dissatisfaction of the local population. As a consequence, distance to the healthcare provider deceptively plays a bigger role than hospital quality and creates very strong public concern just before a hospital closure (Frankfurter Allgemeine, 2013; Thüringer Allgemeine, 2014; WAZ, 2011; WDR, 2015; Westdeutsche Zeitung, 2013). Staff members of the closing hospital often join the local criticism as the closure becomes not only a loss in healthcare access but a loss in the job market as well. However, since information about the potential closure is released early, staff members take action to search for other job opportunities, which will accelerate the closure process. The owner of the hospital has the strong incentive to initiate the market exit as maintaining an inefficient hospital leads to monetary losses. The exit strategy process usually starts with the internal restructuring, the reduction of employees and, in some cases, transforms into a merger or acquisition (Preusker *et al.*, 2014).

3 Data and sampling

Data

The empirical analysis employs a nationally representative sample of patient-level data provided by a large German health insurance company. Data is collected from hospital discharge records based on diagnosis-related group (DRG) reimbursement claims and provides detailed information about patients hospitalised between 2006 and 2012. It includes a wide range of patient characteristics such as age, gender, dates of admission and discharge, place of residence and also includes comprehensive clinical information that was administrated during the hospital spell.

The sample of interest include all patients diagnosed with either an Acute

Myocardial Infarction (*AMI*) or ischaemic or haemorrhagic stroke (*Stroke*). To identify these patients I exploit the information about patient's medical diagnosis classified according to the *World Health Organization's International Statistical Classification of Diseases and Related Health Problems (ICD-10)*.[14] Due to the life-threatening nature of both *AMI* and *Stroke*, patients require immediate medical attention, preferably in units, that have access to a specialised equipment needed for diagnosis and treatment. Thus, travel time to a hospital is particularly important for patients with these conditions. When the hospital market experiences any structural changes such as a reduction in capacity or a hospital closure, emergency patients such as *AMI* or *Stroke*, are among the most sensitive and impaired access to healthcare is likely to reduce their chances for survival and successful recovery (American Heart Association, 2003). Due to these reasons, this group of patients provides close to an ideal base for the empirical setting to study the effects of hospital closures.

I identify and extract the sample of interest that includes all *AMI* and *Stroke* patients admitted to hospitals providing emergency services over the years $2006-2012$. To distinctly describe patients' medical condition prior to the medical emergency, a set of secondary diagnoses[15] was coded for each patient. This allows the analysis to account for patient case-mix by computing the Elixhauser index (Elixhauser *et al.*, 1998).[16] I further restrict the sample to patients' aged > 18 to exclude all younger individuals, particularly newborns, that might have had congenital heart conditions. As one of the objectives of this study is to evaluate access to healthcare I am interested in the first contact the patient receives after the medical emergency only. Therefore I exclude subsequent medical information about transfers to other hospitals.

I complement the sample with several auxiliary datasets. First, to measure the geographical distance to a hospital I use a 5-digit postal code of patient's registered residence and the postal address of each hospital, both

[14] Specifically, to identify patients diagnosed with *AMI* I extract ICD-10 codes: O21 (Acute myocardial infarction), I22 (Subsequent myocardial infarction) and to draw out *Stroke* patients I used codes: I61 (Intracerebral haemorrhage); I63 (Cerebral infarctions); I64 (Stroke, not specified as haemorrhage or infarction); a number of G45 (Transient cerebral ischaemic attacks and related syndromes) group codes: G45.0-; G45.1-; G45.2-; G45.3-; G45.8-; G45.9-. (Note, that for latter codes the international ICD coding has a slight difference from the German specification). In addition, I rely on a following set of secondary diagnosis codes to revise *Stroke* cases. I exclude patients that have one of those secondary diagnoses: C70.0; C70.9; C72.8; C72.9; C79.3; C71.-; S06.-; S07.-; S08.-; S09.-

[15] According to the German Medical Coding guidelines ("Deutsche Kodierrichlinien") the main diagnosis is made after the evaluation of patient's condition mainly responsible for the inpatient or outpatient care, while the secondary diagnosis refers to diseases and/or complaints that either already existed before the evaluation, i.e. diabetes, or was developed after.

[16] The Elixhauser Comorbidity Index (ECI) distinguishes 31 different comorbidities and is often used as a risk-adjustment tool to predict hospital resource use and in-hospital mortality. For a list of comorbidities I include in the analysis, see Table A.1 in Appendix A.

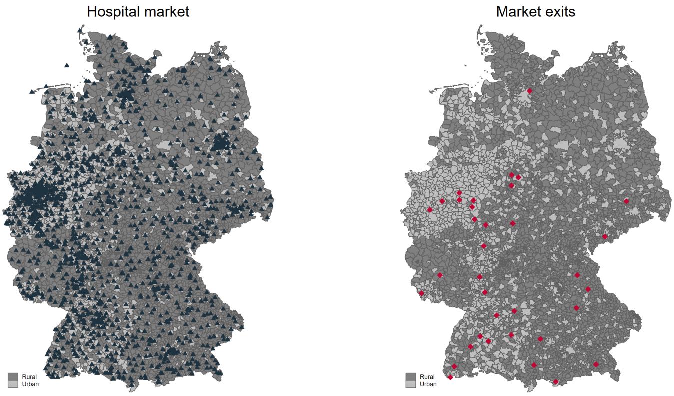
of which are available in the hospital discharge data.[17] Using these I estimate both the travel distance and the travel time for each patient-hospital combination using geocoding API software from Google® and Open Source Routing Machine (OSRM).[18] Some patients in this sample travelled unreasonably far from their residence to receive emergency care, that may not have been a regular event. For instance, a medical emergency might have occurred when an individual was travelling for business or holidays. Thus, I rely on the distribution of the distance travelled and exclude patients who are above the 95^{th} percentile (in this case travelled more than 47.7km), that is approximately 5 % of the sample. Secondly, I augment the study dataset with information from standardised public hospital quality report cards that all hospitals are required by law to publish. The quality report cards include detailed information on numbers of cases and procedures performed for each hospital department, which are particularly important when assessing differences between hospitals. I exploit several quality indicators provided in these quality report cards in the supplementary analysis to provide further evidence for the exclusion restriction. Thirdly, I systematise the reported information about the German hospital market exits in Preusker *et al.* (2014) to identify hospitals that closed over the period of the study. In the robustness analysis I use the categorised reasons for closure to construct an alternative instrument for closure. Finally, I collect publicly available information on the political party composition of local councils in German municipalities published by the German Federal Statistical Offices.[19] Using this information I estimate political party voting shares to construct the instrument for the instrumental variable design explained in Section 4.

[17]This approach follows, e.g., Hentschker and Mennicken (2014, 2018); Mennicken *et al.* (2014) and implicitly assumes that patients travel from the geographic centroid of each 5-digit postal code area corresponding to its geographic center. There are about 8,200 5-digit postal code in Germany with a median size of 27 km^2 and the vast majority below 100 km^2. When interpreting the results from estimation, it is worth noting that there are no obvious reasons why any measurement errors arising from this simplification would be systematically related to the chosen instrument in the empirical strategy.

[18]For a documentation of the latter resource, see http://project-osrm.org/ and Huber and Rust (2016). We exclude a few cases where measuring the distance to a hospital was not possible, such as patients living on an island without a road connection to a hospital or the provided residential postal coded was inaccurate. In total it account for approx. 1% of the sample.

[19]A detailed overview of the municipal politics in Germany is provided in Appendix B.

FIGURE 3.1: Overview of hospital market in Germany



Note.— The figures present an overview of hospital market and market exits in German municipalities. The right panel presents the spatial distribution of all hospitals in the sample that cater emergency care services to patients with *AMI* or *Stroke* conditions; whereas the right panel indicates hospitals that closed during the study period. Each municipality is categorised as urban or rural.

Hospital markets

Figure 3.1 presents the overview of the German healthcare market and market exits by the municipality. The left panel shows the spatial distribution of hospitals offering emergency services to patients with *AMI* or *Stroke* in 2006. Market density is high with about 1,500 hospitals, of which around 67% are located in urban areas. The densest areas are in the state of North Rhine-Westphalia, the most populous state in Germany located in the West, as well as around cities such as Berlin, Stuttgart, Münich and Hamburg. The right panel presents diamond shaped indicators showing the locations of all hospital closures during the study period. A large share of hospitals that closed are in rural areas; however, the majority of closures (57 %) appeared in urban areas.

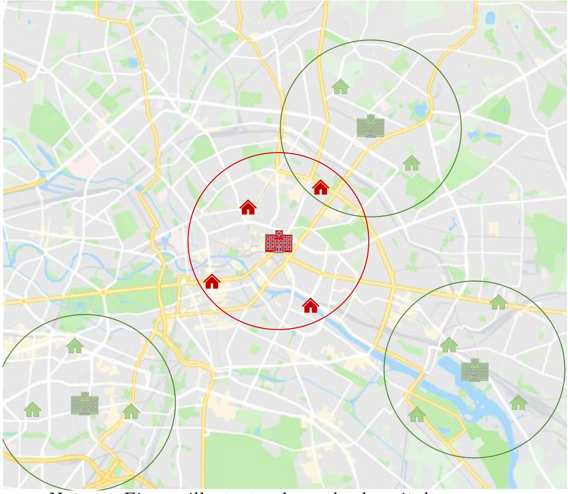
As presented in Figure 3.1, the German hospital market has a very high market density. To identify the effects of hospital closures rather than any other changes in the market structure, the definition of treatment is crucial for the unbiased estimation. Using the information from the calculated patient-hospital distance matrix, I allocate patients to their potential hospital emergency market which is defined by a 15 km radius around the hospital.[20] Figure 3.2 illustrates the market definition. Here a circle represents a hospital market that either closed (highlighted in red) or had no structural changes during the study period (highlighted in green). Patients living in the market with a hospital closure are considered to be treated in this empirical setting.[21] Using a pool of remaining non-treated markets in

[20]Robustness analysis for different specifications of hospital market in terms of the radius was performed and presented in Section 7.

[21]If hospital spatial markets overlap and patient's residence falls into two or more

the sample I perform a Propensity Score Matching technique to achieve a balanced distribution of confounders across treatment groups.[22] Restricting the sample to the matched sample using the Propensity Score Matching technique significantly reduces the size of sample. However, in a dense healthcare market with hospitals that are different in their observed characteristics such as size and ownership type, it is important to eliminate the systematic differences between the treated and non-treated markets. Based on the previous findings, the Propensity score Matching is a more robust approach than a covariate adjustment in the model and reduce potential bias of estimated treatment effects (Austin, 2011; Austin and Mamdani, 2006; Crown, 2014).

FIGURE 3.2: Definition of the hospital emergency market



Note.— Figure illustrates how the hospital emergency market is defined in this study. A red color indicates the closure-affected area, whereas a green color - the non-affected area. The map riffle is irrelevant and chosen only for illustrative purposes.

Table 3.1 presents the descriptive statistics of the matched sample. The final sample comprises of 11,492 patients of which 22 % are defined as treated individuals. On average patients travelled around 5 km to their primary hospital and their travel time was approximately 6 minutes. The risk-adjusted and rescaled in-hospital as well as 30 day mortality was similar.

markets, the closest hospital was chosen as the primary one. For descriptive purposes Figure 3.2 illustrates very densely located hospitals, which is a rare case in this set up. Therefore, a very small amount of patients fell into several markets.

[22] I employ the Nearest Neighbour technique to match for a given treated market with an untreated market that is closest in its propensity score. The matching covariates include a number of hospital market-related characteristics (rural, if hospital in the market has a cardiology or angiology department, if teaching, # of hospital beds, # of doctors, # of nurses) and a number of (averaged) patient-related characteristics to control for patient case-mix in the market (share of male patients, patient's age, shares of each Elixhauser comorbidity). The optimal calliper width is 0.1.

Patients stayed about 10 days in the hospital and only 10% were readmitted due to similar health conditions. The average patient in the sample is a 72 year old male and, based on the Elixhauser Comorbidity index, had about three medical conditions prior to the medical emergency considered in this study. Several indicators related to hospital capacities and specific characteristics provided by the quality report cards are included in the set of controls to account for potential heterogeneity between hospitals. In addition, I consider a set of indicators specific to hospital emergency market. More detailed statistics on the matched and treated hospital markets are provided in Table A.2 in Appendix A.

TABLE 3.1: Descriptive Statistics of matched sample

	Mean	SD	Min	Max
— *Outcomes* —				
Distance to the nearest, km	4.64	3.30	0.03	14.95
Time to the nearest, min	5.84	3.42	0.07	19.82
Death (RA, rescaled)	0.47	0.14	0.02	1.00
Death 30 days (RA, rescaled)	0.48	0.16	0.02	1.00
Length of stay	9.17	8.04	1.00	205.00
Readmission	0.11	0.32	0.00	1.00
— *Treatment & Instrument* —				
Treated	0.22	0.41	0.00	1.00
CDU winning margin	-0.08	0.16	-0.30	0.67
Reason: Economic	0.82	0.39	0.00	1.00
Reason: Hospital plans	0.10	0.30	0.00	1.00
Reason: DRG implementation	0.04	0.19	0.00	1.00
— *Covariates* —				
Age	71.53	13.52	19.00	103.00
If male	0.57	0.49	0.00	1.00
If rural	0.11	0.31	0.00	1.00
EL score	2.70	1.82	0.00	13.00
Beds	655.68	516.57	20.00	2910.00
If university	0.13	0.34	0.00	1.00
If teaching	0.45	0.50	0.00	1.00
If public	0.56	0.50	0.00	1.00
If non-profit	0.37	0.48	0.00	1.00
Cases/doctor	114.20	147.42	0.00	1525.74
Cases/nurse	47.28	32.28	0.00	186.17
Market: Small hospital size	0.35	0.48	0.00	1.00
Market: Middle hospital size	0.33	0.47	0.00	1.00
Market: Large hospital size	0.33	0.47	0.00	1.00
Market: if teaching	0.18	0.39	0.00	1.00
Market: if rural	0.19	0.39	0.00	1.00
Market: average age	71.81	2.25	48.00	84.13
Market: average EL score	2.74	0.34	0.67	6.00
Market: average gender ratio	0.57	0.07	0.00	1.00
Observations	11492			

NOTE.— Table presents descriptive statistics of the matched sample. Here RA abbreviates risk-adjusted, CDU - Christian Democratic Union, DRG - diagnosis related groups, and EL - Elixhauser commorbidity.

4 Econometric framework

Several issues arise when estimating the effect of hospital closures on patient outcomes. First, estimates might suffer from the estimation bias due to patients' sorting into different residential areas. It is likely that patients with worse health conditions might deliberately choose to live closer to a hospital[23] and failure to control for this spatial sorting might result in a downward bias in the estimate of hospital closure. To minimize this bias, I employ a large set of detailed patient- and location-specific characteristics and additionally control for patient case-mix using a set of averaged patient characteristics in the market. A particular advantage of this comprehensive data is a possibility to follow patients over time, that allows to identify hospital transfers and any readmissions after the initial hospital presentation. I explore this feature of the data and estimate the effect of hospital closure on readmission rates. In addition, to account for potential changes and variation over time and I include time fixed effects in the estimation.

In spite of being able to control for a detailed and rich set of observable influences on patient outcomes that might be correlated with a hospital closure, there remains a possibility of unobserved common influences on patient outcomes and the likelihood of hospital closure. For instance, one of the most common reasons for closure is financial instability as a consequence of low patient volume due to poor quality of care, such as, for example, high mortality rates. Estimating the effect of hospital closure on patient outcomes will result in the correlation between the variable of interest and the error term, and a potentially bias and inconsistent estimator. To address this endogeneity problem caused by reverse causality, I adopt two empirical strategies. First, I use the Propensity Score Matching technique explained in the previous section to find a suitable control market that stands as a benchmark to measure differences between treated and non-treated individuals. Second, I employ the Instrumental Variable (IV) empirical strategy to estimate the effect of hospital closure. With a novel instrument in hand, that well predicts the endogenous treatment variable, this approach grants precise and unbiased estimators. The construction of the instrument as well as the validation for the underlying assumptions is explained in detail in Section 5.

Let i be a patient admitted to hospital h at time t. The patient is treated ($\mathbb{1}_{i\in M_j^1}$) if s/he resides in the hospital market M_j that experienced a closure (indicated by M_j^1). Then the effect of hospital closure on outcome Y can be estimated with the following model:

$$Y_{iht} = \rho_0 + \rho_1 \widehat{\mathbb{1}}_{i\in M_j^1} + X_i'\gamma + H_{ht}'\pi + \lambda_t + \mu_d + \epsilon_{iht} \tag{4.1}$$

with the corresponding first stage:

$$\mathbb{1}_{i\in M_j^1} = \phi_0 + \phi_1 Z_{ht}^{CDU} + X_i'\gamma_0 + H_{ht}'\pi_0 + \lambda_t + \mu_d + \omega_{iht} \tag{4.2}$$

[23]See e.g., Chou *et al.* (2014); Currie and Reagan (2003); Grzybowski *et al.* (2011).

where $\widehat{\mathbb{1}}_{i\in M_j^1}$ in (4.1) is the predicted treatment from the first stage estimation and ρ_1 is the estimate of the interest. Herein, I focus on several outcomes Y. First, I consider the effect on geographical healthcare access measured by the shortest distance and travel time to a hospital. Second, to describe the quality aspects, I look at various patient outcomes such as death in-hospital and within 30 days of discharge.[24] Finally, I investigate the efficiency of hospitals by looking at length of stay in days and the readmission rate with a condition that the patient was readmitted within 30 days of discharge and had a similar diagnosis.

The model specification further controls for a vector of patient-specific characteristics X_i such as age, gender, rural status, and a number of Elixhauser comorbidities as well as a vector of hospital-specific characteristics H_{ht} including number of beds; university and teaching status, ownership type and number of cases per doctor and per nurse. I also control for year fixed effects λ_t and admission-day-of-the-week[25] fixed effects μ_d. Let ϵ be uncorrelated random error term. Due to the sampling design that is based on hospital markets, the standard errors will be clustered at the hospital level.

5 Construction of the instrument

The empirical model employs instrumental variable to take into account concerns related to potential endogeneity of hospital closure. Recall that the closure might have influence on the quality of care in the area, but the quality itself may also be a reason for closure. To overcome this limitation, I follow a similar approach as outlined in Bloom *et al.* (2015) and construct an instrument based on the degree of political pressure. The authors argue that politicians loath to deliver policies not popular with the voters (such as the hospital closure), especially in the areas where the political vote winning margin is small. On the other hand, in the areas where one party has a noticeable political advantage against the remaining parties these policies are more likely to be enacted. Using this particular phenomenon, authors adopt the constituency election winning margin as an instrument to instrument the level of hospital competition in the area and evaluate it's effect on management quality.

[24]The outcomes on mortality rates were first risk-adjusted using a logistic regression. The risk adjusters considered in the regression include several patient characteristics such as patient's age, Elixhauser comorbidities, gender, state and urbanity of patient's residence; various hospital-related characteristics such as ownership type, if teaching, if university and several capacity-related measures (number of beds, number of cases per doctor, doctor's specialization level, a number of cases per nurse); and several treatment related characteristics particularly important in the case of a stroke.

[25]To control for average differences in days of the week, I include the admission-day-of-the-week fixed effect into the model. It determines a day of the week, identified as d, i.e. Monday, Tuesday etc., when a patient was admitted to the hospital.

Following this technique, I construct an instrument in a similar way to account for potential endogeneity between hospital quality and market structure. Using the German local municipal government elections results, I calculate shares of votes for major political parties. Policies such as shutdowns of the institutions (that might as well be a major employer in the area) are more likely to happen when the governing party follows market-oriented policy perspective rather than more socialist political ideology. For this reason, I chose to reference the calculation of the winning margin on the condition that the largest centre-right political party in Germany - Christian Democratic Union (CDU) - has a political advantage in the municipality.[26] The winning margin is then constructed as a difference between voting shares of the CDU and the opposition parties, expressing interest in more left orientated political views such as Social Democratic Party (SPD), The Left Party (LINKE) and The Greens (GRUENE) as the following:

$$Z_{ht}^{CDU} = s_{ht}^{CDU} - s_{ht}^{Opposition} \tag{5.3}$$

When the winning margin is positive, the governing party is the CDU and the higher the margin is, more political power the party enjoys in the municipality. To support the relevance assumption for this instrument, I first explore the associations between the defined treatment and the instrument. Similarly as in the UK setting described in Bloom *et al.* (2015) I observe „political punishment" patterns. Table 5.2 shows that treatment is significantly associated with the share of votes both for the CDU (Column (1)) as well as the opposition parties (Column (2)). If an individual resides in the treated market, the share of the CDU votes are significantly and approximately 6 percentage points smaller than among those, who reside in a non-treated market. The difference appears to be even larger for the opposition parties. This provides evidence that a substantive policy such as hospital closure raises public awareness especially related to the political decisions in the municipality.

[26] The reference chosen to calculate the winning margin does not alter the main results and only affects the interpretation of the first stage coefficients.

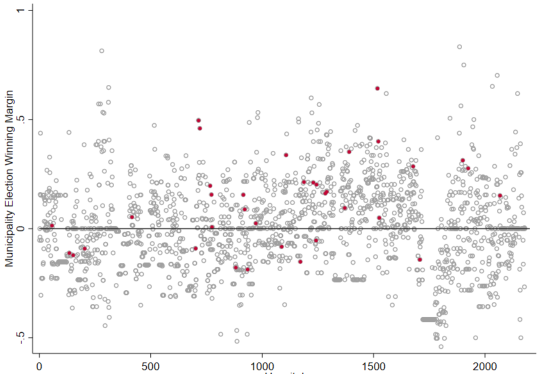
TABLE 5.2: The relevance of the instrument I: Political punishment

	(1) Share CDU	(2) Share Opposition
$\mathbb{1}_{i \in M}$[1]	-0.0581**	-0.296***
	(-2.02)	(-20.83)
Observations	11492	11492

NOTE.— Table presents the estimation results of a linear regression measuring the association between treatment and shares of political votes. The models also control for a set of patient characteristics - age, if male, if rural; Elixhauser commorbidities; hospital characteristics - # of beds, if university, if teaching, if public, if non-profit, # of cases per doctor, # of cases per nurse; as well as year fixed effects.

Further evidence for the relevance of the instrument is presented in Figure 5.3. Figure reports the variation of winning margins across all hospitals in the sample over the study period (numeration of hospitals is random). The solid horizontal line represents political stalemate, when the CDU and the opposition parties evenly divide the votes with no party enjoying a political margin. Yet, this is a particularly rare situation, therefore most of hospitals fluctuate above or below the solid line. All points above the line indicate hospitals in municipalities where the CDU party enjoys the political majority and the opposite below the line. The closer to the line the mark is located, the less political power the party has. Additionally, red points in the figure highlight hospitals that closed during the study period. As suspected, the majority of closures appear in the CDU winning municipalities. A larger portion of those were located in areas where the winning margin is relatively high, highlighting weaker political competition when these substantive policies were implemented.

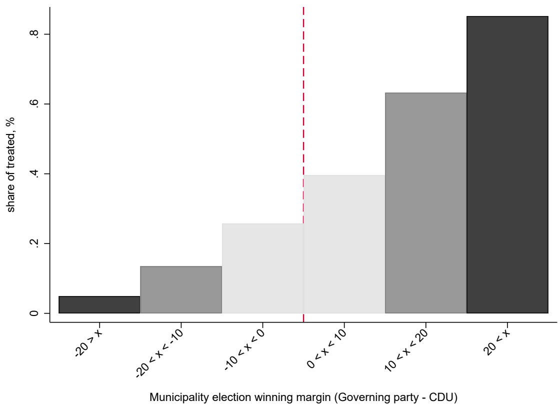
FIGURE 5.3: The relevance of the instrument II: Political winning margin in the location of a hospital



NOTE.— Figure presents the variation in municipality election winning margin for all hospitals that operated in years 2006 – 2012. The numeration of hospital is random. Red coloured circles highlight hospitals that closed during this period.

To further explore the significance of political competition, Figure 5.4 presents the relationship between the winning margin and the treatment. Each column presents a share of individuals by the winning margin divided into intensity of political pressure intervals. The brighter the column, the higher political pressure is observed in the municipality. Columns located to the right of the dashed red line indicate occasions when the CDU had a political advantage against the opposition, whereas columns to the left show the contrary occasions when the CDU had a political disadvantage. It is apparent from this figure that the share of treated individuals is higher when the CDU has a majority of votes and, as a result, the histogram is skewed right. Besides, the scarcity of individuals on the left side of the dashed line shows that the share of treated is smaller when the left-leaning party is in lead.

FIGURE 5.4: Governing party's (CDU) winning margin and the share of treated individuals



NOTE.— Figure presents the share of treated individuals by different winning margin intervals.

Although previously presented evidence supports the relevance of the instrument, it is important to test the exclusion restriction. Recall that in Germany *State Hospital Plans* highly regulate healthcare resources as well as the quality of care. Thus, political parties in the municipality do not have any powers to expand the number of healthcare providers in the area that could affect the geographical healthcare access. Similarly, they do not have any influence neither on hospital capacity nor on the variety of services a hospital provides, both of which could potentially improve the quality of care in the area. As political bodies often sit on the supervisory board of the hospital, the only channel through which politicians could potentially affect the quality of care is the management of the hospital. Table 5.3 provides supportive evidence that this channel is not significant and that the violation of the exclusion restriction is unlikely. Table presents regression estimation results on the multiple quality indicators reported in the hospital quality report cards that are relevant to the patient group of interest. Columns (1) - (3) include quality-related outcomes such as specialist doctors, the proficiency level in surgeries and the proficiency in the diagnostic-related treatments, respectively. The former outcome denotes a share of specialist doctors operating in the hospital, while the proficiency outcomes indicate the number of different services a hospital provides and act as a score from 0, being the lowest, and 7, being the highest possible proficiency.[27] Results show that the instrument does not have any significant effect on any of the quality indicators related to the management of the hospital and provide further support to the assumption that political power does not alter the quality of the services provided in the area.

Table 5.3: Exclusion restriction: Politicians' influence on the quality

	(1) Specialist Doctors	(2) Proficiency in Surgeries	(3) Proficiency in Diagnostics
Winning Margin (CDU)	0.0406	-2.827	0.0406
	(0.03)	(-0.26)	(0.03)
Mean	0.53	0.63	0.92
SD	0.18	1.30	1.57
Observations	2014	2014	2014

Note.— Table presents the estimation results of a linear regression measuring the relationship between the instrument and quality indicators of the hospital such as share of specialist doctors operating in the hospital, number of surgeries and invasive procedures available at the hospital and number of diagnostic procedure performed at the hospital. Each proficiency indicator is a score from 0, being the lowest, and 7, being the highest possible proficiency. Each model additionally accounts for hospital characteristics such as if public, if university, if teaching and hospital fixed effects.

The distribution of the constructed instrumental variable is illustrated in Figure A.1 in Appendix.

6 Results

I first demonstrate that geographical distance to hospital is in fact important for patients with *AMI* or *Stroke* and that worse healthcare access due to a hospital closure could lead to adverse clinical outcomes. Table 6.4 reports estimation results based on the linear regression model that evaluates the associations between the distance patient travelled to the hospital and several patient outcomes. In addition, I allow for potentially non-linear relationships in this setting and express the distance as a second order polynomial

[27]The score of hospital's proficiency in surgery includes a possibility to perform a major coronary surgery including a surgery following any complications of the coronary heart disease, a heart valve surgery, both pacemaker and defibrillator interventions and, lastly, a heart transplantation. The score of hospital's proficiency in diagnostics include a possibility to perform angiography, pulmonary embolectomy, an intervention on the pericard, treatment of health injuries, and other diagnostic and therapeutic treatments for ischemic, pulmonary and other heart diseases. These quality measures indicate that the hospital has the capacity and the capability including specialised angiographers/cardiologists/surgeons and equipment to perform any of diagnostic procedures or surgeries listed.

by including the quadratic curve. Columns (1) - (4) present the estimation results on in-hospital death, death within 30-days of discharge, length of stay and readmission within 30 days of the initial discharge. Reported parameter estimates are then interpreted as the average percentage change in the probability of death if distance travelled increases by one kilometre. Thus, if the patient travels one additional kilometre, the probability of in-hospital death is approximately 0.0014 percentage points higher. Although this result looks small at first glance, recall that a patient on average travels about 5 km to the nearest hospital, if this hospital closed and the distance increased by 5 to 10 kilometre, the patient would face an increased probability of in-hospital death by 0.5 to 1 percentage points. The estimate is even higher in the case of death within 30-days of discharge, signalling that additional complications might arise due to delayed treatment. Distance to the hospital does not seem to play an important role on other patient outcomes such as length of stay and readmission.

TABLE 6.4: Distance effect on patient outcomes

	(1) *Death*	(2) *Death (30-days)*	(3) *Length of stay*	(4) *Readmission*
Distance, km	0.00138**	0.00174**	0.00204	-0.00144
	(2.62)	(2.94)	(0.06)	(-0.67)
(Distance, km)^2	-0.00003*	-0.00003*	0.00043	0.00007
	(-2.10)	(-2.11)	(0.35)	(1.17)
N	11336	11336	11492	11492

NOTE.— Table presents the linear regression model estimating the relationship between distance travelled and patient outcomes. All models also control for a set of patient characteristics: age, if male, if rural, Elixhauser commorbidities; hospital-related characteristics: # of beds, if university, if teaching, if public, # of cases per doctor, # of cases per nurse; as well as year and weekday fixed effects. Outcomes (1) and (2) are risk-adjusted measures and due to additional estimations made beforehand analysis sample is slightly smaller, however this should not affect the main findings. Standard errors are clustered at the hospital level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Following this evidence, I estimate the effects of hospital closures on the geographical healthcare access and patient outcomes. Considering the setting of this study, that focuses on the closure of the nearest medical facility, I anticipate that patient's proximity to the nearest medical care will in fact increase after hospital closure. However, it is essential to examine the extent of this increase and its effects on patient outcomes. Panel B of Table 6.5 reports results from the estimation of the IV model as defined in (4.1) and (4.2) and, for comparative reasons, panel A presents the estimation results of the second stage using the Least Squares. The estimate of the first stage regression reported in Column (1) suggests a significant positive relationship between the selected instrument and the treatment. This result supports the previous discussion in Section 5. The point estimate can be interpreted as one percentage point change in the CDU winning margin, or simply the CDU political power, and is interpreted as 1.3 percentage point change in the likelihood of residing in a closure-affected area. Hence, if the

CDU governing party gains more political power against the opposition, it is more likely that the party will adopt a substantive policy such as closing a hospital in the municipality. The first stage coefficient is highly significant with an *F-statistic* value of around 47, providing further evidence that the instrument is a strong predictor of treatment and supporting the validity of the second stage estimations.[28]

TABLE 6.5: IV estimation. Effect on healthcare access and patient outcomes

		Health care access		*Health outcomes*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	I stage	*Distance*	*Travel time*	*Death*	*Death (30-days)*	*Length of stay*	*Readmission*
			A. OLS Estimation				
$\mathbb{1}_{i \in M^1}$		1.547^{***}	1.451^{***}	-0.006	-0.001	-0.695^{*}	0.027^{*}
		(4.15)	(4.21)	(-1.01)	(-0.21)	(-1.80)	(1.91)
			B. IV Estimation				
Z^{CDU}	1.345^{***}						
$\widehat{\mathbb{1}}_{i \in M^1}$		3.704^{***}	2.603^{***}	-0.008	-0.004	-2.310^{***}	-0.021
		(5.22)	(4.32)	(-0.98)	(-0.39)	(-3.18)	(-1.37)
Observations	11492	11492	11492	11336	11336	11492	11492
F	47.30						

NOTE.— Table presents the estimated effect on healthcare access and patient outcomes using Least Squares (panel A) and IV model (panel B). Here the instrument is the CDU winning margin in the municipality elections against the opposition. All models also control for a set of patient characteristics: age, if male, if rural, Elixhauser commorbidities; hospital-related characteristics: # of beds, if university, if teaching, if public, # of cases per doctor, # of cases per nurse; as well as year and weekday fixed effects. Outcomes (4) and (5) are risk-adjusted measures and due to additional estimations made beforehand analysis sample is slightly smaller, however this should not affect the main findings. Standard errors are clustered at the hospital level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Columns (2) to (3) and (4) to (7) show the results of the second stage estimation on healthcare access and patient outcomes, respectively. As anticipated, patients, living in the area where the hospital closed, face worse healthcare access in terms of the distance and travel time to the hospital. For both outcomes considered the coefficient of $\widehat{\mathbb{1}}_{i \in M^1}$ is highly significant at a significance level of 0.1%. This estimate suggests that residents of closure-affected area travel on average nearly 4 kilometres further (or 3 minutes longer) to the nearest hospital offering emergency care. The estimate of Least Squares is also highly significant, albeit much smaller in magnitude. Hence, not accounting for the potentially endogenous market structure would give induced a downward bias and, consequently, the true effect of hospital closure would have been underestimated. Although the results on the healthcare access confirm that the hospital closure reduces patients' chances of receiving prompt medical care in case of emergency, it may not necessarily result in either worse survival or other health outcomes following the medical event. The estimates reported in Column (4) and (5) indicate that, even though an increase in travel distance is associated with higher mortality rates as shown in Table 6.4, an increase in travel distance due to closure is in fact not critical for these emergency cases. The coefficients signal somewhat lower odds of dying both in-hospital as well as after discharge; however, the estimated effects are not statistically significant. In-

[28] I rely on the evidence by Staiger and Stock (1997), stating that for a strong instrument inference, the *F-statistic* greater than 10 is required.

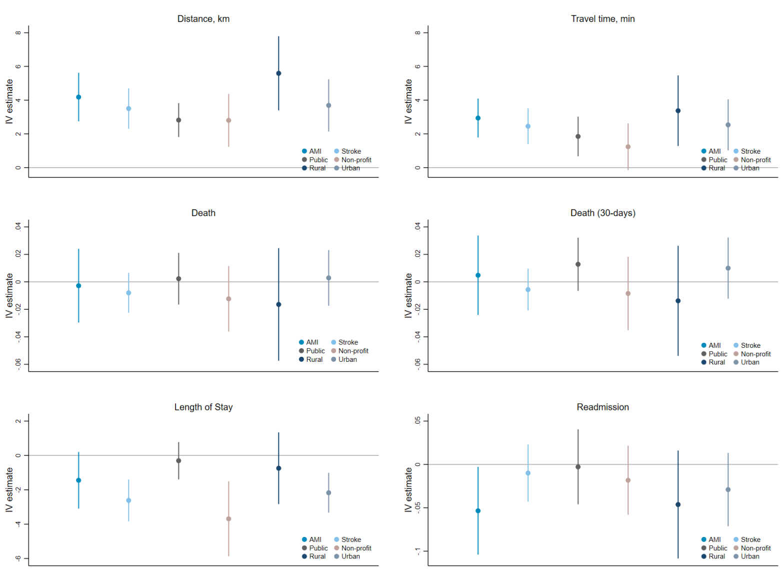
terestingly, as reported in Column (6), patients residing in closure-affected areas has on average shorter length of stay. The estimate suggests that treated individual's hospital stay is more than 2 days shorter than their untreated counterparts. One possible reason for this finding is that, when a hospital closes, the number of patients at the neighbouring hospital increases and, as a result, stimulates more efficient delivery of services in the remaining market. This finding is supported by previous literature showing that the economic pressure arising from competitor closing down leads to gains in efficiency for the remaining healthcare market (Lindrooth *et al.*, 2003). However, these gains in efficiency do not appear to result in any medical complications that would require a readmission, supporting the finding that they do not come at the expense of the quality of care (Column (7)).

7 Heterogeneity and robustness analysis

Finally, I report estimation results from a set of extensions to the main analysis to gauge the robustness of my findings and further assess whether the effects are heterogeneous across specific subgroups of the study sample. I first study heterogeneity with respect to patient's medical condition, the type of hospital and residence location. The findings are illustrated in Figure 7.5, of which each panel shows the estimation results on a set of outcomes considered in the main analysis. Each dot in the figure refers to an estimated parameter ρ_1, that is the point estimate of the second stage of the IV model outlined in (4.1). The solid horizontal line in each panel stands as a reference line indicating the occasion when the effect is zero and insignificant. The analysis sample is split into two sub-samples by each heterogeneous group and indicated by different colors: first, by the medical condition, *AMI* or *Stroke*; second, by admitted hospital ownership status, public or non-profit; and, lastly, by the type of residential location, urban or rural. The estimated effect on healthcare access appears to be insensitive to patient's medical condition: both *AMI* and *Stroke* patients experience longer travel to the nearest hospital by about 4 km (3 min) with slightly larger effect noted for *AMI* patients. However, this significant change in travel time does not result in higher odds of dying for either of these conditions. On the other hand, efficiency gains with respect to the length of stay appear to be mainly driven by the treatment of *Stroke* patients, suggesting that patients with this medical condition could be and are treated quicker in neighbouring hospitals when pressure on the capacity rises. Interestingly, *Stroke* patients are no more likely to be readmitted, confirming that shorter length of stay does not result in any subsequent complications. However, a quality improvement is noted for patients with *AMI*, who are significantly less likely to be readmitted when they receive treatment at the neighbouring rather than at the nearest hospital.

With respect to the ownership type of the hospital, I note no significant differences in healthcare access, odds of dying and readmission results. However, the effect on length of stay appears to be driven by non-profit hospitals, that treat patients residing in closure-affected areas on average 4 days quicker than patients residing in unaffected areas. Non-profit hospitals often have smaller capacities with respect to the number of beds and staff as well as treat a smaller share of patients in the market. Thus, closing a neighbouring hospital seems to place a higher pressure on these hospitals that respond by providing medical services more efficiently and, relying on the results on other outcomes, effectively. With regards to the type of residential location, the effect on healthcare access is as expected larger for patients living in rural areas; however, not surprisingly, the efficiency gains are driven only by urban areas, where larger hospital complexes are often located.

FIGURE 7.5: Heterogeneity analysis by different subgroups



NOTE.— Figure presents the IV estimation results on a set of sub-samples. Each dot denotes a point estimate (and its 95% confidence interval) of the second stage of the IV model. The instrument is the CDU winning margin in the municipality elections against the opposition. All models also control for a set of patient characteristics: age, if male, if rural, Elixhauser commorbidities; hospital-related characteristics: # of beds, if university, if teaching, if public, # of cases per doctor, # of cases per nurse; as well as year and weekday fixed effects. Standard errors are clustered at the hospital level.

Next, to provide additional evidence for the reliability of the instrument I perform several robustness analyses. First, one possible issue with the construction of the instrument could arise from the fact that the instrument is constructed at the market level, whereas the analysis of treatment effects is at the patient level. To provide the support that this does not cause any problems, I estimate the same IV model defined in (4.1) and (4.2) aggregated at the market level. Herein, each outcome denotes the average outcome in the hospital market and models also control for patient case-mix in the area determined by averaged patient characteristics. Results are presented in Table 7.6. Despite the aggregated sample, results on the first stage reported in Column (1) again support the relevance of the instrument with an *F-statistic* value of 56. The estimated coefficient on other patient outcomes are similar in terms of both statistical significance and effect size. In contrast to the results from the main analysis, the effect size is smaller and insignificant for length of stay. This is not unexpected and is likely a result of removing some of the variation in the outcomes, and should not be interpreted as contradicting the main finding.

TABLE 7.6: Robustness analysis I. Aggregated analysis

		Health care access		*Health outcomes*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	I stage	*Distance*	*Travel time*	*Death*	*Death (30-days)*	*Length of stay*	*Readmission*
Z^{CDU}	1.015***						
	(7.52)						
$\widehat{\mathbb{1}_{i\in M^1}}$		3.988***	2.158***	-0.000	0.015	0.548	-0.002
		(4.90)	(2.89)	(-0.01)	(0.95)	(0.62)	(-0.06)
Observations	179	179	179	171	171	179	179
F	56.62						

NOTE.— Table presents IV estimation results on the aggregated to hospital market level sample. The instrument is the CDU winning margin in the municipality elections against the opposition. All models also control for a set of averaged patient characteristics at the market level: age, if male, if rural, dummies for Elixhauser commorbidities; and hospital characteristics - # of beds, if university, if teaching, if public, # of cases per doctor, # of cases per nurse; as well as year and weekday fixed effects. Outcomes (4) and (5) are risk-adjusted measures and due to additional estimations made beforehand analysis sample is slightly smaller, however this should not affect the findings. Standard errors are robust. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Second, another potential problem with the constructed instrument might be related to the main rationale of the significance of political pressure on hospital closure decisions. It is likely that, if a hospital experiences financial difficulties, politicians might be under pressure to act to reduce current and any future monetary losses. I employ the information provided in the report by Preusker *et al.* (2014) about the reason for hospital closure and augment the instrument used in the main analysis by interacting with a dummy indicator variable for whether the main reason for closure was financial. This allows me to identify those areas where the political pressure in the market is only driven by economic incentives. The instrument is then defined as the following

$$\ddot{Z}_{ht}^{CDU} = Z_{ht}^{CDU} \times \mathbb{1}_{economic} \tag{7.4}$$

where $\mathbb{1}_{economic}$ is a dummy variable indicating whether a hospital closed due to economic insolvency or other similar reasons. This specification replaces the instrument employed in the first stage (4.2) and the corresponding estimation results are shown in Table 7.7. Based on the first stage *F-statistic*, the alternative specification of the instrument is again highly relevant when predicting the treatment. The estimated coefficients from the second stage are in line with the main results discussed in Section 6 and are only slightly higher for healthcare access. This finding gives additional credibility to the selected instrument if one suspects that political pressure could be driven by economics only.

TABLE 7.7: Robustness analysis II. Alternative specification of the instrument

		Health care access		*Health outcomes*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	I stage	*Distance*	*Travel time*	*Death*	*Death (30-days)*	*Length of stay*	*Readmission*
$\ddot{Z}^{CDU}$	1.187***						
	(5.99)						
$\widehat{\mathbb{1}_{i\in M}1}$		4.714***	3.377***	-0.007	-0.002	-2.334***	-0.0230
		(5.68)	(4.98)	(-0.67)	(-0.15)	(-2.67)	(-1.20)
Observations	11492	11492	11492	11336	11336	11492	11492
F	35.89						

NOTE.— Table presents IV estimation results using an alternative specification of the instrument. The instrument here is the interaction between the CDU winning margin against the opposition and the dummy variable indicating if the reason for closure is related to economic insolvency. All models also control for a set of patient characteristics: age, if male, if rural, Elixhauser commorbidities; hospital-related characteristics: # of beds, if university, if teaching, if public, # of cases per doctor, # of cases per nurse; as well as year and weekday fixed effects. Outcomes (4) and (5) are risk-adjusted measures and due to additional estimations made beforehand analysis sample is slightly smaller, however this should not affect the findings. Standard errors are clustered at the hospital level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, I study whether the main estimation results are sensitive to the definition of the hospital emergency market. In the main specification, I assumed that patients residing within a 15 km radius of a hospital are referred to that hospital. While this is likely the case in more populated areas, it might not necessarily reflect reality in less populous areas where small hospitals do not have the capacity to treat patients with severe medical conditions such as *AMI* or *Stroke*. To investigate whether the definition of the hospital emergency market alters the main findings, I estimate the IV model specified in (4.1) and (4.2) using 25 km and 50 km radiuses. Table 7.6 presents estimated coefficients using both definitions in Panel A and B, respectively. Note, that changing the definition of the market enlarges the geographical area in the study, thus the number of patients considered in each model increases with increasing hospital market catchment areas. I find that the relevance of the instrument is insensitive to the definition of the market and the political pressure still plays an important role. However, the estimated coefficients, albeit significant, are slightly smaller in size which is expected result when the market expands.

TABLE 7.8: Robustness analysis III. Different definitions of a hospital market

		Health care access		*Health outcomes*			
	(1) *I stage*	(2) *Distance*	(3) *Travel time*	(4) *Death*	(5) *Death (30-days)*	(6) *Length of stay*	(7) *Readmission*
				A. 25 km radius			
Z^{CDU}	0.819***						
	(5.00)						
$\mathbb{1}_{i \in M^1}$		2.519***	2.098**	-0.006	-0.007	-1.392*	-0.046*
		(2.74)	(2.38)	(-0.81)	(-0.89)	(-1.75)	(-1.71)
Observations	23354	23354	23354	23250	23250	23354	23354
F	25.04						
				B. 50 km radius			
Z^{CDU}	0.982***						
	(10.11)						
$\mathbb{1}_{i \in M^1}$		1.357**	1.200*	-0.004	-0.002	-1.368**	-0.012
		(2.16)	(1.94)	(-0.93)	(-0.32)	(-2.26)	(-1.20)
Observations	67371	67371	67371	67141	67141	67371	67371
F	102.30						

NOTE.— Table presents IV estimation results using different definitions of hospital market. The instrument is the CDU winning margin in the municipality elections against the opposition. All models also control for a set of patient characteristics: age, if male, if rural, Elixhauser commorbidities; hospital-related characteristics: # of beds, if university, if teaching, if public, # of cases per doctor, # of cases per nurse; as well as year and weekday fixed effects. Outcomes (4) and (5) are risk-adjusted measures and due to additional estimations made beforehand analysis sample is slightly smaller, however this should not affect the findings. Standard errors are clustered at the hospital level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

8 Summary and concluding remarks

In this paper I study the effects of hospital closures on geographical healthcare access and clinical patient outcomes. I employ comprehensive administrative hospital discharge data that provide detailed information about patients, their medical condition, and treatment received at the hospital. This data is supplemented with several auxiliary datasets. First, I link the data with public hospital quality report cards and employ a large set of various hospital-related characteristics to account for potentially unobserved heterogeneous effects between different hospitals. I also collect and systematize information about all hospital market exits during the study period and identify them in the data. To construct an instrumental variable that corrects for potentially endogenous market structure when studying the effects of hospital closures on various health outcomes, I collect publicly available data on political party composition of local councils in the German municipalities. Local politicians who undertake substantive policies such as hospital closure are often „punished" by voters. I exploit this feature and estimate a measure of concentration in political power in the market. I condition on the largest centre to centre-right political party winning and estimate the voting margin to instrument for the treatment defined as individuals residing in a closure-affected areas. I exploit this comprehensive linked dataset and I apply the Instrumental Variable approach to study the effects of hospital closures on geographical healthcare access expressed in distance and travel time to the hospital and several patient clinical outcomes such as death, length of stay and readmission. I find that political power in the local area plays a substantial role in determining the future of hospitals and, although this did not have any effect on clinical quality and the variety of services provided in the area, it is a significant predictor of hospital closures. Patients living in closure-affected areas on average travel further to access care, but this does not result in reduced survival for severe acute conditions such as acute myocardial infarction or stroke. It is important to note that the effects on the mortality could been impinged by the lack of information on out-of-hospital mortality, which I leave for future research. However, the results on other clinical outcomes provide compelling evidence that longer travel times due to closure do not result in additional readmissions due to any medical complications following hospital treatment. To the contrary, closing a hospital stimulates efficiency gains as patients are treated more rapidly at neighbouring hospitals which does not come at the expense of the quality of care.

My findings contribute to the existing literature on healthcare consolidation policies. In line with previous findings I provide empirical evidence that the hospital closure has a negative effect on geographical healthcare access. However, most of previous literature relied on the strong assumption that hospitals provide universal care and concentrated on various patient groups whose choice of hospital might have relied on their personal preferences (Burkey *et al.*, 2017; Hentschker and Mennicken, 2014; Mennicken *et al.*, 2014). Thus, the findings could have underestimated the effects of hospital closure policies. To complement the existing literature, I consider the most vulnerable group of patients – those with *AMI* and *Stroke*, who due to their critical medical condition requiring emergency care, will not choose their preferred hospital. I additionally select only those hospitals that are equipped with specialised equipment for treating these emergency patients. Using very detailed information about market exits I am able to identify all hospital closures over the study period and exploit the effects on patient outcomes rather than exploiting policy-induced variation in distance due to closures that a large body of literature lies at the heart of (Avdic, 2016; Blondel *et al.*, 2011; Buchmueller *et al.*, 2006; Ravelli *et al.*, 2011). While this measure is relevant for more concentrated markets such as the U.S. or Sweden, it is less informative in a market with high hospital density. Thus, in this paper I employ an alternative empirical approach and using a measure of concentration of political power I provide compelling evidence for the importance of political decisions in hospital markets.

In conclusion, these findings reveal that, in times of great consolidation of health systems, local politics is an important channel that could mediate potentially adverse effects on social welfare. This channel offers a broad scope for communication to reduce public concerns when a hospital forfeit of its future. As my results suggest, even during one of the strongest periods of healthcare consolidation in Germany, this phenomenon did not result in any adverse clinical outcomes and policy-makers should only be considered with closing hospitals in less densely populated areas.

References

American Heart Association (2003). *Heart and Stroke Facts*. Tech. rep., American Heart Association.

Augurzky, B. and Schmitz, H. (2010). Is there a Future for Small Hospitals in Germany. *Ruhr Economic Papers*, (198), 1–17.

Austin, P. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, **46**, 399–424.

Austin, P. C. and Mamdani, M. M. (2006). A comparison of propensity score methods: a case-study estimating the effectiveness of post-ami statin use. *Statistics in Medicine*, **25** (12), 2084–2106.

Avdic, D. (2016). Improving efficiency or impairing access? Health care consolidation and quality of care: Evidence from emergency hospital closures in Sweden. *Journal of Health Economics*, **59**, 44–60.

—, Moscelli, G., Pilny, A. and Sriubaité, I. (2019). Subjective and objective quality and choice of hospital: Evidence from maternal care services in Germany. *Journal of Health Economics*, **68**.

Benham, L., Maurizi, A. and Reder, M. (1968). Migration, location and remuneration of medical personnel: physicians and dentists. *The Review of Economics and Statistics*, **50** (3), 332–347.

Bindman, A. B., Keane, D. and Lurie, N. (1990). A public hospital closes. Impact on patients' access to care and health status. *JAMA*, **264** (22), 2899–2904.

Blondel, B., Drewniak, N., Pilkington, H. and Zeitlin, J. (2011). Out-of-hospital births and the supply of maternity units in France. *Health & Place*, **17** (5), 1170–1173.

Bloom, N., Propper, C., Seiler, S. and van Reenen, J. (2015). The Impact of Competition on Management Quality: Evidence from Public Hospitals. *Review of Economic Studies*, **82**, 457–489.

Buchmueller, T. C., Jacobson, M. and Wold, C. (2006). How far to the hospital? The effect of hospital closures on access to care. *Journal of Health Economics*, **25**, 740–761.

Bünnings, C., Schmitz, H., Tauchmann, H. and Ziebarth, N. R. (2019). The role of prices relative to supplemental benefits and service quality in health plan choice. *Journal of Risk and Insurance*, **86** (2), 415–449.

Burkey, M. L., J., B., Eiselt, H. A. and Toyoglu, H. (2017). The impact of hospital closures on geographical access: Evidence from four southeastern states of the United States. *Operations Research Perspectives*, **4**, 56–66.

Busse, R. and Blümel, M. (2014). Germany: Health system review. European Observatory on Health Systems and Policies. *Health Systems in Transition*, **16** (2), 1–296.

Capps, C., Dranove, D. and Lindrooth, R. C. (2010). Hospital closure and economic efficiency. *Journal of Health Economics*, **29** (1), 84–109.

Chou, S., Deily, M. E. and Li, S. (2014). Travel Distance and Health Outcomes for Scheduled Surgery. *Medical Care*, **52** (3), 250–257.

Ciliberto, F. and Lindrooth, R. C. (2007). Exit from the Hospital Industry. *Economic Inquiry*, **45** (1), 71–81.

Clark, A. E. and Milcent, C. (2011). Public employment and political pressure: The case of French hospitals. *Journal of Health Economics*, **30** (5), 1103–1112.

Coenen, M., Haucap, J. and Herr, A. (2012). Regionalität: Wettbewerbliche Überlegungen zum Krankenhausmarkt. In J. Klauber, M. Geraedts, F. Friedrich and W. J. (eds.), *Krankenhaus-Report 2012 – Schwerpunkt: Regionalität*, Schattauer. Stuttgart, pp. 149–1963.

Countouris, M., Gilmore, S. and Yonas, M. (2014). Exploring the impact of a community hospital closure on older adults: A focus group study. *Health & Place*, **26**, 143–148.

Crown, W. (2014). Propensity-score matching in economic analyses: Comparison with regression models, instrumental variables, residual inclusion, differences-in-differences, and decomposition methods. *Applied health economics and health policy*, **12**.

Cuellar, A. E. and Gertler, P. J. (2003). Trends in hospital consolidation: The formation of local systems. *Health Affairs*, **22** (6), 77–87, pMID: 14649434.

Currie, J. and Reagan, P. B. (2003). Distance to hospital and children's use of preventive care: Is being closer better, and for whom? *Economic Inquiry*, **41** (3), 378–391.

Cutler, D. M. and Horwitz, J. (2007). Converting hospitals from not-for-profit to for-profit status: Why and what effects? In D. M. Cutler (ed.), *The changing hospital industry: Comparing not-for-profit and for-profit institutions*, University of Chicago Press, pp. 45–90.

Deily, M. E., McKay, N. L. and Dorner, F. H. (2000). Exit and Inefficiency: The Effects of Ownership Type. *The Journal of Human Resources*, **35** (4), 764–747.

Deutsche Krankenhausgesellschaft DKG (2014). Bestandsaufnahme zur Krankenhausplanung und Investitionsfinanzierung in den Bundesländern. In *Dezernat II - Krankenhausfinanzierung und -planung*, Bundesverband der Krankenhausträger in der Bundesrepublik Deutschland.

Deutsche Krankenhausgesellschaft DKG (2018). Bestandsaufnahme zur Krankenhausplanung und Investitionsfinanzierung in den Bundesländern. In *Dezernat II - Krankenhausfinanzierung und -planung*, Bundesverband der Krankenhausträger in der Bundesrepublik Deutschland.

Dor, A. and Friedman, B. (1994). Mergers of not-for-profit hospitals in the 1980s: Who were the most likely targets? *Review of Industrial Organization*, **9** (4), 393–407.

Dranove, D. (1998). Economies of scale in non-revenue producing cost centers: Implications for hospital mergers. *Journal of Health Economics*, **17** (1), 69–83.

— and Lindrooth, R. (2003). Hospital consolidation and costs: Another look at the evidence. *Journal of Health Economics*, **22** (6), 983–997.

— and Shanley, M. (1995). Cost reductions or reputation enhancement as motives for mergers: The logic of multihospital systems. *Strategic Management Journal*, **16** (1), 55–74.

Elixhauser, A., Steiner, C., Harris, D. R. and Coffey, R. (1998). Comorbidity Measures for Use with Administrative Data. *Medical Care*, **36** (1), 8–27.

Engjom, H., Morken, N.-H., Norheim, O. and Klungsøyr, K. (2014). Availability and access in modern obstetric care: a retrospective population-based study. *BJOG: An International Journal of Obstetrics & Gynaecology*, **121** (3), 290–299.

Frankfurter Allgemeine (2013). Kampf um die Grundversorgung. Available at: http://www.faz.net/aktuell/schliessung-von-krankenhaeusern-kampf-um-die-grundversorgung-12133920.html.

Freier, R. and Thomasius, S. (2016). Voters prefer more qualified mayors, but does it matter for public finances? Evidence for Germany. *International Tax and Public Finance*, **23**, 875–910.

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

Gaynor, M. (2011). Examining the Impact of Health Care Consolidation. Statement before the Committee on Energy and Commerce, Oversight and Investigations Subcommittee, U.S. House of Representatives.

Greifeneder, S. (2019). Germany. In S. Ellson (ed.), *The Healthcare Law Review (3rd edition)*, London, UK.: Law Business Research Ltd., pp. 45–53.

Grzybowski, S., Stoll, K. and Kornelsen, J. (2011). Distance matters: a population based study examining access to maternity services for rural women. *BMC Health Services Reseach*, **11** (147).

Hansmann, H., Kessler, D. and McClellan, M. B. (2007). Ownership form and trapped capital in the hospital industry. In *NBER Chapters: The Governance of Not-for-Profit Organizations*, National Bureau of Economic Research, Inc., pp. 45–70.

Harrison, T. D. (2007). Consolidations and closures: An empirical analysis of exits from the hospital industry. *Health Economics*, **16** (5), 457–474.

Hentschker, C. and Mennicken, R. (2014). The Volume-Outcome Relationship and Minimum Volume Standards – Empirical Evidence for Germany. *Health Economics*, **24** (6), 644–658.

— and — (2018). The Volume-Outcome Relationship Revisited: Practice Indeed Makes Perfect. *Health Services Research*, **53** (1), 15–34.

Ho, V. and Hamilton, B. H. (2000). Hospital mergers and acquisitions: does market consolidation harm patients? *Journal of Health Economics*, **19**, 767–791.

Hsia, R. Y., Kanzaria, H. K., Srebotnjak, T., Maselli, J., McCulloch, C. and Auerbach, A. D. (2012). Is emergency department closure resulting in increased distance to the nearest emergency department associated with increased inpatient mortality? *Annals of Emergency Medicine*, **60** (6), 707–715.e4.

Huber, S. and Rust, C. (2016). Calculate Travel Time and Distance with OpenStreetMap Data using the Open Source Routing Machine (OSRM). *The Stata Journal*, **16** (2), 416–423.

Huckman, R. S. (2006). Hospital integration and vertical consolidation: An analysis of acquisitions in New York State. *Journal of Health Economics*, **25**, 58–80.

Karmann, A. and Roesel, F. (2017). Hospital Policy and Productivity – Evidence from German States. *Health Economics*, **26**, 1548–1565.

Krause, M. (2019). *Communal fees and election cycles: Evidence from German municipalities*. ifo Working Paper Series 293, ifo Institute - Leibniz Institute for Economic Research at the University of Munich.

Krishnan, R., Joshi, S. and Krishnan, H. (2004). The influence of merges on firms' product-mix strategies. *Strategic Management Journal*, **25**, 587–611.

Kuhn, M. and Ochsen, C. (2019). Population change and the regional distribution of physicians. *The Journal of the Economics of Ageing*, **14**.

Levitt, S. D. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *The American Economic Review*, **87** (3), 270–290.

Linden, A. (2006). What Will It Take for Disease Management to Demonstrate a Return on Investment? New Perspectives on an Old Theme. *The American Journal of Managed Care*, **12**, 217–222.

Lindrooth, R. C., Lo Sasso, A. T. and Bazzoli, G. J. (2003). The effect of urban hospital closure on markets. *Journal of Health Economics*, **22**, 691–712.

Mark, T. L. (1999). Analysis of the rationale for, and consequences of, nonprofit and for-profit ownership conversions. *Health Services Research*, **34** (1), 83–101.

Mennicken, R., Kolodziej, I. W., Augurzky, B. and Kreienberg, R. (2014). Concentration of gynaecology and obstetrics in Germany: Is comprehensive access at stake? *Health Policy*, **118** (3), 396–406.

OECD (2019). *Health at a glance: Europe 2019. OECD Indicators*. Tech. rep., OECD Publishing, Paris.

Pilny, A. (2014). *Mergers and Acquisitions in the German Hospital Market – Who are the Targets?* Tech. Rep. 518, RWI - Leibniz-Institut für Wirtschaftsforschung, Ruhr-University Bochum, TU Dortmund University, University of Duisburg-Essen.

— (2017). Explaining differentials in subsidy levels among hospital ownership types in Germany. *Health Economics*, **26**, 566–581.

—, Wübker, A. and Ziebarth, N. R. (2017). Introducing Risk Adjustment and Free Health Plan Choice in Employer-Based Health Insurance: Evidence from Germany. *Journal of Health Economics*, **56**, 330–351.

Preusker, U. K., Müschenich, M. and Preusker, S. (2014). *Darstellung und Typologie der Marktaustritte von Krankenhäusern Deutschland 2003 - 2013*. Tech. rep., Preusker Health Care OY.

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.-C., Saunders, L. D., Beck, C. A., Feasby, T. E. and Ghali, W. A. (2005). Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care*, **43** (11), 1130–1139.

Ravelli, A., Jager, K., de Groot, M., Erwich, J., Rijninks-van Driel, G., Tromp, M., Eskes, M., Abu-Hanna, A. and Mol, B. (2011). Travel time from home to hospital and adverse perinatal outcomes in women at term in the Netherlands. *BJOG: An International Journal of Obstetrics & Gynaecology*, **118** (4), 457–465.

Schmid, A. and Varkevisser, M. (2016). Hospital merger control in Germany, the Netherlands and England: Experiences and challenges. *Health Policy*, **120** (1), 16–25.

Schmitt, M. (2017). Do hospital mergers reduce costs? *Journal of Health Economics*, **52**, 74–94.

Schulten, T. (2006). Liberalisation, privatisation and regulation in the German healthcare sector/hospitals. *Country reports on liberalisation and privatisation processes and forms of regulation. Deliverable 1 for the Project „Privatisation of Public Services and the Impact on Quality, Employment and Productivity (PIQUE)”.*

Shen, Y.-C. (2003). Changes in hospital performance after ownership conversions. *Inquiry*, **40** (3), 217–234.

Skinner, J. (1994). What Do Stochastic Frontier Cost Functions Tell Us about Inefficiency. *Journal of Health Economics*, **13** (3), 323–328.

Sloan, F. A., Ostermann, J. and Conover, C. J. (2003). Antecedents of hospital ownership conversions, mergers, and closures. *Inquiry*, **40** (1), 39–56.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, **65** (3), 557–586.

Succi, M. J., Lee, S. D. and Alexander, J. A. (1997). Effects of Market Position and Competition on Rural Hospital Closures. *Health Services Research*, **31**, 679–699.

Thüringer Allgemeine (2014). Krankenhaus in Schleusingen wird geschlossen. Available at: http://www.thueringer-allgemeine.de/web/zgt/leben/detail/-/specific/Krankenhaus-in-Schleusingen-wird-geschlossen-1381314980.

Tiemann, O., Schreyögg and Busse, R. (2011). Hospital ownership and efficiency: A review of studies with particular focus on Germany. *Health Policy*.

Town, R., Wholey, D., Feldman, R. and Burns, L. (2006). The welfare consequences of hospital mergers. *NBER Working Paper; No. 12244*.

Vogt, V. (2016). The contribution of locational factors to regional variations in office-based physicians in Germany. *Health Policy*, **120**, 198–204.

Wasem, J., Gress, S. and Okma, K. G. (2004). The role of private health insurance in social health insurance countries. In R. B. Saltman, R. Busse and J. Figueras (eds.), *Social health insurance systems in western Europe*, New York: World Health Organization on behalf of the European Observatory on Health Systems and Policies, pp. 227–247.

WAZ (2011). Das Krankenhaus im Wimbern wird geshlossen. Available at: http://www.derwesten.de/staedte/menden/das-krankenhaus-in-wimbern-wird-geschlossen-id5151176.html.

WDR (2015). Die St. Anna-Klinik in Wuppertal Elberfeld wird geschlossen. Die Klapperstorch wird hier nicht mehr landen. Available at: http://www1.wdr.de/studio/wuppertal/themadestages/Anna104.html.

Westdeutsche Zeitung (2013). Willich ohne Krankenhaus: Katharinen-Hospital wird geschlossen. Available at: http://www.wz-newsline.de/lokales/kreis-viersen/willich/willich-ohne-krankenhaus-katharinen-hospital-wird-geschlossen-1.1446813.

Williams, D., Hadley, J. and Pettengill, J. (1992). Profits, Community Role, and Hospital Closure: An Urban and Rural Analysis. *Medical Care*, **30** (2), 174–187.

World Health Organization, W. S. O., World Health Federation (2011). Tech. Report. In S. Mendis, P. Puska and B. Norrving (eds.), *Global Atlas on Cardio-vascular Disease Prevention and Control: Policies, Strategies, and Interventions*, World Health Organization, Geneva.

Zuckerman, A., Welch, W. P. and Pope, G. C. (1990). A geographic index of physician practice costs. *Journal of Health Economics*, **9**, 39–69.

Appendix A: Additional tables and figures

TABLE A.1: Classification of Elixhauser Comorbidities

Variable	Comorbidity
el1	Congestive heart failure
el2	Cardiac arrhythmias
el3	Vascular disease
el4	Pulmonary circulation disorders
el5	Peripheral vascular disorders
el6	Hypertension, uncomplicated
el7	Hypertension, complicated
el8	Paralysis
el9	Other neurological disorders
el10	Chronic pulmonary disease
el11	Diabetes, uncomplicated
el12	Diabetes, complicated
el13	Hypothyroidism
el14	Renal failure
el15	Liver disease
el16	Peptic ulcer disease (excluding bleeding)
el17	AIDS/HIV
el18	Lymphoma
el19	Metastatic cancer
el20	Solid tumor without metastasis
el21	Rheumatoid arthritis/collagen vascular diseases
el22	Coagulopathy
el23	Obesity
el24	Weight loss
el25	Fluid and electrolyte disorders
el26	Blood loss anemia
el27	Deficiency anemia
el28	Alcohol abuse
el29	Drug abuse
el30	Psychoses
el31	Depression

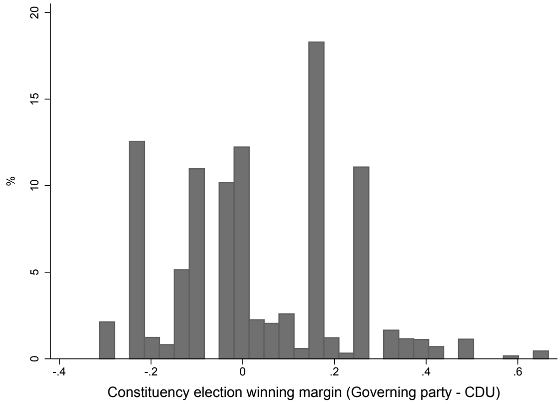
NOTE.— Table presents all Elixhauser comorbidities. Detailed classification of Elixhauser Comorbidities with respective ICD-9 and ICD-10 codes can be found in Quan *et al.* (2005).

TABLE A.2: Descriptive Statistics of treated and control samples

	Treated		Control		Difference	
	mean	sd	mean	sd	b	t
Market: age	70.79	4.97	70.61	5.38	-0.18	(-0.24)
Market: male	0.60	0.15	0.59	0.17	-0.01	(-0.54)
Market: if rural	0.65	0.48	0.49	0.50	-0.16*	(-2.13)
Market: Elixhauser	2.52	0.76	2.73	0.68	0.21	(1.96)
Market: # beds	113.58	102.96	116.41	91.86	2.83	(0.19)
Market: Small hospital size	0.30	0.46	0.20	0.40	-0.10	(-1.60)
Market: Middle hospital size	0.32	0.47	0.41	0.49	0.08	(1.17)
Market: Large hospital size	0.38	0.49	0.40	0.49	0.02	(0.26)
Market: If university	0.00	0.00	0.00	0.00	0.00	(.)
Market: If teaching	0.20	0.41	0.16	0.37	-0.04	(-0.71)
Market: If public	0.42	0.50	0.59	0.49	0.17*	(2.34)
Market: If non-profit	0.11	0.31	0.00	0.00	-0.11**	(-3.20)
Market: Cases/doctor	235.84	72.06	369.10	437.44	133.25*	(2.44)
Market: Doctor's specialization level	0.53	0.20	0.55	0.20	0.02	(0.65)
Market: Cases/nurse	69.16	18.01	69.50	28.69	0.34	(0.09)
Observations	93		86		179	

NOTE.— Table presents the descriptive statistics of hospital markets by treated and control groups. All statistics are aggregated to mean values in the study period and present an average patient as well as hospital in the market.

Figure A.1: The distribution of the instrument



Note.— Figure presents the distribution of the instrument that is the CDU winning margin in the municipality elections against the opposition.

Appendix B

How are hospitals financed?

The *Hospital Financing Act 1972* (Krankenhausfinanzierungsgesetz, KHG) set up a dual-financing framework for hospitals (See Figure B.1). This concept distinguishes operational costs, such as expenditure for patient care and medical supply, from investment costs, such as new buildings and equipment. While operational costs are mainly reimbursed by statutory and private health insurers,[29] investments in capital are secured by the federal state; thus the concept lays basis for several independent decision-makers in the financing of a hospital. Here *the National Association of Health Insurers (GKV-Spitzenverband)* acts as a consulting party with respect to hospital financing, whereas the federal state designs the investment plan and makes decisions about the type and the size of funding hospitals receive (Karmann and Roesel, 2017; Pilny, 2017; Preusker *et al.*, 2014). Such conditions are described in the *State Hospital Plans* (Landkrankenhausplan). The *State Hospital Plans* set region-specific aims that follow the main national goals to ensure efficient, high quality and, in the future, economically independent hospitals. All hospitals included in the *State Hospital Plans* are entitled to receive individual grants, chiefly for long-term investments in new capital, and lump-sum grants, that cover the short-term assets and small scale buildings. In 2009, the *Hospital Financing Reform Act* (KHRG) complemented the existing funding model with additional financial aid on merit basis with the federal government deciding whether and how to distribute the additional investment (Busse and Blümel, 2014).

Hospitals that become dependent on the federal benefits are highly restricted by German Healthcare Law.[30] They have an interest in providing high quality of services to attract more patients in order to maintain the financial support from the state. Making the provision of healthcare effective and efficient ensures market stability and profitability. The quality of hospitals is controlled by the Federal Joint Committee (Gemeinsamer Bundesausschuss, G-BA)[31] which was founded in 2004 through *Health Modernization Act*. The G-BA defines the hospital performance quality criteria that are relevant for hospital planning and which form the basis for each *State Hospital Plan*. Hospitals are obliged to submit quarterly quality information to the *Institute for Quality Assurance and Transparency in Healthcare* (Institut für Qualitätssicherung und Transparenz im Gesundheitswesen, IQTIG), which is evaluated and published online biannually. Using this information,

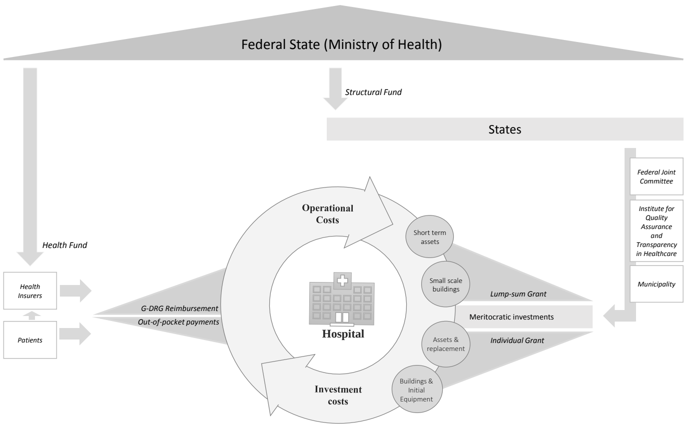
[29]Since 2004 hospital reimbursement system for inpatient care is based on patient classification system German Diagnosis-related groups (G-DRG).

[30]The German healthcare Law is summarized in Greifeneder (2019) and outlined in https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen.html.

[31]G-BA is the highest decision-making body of the joint self-government of physicians, dentists, hospitals and health insurance funds in Germany.

the G-BA formulates the assessment criteria for hospital performance, that federal states ought to incorporate into their hospital plans. Hospitals that do not comply with this criteria are excluded from the plan and, in most cases, close (Busse and Blümel, 2014; Preusker *et al.*, 2014).

FIGURE B.1: Hospital financing



Note.— Figure presents the financial structure of hospitals in Germany.

To promote healthy competition in the market, the KHG regulates with respect to the variety of hospital ownership types in the state territories (Karmann and Roesel, 2017). The German hospital market has a multi-ownership structure, that includes three different types of ownership: private for-profit, private non-profit and public. In 2012, Germany had approximately 2000 hospitals with total of 500,000 beds of which 48% were public, 34% private non-profit and 18% private for-profit (Busse and Blümel, 2014). Management of the hospital differs by ownership type; however, according to the law, all hospitals are entitled to receive subsidies from the federal state regardless of their ownership type, ensuring the equality in the granting process (Deutsche Krankenhausgesellschaft DKG, 2014). Evidence from Pilny (2017) shows that private for-profit hospitals receive the least financial support as profitable hospitals can often rely on their own profits and do not need to receive benefits from the federal state. Meanwhile, public and non-profit hospitals do often rely on these subsidies and are mainly funded by the federal state and health insurance generated funds (Pilny, 2017; Tiemann *et al.*, 2011). Public hospitals are owned by public entities such as local or regional governments (Tiemann *et al.*, 2011), thus are highly dependent on various politics at both federal and municipal level. Even though the federal state decides on the number of hospitals and hospital beds, municipal decision-bodies are responsible for ensuring the stationary medical supply in their territory. The city or county councils are usually the owners of public hospitals with mayors and county commissioners leading or participating in hospital supervisory boards. They also oversee existing hospital finances and, even though major investments come from the state, they often need to cover monetary losses in case of financial insufficiency. Thus, municipal politics play an important role in the management of hospitals; however, their power and authority differs from state to state (Busse and Blümel, 2014; Deutsche Krankenhausgesellschaft DKG, 2018). Non-profit hospitals are in most cases owned by the church or other welfare institutions. Even though they are less directly dependent on local politicians, they are still exposed to local policy changes due to close cooperation with the municipality leaders. As hospitals of these ownership types are highly dependent on the federal state funding (Pilny, 2017), I will focus on public and non-profit hospitals in this paper.

Municipal Politics

In Germany, the federal system is comprised of three tiers of governments. The cornerstone of the German political system is federal assembly (*Bundestag*) that has the widest political powers and is responsible for the enactment of all legislations. The interest of all 16 federal states in Germany is represented by the federal council (*Bundesrat*). The lowest tier of administration is local governments in each municipality that account for approximately 12,500 municipal governments. Municipal governments are responsible for all administrative tasks on local matters and the execution of any legislative assignments made by the federal and state governments. Municipal political bodies have little influence in federal or state level legislation, they are responsible for executing given tasks and often only have discretion on how to complete them (Freier and Thomasius, 2016; Krause, 2019).

Local municipal governments are typically elected every 5 years.[32] The election cycle varies from state to state and is not controlled by individual municipality. During an election a local council is chosen to represent the interests of the municipality. Local councils share a joint responsibility for all municipal affairs with the mayor or the executive who is sometimes elected at a separate election as an individual candidate. While the mayoral elections follow a majoritarian electoral system, the local council elections are conducted as a plurality voting system and the council is elected according to a proportional representation system (Freier and Thomasius, 2016; Krause, 2019).

The composition of politics at the local level is mainly shaped by six major political parties that are currently active in Germany. These are the two largest political parties: the Christian Democratic Union (*CDU*) that follows a centre-right political ideology and the Social Democratic Party (*SPD*) that has a centre-left political interest. In the recent elections of

[32] In the state of Bavaria the election is held each 6 years.

the *Bundestag* the populist party, the Alternative for Germany (*AfD*), that follows a right-wing to far-right political ideology joined these major political parties and became the third largest political party at the federal election. The other active political parties are: the Free Democratic Party (FDP) on the centre to centre-right, the Greens (Gruene) on the centre left and the Die Linke (LINKE), on the left wing.[33]

[33] More detailed information about each political party in Germany can be found at the official websited of *Bundestag*: https://www.bundestag.de/parlament/fraktionen.

Essay 4: Economic Consequences of Road Traffic Injuries. Application of the Super Learner Algorithm[*]

1 Introduction

The design of contracts between payers and providers of healthcare is an important factor in the cost and quality of healthcare delivery. Providers have a strong influence over the treatment decisions and payers can only imperfectly observe potential costs and the quality of provided healthcare (Eggleston, 2000). For example, when the contract is based on a fixed lump-sum amount, it gives an incentive for cost-saving, but encourages the avoidance of high-cost clients and lowers unobserved quality. On the other hand, when the payment is fixed and closely aligns with actual costs, the provider will have a lower risk for underpayment and lower incentives for patient selection or quality reduction (Cucciare and o'Donohue, 2006; Ellis *et al.*, 2018).[1] The efficiency of these contracts require adequate cost-sharing that is only feasible if the payer and the provider share the information on risk such as actual costs and patient outcomes (Ellis and McGuire, 1993).

This paper contributes to the literature on predicting healthcare costs and patient outcomes in an environment where a single payer contracts with multiple providers. Predicting healthcare costs has long been at a fo-

[*]The author thanks to Anthony Harris and Andrew M. Jones for a substantial contribution in supervising this project and to Belinda Gabbe for providing insightful feedback on the injury population. Valuable comments by participants of the Virtual Essen Health Economics Seminar Series are greatly appreciated. Financial support from the DAAD/Go8 funding by the "Bundesministerium für Bildung und Forschung (BMBF)", German Federal Ministry of Education and Research is gratefully acknowledged. The Victorian State Trauma Registry is a State Government of Victorian Department of Health and Human Services and Transport Accident Commission funded initiative.

[1]For studies analysing the provider incentives and the optimal reimbursement system in markets characterised by supply-side incentives, see e.g. Dranove (1987); Ellis and McGuire (1986, 1988, 1990); Frank and Lave (1989); Lave (2003); McClellan (1997); Newhouse *et al.* (1997); Newhouse (1996).

cal point of health economics literature. Modelling such outcomes comes with a number of statistical challenges, because the statistical profiles are often characterized by non-normal distributions. The distribution of these data is asymmetric, strongly skewed and exhibit a particularly long tail representing patients with high costs and poor outcomes (Jones, 2011). Researchers have made great efforts in building statistical models to accurately predict health care costs.[2] As medical information has become more and more detailed, the focus has shifted towards a variety of new statistical approaches utilised in the field of "big data". Data science methods such as supervised Machine Learning (ML) offer functional flexibility and the ability to fit difficult data patterns without imposing prior assumptions. Several data science techniques, such as testing out-of-sample performance, have already been adopted in the modelling literature (Bertsimas *et al.*, 2008; Jones and Lomas, 2016; Jones *et al.*, 2014). The main advantages of ML methods are the ability to uncover complex structures not known/specified in advance (Mullainathan and Spiess, 2017) and to account for potential multicollinearity when controlling for a large set of covariates (James *et al.*, 2013). Thus, they enable fitting very flexible functional forms without overfitting the data and can perform better at out-of-sample predictions than standard regression analysis (Chu and Zhang, 2003). Researchers have also used such methods in efforts to predict healthcare resource use, service utilisation and various clinical outcomes (e.g., Arandjelović, 2015; Bertsimas *et al.*, 2008; Burnham *et al.*, 2018; Einav *et al.*, 2016; Kan *et al.*, 2019; Lahiri, 2014; Pyrkov *et al.*, 2018; Rose, 2016).

In this paper we employ a ML based algorithm – the Super Learner – to predict healthcare costs and patient outcomes. The Super Learner algorithm is based on multiple parametric and non-parametric statistical models and selects an optimal weighted combination of them to find the best predictive model. Proposed by van der Laan *et al.* (2007) the algorithm has demonstrated significant potential in research related to healthcare that predicts various clinical patient outcomes (Kessler *et al.*, 2014; Pirracchio *et al.*, 2015). Although still in its infancy, the Super Learner algorithm has been used in the economic context for risk adjustment in the health insurance markets (Park and Basu, 2018; Rose, 2016; Shrestha *et al.*, 2018) and for predicting the unprofitability of health insurance enrollees (Rose *et al.*, 2017). To this end, we employ a comprehensive insurance claims data provided by the statutory insurance company Transport Accident Commission

[2] In the context of risk adjustment for both provider and insurance health plan payment, researchers have applied various parametric statistical models (see, e.g., Curtis *et al.*, 2014; Dixon *et al.*, 2011; Duan, 1983; Iezzoni, 2012; Jones, 2000; Jones *et al.*, 2014; Manning *et al.*, 2005; Shuman *et al.*, 1972) and a variety of semi-parametric models (see, e.g., Deb and Burgess, 2003; Gilleskie and Mroz, 2004; Jones, 2011; Jones *et al.*, 2014, 2015; Manning *et al.*, 2005; McDonald *et al.*, 2013; Mullahy, 2009). While traditional parametric models are easy to interpret, they often suffer from problems caused by the presence of significant correlations between the selected covariates (James *et al.*, 2013). The performance of semi-parametric models is mixed (Jones and Lomas, 2016).

(TAC) offering compulsory third party insurance for Victorians who were injured in a traffic incident. We link the claimant information to a rich patient-level dataset, the Victorian State Trauma Registry (VSTR), that provides detailed information about all major trauma patients in Victoria. Major trauma is the most complex type of injury that has a potential to cause death or prolonged disability. Patients with a major trauma are at risk for high long-term costs and often describe a patient cohort that is more complex than the average.

We contribute to the literature in the following ways. First, we adapt advanced statistical methods to improve the predictive power of traditional regression-based approaches and contribute to this emerging evidence of the application of the Super Learner in the economic context. Second, we predict not only costs but also patient outcomes that are relevant for risk-sharing between the payer and the provider as they inform payers about the quality of healthcare. We use the concept of net benefits as proposed by Stinnett and Mullahy (1998) to estimate the value of purchased services. Net benefits are calculated as the monetary value of patient lifetime outcomes and expressed in the Quality of Life Years (QALYs) after treatment. In addition, since the participation in paid employment after a traffic incident is an important indicator of well-being both in terms of income and mental health, we also predict return to work.

This paper is organised as follows. Section 2 explains the institutional context including the information about the Victorian State Trauma System and the TAC insurance in the state of Victoria. Section 3 describes the data, while Section 4 outlines the methodology used to estimate patient outcomes and introduces the prediction methods in detail. Section 5 reports prediction results and evaluates them based on a number of evaluation criteria, Section 6 discusses potential prediction errors and Section 7 concludes.

2 Institutional Context

Victoria is the second most populous state in Australia with a population of approximately 6.5 million.[3] The state operates a regionalized trauma system to ensure that injured patients receive the best possible medical treatment and specialized hospital care. The system is categorized into three levels of care: three major trauma services (MTS) (two adult and one paediatric), that provide definite care to most of the state's trauma patients either through primary triage or secondary transfer; metropolitan trauma services (MeTS) and regional trauma services are the second level of care that also provide immediate care when MTS cannot be reached in time; and Metropolitan Primary Care Services offer the third level of care (DHHS, Feb 2014). Around 80% of trauma cases receive care at a designated

[3] as of September 2018, The Australian Bureau of Statistics, 3101.0 - Australian Demographic Statistics Catalogue.

trauma centre for major trauma services and nearly 90% of road injuries are treated at a MTS (VSTR, 2014). The trauma system is monitored using a population-based trauma registry that collects data about all major trauma patients irrespective of the admitting hospital. The registry is used for research purposes in order to continuously enhance the quality of trauma management and improve patient outcomes after an injury (see, e.g. DHHS, Feb 2014; DHHS, Jul 2014; VSTR, 2014; Beck *et al.* (2016); Gabbe *et al.* (2014, 2012)).

Multiple sources of funding exist in Australia to finance the treatment of trauma. While Australia's publicly funded universal healthcare insurance scheme (Medicare) provides healthcare coverage for all Australian citizens and permanent residents, nearly all care for road injuries in the state of Victoria is funded by a publicly-owned organisation, the Transport Accident Commission (TAC). The TAC operates on a "no-fault" basis and provides financial and rehabilitation support for Victorians who were injured in transport incidents. The compensation covers out-of-pocket medical and non-medical costs[4] and life-back-on-track expenses including income assistance, rehabilitation, return to work programs, travel and funeral costs as well as costs for specialised equipment such as wheelchairs and modified vehicles to support patients who acquired disability due to injury (TAC, 2018). The TAC collects funds via a TAC charge, a component of the vehicle registration yearly fee. The TAC charge for each insuree varies by several characteristics: vehicle-related such as the type and use of the vehicle (i.e., passenger vehicle, goods vehicle, motorcycle etc.); the level of risk for traffic incident in the area (postcode where the vehicle is kept); and any eligible discounts the owner of the vehicle has, e.g. social security recipient. pensioner or concession card.

Both public and private healthcare providers are eligible to seek reimbursement from the TAC if they provide TAC approved services to a patient who experienced a road traffic-related injury.[5] The reimbursement of public hospitals follows the Activity Based Funding National Framework based on the Australian National Diagnosis Related Group (AN-DRG). The payment per each activity unit is specific for TAC patients.[6] In the case of

[4]For example, non-medical costs could be related to travelling to medical appointments, accommodation, any legal or administrative costs associated with the claims reimbursement and disability-related health support.

[5]The full list of approved treatments and services is available in the official website of TAC https://www.tac.vic.gov.au/providers/working-with-the-tac/what-we-can-pay-for. These services can be provided without a prior approval if the treatment takes place within the first 90 days of the injury. In the case of the service being provided after that time, the healthcare provider has to contact the TAC for approval.

[6]There is a slight difference in the price per unit applicable for TAC patients. For example, the most recently published *Policy and funding guidelines 2019-20* by the Department of Health and Human Services indicates that the price per activity unit for TAC patients is $5,843, while for other patients treated in metropolitan and regional hospital is $5,029, that accounts for approx 14% higher negotiated price per unit for TAC patients.

private healthcare providers, the TAC has several contracted and arrangement partners that have binding contracts or arrangements with the TAC about the provision and reimbursement for services payable by TAC. If the provider does not have a contract or an arrangement with the TAC, health and service providers are reimbursed on a fee-for-service basis. If the fee for service is above the TAC payable fee, the additional costs are usually covered by the consumer.

3 Data

The empirical analysis uses data from the VSTR that includes information on all major trauma patients in Victoria.[7] It provides a wide range of patient characteristics such as age, gender, socio-economic status as well as comprehensive medical and non-medical information about patient injuries. The registry records various characteristics at the scene of the incident such as cause, place and mechanism of the injury and allows us to control for potentially important non-medical differences between injuries. The clinical information is recorded at the time of admission and during the hospital stay and provides comprehensive information about the patient's treatment and recovery.

After discharge from the hospital each patient is followed up by a telephone interview at 6, 12 and 24 months after injury.[8] An interviewer collects detailed information about patients' recovery, level of physical function and return to work. The follow up interview additionally includes the 3-level EuroQol five dimensions questionnaire (EQ-5D-3L).[9] The EQ-5D-3L instrument comprises five dimensions based on patient's mobility, self-care, usual daily activities, experience of pain or discomfort and anxiety or depression. Each dimension is ranked in order of increasing severity according to: *no problems*, *some problems* or *extreme problems*. Using this information, we estimate a health utility score for each patient that represents patient's health state at 6, 12 and 24 months after the injury occurred. To identify post-discharge deaths the registry is linked with the state's deaths register.

Our sample of interest is restricted to patients who experienced a major trauma as a result of a road traffic crash in Victoria and are aged above 15 years. The restriction leaves us with 11,625 major trauma patients in 2009–2017. We link this sample with the insurance claims data that includes

[7] Using the ICD-10-AM information, major trauma is defined if any of the following criteria are met (i) Death (at scene of injury or in-hospital) related to injury; (ii) an Injury Severity Score >12; (iii) admission to an intensive care unit (ICU) for > 24h and requiring mechanical ventilation for at least part of their ICU stay; and (iv) urgent surgery is performed.

[8] The response rate of the follow up study at 6 and 12 months was around 70% for full data collection; 10% partial data collection and around 5% reported death before the study; at 24 months the respective figures were 65%, 20% and 7% respectively.

[9] For a detailed documentation about EQ-5D-3L questionnaire see https://euroqol.org/eq-5d-instruments/; and more information on use of EQ-5D-3L in injury setting see Derrett *et al.* (2009).

information on around 8 million instances of claims paid to these patients. This provides information about all inpatient and outpatient costs and price and information of all additional services used outside the hospital. Descriptive statistics for the sample are presented in Table A.1 in the Appendix.

4 Methods

4.1 Outcomes

Direct costs of the injury

Using detailed information from the TAC insurance claims data, for each patient i we compute the *total direct costs of injury (D_i)* attributed to the initial treatment as well as all subsequent costs a patient had within 24 months after discharge from hospital. Here D_i is a function of both medical, M_i, (ambulance care, inpatient care hospital stay, outpatient care, rehabilitation and prescription drugs) and non-medical, N_i, (travel, accommodation, legal, administrative and disability related health support) expenses, but does not include any impairment payments or loss of earnings.[10]

$$D_i = M_i + N_i \tag{4.1}$$

Net benefits of treatment

To better understand the quality of healthcare providers, we follow the framework developed by Stinnett and Mullahy (1998). We calculate the utility gained from provided care as the difference between the Quality-adjusted-life-years (QALYs) with treatment and the expected QALYs without treatment. For each patient i the *Net Health Benefit* (NHB) is defined as the following:

$$NHB_i = QALY_i^{CARE} - QALY_i^{NON} \tag{4.2}$$

where NHB is expressed in units of "benefit" gained from treatment such as QALYs. $QALY^{CARE}$ is the benefit gained from the treatment, whereas $QALY^{NON}$ is the outcome without treatment. The QALYs is a measure of disease burden expressed in the number of years lived in perfect health. It is estimated using the patient's health state and the number of years lived in the given state. We employ the patient's health utility score based on the EQ-5D recorded in the follow up study as described in Section 3 and multiply with the patient's expected life years. We rely on the information about the Australian life expectancy by age and gender provided

[10] We exclude any expenses related to impairment annuity, loss of earnings or death benefits paid to a spouse or other family member to avoid double counting with health-related quality of life used to calculate net benefits of treatment below.

by the Australian Institute of Health and Welfare.[11] A number of studies have concluded that an individual who had a traumatic injury is likely to suffer from long-term consequences from physical and psychologic impairment, acquired disability or Post-traumatic Stress Disorder and the highest risk for long-term effects was found in patients who were hospitalised or required critical medical care (Holbrook *et al.*, 2001a, 2005, 2001b; Holtslag *et al.*, 2007; Sluys *et al.*, 2005). Following this evidence, we estimate lifelong NHB for patients who encountered a life-threatening injury and, due to the severity and extensiveness of their injury, were admitted to the Intensive Care Unit (ICU). Using the information recorded during the in-hospital stay, we select a sample of patients who spent at least 1 day in the ICU and were mechanically ventilated. If patients did not receive this medical intervention, the counterfactual state would likely be death. Thus, it is reasonable to assume that their $QALY^{NON}$ would be equal to zero and all units of "benefit", $QALY_i^{CARE}$, would be gained from the medical intervention. Considering that patients have positive time preference we follow the recommendation by the National Institute for Health and Care Excellence and discount NHB to current values at a rate of 3.5 % per year (Whitehead and Ali, 2010).

To express NHB in monetary terms we compute a *Net Monetary Benefit* (hereafter: NMB_i) including the direct costs. For each QALY gained from treatment, Huang *et al.* (2018) assume that an individual is willing to pay approximately A\$67,000[12] for a sustained health improvement. NMB_i is then defined by

$$NMB_i = NHB_i * WTP^{QALY} - D_i \tag{4.3}$$

with the patient's Willingness to Pay for each QALY, WTP^{QALY}, gained from treatment.

Return to work

Health-related quality of life may not capture all of the utility related to paid work, thus we consider the return to work (RTW) as an alternative outcome for prediction.[13] Individuals who are unable to return to work after an injury experience greater physical difficulties and poorer mental health

[11] The life expectancy tables can be found here: https://www.aihw.gov.au/reports/life-expectancy-death/deaths-in-australia/contents/age-at-death [Accessed 14.04.2020]. The data were collected from the AIHW 17 Jul 2019 report. We count the expected years of life subject to patient's age and gender. As the patient's age was recorded at the time of injury, we adjust it by adding two years in line with the time when the follow up study was conducted.

[12] The selected willingness to pay for one QALY is a high value in comparison with other international estimates. However, in this paper we are primarily interested in the relative contribution of risk factors, thus any change in the constant monetary value for QALY does not have an impact on the interpretation of the estimation results.

[13] To date the most common statistical approaches to predict the return to work have been logistic regression and Cox proportional hazard models (see, e.g., Ip *et al.*, 1995; Kong *et al.*, 2011; Nielsen *et al.*, 2010; Van Patten *et al.*, 2016).

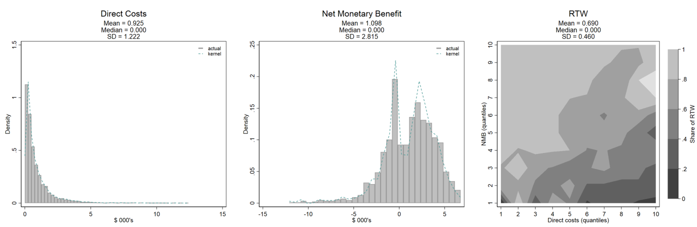
(Hoffman *et al.*, 2007; Iles *et al.*, 2008), thus understanding the main barriers resuming employment could potentially mitigate patients' economic losses in the long run. As the most prevalent group among road-traffic injuries are of working age who have many years to participate in the labour market, RTW is a useful indicator of the potential health and well-being consequences of road injuries.

We model a binary response outcome - patient's likelihood to return to work within one-year of the injury. This information was collected in the questionnaire in the follow up study subject to the condition that patient worked (for income) before the injury. The working population in this study accounts for around 70% of all road-traffic injuries.

Observed Outcomes

Figure 4.1 presents all outcomes considered in this study. *Direct Costs* features a non-normal and positively skewed distribution and a particularly heavy tail. This is typical of healthcare spending where the vast majority of patients exhibit low costs and a few have extremely high costs more than 10 times higher than the population average. The middle panel illustrates the distribution of *Net Monetary Benefit*. Recall, that this outcome is estimated for a sub-sample of patients who were admitted to the ICU.[14] Similar to *Direct Costs*, this distribution is non-normal. With a significant proportion of patients having a positive net benefit from treatment, the outcome is skewed left and has a lower kurtosis than the *Direct Costs*. Patients in the left tail either died soon after the discharge or reported very poor outcomes are classified by the EQ5D as "worse than death". The distribution has several modes that adds an additional complexity to the modelling.

FIGURE 4.1: The Presentation of Prediction Outcomes



NOTE.— Figure presents the empirical distribution of *Direct costs* of injury on the left panel, the distribution of *Net Monetary Benefit* in the middle panel (both expressed in 100 thousands AU$) and the interaction of the former and the latter with *Return to Work* on the right panel. *Return to Work* is averaged over five quantiles.

The right panel of Figure 4.1 illustrates the variation in the binary outcome, *RTW*, with respect to the *Direct Costs* (presented on the *x axis*)

[14] This sub-sample has slight differences in estimated *Direct costs*. While the distributional properties and extreme values of treatment costs are statistically similar, the costs of care for these patients were on average higher due to an expensive treatment at the ICU.

and *Net Monetary Benefit* (presented on the *y axis*) using a sub-sample of patients who worked prior to the injury. Patients who were less likely to return to work within one-year of injury are presented by the darker shaded area while those who were more likely to return to work by the brighter areas. Figure 4.1 shows that in most cases patients who return to work have higher net benefits of treatment irrespective of their treatment costs suggesting that gains in QALYs outweigh higher costs. These patients presumably recover well or have sufficient support to return to work and other social activities. Patients with a more significant and long-lasting disability have the lowest probability of returning to work with low utility and high costs. However, some patients are situated away from this pattern and while they return to work, they still have comparatively low utility and high costs. Predicting the pattern for these groups is the challenge in this paper.

4.2 Prediction methods

To perform the predictions we utilise an ensemble ML framework – the Super Learner algorithm. The Super Learner utilises various selected algorithms and builds a prediction function as a weighted combination of them. This makes the Super Learner a very versatile algorithm that often outperforms any single algorithm (van der Laan and Rose, 2011).

To allow for flexibility of the prediction function we consider both parametric as well as non-parametric statistical models. We set up the Super Learner based on the following menu of six prediction algorithms. For each algorithm we consider a comprehensive collection of predictors to find the best performing prediction function. We choose predictors based on risk factors associated with the severity of injury and the cost of routine treatments that lead to poor health and labour market outcomes. The full set of covariates before regularization contains various patient demographic characteristics (age, gender, residential region, SES quintiles); clinical treatment-related characteristics (Injury Severity Score [ISS], Glasgow Comma Scale [GCS], number of days in ICU, number of ventilated hours, number of commorbidities as well as the comorbidity index); injury-specific controls (injury group, mechanism, activity, cause and place); health related behavioural covariates (if alcohol/drug/substance use, if any mental issues, if mood or neurotic disorders); admitted hospital, year and month fixed effects and a large set of binary main diagnosis variables.[15] To construct the Super Learner, we first employ a regression model.[16] Due to the large number of intercorrelated covariates, least squares estimators might suffer from high variance and over prediction. We supplement our menu with several other algorithms based on regularisation methods. There are two types of regular-

[15]The full set of covariates is shown in Table A.1.

[16]We employ linear regression models when modelling the continuous outcomes: *Direct Costs* and the *Net Monetary Benefit*; and logistic regression models when modelling the binary response outcome - *RTW*.

ization methods: *L1-regularization* augments the OLS loss function with a tuning parameter for all non-zero coefficients that penalizes the sum of coefficients' absolute values, whereas *L2-regularization* introduces a penalty for the sum of squared coefficients. While a high L2-penalty shrinks covariates towards zero, a very high L1-penalty sets them to be zero and in this way drops the covariate from the best fitting model (Tibshirani, 1996; Tikhonov *et al.*, 2013; Tikhonov and Arsenin, 1977; Zou and Hastie, 2005). We add two regularisation algorithms: lasso which is a penalized regression with a tuning parameter λ chosen via an internal 10-fold cross validation; and an elastic net regularized regression method with α and λ values selected via an internal 10-fold cross validation. In addition, we consider unpenalized lasso regression, that is a linear regression with a sub-set of covariates that are selected in the first step using L1-regularization.

To better map non-linear relationships of the predicted outcomes we supplement our menu with several non-parametric statistical models. We first set up a tree-structured model, a common machine-learning data structuring approach, that can be visualised as a tree-like diagram. A Decision Tree splits the data into a set of subsamples (tree branches) by a given characteristic (predictor) that minimise the sum of squared residuals and best predict the outcome. Each branch could either lead to another subtree or have a leaf/terminal node with an assigned decision label, that is the predicted outcome. By applying this data splitting method nearly each observation in the data can be assigned to a different tree branch, but such procedure would lead to overfitting. To boost the accuracy and stability of the tree, the tree is pruned by setting constraints on the model parameters (Biggs *et al.*, 1991; Breiman, 2001; Breiman *et al.*, 1984; Maimon and Lior, 2014; Mola, 1998; Scornet *et al.*, 2014). We set a constraint of at least 50 observations in the terminal node and estimate a decision tree model. Additionally, we supplement the analysis by introducing a Random Forest, an ensemble learning method that is based on growing multiple decision trees (Maimon and Lior, 2014). Random forest averages a number of decision trees over many subsamples using a bootstrapping method. Within each bootstrap sample, the algorithm employs a random number of predictors to decide each split in each tree and, thus, reduce the correlation between samples. Similarly, we constrain each tree to have at least 50 observations in the terminal node and grow 500 random trees to estimate bootstrapped standard errors.

Employing this diverse set of algorithms, we follow the strategy outlined in Rose *et al.* (2017) and specify the Super Learner algorithm as follows:

$$\Psi(P_0) = \alpha_1 \hat{\psi}_{reg} + \alpha_2 \hat{\psi}_{L1reg} + \alpha_3 \hat{\psi}_{lasso} + \alpha_4 \hat{\psi}_{enet} + \alpha_5 \hat{\psi}_{tree} + \alpha_6 \hat{\psi}_{forest} + \epsilon \quad (4.4)$$

and estimate it using least squares method.

4.3 Performance evaluation

Metrics

We employ several statistical metrics to evaluate the performance of each algorithm including the Super Learner. When modelling the continuous outcomes D_i and NMB_i we estimate the coefficient of determination, R^2, that evaluates the proportion of the variance explained by the selected set of covariates, and the mean squared error (MSE) measuring the prediction error (Wooldridge, 2020). These metrics are defined as follows:

$$MSE(\psi^j) = \frac{1}{J}\sum_j (y_i - \hat{\psi}_i^j)^2 \tag{4.5}$$

$$R^2(\psi^j) = \frac{\sum_{i,j}(y_i - \hat{\psi}_i^j)^2}{\sum_i (y_i - \bar{y})^2} \tag{4.6}$$

for each outcome y of patient i predicted by algorithm j. These metrics are evaluated based on the cross validation re-sampling procedure outlined below.

The prediction of a qualitative response, such as a binary outcome RTW, is known as a classification exercise in the ML literature. To evaluate the performance of the classification we set up a confusion matrix that represents a number of true positives (TP), a number of false positive (FP), a number of true negatives (TN) and a number of false negatives (FN) as outlined in the matrix below:

TABLE 4.1: Confusion matrix

		Observed	
		$Positive(1)$	$Negative(0)$
$Classified$	$Positive(1)$	TP	FP
	$Negative(0)$	FN	TN

Based on the confusion matrix we estimate two accuracy measures for a binary classifier: the *Sensitivity* - a true positive rate that is a rate of correctly classified positive outcomes and the *Specificity* - a true negative rate or a rate of correctly classified negative outcomes defined as the following

$$Sensitivity = \frac{TP}{TP+FN} \tag{4.7}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4.8}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.9}$$

We plot a *receiver-operating-characteristics* (ROC) curve that evaluates the performance of the classifier summarized over various classifying thresholds. Due to the unbalanced structure of our selected binary response outcome we chose a set of classifying thresholds that fluctuates around the mean value of the outcome and calculate the accuracy measures using each of these thresholds. The ROC curve plots *Sensitivity* on the vertical axis against $(1 - Specifity)$ on the horizontal axis and represents the overall performance of the prediction by the area under the ROC curve. The larger the size of the area, the better the prediction performance. Using this, we classify the final prediction using a threshold with the largest area under the ROC curve. Additionally, we consider the *Accuracy*, that describes the prediction accuracy in percentages (James *et al.*, 2013; Maimon and Lior, 2014).

Cross-Validation

We perform the following procedure to validate the prediction performance. First, we randomly divide the sample into two parts: *a training sample* that is used to fit each of the algorithms and *a validation sample* used to predict and validate the predictions. This re-sampling procedure is based on 60:40 % split. Second, we implement 10-fold cross-validation when fitting regularized regressions and non-parametric models. We partition the training sample into smaller training and validation sets and repeat the process 10 times (folds), with each of the randomly selected validation sample used only once to evaluate the prediction. The results from all validation folds are then averaged (James *et al.*, 2013; van der Laan and Duboit, 2003).

All results in this paper are presented for the validation sample.

5 Prediction results

In this section we discuss prediction results using the set of algorithms outlined in Section 4. We first employ the training sample via 10-fold cross-validation to fit each single algorithm and then obtain predictions using a leave-out sample to validate the prediction performance. Results for both continuous outcomes are shown in Figure 5.2 and Figure 5.3. The upper panel illustrates the distribution of predicted values, whereas the bottom panel reports statistical measures for goodness of fit to evaluate the performance of each algorithm.

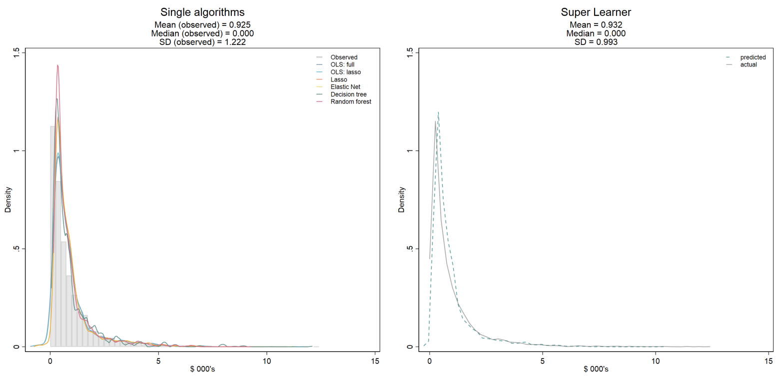
As shown in the upper left panel of Figure 5.2, all single algorithms have captured the positive skewness of the cost data, but several of them (particularly the OLS regression with a full set of covariates) (mis)predict negative values for patients with very low treatment costs. Only non-parametric models such as the Regression Tree and the Random Forest perform better in this particular feature by predicting only positive values.[17] However, only the Regression tree performs well when describing the long tail of the distribution that represents patients with very high costs. In the case of the Random Forest, a poor prediction of high costs comes at the expense of higher predicted levels of low costs patients visualized by a spike at low values. The bottom panel of Figure 5.2 reports statistical metrics as defined in (4.5) and (4.6) to evaluate the overall performance. Based on these measures, all single algorithms, except for the Regression Tree, perform similarly with an R^2 value equal to 0.64 and a MSE of 0.52. The Regression tree, while performing better in predicting the tail, failed to accurately predict low costs patients leading to a significantly lower R^2 value of 0.54 and a higher MSE of 0.68. The Elastic Net is the best performing single algorithm with R^2 value equal to 0.65 and a MSE of 0.52. The Super Learner algorithm has remarkably outperformed all single algorithms considered in this paper with predictions shown in the upper right panel of the Figure 5.2 with R^2 value equal to 0.79 and MSE equal to 0.49.

A more perceptible difference in the performance of single algorithms is noted in Figure 5.3 that presents the prediction results for *Net Monetary Benefit*. Here parametric models perform better in quantifying the left tail of the distribution. However, particularly the Lasso regression and the Elastic Net predict significantly higher levels of patients with low positive net benefits ranging from 0 to around \$4000. Non-parametric models are better of predicting the bimodal shape of the distribution. Based on the goodness of fit measures reported in the bottom panel, the Random Forest is the best performing single algorithm with a R^2 value equal to 0.45 and a MSE of 4.39. A similar performance is noted for the Lasso regression (R^2: 0.44 and MSE: 4.46) and the Elastic Net (R^2: 0.44 and MSE: 4.45). Similarly as in the case of *Direct Costs* the Super Learner had best overall performance in predicting the distribution of the *Net Monetary Benefit* with the highest R^2 equal to 0.54 and the lowest MSE value equal to 4.18. For more detailed evaluation of each algorithm, refer to the additional evidence provided in the Appendix.[18]

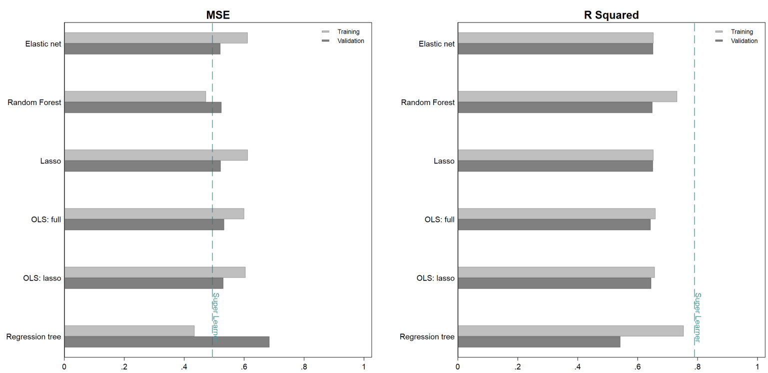
[17] Descriptive statistics of all predictions can be found in Table A.2 in the Appendix.

[18] Additional graphical evidence for the performance of each single algorithm as well as the Super Learner is presented in the Appendix. We show empirical distributions of predicted values in Figure A.1 and Figure A.2; quantile-quantile scatter plots that plots the quantiles of predicted values against the quantiles of the observed values in Figure A.4 and Figure A.5 and prediction errors for each observation in the scatter plots presented in Figure A.6 and Figure A.7.

FIGURE 5.2: Outcome: *Direct Costs*



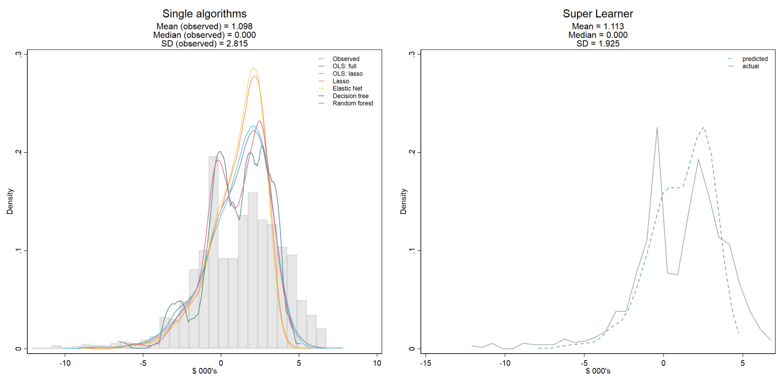
The distribution of observed outcomes and predictions



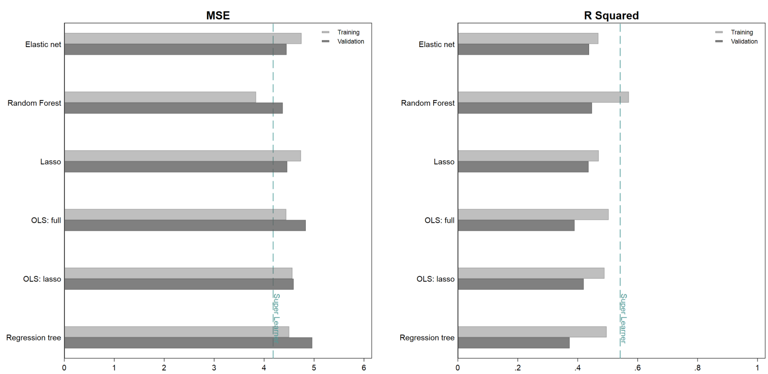
The Goodness of fit: Mean Squared Error and R^2

NOTE.— Figure presents the prediction results for the outcome *Direct Costs*: the empirical distributions (restricted to the validation sample only) of predicted values in the upper panel and statistical measures on Goodness of fit in the bottom panel. Both MSE and R^2 statistical measures to evaluate the Super Learner algorithm are estimated using a reduced form model and the validation sample.

FIGURE 5.3: Outcome: *Net Monetary Benefit*



The distribution of observed outcomes and predictions



The Goodness of fit: Mean Squared Error and R^2

NOTE.— Figure presents the prediction results for the outcome *Net Monetary Benefit*: the empirical distributions (restricted to the validation sample only) of predicted values in the upper panel and statistical measures on Goodness of fit in the bottom panel. Both MSE and R^2 statistical measures to evaluate the Super Learner algorithm are estimated using a reduced form model and the validation sample. Analysis sample is restricted to patients admitted to the ICU.

Table 5.2 presents estimated contributions of each single algorithm as specified in (4.4). From the left to the right each column reports results from a full (indicated by *I*) and a reduced (indicated by *II*) form of the Super Learner for each outcome discussed in Section 4.1. The reduced form model is an equivalent to a model specified in (4.4) with a preceding step of selecting a subset of potential predictors using the L1-regularization method. Recall, that the specification of the Lasso regression is conceptually similar to a linear regression, but, in contrast to the traditional least-squares estimation, it augments the loss function and introduces a penalty for model parameters. Moreover, the Elastic Net is a related technique that has a flexibility to generate close to zero coefficients along with a variable selection when zero-valued coefficients are eliminated from the model. Thus, in practice, these models often make statistically similar predictions when selected tuning parameters for optimum performance cause one algorithm to resemble the other. Similarly as depicted in Figure 5.2 and Figure 5.3, we observe a strong collinear relationship[19] between these algorithms in Table 5.2 and, as a result, we perform a reduced form of the Super Learner with selected predictors via L1-regularization to avoid over-fitting the data.

Table 5.2: Estimation of the Super Learner

	(1) Direct costs	(2) Direct costs	(3) NMB	(4) NMB	(5) RTW	(6) RTW
	I	II	I	II	I	II
Lasso	-1.095*		-0.752		27.12***	
	(-2.31)		(-1.64)		(6.80)	
Elastic net	1.024*		0.872	0.182	-48.21***	
	(2.08)		(1.80)	(0.88)	(-11.61)	
OLS/Logit: full	0.205	0.214	-0.213		2.086*	
	(1.50)	(1.57)	(-1.59)		(1.97)	
OLS/Logit: lasso	0.331	0.267	0.548*	0.278	8.219***	
	(1.76)	(1.92)	(2.51)	(1.75)	(5.89)	
Decision tree	0.0665**	0.0736**	0.0465	0.0583	-0.491	
	(2.70)	(3.02)	(0.68)	(0.86)	(1.75)	
Random forest	0.475***	0.449***	0.5568***	0.542***	11.88***	1.464***
	(10.44)	(11.36)	(5.94)	(5.72)	(19.55)	(28.01)
Observations	4650	4650	1379	1379	2948	2948
R^2	0.79	0.79	0.54	0.54		
MSE	0.494	0.494	4.170	4.179		
Accuracy					0.778	0.759

Note.— The predictions of *Direct Costs* and *Net Monetary Benefit* are based on a linear specification of lasso, Elastic Net and OLS regressions, while the predictions of *RTW* are estimated using a logistic regression specification. All estimations performed on validation sample. Columns *I* report the full specification, while Columns *II* report estimation results from reduced form models with a preceding step of the L-1 regularization to select the predictors. Models with *RTW* outcome restricted to a subsample of patient who worked prior to the injury and models with *Net Monetary Benefit* are restricted to patients who were admitted to the ICU during their hospital stay. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

With respect to the outcome *Direct Costs* reported in Columns (1) and (2), we note that the reduced form model does not significantly alter the estimated contributions and does not affect the overall prediction performance with a steady R^2 value equal to 0.79 and a MSE of 0.49. The contribution of the Random Forest algorithm is the highest in magnitude with an estimate

[19]Further evidence on the collinearities between the predicted values are shown in the correlation matrixes presented in Figure A.9, Figure A.10, Figure A.11 and in the descriptive statistics of predictions reported in Table A.2 in the Appendix.

of 0.45, following by the OLS specification with lasso selected covariates and the OLS specification with a full set of covariates, respectively. The lowest contribution is estimated for the Regression Tree. Next, Columns (3) and (4) show estimation results on the outcome *Net Monetary Benefit* using a subsample of patients admitted to the ICU units. Unlike in the case of *Direct Costs*, in addition to the Lasso regression the prior L1-regularization suggests to eliminate the OLS full specification. It noticeably reduces the contribution of the Elastic Net and the OLS specification with lasso selected covariates, but only slightly alters the contribution of the Random forest, that is also the highest in magnitude.

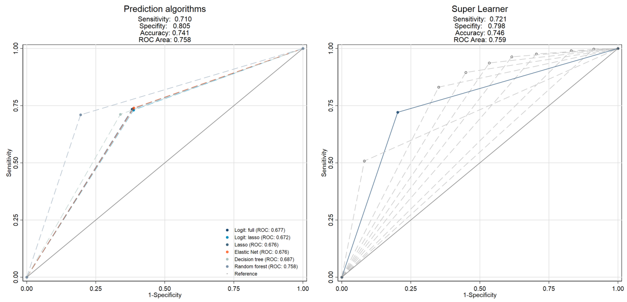
Lastly, Columns (5) and (6) in the Table 5.2 outline the Super Learner specification for classifying the binary response outcome *RTW*. Similarly, as in the case of previously discussed outcomes, the Super Learner is likely affected by high collinearities between single algorithms that is reflected by reversed signs of algorithms' contributions. It is expected, that due to high positive correlation, as in the case of the parametric models such as the logistic regression, Lasso and the Elastic Net, one algorithm withdraws the contribution from the other. Thus, using similar techniques, we perform L1-regularization and re-estimate the Super Learner logistic regression in the reduced form. L1-regularization here suggests that selecting the Random Forest with a contribution of 1.464 leads to the best classifying model. This result goes in line with the existing literature on the performance of the Random Forest in the classification that has been demonstrated to often have improved prediction accuracy in comparison to other supervised learning methods (Breiman, 2001; Kuhn and Johnson, 2016; Svetnik *et al.*, 2003).

Additional evidence on the classification performance by each single algorithm as well as the Super Learner is presented in Figure 5.4. Recall, that the ROC curve plots the *Sensitivity* on the y-axis against (*1-Specificity*) on the x-axis and the main measure of interest is the calculated area under the ROC curve. The larger the area, the better overall classification performance. The left panel of Figure 5.4 reports these measures for each single algorithm with a selected classification rule that leads to the largest area under the ROC curve, accordingly.[20] These results, again, demonstrate that the Random Forest algorithm outperforms other single algorithms and provide further support to the L1-regularization performed in the reduced form specification. Only a slight improvement in the classification performance is noted by the Super Learner specification. In comparison to the Random Forest the estimated area under the ROC curve is nearly the same as in the case of the Super Learner; however, the Accuracy reported in the panel

[20]To find the best fit for each single algorithm we perform a classification with a number of different thresholds fluctuating around the mean of the outcome. For more detailed evidence on the performance using various selected classification rules refer to Figure A.8 in the Appendix.

header shows that the Super Learner has a slightly better prediction performance with an accuracy rate of approximately 75%. For more detailed information on the empirical distribution of predicted probabilities, refer to Figure A.3 in the Appendix.

FIGURE 5.4: The ROC curves



NOTE.— Figure presents the Goodness of Fit measures for the prediction of *RTW*. In the left panel each line indicates the Specificity/Sensitivity measures for each single algorithm using a classification rule that leads to the largest area under the ROC curve. The panel header reports the statistical metrics in detail for the best performing single algorithm. The right panel presents the prediction results of the Super Learner using several selected thresholds that fluctuate around the mean value of the outcome. The blue line indicates the metrics using the classification rule with the best performance and the panel header reports its statistical evaluation metrics. The results are shown on the validation sample and are further restricted to a sub-sample of patients who worked prior to the injury

6 Prediction errors by different groups

We next study the extent of prediction errors from the Super Learner. Using the reduced form specification presented in Table 5.2 in Section 5, we analyse the differences between the observed and predicted outcomes for different groups of patients. This provides a better understanding who are at risk for high costs and worse outcomes and may form a group of interest for the payer. Figure 6.5 and Figure 6.6 provide graphical illustration for outcomes *Direct Costs* and *Net Monetary Benefit*, respectively.

The upper panel of Figure 6.5 reports average injury costs by selected injury & treatment-related characteristics and their corresponding average prediction error. Treatment costs are on average higher at the MTS, that is expected as this type of hospitals offers the highest level of trauma care in Victoria and, in most cases, treat very severely injured patients. A vast majority of patients (around 90%) are treated at the MTS with a considerable variation of type of injuries as well as patients characteristics providing enough evidence to make accurate predictions. The Super Learner demonstrates a high performance in predicting treatment costs for MTS, but exhibit a positive prediction error for patients treated in hospitals offering lower levels of trauma care, often located in more regional and rural areas. Regional variations in patients' clinical and socio-demographic characteristics as well as the quality of care provided are likely determinants of such differences in prediction errors. A similar pattern we observe when looking at the type of residential location and the socioeconomic status shown in the bottom panel of Figure 6.5. While costs of care are on average lower for patients living in lower socioeconomic status households and in metropolitan areas, the prediction error is much the same as for patients residing in more remote areas and does not exceed an overprediction of AU$ 2,000. More significant errors we note for patients living in higher socioeconomic status households and for those who were injured in Victoria but permanently residing outside the state and likely have more unobserved characteristics in the registry.

One of the most complex groups of patients to predict costs of care is patients who experience spinal cord injury, the most expensive road traffic injury. The variation in treatment costs for these patients is substantial, with some patients having very high costs and others – significantly lower. This is partially driven by lower chances of survival during hospital treatment as well as after discharge that causes total treatment costs to be lower than the algorithm predicts. In addition, age is also a significant factor as older individuals often have lower costs because of their lower chances of successful recovery that often result in assisted living without long-lasting and expensive rehabilitation services. For similar reasons the prediction error is high in absolute terms for patients with an isolated head injury. However, in this case, the prediction error is negative. With additional and more detailed clinical information about the extent and severity of the injury these errors could be addressed in the risk adjustment. This result signals the importance of discussed characteristics in the prediction of treatment costs, in particular for the youngest and the oldest groups of patients with the most severe injuries such as the spinal cord injury and those who are treated and reside in more rural and remote areas.

A comparison of averages in *Net Monetary Benefit* and it's respective prediction errors are presented in Figure 6.6 and tells a similar story; patients with spinal cord injuries have the highest benefits from treatment, but is one of the most complex groups of patients to obtain accurate predictions of. In addition, residents living in regional and remote areas as well as outside Victoria have higher prediction error. Unlike in the case of *Direct Costs*, the error is mostly positive, meaning that the algorithm predicts greater benefits from treatment than they are observed. It is an important result for the payer informing about target groups of patients such as residents of regional and remote areas. These patients are more cost-burdened and, in addition, have lower observed benefits from treatment than residents living in metropolitan areas. The benefits from treatment are often overpredicted, thus it is important to consider this group of patients when applying recoupment adjustment to hospital payments. Moreover, it is worth noting, that the patient group with the lowest and negative benefits from treatment are patients with chest and abdominal injuries, suggesting the urgency to target this group for potential improvements in their recovery process. These results, again, illustrate a significant variation of costs and benefits among different groups of patients defined by their type of injury and residential location. However, considering that only a moderate proportion (54%) of variance in the *Net Monetary Benefit* is explained using the Super Learner, we acknowledge the need for further observable characteristics to improve the accuracy of the prediction.

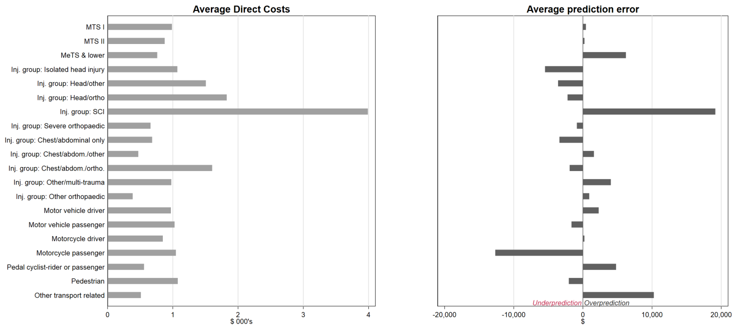
Similar to the error analysis of the continuous outcomes, Table 6.3 reports details about the classification errors made by the Super Learner when classifying *RTW*. Based on the classification error rate, the algorithm performs better in classifying a negative outcome. From a total of 944 patients who did not return to work, the algorithm correctly classified 753, implying a misclassification of every 5th patient. From a total of 2004 patients who returned to work, the algorithm suggested 1445 positive outcomes with an approximate error rate of 28%. The Super Learner appears to be more sensitive to the prediction of positive outcome that is more common for these patients. Two out of three patients successfully return to work after the injury, but the prediction algorithm suggests a slightly lower success rate. Among the group of misclassified patients approximately two-thirds reside in metropolitan areas and are the youngest group of patients aged 15-24 years (results not reported in the table). More commonly, the misclassification of a negative outcome is made for patients with severe and moderate orthopaedic injuries, while a positive outcome is more incorrectly specified for patients with head, chest and abdominal injuries. One possible reason for this result could be that treating patients with orthopaedic injuries require longer rehabilitation care, even though they are not among the most costly patient groups. On the other hand, patients with head, chest and abdominal injuries are one of the most expensive injuries, suggesting a higher severity of a clinical case and poorer outcomes than otherwise expected. Interestingly, we do not observe any differences in misclassification across socioeconomic status.

TABLE 6.3: Prediction errors in *RTW*

	(1) Observed: 0	(2) Observed: 1	(3) Total
Classified: 0	753	559	1312
Classified: 1	191	1445	1636
Total	944	2004	2948
Error rate	**0.20**	**0.28**	**0.25**

NOTE.— Table presents the classification errors made by the Super Learner algorithm. Error rate denotes a share of misclassified outcomes. Estimation performed on validation sample and restricted to a subsample of patient who worked prior to the injury.

Figure 6.5: Prediction errors in *Direct Costs*

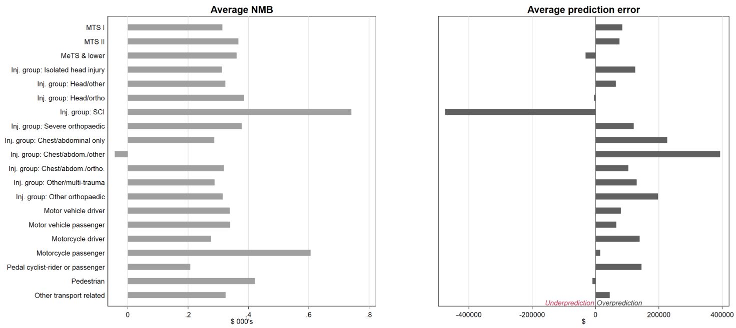


by injury & treatment-related characteristics

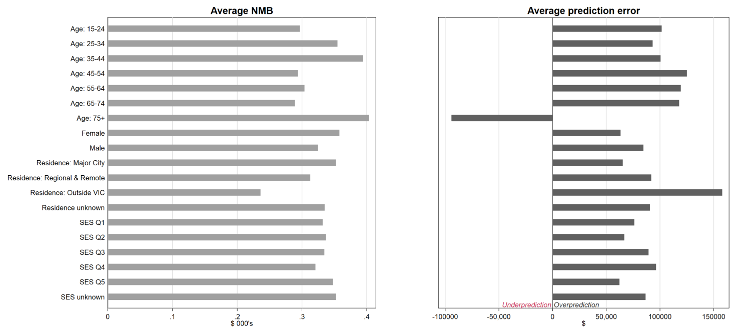
by patient characteristics

Note.— Figure presents average prediction errors by selected groups of patients that are made by the Super Learner algorithm in the prediction of *Direct Costs*. Here MTS refers to the Major Trauma Services - the highest level of trauma care in Victoria. Patients considered in this analysis were admitted to two different MTS, that were de-identified using indicator I and II. MeTS refers to Metropolitan Trauma services; lower levels of care include regional trauma services and rural healthcare services. SES refers to the Social Economic Status as defined buy the Index of Relative Socio-Economic Advantage and Disadvantage. All prediction errors are reported on the validation sample.

Figure 6.6: Prediction errors in *Net Monetary Benefit*



by injury & treatment-related characteristics



by patient characteristics

Note.— Figure presents average prediction errors by selected groups of patients that are made by the Super Learner algorithm in the prediction of *Net Monetary Benefit*. Here MTS refers to the Major Trauma Services - the highest level of trauma care in Victoria. Patients considered in this analysis were admitted to two different MTS, that were de-identified using indicator I and II. MeTS refers to Metropolitan Trauma services; lower levels of care include regional trauma services and rural healthcare services. SES refers to the Social Economic Status as defined buy the Index of Relative Socio-Economic Advantage and Disadvantage. All prediction errors are reported on the validation sample. Analysis sample is restricted to patients admitted to the ICU.

7 Conclusion

In this paper we employ supervised machine learning algorithms to construct a powerful prediction model for healthcare costs and patient outcomes in the context of road traffic injuries. We employ a comprehensive patient-level dataset of Victorian State Trauma Registry that records all major trauma patients in Victoria. We link this dataset to detailed insurance claims data provided by the Transport Accident Commission and compute healthcare costs for each patient who suffered a major trauma in a road-traffic related injury. In addition to predicting healthcare costs in traditional provider reimbursement frameworks, we consider the societal value of health and work in recovery from injury that provide a better understanding of the quality of healthcare provider. First, we estimate the net monetary benefit gained from treatment that relies on the concept of Quality Adjusted Life Years used in the cost-effectiveness literature (Stinnett and Mullahy, 1998). Second, as the paid employment is an important factor for patient well-being we predict patient's probability to return to work after suffering from a major trauma and inform about potential losses in labour market. We utilise both parametric and non-parametric statistical models to construct an ensemble machine learning framework - the Super Learner - and predict the economic consequences of road traffic injury: *Direct Costs*, *Net Monetary Benefit* and *Return to Work*.

Our findings demonstrate that the Super Learner is effective and performs well in predicting all outcomes considered in this paper. In addition to high overall performance in predicting outcomes for patients with a mild and a moderate severity of an injury, it performs well in describing patients with uncommon characteristics and is able to classify patients with the highest healthcare costs and the lowest net benefits gained from treatment. The algorithm only slightly outperforms the Random Forest prediction of a binary response outcome that is often referred to as the best performing classification algorithm in the machine learning literature. We extend our prediction analysis by examining in detail the Super Learner's performance by different groups of patients. This analysis reveals further information about sensitive groups and has a strong relevance for future budget planning and reimbursement for healthcare providers. Injury groups such as a spinal cord injury and chest and abdominal injuries are among the most complex groups to get an accurate prediction of potential future costs. This indicates a need of particularly detailed information about the treatment of these patients to ensure an adequate remuneration. Average cost and net benefits from treatment vary widely across injury types and patient characteristics but in a way that is largely predictable. The algorithms used here predict over half of the variation in cost and net benefits suggesting that adjustment to capitation or prospective payments are feasible. However, we acknowledge the limitation of our study when estimating net benefits due to the lower response rate of the follow up study (around 70%). This may slightly alter our results, particularly from the comparison by different groups. We leave this to investigate for future research using the data with higher response rates.

Machine learning offers powerful tools to predict patient healthcare costs and with a comprehensive set of controls considered in this paper explained nearly 80% of the variation. In addition to accurately predicted costs, these methods had a considerable performance in predicting patient outcomes. This sheds light on the use of future healthcare services and the quality of healthcare providers and provides a crucial information for the payer in designing contracts for healthcare providers. How this kind of information can be incorporated in practice and can structure provider incentives to supply care of a chosen quality and price is an important part of the research agenda in value-based healthcare.

References

Arandjelović, O. (2015). Prediction of health outcomes using big (health) data. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2543–2546.

Beck, B., Bray, J., Cameron, P., Cooper, D. and Gabbe, B. (2016). Trends in severe traumatic brain injury in Victoria, 2006–2014. *The Medical Journal of Australia*, **204**, e1–e6.

Bertsimas, D., Bjarnadóttir, M., Kane, M., Kryder, J., Pandey, R., Vempala, S. and Wang, G. (2008). Algorithmic prediction of health care costs and discovery of medical knowledge. *Operations Research*, **56** (6), 1382–1392.

Biggs, D., Ville, B. D. and Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, **18** (1), 49–62.

Breiman, L. (2001). Random forests. *Machine Learning*, **45**, 5–32.

—, Friedman, J., Stone, C. and Olshen, R. (1984). *Classification and Regression Trees*. The Wadsworth and Brooks/Cole Statistics: Probability Series, Taylor & Francis.

Burnham, J. P., Lu, C., Yaeger, L. H., Bailey, T. C. and Kollef, M. H. (2018). Using wearable technology to predict health outcomes: a literature review. *Journal of the American Medical Informatics Association*, **25** (9), 1221–1227.

Chu, C.-W. and Zhang, G. P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics*, **86** (3), 217–231.

Cucciare, M. and o'Donohue, W. (2006). Predicting Future Healthcare Costs: How Well Does Risk-Adjustment Work? *Journal of Health Organization and Management*, **20**, 150–162.

Curtis, K., Lam, M., Mitchell, R., Black, D., Taylor, C., Dickson, C., Jan, S., Palmer, C. S., Langcake, M. and Myburgh, J. (2014). Acute costs and predictors of higher treatment costs of trauma in New South Wales, Australia. *Injury*, **45** (1), 279–284.

Deb, P. and Burgess, J. (2003). A quasi-experimental comparison of econometric models for health care expenditures. *Economics Working Paper Archive at Hunter College*, **212**.

Department of Health and Human Services (Feb 2014). *Trauma towards 2014 - Review and future directions of the Victorian State Trauma System*. Tech. rep., Victoria State Government.

Department of Health and Human Services (Jul 2014). *Victorian State Trauma System and Registry Report 2014-15*. Tech. rep., Victoria State Government.

Derrett, S., Black, J. and Herbison, G. (2009). Outcome after injury – a systematic literature search of studies using the EQ-5D. *Journal of Trauma*, **67** (4), 883–890.

Dixon, J., Smith, P., Gravelle, H., Martin, S., Bardsley, M., Rice, N., Georghiou, T., Dusheiko, M., Billings, J., Lorenzo, M. D. and Sanderson, C. (2011). A person based formula for allocating commissioning funds to general practices in England: development of a statistical model. *BMJ*, **343**.

Dranove, D. (1987). Rate-setting by diagnosis related groups and hospital specialization. *RAND Journal of Economics*, **18** (3), 417–427.

Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, **78**, 605–610.

Eggleston, K. (2000). Risk selection and optimal health insurance-provider payment systems. *The Journal of Risk and Insurance*, **67** (2), 173–196.

Einav, L., Finkelstein, A., Kluender, R. and Schrimpf, P. (2016). Beyond Statistics: The Economic Content of Risk Scores. *American Economic Journal: Applied Economics*, **8** (2), 195–224.

Ellis, R. P., Martins, B. and Rose, S. (2018). Chapter 3 - risk adjustment for health plan payment. In T. G. McGuire and R. C. van Kleef (eds.), *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets*, Academic Press, pp. 55–104.

— and McGuire, T. G. (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*, **5** (2), 129–151.

— and — (1988). Insurance principles and the design of prospective payment systems. *Journal of Health Economics*, **7** (3), 215–237.

— and — (1990). Optimal payment systems for health services. *Journal of Health Economics*, **9** (4), 375–396.

— and — (1993). Supply-Side and Demand-Side Cost Sharing in Health Care. *Journal of Economic Perspectives*, **7** (4), 135–151.

Frank, R. G. and Lave, J. R. (1989). A comparison of hospital responses to reimbursement policies for medicaid psychiatric patients. *The RAND Journal of Economics*, **20** (4), 588–600.

Gabbe, B., Lyons, R., Fitzgerald, M., Judson, R., Richardson, J. and Cameron, P. (2014). Reduced population burden of road transport–related major trauma after introduction of an inclusive trauma system. *Annals of surgery*, **261**.

—, Simpson, P., Sutherland, A., Wolfe, R., Fitzgerald, M., Judson, R. and Cameron, P. (2012). Improved functional outcomes for major trauma patients in a regionalized, inclusive trauma system. *Annals of Surgery*, **255**, 1009–1015.

Gilleskie, D. and Mroz, T. (2004). A flexible approach for estimating the effects of covariates on health expenditures. *Journal of Health Economics*, **23**, 391–418.

Hoffman, B., Papas, R., Chatkoff, D. and Kerns, R. (2007). Meta-analysis of psychological interventions for chronic low back pain. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, **26** (1), 1–9.

Holbrook, T., Hoyt, D. and Anderson, J. (2001a). The impact of major in-hospital complications on functional outcome and quality of life after trauma. *The Journal of trauma*, **50**, 91–95.

—, —, Coimbra, R., Potenza, B., Sise, M. and Anderson, J. (2005). Long-term posttraumatic stress disorder persists after major trauma in adolescents: New data on risk factors and functional outcome. *The Journal of trauma*, **58**, 764–769.

—, —, Stein, M. and Sieber, W. (2001b). Perceived threat to life predicts posttraumatic stress disorder after major trauma: Risk factors and functional outcome. *The Journal of trauma*, **51**, 287–292.

Holtslag, H., Beeck, E., Lindeman, E. and Leenen, L. (2007). Determinants of long-term functional consequences after major trauma. *The Journal of trauma*, **62**.

Huang, L., Frijters, P., Dalziel, K. and Clarke, P. (2018). Life satisfaction, qalys, and the monetary value of health. *Social Science & Medicine*, **211**, 131–136.

Iezzoni, I. (2012). *Risk Adjustment for Measuring Healthcare Outcomes*. IL: Health Administration Press, 4th edn.

Iles, R., Davidson, M. and Taylor, N. (2008). Psychosocial predictors of failure to return to work in non-chronic non-specific low back pain: A systematic review. *Occupational and Environmental Medicine*, **65**, 507–517.

Ip, R. Y., Dornan, J. and Schentag, C. (1995). Traumatic brain injury: factors predicting return to work or school. *Brain Injury*, **9** (5), 517–532.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Jones, A. M. (2000). Health econometrics. In A. J. Culyer and J. P. Newhouse (eds.), *Handbook of Health Econometrics*, vol. 1, Elsevier: Amsterdam, pp. 265–344.

— (2011). Models for health care. In M. P. Clements and D. F. Hendry (eds.), *Oxford Handbook of Economic Forecasting*, Oxford: Oxford University Press, pp. 625–654.

— and Lomas, J. (2016). A quasi-monte-carlo comparison of parametric and semiparametric regression methods for heavy-tailed and non-normal data: an application to healthcare costs. *Journal of the Royal Statistical Society*, **179** (Part 4), 951–974.

—, — and Rice, N. (2014). Applying beta-type size distributions to healthcare cost regressions. *Journal of Applied Econometrics*, **29**, 649–670.

—, — and — (2015). Healthcare cost regressions: going beyond the mean to estimate the full distribution. *Health Economics*, **24**, 1192–1212.

Kan, H. J., Kharrazi, H., Chang, H.-Y., Bodycombe, D., Lemke, K. and Weiner, J. (2019). Exploring the use of machine learning for risk adjustment: A comparison of standard and penalized linear regression models in predicting health care costs in older adults. *PLOS ONE*, **14**, e0213258.

Kessler, R. C., Rose, S., Koenen, K. C., Karam, E. G., Stang, P. E., Stein, D. J., Heeringa, S. G., Hill, E. D., Liberzon, I., McLaughlin, K. A., McLean, S. A., Pennell, B. E., Petukhova, M., Rosellini, A. J., Ruscio, A. M., Shahly, V., Shalev, A. Y., Silove, D., Zaslavsky, A. M., Angermeyer, M. C., Bromet, E. J., de Almeida, J. M. C., de Girolamo, G., de Jonge, P., Demyttenaere, K., Florescu, S. E., Gureje, O., Haro, J. M., Hinkov, H., Kawakami, N., Kovess-Masfety, V., Lee, S., Medina-Mora, M. E., Murphy, S. D., Navarro-Mateu, F., Piazza, M., Posada-Villa, J., Scott, K., Torres, Y. and Carmen Viana, M. (2014). How well can post-traumatic stress disorder be predicted from pre-trauma risk factors? An exploratory study in the WHO World Mental Health Surveys. *World Psychiatry*, **13** (3), 265–274.

Kong, W., Tang, D., Xiaoyuan, L., Yu, I., Liang, Y. and He, Y. (2011). Prediction of return to work outcomes under an injured worker case management program. *Journal of Occupational Rehabilitation*, **22**, 230–240.

Kuhn, M. and Johnson, K. (2016). *Applied Predictive Modeling*. 5th edition, Springer Science+Business Media New York 2013.

Lahiri, B. (2014). *Predicting Healthcare Expenditure Increase for an Individual from Medicare Data*. Retrieved from: https://www.academia.edu/7836580/Predicting_Healthcare_Expenditure_Increase_for_an_Individual_from_Medicare_Data [accessed 21.04.2020].

Lave, J. R. (2003). Developing a medicare prospective payment system for inpatient psychiatric care. *Health Affairs*, **22** (5), 97–109.

Maimon, O. and Lior, R. (2014). *Data Mining With Decision Trees: Theory And Applications (2nd Edition)*. Series In Machine Perception And Artificial Intelligence, World Scientific Publishing Co. Pte. Ltd: Singapore.

Manning, W. G., Basu, A. and Mullahy, J. (2005). Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*, **24** (3), 465–488.

McClellan, M. (1997). Hospital reimbursement incentives: An empirical analysis. *Journal of Economics & Management Strategy*, **6** (1), 91–128.

McDonald, J. B., Sorensen, J. and Turley, P. A. (2013). Skewness and kurtosis properties of income distribution models. *Review of Income and Wealth*, **59** (2), 360–374.

Mola, F. (1998). Classification and regression trees software and new developments. In A. Rizzi, M. Vichi and H.-H. Bock (eds.), *Advances in Data Science and Classification*, Springer Berlin Heidelberg, pp. 311–318.

Mullahy, J. (2009). Econometric modeling of health care costs and expenditures: A survey of analytical issues and related policy considerations. *Medical Care*, **47** (7), S104–S108.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, **31**, 87–106.

Newhouse, J., Buntin, M. and Chapman, J. (1997). Risk adjustment and medicare: Taking a closer look. *Health affairs (Project Hope)*, **16**, 26–43.

NEWHOUSE, J. P. (1996). Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature*, **34** (3), 1236–1263.

NIELSEN, M. B., MADSEN, I. E. H., BÜLTMANN, U., CHRISTENSEN, U., DIDERICHSEN, F. and RUGULIES, R. (2010). Predictors of return to work in employees sick-listed with mental health problems: Findings from a longitudinal study. *European Journal of Public Health*, **21**, 806–811.

PARK, S. and BASU, A. (2018). Alternative evaluation metrics for risk adjustment methods. *Health Economics*, **27** (6), 984–1010.

PIRRACCHIO, R., PETERSEN, M. L., CARONE, M., RIGON, M. R., CHEVRET, S. and VAN DER LAAN, M. J. (2015). Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *The Lancet. Respiratory Medicine*, **3** (1), 42–52.

PYRKOV, T. V., SLIPENSKY, K., BARG, M., KONDRASHIN, A., ZHUROV, B., ZENIN, A., PYATNITSKIY, M., MENSHIKOV, L., MARKOV, S. and FEDICHEV, P. O. (2018). Extracting biological age from biomedical data via deep learning: too much of a good thing? *Scientific Reports*, **8** (5210).

ROSE, S. (2016). A machine learning framework for plan payment risk adjustment. *Health Services Research*, **51**, 2358–2374.

—, BERGQUIST, S. L. and LAYTON, T. J. (2017). Computational health economics for identification of unprofitable health care enrollees. *Biostatistics*, **18** (4), 682–694.

SCORNET, E., BIAU, G. and VERT, J.-P. (2014). Consistency of random forests. *The Annals of Statistics*, **43**.

SHRESTHA, A., BERGQUIST, S., MONTZ, E. and ROSE, S. (2018). Mental health risk adjustment with clinical categories and machine learning. *Health Services Research*, **53** (S1), 3189–3206.

SHUMAN, L. J., WOLFE, H. and HARDWICK, C. P. (1972). Predictive hospital reimbursement and evaluation model. *Inquiry*, **9** (2), 17–33.

SLUYS, K., HÄGGMARK, T. and ISELIUS, L. (2005). Outcome and quality of life 5 years after major trauma. *The Journal of Trauma*, **59**, 223–232.

STINNETT, A. A. and MULLAHY, J. (1998). Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analaysis. *Medical Decision Making*, **18** (2), S68–S80.

SVETNIK, V., LIAW, A., TONG, C., CULBERSON, J. C., SHERIDAN, R. P. and FEUSTON, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, **43** (6), 1947–1958.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58** (1), 267–288.

TIKHONOV, A., GONCHARSKY, A., STEPANOV, V. and YAGOLA, A. (2013). *Numerical Methods for the Solution of Ill-Posed Problems*. Mathematics and Its Applications, Springer Netherlands.

TIKHONOV, A. N. and ARSENIN, V. Y. (1977). *Solutions of Ill-Posed Problems*. Winston & Sons, Washington.

Transport Accident Commision (2018/19). *TAC Annual Report*. Retrieved from: https://www.tac.vic.gov.au/about-the-tac/media-and-events/news-and-events/2019/tac-annual-report-tabled-in-parliament [accessed 14.04.2020], Victoria State Government.

van der Laan, M. J. and Duboit, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: finite sample oracle inequalities and examples. *Technical Report 130: Division of Biostatistics, University of California, Berkeley*.

—, Polley, E. C. and Hubbard, A. E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, **6**.

— and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Berlin, Heidelberg, New York: Springer.

Van Patten, R., , Z. C., Merz, Mulhauser, K. and Fucetola, R. (2016). Multivariable prediction of return to work at 6-month follow-up in patients with mild to moderate acute stroke. *Archives of Physical Medicine and Rehabilitation*, **97** (12), 2061–2067.

Victorian State Trauma Registry (Apr 2014). *Victorian State Trauma Registry Special Focus Report. Review of the Case Review Group Indicators – Addendum to Report*. Retrieved from: https://trauma.reach.vic.gov.au/sites/default/files/VSTS%20guideline%20Ver%202.0.pdf [accessed 14.04.2020], Victoria State Government.

Whitehead, S. J. and Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*, **96** (1), 5–21.

Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach*. 7th edition, Boston: Cengage.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, **67** (5), 768–768.

Appendix: Additional tables and figures

TABLE A.1: Descriptive statistics of prediction covariates I

	mean	sd	min	max
— Sample restriction variables —				
Training sample	0.60	0.49	0	1
If admitted to the ICU	0.30	0.46	0	1
If worked prior to the incident	0.67	0.47	0	1
— Outcomes —				
Direct costs of injury, $ 000's	0.95	1.29	0	17
Net Monetary Benefit, $ 000's	2.50	2.57	-17.45	6.75
Return to Work within 1 year	0.68	0.47	0	1
— for the computation of outcomes —				
Net Health Benefit	0.25	0.32	-0.22	1.43
Quality-Adjusted Life years	24.10	17.99	0	67
Quality-Adjusted Life years (discounted)	5.15	2.92	0	10
— Patient demographics —				
If male	0.68	0.47	0	1
Age: 15-24	0.21	0.41	0	1
Age: 25-34	0.18	0.38	0	1
Age: 35-44	0.16	0.36	0	1
Age: 45-54	0.14	0.35	0	1
Age: 55-64	0.11	0.32	0	1
Age: 65-74	0.09	0.28	0	1
Age: 75+	0.11	0.31	0	1
Education: Tertiary	0.46	0.50	0	1
Education: Secondary	0.42	0.49	0	1
Education: Primary	0.03	0.16	0	1
Education: Other	0.01	0.10	0	1
Education: Unknown	0.09	0.28	0	1
Marital Status: Single - Never married	0.15	0.35	0	1
Marital Status: Currently married	0.17	0.37	0	1
Marital Status: Separated	0.02	0.12	0	1
Marital Status: Divorced	0.02	0.15	0	1
Marital Status: Widowed	0.03	0.16	0	1
Marital Status: Living with partner (defacto relationship)	0.06	0.23	0	1
Marital Status: Partnered but not living together	0.03	0.18	0	1
Marital Status: Other	0.00	0.01	0	1
Marital Status: Unknown	0.53	0.50	0	1
Type of residence: Major City	0.71	0.45	0	1
Type of residence: Regional & Remote	0.24	0.43	0	1
Type of residence: Outside VIC	0.03	0.16	0	1
Type of residence: Unknown	0.02	0.14	0	1
Region: Barwon South West	0.09	0.28	0	1
Region: Gippsland	0.06	0.23	0	1
Region: Grampians	0.04	0.19	0	1
Region: Hume	0.06	0.24	0	1
Region: Loddon Mallee	0.04	0.20	0	1
Region: Eastern Metro	0.14	0.34	0	1
Region: Northern Metro	0.16	0.37	0	1
Region: Southern Metro	0.23	0.42	0	1
Region: Western Metro	0.14	0.35	0	1
Region: Overseas	0.00	0.04	0	1
Region: Unknown in Victoria	0.01	0.09	0	1
Region: Unknown outside Victoria	0.00	0.01	0	1
Region: New South Wales	0.01	0.12	0	1
Region: Queensland	0.00	0.06	0	1
Region: South Australia	0.00	0.06	0	1
Region: Western Australia	0.00	0.04	0	1
Region: Tasmania	0.00	0.03	0	1
Region: Northern Territory	0.00	0.03	0	1
Region: Australian Capital Territory	0.00	0.03	0	1
SES: Q1	0.18	0.38	0	1
SES: Q2	0.17	0.37	0	1
SES: Q3	0.19	0.39	0	1
SES: Q4	0.20	0.40	0	1
SES: Q5	0.25	0.43	0	1
SES: Unknown	0.02	0.14	0	1
Observations	11625			

Continued on next page

TABLE A.1: Descriptive statistics — *Continued from previous page*

	mean	sd	min	max
— *Clinical treatment-related* —				
If patient died in hospital	0.04	0.19	0	1
ISS < 12	0.28	0.45	0	1
ISS > 12	0.53	0.50	0	1
ISS unknown	0.19	0.39	0	1
CCI = 0	0.71	0.45	0	1
CCI = 1	0.22	0.41	0	1
CCI > 1	0.07	0.26	0	1
Days in ICU	2.22	5.48	0	140
Hours ventilated	29.56	99.43	0	3089
— *Injury-related characteristics* —				
MTS I	0.52	0.50	0	1
MTS II	0.41	0.50	0	1
MeTS & lower	0.07	0.26	0	1
Unintentional	0.97	0.16	0	1
Intentional-self harm	0.01	0.11	0	1
Assault/Maltreatment	0.00	0.06	0	1
Intent cannot be determined	0.00	0.07	0	1
Intentional-other	0.01	0.08	0	1
AIS: Upper extremity	0.75	0.94	0	4
AIS: Lower extremity	1.16	1.32	0	5
AIS: Thorax	0.47	0.50	0	1
AIS: Head	0.96	1.49	0	6
AIS: Spine	0.86	1.18	0	6
AIS: Face	0.39	0.72	0	4
AIS: Abdominal pelvis	0.50	1.08	0	6
AIS: Neck	0.07	0.42	0	5
AIS: External burns	0.10	0.32	0	5
Inj. group: Isolated head injury	0.02	0.14	0	1
Inj. group: Head/other	0.03	0.17	0	1
Inj. group: Head/ortho	0.12	0.33	0	1
Inj. group: SCI	0.01	0.11	0	1
Inj. group: Severe orthopaedic injuries	0.24	0.43	0	1
Inj. group: Chest/abdominal injuries only	0.01	0.11	0	1
Inj. group: Chest/abdo/other	0.00	0.05	0	1
Inj. group: Chest/abdo/ortho	0.05	0.21	0	1
Inj. group: Other/multi-trauma	0.27	0.45	0	1
Inj. group: Other orthopaedic injuries	0.24	0.43	0	1
Cause: Motor vehicle driver	0.36	0.48	0	1
Cause: Motor vehicle passenger	0.14	0.34	0	1
Cause: Motorcycle driver	0.26	0.44	0	1
Cause: Motorcycle passenger	0.01	0.09	0	1
Cause: Pedal cyclist-rider or passenger	0.08	0.26	0	1
Cause: Pedestrian	0.15	0.36	0	1
Cause: Other transport related circumstance	0.01	0.10	0	1
Place: Home	0.01	0.10	0	1
Place: Residential Institution	0.00	0.02	0	1
Place: School, public admin area	0.00	0.03	0	1
Place: Medical hospital	0.00	0.05	0	1
Place: Athletics and sports area	0.01	0.07	0	1
Place: Road, street, or highway	0.91	0.29	0	1
Place: Trade or service area	0.01	0.11	0	1
Place: Industrial or constructional area	0.00	0.02	0	1
Place: Farm	0.01	0.09	0	1
Place: Place for recreation	0.01	0.08	0	1
Place: Other specified place	0.03	0.18	0	1
Place: Place unknown	0.01	0.10	0	1
Activity: Sports	0.02	0.15	0	1
Activity: Leisure	0.06	0.23	0	1
Activity: Working for Income	0.01	0.10	0	1
Activity: Education	0.00	0.01	0	1
Activity: Other Work	0.00	0.07	0	1
Activity: Being Nursed	0.00	0.02	0	1
Activity: Vital activity	0.01	0.07	0	1
Activity: Other activity	0.61	0.49	0	1
Activity: Activity unknown	0.29	0.46	0	1
Observations	11625			

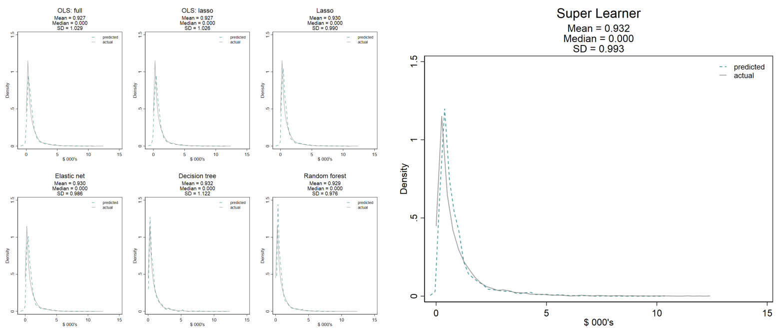
Continued on next page

TABLE A.1: Descriptive statistics — *Continued from previous page*

	mean	sd	min	max
— Health-related behaviour —				
Alcohol condition	0.06	0.23	0	1
Drug conditions	0.02	0.15	0	1
Substance use condition	0.08	0.26	0	1
Any Mental condition	0.10	0.29	0	1
Mood disorders	0.01	0.11	0	1
Neurotic disorder conditions	0.01	0.11	0	1
— Insurance coverage-related characteristics —				
TAC indicator for catastrophic injury	0.05	0.22	0	1
TAC division: Independence	0.08	0.28	0	1
TAC division: Rapid Recovery	0.87	0.34	0	1
TAC division: Supported Recovery	0.05	0.21	0	1
TAC division: Unknown/Other	0.00	0.05	0	1
Vehicle premium risk zone: high	0.39	0.49	0	1
Vehicle premium risk zone: medium	0.17	0.38	0	1
Vehicle premium risk zone: low	0.21	0.40	0	1
Vehicle premium risk zone: unknown	0.24	0.43	0	1
TAC premium insurance class: Passenger vehicle	0.47	0.50	0	1
TAC premium insurance class: Goods vehicle	0.08	0.27	0	1
TAC premium insurance class: Motorcycles	0.18	0.38	0	1
TAC premium insurance class: Other	0.03	0.17	0	1
TAC premium insurance class: Unknown	0.24	0.43	0	1
— Time —				
Year 2009	0.10	0.30	0	1
Year 2010	0.10	0.30	0	1
Year 2011	0.12	0.32	0	1
Year 2012	0.11	0.31	0	1
Year 2013	0.11	0.32	0	1
Year 2014	0.11	0.31	0	1
Year 2015	0.12	0.32	0	1
Year 2016	0.12	0.33	0	1
Year 2017	0.11	0.31	0	1
Observations	11625			

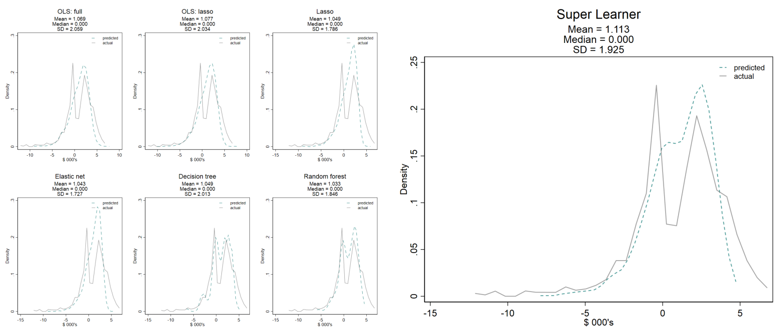
NOTE.— Table presents descriptive statistics of variables used to restrict the sample, generate the prediction outcomes and covariates included in the prediction models. In most cases all models include a set of dummy indicators divided into groups: patient demographics, clinical treatment-related, injury-related characteristics, health-related behaviour and insurance coverage-related characteristics. Here SES is the Social Economic Status as defined by the Index of Relative Socio-economic Advantage and Disadvantage; MTS refers to Major Trauma Services, MeTS - Metropolitan Trauma Services, lower levels of care include Regional Trauma Services and Rural Healthcare Services; ISS refers to the Injury Severity Score; CCI - Charlson Comorbidity Index and AIS - the Abbreviated Injury Scale. In addition, in all prediction models we an extensive set of dummy indicators for the main diagnosis. This is not reported in the table.

FIGURE A.1: Distribution of Predicted values: Outcome *Direct Costs*



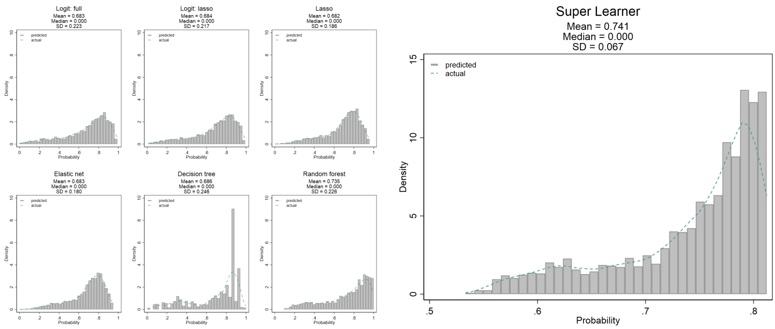
NOTE.— Figure presents the Kernel Density estimation of predicted values of the *Direct costs* within two years of the injury by each single algorithm and the Super Learner. Grey solid line refers to the observed values of the outcome, while the dashed green line to the predicted values. The subtitles report the main statistical measures of the predictions. Statistics shown on the validation sample.

FIGURE A.2: Distribution of Predicted values: Outcome *Net Monetary Benefit*



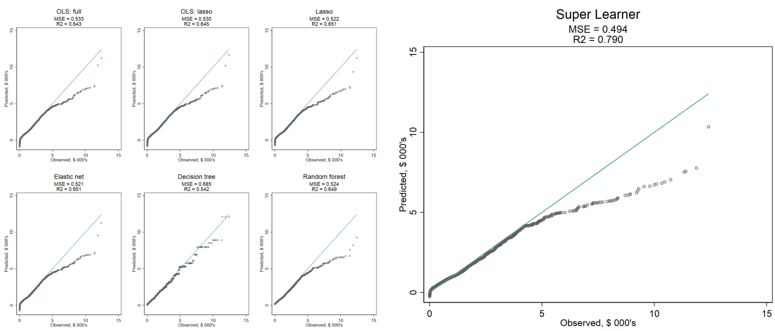
NOTE.— Figure presents the Kernel Density estimation of predicted values of the *Net Monetary Benefit* following 2 years after the injury by each single algorithm and the Super Learner. Grey solid line refers to the observed values of the outcome, while the dashed green line to the predicted values. The subtitles report the main statistical measures of the predictions. Statistics shown on the validation sample and are further restricted to a sub-sample of patients who were admitted to the ICU.

FIGURE A.3: Distribution of Predicted values: Outcome *RTW*



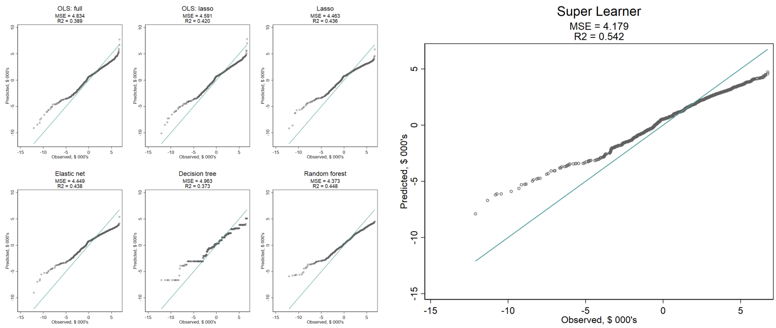
NOTE.— Figure presents empirical distributions of predicted values of the *RTW* by each single algorithm. The subtitles report the main statistical measures of the predictions on the validation sample. Statistics shown on the validation sample and are further restricted to a sub-sample of patients worked prior to the injury.

FIGURE A.4: Quantile-Quantile plot: Outcome *Direct Costs*



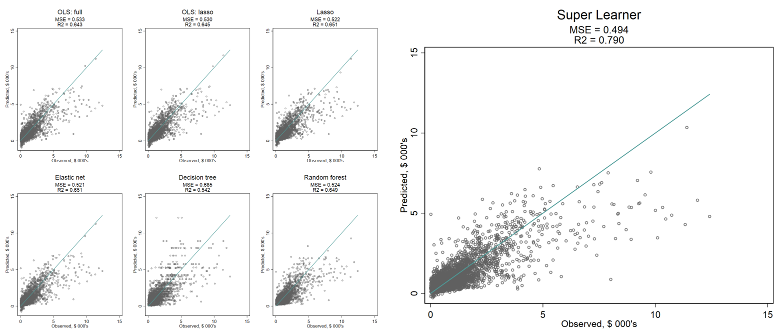
NOTE.— Figure presents the Quantile - Quantile plot that plots quantiles of the observed values against the predicted values for the outcome *Direct Costs*. The subtitle reports the Goodness of Fit statistical measures of the predictions on the validation sample. Quantiles estimated on the validation sample.

FIGURE A.5: Quantile-Quantile plot: Outcome *Net Monetary Benefit*



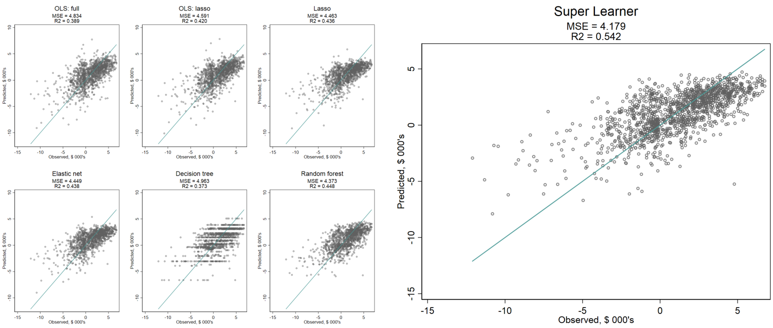
NOTE.— Figure presents the Quantile - Quantile plot that plots quantiles of the observed values against the predicted values for the outcome *Net Monetary Benefit*. The subtitle reports the Goodness of Fit statistical measures of the predictions on the validation sample. Quantiles estimated on the validation sample and further restricted to a sub-sample of patients who were admitted to the ICU.

FIGURE A.6: Prediction Error for each observation: Outcome *Direct costs*



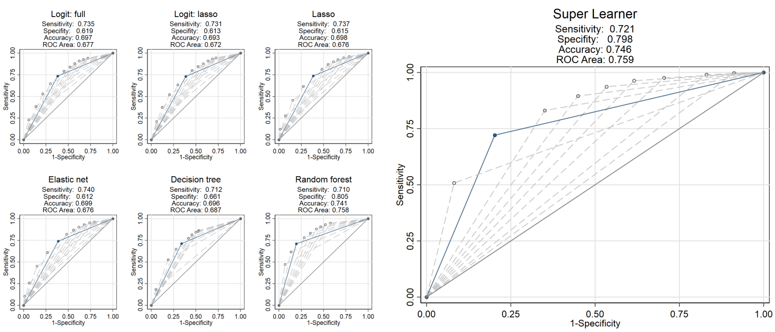
NOTE.— Figure presents the prediction error for each observation of the outcome *Direct Costs*. The reference line denotes a perfect prediction. The observations shown on the validation sample.

FIGURE A.7: Prediction Error for each observation: Outcome *Net Monetary Benefit*



NOTE.— Figure presents the prediction error for each observation of the outcome *Net Benefit*. The reference line denotes a perfect prediction. The observations shown on the validation sample and further restricted to a sub-sample of patients who were admitted to the ICU.

FIGURE A.8: Distribution of Predicted values: Outcome *RTW*



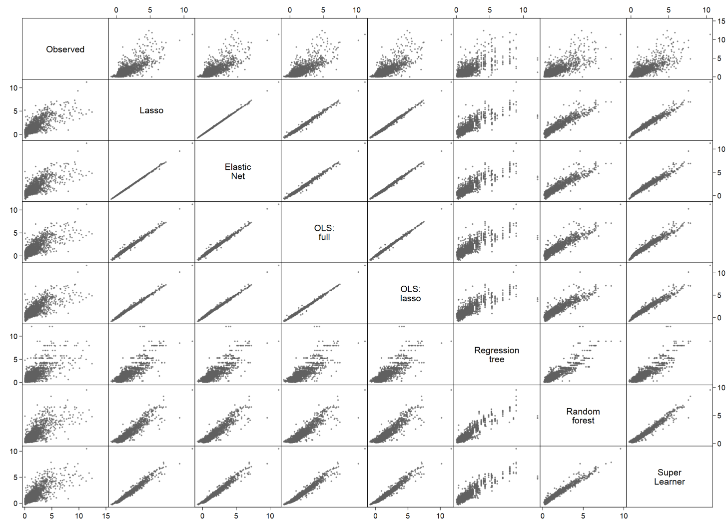
NOTE.— Figures present the Goodness of Fit measures for the prediction of *RTW* by each single algorithm (on the left panel) and the Super Learner (on the right panel). In each graph each line represents the Specificity/Sensitivity statistical measures for different selected classification thresholds that fluctuate around the mean of the outcome. The blue lines indicate the best prediction performed by each algorithm according to the estimated area under the ROC curve and the subtitle reports its statistical metrics in detail. The results are shown on the validation sample and are further restricted to a sub-sample of patients who worked prior to the injury.

TABLE A.2: Descriptive statistics of predictions

	mean	sd	min	median	max
—Direct costs—					
Observed	0.92	1.22	0.00	0.51	12.41
Lasso	0.93	0.99	-0.76	0.62	11.23
Elastic Net	0.93	0.99	-0.75	0.62	11.26
OLS: full	0.93	1.03	-0.93	0.62	11.23
OLS: lasso	0.93	1.03	-0.85	0.62	11.69
Regression tree	0.93	1.12	0.05	0.57	12.11
Random forest	0.93	0.98	0.16	0.59	9.27
Super Learner	0.93	0.99	-0.27	0.60	10.35
Observations	4650				
—Net Monetary Benefit—					
Observed	1.10	2.81	-12.11	1.39	6.75
Lasso	1.05	1.79	-9.14	1.45	5.87
Elastic Net	1.04	1.73	-9.00	1.41	5.40
OLS: full	1.07	2.06	-9.09	1.38	7.74
OLS: lasso	1.08	2.03	-10.11	1.43	7.84
Regression tree	1.05	2.01	-6.61	1.51	5.10
Random forest	1.03	1.85	-5.89	1.26	4.53
Super Learner	1.11	1.93	-7.89	1.42	4.73
Observations	1379				
—RTW—					
Observed	0.69	0.46	0.00	1.00	1.00
Lasso	0.68	0.19	0.02	0.73	0.97
Elastic Net	0.68	0.18	0.02	0.73	0.97
Logit: full	0.69	0.22	0.00	0.75	0.99
Logit:: lasso	0.69	0.22	0.01	0.75	0.99
Regression tree	0.69	0.24	0.00	0.80	1.00
Random forest	0.74	0.23	0.10	0.82	1.00
Super Learner	0.74	0.07	0.54	0.77	0.81
Observations	2948				

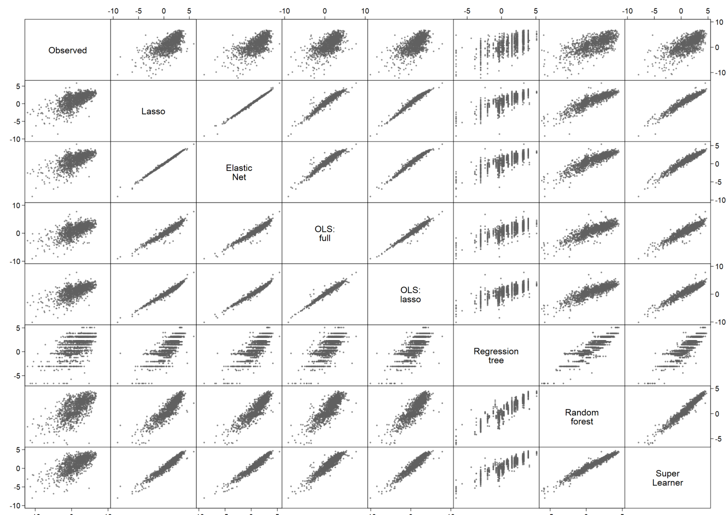
NOTE.— Table presents the descriptive statistics of predictions made using single algorithms and the Super Learner. The predictions of *Direct Costs* and *Net Monetary Benefit* are based on a linear specification of lasso, elastic net and OLS regressions, while the predictions of *RTW* are estimated using a logistic regression specification. All estimations performed on validation sample. Models with *RTW* outcome restricted to a subsample of patient who worked prior to the injury and models with *Net Monetary Benefit* are restricted to patients who were admitted to the ICU during their hospital stay.

FIGURE A.9: Correlation Matrix: Outcome *Direct costs*



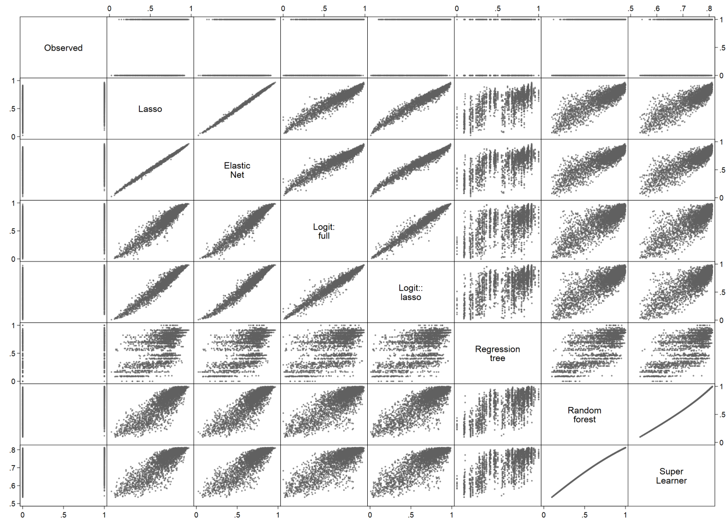
NOTE.— Figure presents the correlation matrix of predictions made by each single algorithm in the prediction of the outcome *Direct costs*. The results are shown on the validation sample.

FIGURE A.10: Correlation Matrix: Outcome *Net Monetary Benefit*



NOTE.— Figure presents the correlation matrix of predictions made by each single algorithm in the prediction of the outcome *Net Monetary Benefit*. Sample is restricted to patients who were admitted to the ICU during their hospital stay and the results are shown on the validation sample.

FIGURE A.11: Correlation Matrix: Outcome *RTW*



NOTE.— Figure presents the correlation matrix of predictions made by each single algorithm in the prediction of the outcome *Net Monetary Benefit*. Sample is restricted to patients who worked prior to the injury and the results are shown on the validation sample.