Essays on Physician Innovation and Practice Style in Healthcare Markets

by

Ivan Nikolaev Badinski

B.A., Williams College (2014)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2021

Author .....
Department of Economics
December 13, 2020

Certified by .....
Amy Finkelstein
John & Jennie S. MacDonald Professor of Economics, MIT
Thesis Supervisor

Certified by .....
Heidi Williams
Charles R. Schwab Professor of Economics, Stanford University
Thesis Supervisor

Certified by .....
Tobias Salz
Castle Krob Career Development Assistant Professor of Economics, MIT
Thesis Supervisor

Accepted by .....
Amy Finkelstein
John & Jennie S. MacDonald Professor of Economics, MIT
Chairman, Department Committee on Graduate Theses

Essays on Physician Innovation and Practice Style in Healthcare Markets

by

Ivan Nikolaev Badinski

Submitted to the Department of Economics
on December 13, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis consists of three chapters on technological innovation, diffusion, and practice style among physicians.

The first chapter investigates the effect of age on a surgeon's propensity to use new medical procedures. I identify a large number of medical technologies undergoing diffusion by a well-defined risk set of physicians by exploiting the parent-descendant relationship among ICD9-CM inpatient procedure codes where each new code has a well-defined antecedent. I find that surgeons that are ten years older at the time of new-code approval are sixteen percent less likely to use this code. Evidence from the diffusion of new pharmaceuticals, diagnostic codes, and minimally invasive procedures suggests that this effect may be driven by skill acquisition costs rather than information frictions.

In the second chapter, I study the impact of market size on the development of novel surgeries, an important domain of medical innovation where intellectual property rights, approval regulation, and financial incentives play only a minor role. Using the codification of ICD9 CM procedure codes as a novel measure of new-surgery development, I investigate the behavior of surgical innovation and compare it to pharmaceuticals where traditional innovative institutions are salient. I find that the two processes follow very different aggregate trends. Despite this difference, I estimate a positive and significant elasticity of surgical innovation with respect to potential market size by leveraging quasi-exogenous changes in potential market size due to shifting US demographics.

The third chapter, joint with Amy Finkelstein, Matthew Gentzkow, Peter Hull, and Heidi Williams, studies the role of physician practice style in Medicare geographic spending variation. We estimate a model that allows for variation in patient demand, physician treatment intensity, and regional supply-side factors, as well as patient-physician sorting. The model is identified by quasi-experimental migration of Medicare patients and physicians and their matching within regions. We find that physicians vary greatly in their treatment intensity. Our baseline decomposition suggests that about 30 percent of regional variation in health care utilization is explained by differences in average physician treatment intensity, 20 percent by other area supply factors, and 50 percent by differences in patient demand.

JEL Classifications: O33, I11, I18

Thesis Supervisor: Amy Finkelstein
Title: John & Jennie S. MacDonald Professor of Economics, MIT

Thesis Supervisor: Heidi Williams
Title: Charles R. Schwab Professor of Economics, Stanford University

Thesis Supervisor: Tobias Salz
Title: Castle Krob Career Development Assistant Professor of Economics, MIT

Acknowledgments

My time at MIT has been a truly life-changing experience and one that I am deeply grateful for. I have had the wonderful opportunity to work with, talk to, and learn from some of the brightest minds of our time. I am thankful not only for the help everyone here has given me in improving my abilities, in exploring my interests, and in growing as a person, but also for showing me that being a better researcher also means being kinder and more compassionate to oneself and others.

Looking back at these past few years, I cannot help but appreciate how long, at times difficult, and overall incredibly rewarding a journey this has been. It is hard to express all the myriad ways in which I have changed since my first steps on this path, but one lesson shines brighter than the rest – we do not face challenges alone, we do not overcome trials by ourselves, and we do not achieve victories on our own. I wish to thank everyone who has touched and enriched my time here and helped me keep taking the next step.

I want to thank my advisors, Amy Finkelstein, Heidi Williams, and Tobias Salz, for their truly incredible mentorship over all these years.

I want to thank Amy for her singular focus and deep commitment to helping me improve and push my research work to the next level. I want to thank her for making sure that we meet regularly, for taking me on as a research assistant, for making such consistent effort in bringing the graduate students together for breakfast research conversations and board game fun, and, last but not least, for showing me that excellency and enjoyment of research go hand in hand.

I want to thank Heidi for being so kind, compassionate, caring, and deeply, deeply inspiring as a researcher. I want to thank her for taking me on as a research assistant in my first year after having had only a single conversation with me before I had even started my PhD program. I want to thank her for always giving me thoughtful feedback on my work, for taking the time for fun research-themed walks, for maintaining her high level of involvement with my work despite moving across the country, and for reaching out to help me through the few times I was feeling down and discouraged.

Finally, I want to thank Tobias for being an absolute rock in terms of supporting me through the most stressful times of my dissertation work, for being a shining example of a truly excellent researcher who has wide-ranging interests and always offers insightful comments, and for making our research meetings and calls incredibly fun and enjoyable. I want to thank him for always being available to talk even on short notice, for being willing to chat about even the most half-baked of ideas, and for bringing a combination of sincerity and intellectual intensity to our discussions that have kept reminding me again and again that the challenges and rewards of academic work form flip sides of the same coin.

I want to thank my co-authors Matthew Gentzkow and Peter Hull for showing me how fun collaborative research can be. I want to thank Matt for the enthsiasm and truly impressive intellectual rigor that he brings to our project calls and that have been a regular source of inspiration for years. I want to thank Peter for the thoughtfulness, reliability, friendliness, and patience that he has consistently invested in our work together. Working with someone as smart and dedicated to academic excellence as him has been an outstanding opportunity to learn and improve.

I also want to thank Jon Gruber for his infective energy, incisive comments, and enthusiasm for public and health economics.

I want to thank my fellow graduate students for making this journey a truly unique experience. This includes, among many others, Tamar Oostrom, Anton Popov, Carolyn Stein, Mayara Felix, Michael Wong, Mikel Petri, Joe Hazell, Patricia Anghel, Soomi Kim, Jeremy Yang, Layne Kirshon, Ben Deaner, Tim Simmons Justicz, Sophie Sun, Ljubica Ristovska, Allan Hsiao, Olivia Kim, Martina Uccioli, Pedro Tepedino, Jon Cohen, Sean Wang, Anna Russo, Abby Ostriker, Alden Cheng, Hector Blanco, Aaron Goodman, Claire Lazar, Pari Sastry, and Aileen Devlin. Our lunches, practice talks, trivia nights, and random discussions of Eurasian geopolitics, utilitarian philosophy, true-crime podcasts, food-themed anime, and the unfortunate final season of a certain medieval fantasy show among others have infused graduate school with warmth and joy.

I want to thank Tamar, Carolyn, and Anton for being the best possible office-mates one can ask for. I want to thank Tamar for being an outstanding conversation and commiseration partner as well as for making late evenings in the office as close to enjoyable as possible. I want to thank Carolyn for introducing me to the joys of slow-cooking. I want to thank Anton for many amazing board/video game nights and excellent insight into the inner workings of the European League community.

I have also greatly benefited from the advice of older students who have generously shared their own experience and insights. This includes, among others, Colin Gray, Ryan Hill, Jane Choi, Sarah Abraham, Maddie McKelway, Jon Petkun, Pablo Azar, Alonso Bucarey, David Colino, Jetson Leder-Luis, Elizabeth Setren, Seth Blumberg, Michael Stepner, Will Rafey, Alex Olssen, and Ray Kluender.

I also want to thank Paul Friedrich, Ken Jung, and Isaac Norwich for their camaraderie and dedication in working together on various projects. I want to thank Paul for introducing me to the endless joys of svn, for organizing awesome trivia nights, and for always being a ray of sunshine in the office. I want to thank Ken for having me on his radio show and introducing me to his wonderful music. I want to thank Isaac for his endless commitment to our project and admirable effort in battling the dragons of claims data.

I want to thank Maryam, Paul, Zainab, Yasmin, Alex, Eli, Jordan, Zanele, Ashley, and Bethany for always taking the time to talk and discuss the things that push us and the things that matter to us. Similarly, I want to thank Evan, Kate, Jose, Greg, Allegra, Jonathan, Max, Michael, Michaela, and Mary for all the support and wonderful conversations throughout the years.

I want to thank my Williams College professors who have played a colossal role in shaping my thinking and interests and who have done so much to support me and give me all the resources I could possibly imagine wanting in my endeavors. I want to thank Steven Nafziger for being an incredible undergraduate thesis advisor and giving me my first RA position ever, Jon Bakija for being an awesome mentor and co-author, Frank Morgan for inspiring in me an appreciation for the beauty of mathematics and for teaching me real analysis outside of the regular school year, and Steven Miller for always reaching out, keeping in touch and making math classes fun. I also want to thank Keith McPartland, Melissa Barry, and Bojana Mladenovic for showing me how amazing and personally relevant philosophy can be.

I have been fortunate in having had the chance to meet and befriend many amazing individuals outside of MIT who have always helped and inspired me, and who have given

my life so much warm joy. These include my Williams friends who have made my stay in this country an absolute blast – Rahul Sangar, Philippe Demontigny, Jesse Sardell, Wade Phenicie, Jirapat Samranvedhya, Kamuela Lau, Craig Burt, Jamie Baik, Catherine Pang, Annie Jeong, Jensen Pak, Donnie Kost, Khan Shairani, Daniel Yoo, Jennifer Chan, Meilu McDermott, Zach McKenzie, Steve Im, Olivia Wang, Maikhanh Nguyen, Sam Austin, Kyle Martin, Lizzie Chan, Phonkrit Tanavisarut, Chris Alder, Keelia Willison Alder, and many, many others.

These also include my Bulgarian friends who have always been a source of energy and strength and who have kept me up to date on the latest music trends back home - Victoria Angelova, Kamen Velichkov, Georgi Klissurski, Martin Angelov, Kliment Minchev, Zara Rancheva, Zeynep Yozgyur, Mila Topalova, Maria Velkova, Magdalena Staneva, Stasi Georgieva, Ivet Vidinova, Bojidar Vidinov, Nguyen Le, Vicky Dotcheva, Maria Boshnacova, Maria Tsvetkova, Maria Raycheva, Dean Bezlov, Dobri Dobrev, Anna Antonova, Angel Shotev, Simona Petrova, Tsvetelina Paunova, Pavel Atanassov, Kalina Tsoneva, Pavel Pavlov, Martin Georgiev, Smilen Stoychev, Anna Stoycheva, Martin Popov, Yana Mats, Svetlana Nencheva, Nediyana Daskalova, and many others.

I want to thank my former AG co-workers for being absolutely wonderful people who have consistently given me a reason to step outside of my office for an enjoyable meal or activity – Ju Li, Anne Cai, Frank Qian, and Khanh Ngo.

I want to thank Todor Ivanov, Maya Ivanov, Maxim Ivanov, and Denitza Anderson for helping me get settled and generously assisting me with all the unforeseen challenges of life in a foreign country.

I want to thank my partner, Jenny Qiu, for being a special source of light and love in my life. I never thought I'd find someone who understands my non-diminishing marginal utility for cookies, but in this and in many other things, she has proven to be a source of wonder and delight.

Finally, I want to thank my family for being the absolute bedrock of my life. Words cannot begin to express the amount of love, care, affection, and support they have given me. My father, Nikolay, has been the deepest source of inspiration for me with his absolute dedication to our family, kind nature, extreme smarts, and willingness and ability to understand any problem I am facing and guide me through it through dialogue and empathy. My mother, Tzvetanka, has always been a true hero to me with her wholehearted care, deep compassion, and all-encompassing unconditional love for all of us. The same kindness, dedication, and love have been a hallmark of my entire family – my sister Ani, my uncles Goro and Lubo, my aunts Mimi and Sashka, my cousins Teddy, Emo, and Stanislav, and my grandparents – Ivan, Nadya, Mircho, and Mimi. I want to give my sister Ani and my cousin Teddy a special shout-out for being the most excellent hide-and-seek partners growing up.

To all of you and to everyone else not mentioned here who has shared my path and helped me go forward - a huge and sincere "Thank you."

Contents

List of Figures

List of Tables

1 Chapter 1
The Age of Modern Medicine

1.1 Introduction

The successful adoption of cost-effective technologies is commonly touted as an essential component of achieving better value at lower cost in the healthcare sector (Cutler (2005), Chandra and Skinner (2012), Skinner and Staiger (2012)). Although various institutional and financial factors underlying technological diffusion have received much attention in healthcare economics, the role of physician demographics in this process has remained relatively underexplored. This is surprising because concerns about the impact of an aging labor force on workers' ability to keep up with the technological frontier have been gaining in prominence (Deming and Noray (forthcoming)). A natural implication of these ideas is that the impact of physician age on healthcare technology adoption is an important, yet unanswered question.

This paper explores the differential adoption patterns of younger and older physicians of multiple technologies in an important domain of medical innovation - surgical procedures. To do so, I employ a novel measure of technological diffusion that uses the branching structure of annual

revisions to ICD-9-CM procedure codes. The evolution of codes creates a linkage of each new code to an older one that serves as its technological substitute. This feature of the system allows me to study the simultaneous diffusion of multiple technologies by consistently identifying each diffusing innovation's older technological substitutes and agents that are "at risk" for adoption. These are key challenges to the study of diffusion more broadly and overcoming them allows me to confront the data with a model explaining the effects of aging based on the latest insights in psychology and economics.

I find a negative relationship between age and surgical procedure adoption. My analyses show that a 10-year-increase in age is associated with a 16% (1.8 percentage points) decrease in the propensity to use the new procedure codes. The results are invariant to the addition of various medical and demographic controls for the physicians and their patient mix to the specifications. This suggests that patient selection is not a driver of the results and is consistent with the idea that innovations within the same category are likely to be used by phyisicians with a similar skill-set on patients with a similar medical condition and outlook. Lastly, I perform a simple counterfactual exercise implying that the gains from removing the age-based barriers to adoption are likely to be large. In particular, my estimates suggest that helping older physicians achieve the adoption propensity of the youngest group of doctors will increase technology adoption rates by 19%.

I document two additional patterns that shed light on the properties of the likely root cause of these differences. First, I test whether the differences between young and old are due to differential adoption (extensive-margin) or differential use conditional on adopting (intensive margin). I find suggestive evidence that extensive-margin frictions are likely to be important in this context. Second, I investigate the persistence of the difference by examining the evolution of this gap over time and document that it remains significant years after the introduction of a new code. This indicates that the underlying mechanisms are not transient in nature and may lead to long-term differences in technological capabilities between the young and the young.

I interpret the patterns in the data through an economic model of adoption nesting the dominant paradigm of the cognitive effects of aging within psychology. The model features a physician agent making adoption decisions driven by a central trade-off - fixed costs of technology adotion incurred in the present versus benefits to patients in the future. The model indicates that there are two channels for the impact of age on this decision. First, higher age decreases in the value of

adoption through lower expected remaining practice life. Second, increases in age affect the costs of adoping new technology through an underlying substitution between practical experience and ability to deal with novel situations.

I continue by presenting empirical evidence for the two channels in the conceptual framework. First, I use the model to derive a simple empirical test of the career life expectancy channel that relies on including an interaction between arrival rate and age in my specifications. Second, I perform an analogous analysis in two settings where fixed adoption costs are likely to be lower - new diagnostic code use and novel pharmaceutical drug presrciptions - and one setting where the same costs are likely to be relatively high - minimally invasive procedures. Consistent with the model's predicitons, I find that there is no age-based gradient in the first two settings, but a relatively steeper gradient in the latter setting.

This result is not simply of theoretical interest, but has important policy implications. In particular, there is an active policy concern for the maintenance and expansion of physician skills after graduating from medical school (Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) (2020)). Previous research in technology diffusion has suggested that information frictions could be important barriers for the spread of new innovations (Agha and Molitor (2018)). My study suggests that these frictions are not likely to play a big role in the current setting and that skill-acquisition costs are more important barriers to diffusion. Together with the counterfactual results from above, they strongly imply that policies that subsidize physician investment in new skill acquisition are the only method to achieve large and lasting gains in effective technology diffusion.

1.2 Related literature

This paper is closely related to a prolific literature on technological diffusion in healthcare. Technological diffusion has been traditionally seen as an important factor behind both the large gains in welfare and increases in costs in healthcare markets over the past few decades (Cutler (2005), Chandra and Skinner (2012), Skinner and Staiger (2012)). Additionally, popular models of diffusion in the sector imply that small differences in provider adoption propensities can have large effects on steady-state outcomes and costs (Skinner and Staiger (2012)). This has led to much

attention devoted to the causes and consequences of technological diffusion in empirical work (Baker and Phibbs (2002), Skinner and Staiger (2012), Clemens and Gottlieb (2014)). The popularly explored determinants of technological adoption include health insurance expansions (Weisbrod (1991), Finkelstein (2007)), health insurance financial incentives (Baker and Phibbs (2002), Acemoglu and Finkelstein (2008), Clemens and Gottlieb (2014)), and opinion leadership (Agha and Molitor (2018)).

Physician demographic determinants of technological diffusion, on the other hand, remain underexplored. Most of the empirical evidence on this topic comes from case studies in the diffusion of prescription pharmaceuticals or surgical procedures that looks at the correlation between adoption behavior and a variety of physician or practice characteristics.[1] [2]While such studies constitute important forays into the issue, they offer conflicting evidence on the magnitude and direction of the age-adoption profile. Additionally, extrapolating to other domains of innovation in medicine or even other prescription medications or procedures is difficult.

On the one hand this is surprising. Recent research in labor and financial economics has raised concerns that the aging of the labor force may lead to a decreased ability of workers to keep up with the technological frontier. For instance, Deming and Noray (forthcoming) find that the wage premium commanded by STEM workers relative to their non-STEM counterparts declines with age. The authors explain this pattern as a result of the inability of older workers to keep up with the fast changes in the technological frontier of these fields.

Furthermore, changes in the ability to learn and adapt with time have been the focus of other influential studies in economics. For instance, Agarwal et al. (2009) find that consumer financial mistakes follow a hump-shaped pattern with age, where financial skill peaks in middle-age and then quickly deteriorates afterwards. Similarly, Korniotis and Kumar (2009) find that while investment knowledge increases with age, investment skill declines. Both studies lean on a rich literature in

[1]Examples in the case of medical procedures include Escarce (1996), Beckelis et al. (2017), Artis et al. (2006), Forte et al. (2010), and Gratwohl et al. (2010). Examples in the case of drugs include Steffensen et al. (1999), Mark et al. (2002), Johannesson and Lundin (2002), Kozyrskyj et al. (2007), Bourke and Roper (2012), and Huskamp et al. (2013).

[2]I am aware of only one study in this space - Glass and Rosenthal (2004) - which analyzes the diffusion of a large number of similar technologies (new pharmaceuticals) in order to draw more general conclusions about the behavior of healthcare providers in this space. The authors investigate the diffusion of 32 new pharmaceuticals (some first in class and some follow-on drugs) and relate it to the demographic characteristics of prescribing physicians using logistic regression.

psychology (see, e.g., Samanez-Larkin and Knutson (2015) for a review) to explain these patterns as the result of the interplay of practical experience (crystallized intelligence) and ability to deal with novel situations (fluid intelligence). That is, while older workers may have more practical experience that helps with a task, they are also less likely to engage with complex tasks that require novel learning. This body of knowledge strongly suggests that demographic aging effects are likely to be very important in healthcare where technology moves at a rapid pace.[3]

On the other hand, the lack of work in this area in healthcare economics faces unique challenges. One such difficulty is the measurement of technological diffusion consistently for a broad set of technologies. As mentioned above, while case studies offer an interesting look into the problem, their very nature precludes broad conclusions. Measuring many new technologies well using conventional data sources in this sector is challenging because one needs to identify consistently a large number of innovations, their substitutes, and the people who are likely to use them.[4]This study aims to overcome these difficulties by employing a novel measure of technology diffusion that uses the evolution of procedural codes to allow a glimpse into the link between physician age and technology adoption decisions across multiple innovations.

1.3 Institutional background

The central method of measuring innovation in this study is the International Classification of Disease, 9th Revision, Clinical Modification (ICD-9-CM). This is a system of about 3,500 four-digit procedural codes which is used by virtually all hospitals in the US to document inpatient procedures for financial and analytical purposes. The system has a hierarchical organization where

[3]In a different context where adaptability and learning are important, a body of papers in labor economics has investigated the relationship between individual inventiveness as measured by great inventions or patent applications and age. Studies in this space look at the relationship, either in the cross-section or a panel setting, between individual inventor age and scientific creativity as measured by e.g., patent applications. Examples include Henseke and Tivig (2008), Feyer (2008), Mariani and Romanelli (2007), Schettino et al. (2008), and Hoisl (2007). The typical finding in these studies is that individual creativity has an inverted U-shape and peaks around 50 years of age (Frosch (2011)). Another strand has looked at the age composition of specific firms or industries and its relationship to firm performance as measured by output or propensity to adopt new technologies. Examples include Hellerstein et al. (1999), Crepon et al. (2003), annd Meyer (2009). The typical finding in these studies is, again, that middle-aged workers tend to be more productive than both their younger and older counterparts.

[4]Traditionally, this has only been possible in the case of pharmaceuticals where a large number of trackable new technologies is available. However, pharmaceuticals represent a case where the likely adoption barriers are lack of information rather than the need for skill acquisition and learning. They also represent a case where profit-maximizing firms have an incentive to keep physicians informed of the existence and applicability of the product, so that the system's institutions are built to counteract some of the effects of interest in this study (Scott-Morton and Kyle (2011)).

similar procedures that affect the same condition, organ, and system form sequences that are numerically close to each other. Close groups of 4-digit codes refer to the same kind of operation performed using different methods or approaches, [5] procedures starting with the same two digits typically affect the same organ or medical condition,[6] and close clusters of two starting digits refer to entire systems (United States General Accounting Office (GAO) (2002)).[7] The preservation of this hierarchical structure has been an implicit goal as the system evolves due to changes in medical practices and knowledge.

I identify new and old surgical procedures and define the risk set of doctors using the evolution history of ICD-9-CM inpatient procedure codes. Every year, the CMS, the federal agency responsible for maintaining the system, issues revisions, including code additions, deletions, and edits. For each code addition, the agency documents the kind of procedure it designates and the parent code that was previously used to denote instances of this procedure. I define the procedures falling under the newly added code as the "new technology" and those that continue to fall under the parent code, which was previously used to designate the procedures that now fall under the new code, as the "old technology". The set of doctors who were observed using the parent code in the year of codification is the new technology's relevant risk set.

A simple example may help clarify my data construction procedure. In 2008, the CMS introduced ICD-9-CM procedure code 53.62, "laparoscopic incisional hernia repair with graft or prosthesis" to differentiate it from its parent code 53.61, "other open incisional hernia repair with graft or prosthesis". Prior to the introduction of code 53.62, 53.61 was used to denote both open and laparoscopic incisional hernia repairs.[8] In my analysis, 53.62 is an example of a "new procedure" and 53.61 is an example of its substitute "old procedure." The risk set of interest in this example is the set of doctors who were observed using code "53.61" in 2008. I then compare the propensity of younger physicians in this set to use the code over the next 5 years relative to that of older doctors.

[5]For example, codes in the 53.5-53.7 refer to various kinds of ventral and diaphragmatic hernia repair: 53.51 denotes "incisional hernia repair", 53.61 denotes "open ventral hernia repair", and 53.71 stands for "laparoscopic repair of diaphragmatic hernia"

[6]For instance, codes starting with "53" refer to hernia repair in general, "52" - to operations of the pancreas, etc.

[7]For example, codes beginning with "01" to "05" refer to operations on the nervous sytem, codes beginning with "06" to "07" refer to operations on the endocrine system and so on.

[8]In the case of many laparoscopic procedures, this was accomplished by using the parent code and a "laparoscopy" code modifier 54.21.

This example procedure's technological profile illustrates the thought experiment behind my empirical strategy. In particular, an incisional hernia is a spilling of internal tissue through its surrounding abdominal muscle wall due to a weakness from a previously existing incision. "Open hernia repair" refers to the traditional open approach that involves fully "opening" the abdomen to repair the bulging. "Laparoscopic hernia repair" refers to a novel minimally invasive approach that conducts the repair through two small incisions at the side of the patient's abdomen (American College of Surgeons (2012)). [9] In this case, the innovation, laparoscopy, is documented by the new code, and its older technological substitute, the open approach, is denoted by the old code. I identify the physicians that use the procedure at the point when the new code was introduced and then follow the adoption propensity among them conditional on their age in 2008.

The same example also illustrates some of the subtler points in interpreting the results from my analysis. Specifically, the definition of technological space that I use, pairs of parent and descendant codes, is typically narrower than definitions used by other authors writing about individual procedures. For instance, in a 2013 article in *Surgical Endoscopy*, Colavita et al. explore the rates and outcomes of open and laparoscopic ventral hernia repair in a nationally representative dataset of inpatient stays, the Nationwide Inpatient Sample. They define "open ventral hernia repair" using codes "53.61" and "53.69" and "laparoscopic ventral hernia repair" using codes "53.62" and "53.63". Procedure code "53.63", "other laparoscopic repair of other hernia of anterior abdominal wall with graft or prosthesis" is the descendant for parent code "53.69", "other and open repair of other hernia of anterior abdominal wall with graft or prosthesis." In my analysis, the pairs "53.61"-"53.62" and "53.69"-"53.63" are separate technologies due to the way these procedures are classified in the ICD9-CM.

The narrower definition of technological space I use may thus be interpreted as that of component technologies. In particular, the technologies I use here are in some sense the subdivisions of broader categories of surgical procedures. The magnitude of the estimates and the baseline diffusion levels in the analysis sample may or may not, therefore, match those of broader categories in the medical literature. In the case of hernia repair here, the baseline diffusion rates of laparoscopy

[9] Many authors have commented on the benefits of laparoscopy in terms of patient recovery time and experience of post-operative pain as well as on the relatively low take-up of the procedure due to the difficulty of learning it. In particular, the laparoscopic approach essentially requires re-learning the procedure as the surgeon's access and field of view are significantly altered by the instruments she has to use (Edwards and Bailey (2000), Sood et al. (2015), Bansal et al. (2016)).

in the two pairs of codes - 21% and 23% respecitvely - are close to each other and to the 18% laparoscopy rate identified by Colavita et al. (2013). However, this may not necessarily be true in all cases.[10] In the appendix, I consider some of the ways in which this narrower definition may be important for the estimation and interpretation of the results.

In the full analysis, I stack all technologies that satisfy the sample restrictions (broadly, new-code additions where the previous code has undergone only a single revision in the sample period) and compare the propensity of younger relative to older physicians to use the new code. As outlined above, the advantage of this approach is that it can be consistently applied to a large number of diffusing technologies to analyze their adoption path. For each surgical procedure of interest, the system provides a method of determining the set of procedures which are likely to act as its technological substitutes as the procedures that continue being designated under the parent code. It also identifies the group of agents for whom this new technology is likely to be relevant as the physicians who have already gained the human capital necessary to specialize in the treatment of a particular organ or condition. To the best of my knowledge, this is the first study in health economics outside of the realm of prescription pharmaceuticals where these challenges to measuring the large-scale diffusion of technology have been overcome.

1.4 Data and Empirical Framework

The primary dataset used in this project is comprised of the inpatient stays of a 20% sample of US Medicare beneficiaries between 2001 and 2013. These data contain information on the patient's diagnoses and existing medical conditions, all services provided, length of stay, a variety of demographic characteristics, and the identities of the operating and attending physicians. The advantage of these data in investigating the relationship between age and adoption is that they contain detailed information on a pysician's surgical history at the individual patient level as well as relevant patient characteristics. This allows me to measure with precision a physician's adoption decisions and evaluate the role of patient selection, an important concern in healthcare contexts.

[10] Additionally, recall that new and old surgeries are not perfect substitutes and there may be some fuzziness in how a procedure is coded. Hence, depending on the authors' purpose, it is possible that a study may define the relevant technological space as a combination of parent codes and only some of their descendant codes or vice versa depending on the authors' purpose (see, e.g., Panaich et al. (2016)'s study on the use of atherectomy).

1.4.1 Analysis sample

As described in the institutional setup, I define the set of technological substitutes for each new procedure as the procedure or set of procedures which were designated under the new procedure's parent code prior to codification. In order to obtain a set of procedure swith well-defined risk-sets, I restrict to instances of code revisions that involve only a single parent code where both the parent and descendant codes are still being used at the end of my sample window. This drops revisions where the new procedures were designated under multiple parent codes prior to codification and instances where the parent code was completely deleted and split off into multiple descendant categories. As shown in the appendix, single-parent codification events represent the lion's share of ICD9-CM. In order to ensure that I observe the entire diffusion path of a new code in the primary patient sample, I also restrict to revisions that occur in or after 2001. This gives me 85 instances of code revisions. I extract all instances when the primary procedure of a Medicare inpatient stay was documented using one of these codes.

Next, I turn to the definition of the set of doctors that are at risk for adoption for each new procedure. I define the risk set as those physicians that are observed performing the parent code in the year of the introduction of the new code. In the case of hernia repair for example, I look at physicians that show up as operating physicians in the inpatient file on claims where the primary procedure is 53.61, ventral hernia repair, in 2008. I then follow these physicians after the introduction of 53.62, lap ventral hernia repair, in 2008. The outcome of interest is the adoption of 53.62 post 2008 and the independent variable of interest - the age of these operating physicians in 2008. I obtain age information from the AMA Physician Masterfile, a database containing information on the education, career, and demographics of 1.4 million current and past physicians (American Medical Association (2020)). The general analysis then stacks all 85 revisions and measures the effect of age on adoption as a single parameter.

In what follows, I will call the set of codes used in each of these 85 revisions a "treatment category". For instance, the ventral hernia repair treatment category contains the codes "53.61" and "53.62", where the former is the parent code and the latter is the descendant code. A specific treatment category always contains a single parent code (by sample construction), but may contain multiple descendant codes.

This set of treatment categories spans a wide variety of body systems and medical conditions. Table A1 in the appendix summarizes the number of code changes and percent of sample observations covered by each aggregate category of ICD9-CM procedure codes (in the hierarchical structure introduced above, these are sequences of codes defined by their first two digits). The digestive, cardiovascular, and muscoskeletal systems are the most important body systems in terms of their influence on the sample. The table also shows that there is a substantial amount of variation in the number of observations per treatment category as some procedures have many more patients than others. [11]

Table 1.1 contains summary statistics for the sample. Panel A describes the sample at the procedure level. As this is Medicare, patients tend to be older, on average 74 years old, with roughly equal numbers of men and women and about 10% of patients having chronic conditions in the year prior to their visit to the doctor in my sample. Panel B contains statistics at the doctor level. The physicians are 53 years old on average, with physicians mostly men, reflecting the gender imbalance in the surgical professions. I observe each doctor for 3.5 years on average, indicating that I have a decent number of physicians that appear more than once in the observed procedures. In total I observe 179,000 procedures by 27,000 doctors in 85 different treatment categories.

Figure 1.1 offers a first look into the doctor age distribution in this sample. Figure 1.1 is a histogram of the age of physicians at the time the new code in the relevant treatment category is introduced. The distribution is roughly bell-shaped, with a range of ages and a large mass between 40 and 60 years old. This indicates that there is a lot of variation in the main explanatory variable of interest and that the variation is coming from physicians at all ages and especially from those that are in the prime years of their careers.

Given that the point of codification is a particular point in the diffusion of a new technology

[11] There is a significant number of codes in "miscellaneous" or "not elsewhere classified" categories. The official nomenclature behind these categories, however, is slightly misleading as the space constraints in the system have forced the CMS to put some codes more naturally classified as part of other systems into the miscellaneous categories. the codes in the "not elsewhere classified" category, which account for a large share of the observations in the sample, are six new procedures for cardiac resynchronization pacemakers, some of which have defibrillation capabilities. Normally, these procedures would be entered in the three-digit sequences 37.7 and 37.8, but those were completely full at the time the CMS made its decision. As a result, the CMS was forced to assign the new codes to the sequence beginning with "00", which is outside of their natural hierarhichal place. The issues with space are one of the major reasons why the US has chosen to transition to the newer ICD10-PCS as of October 2015 (United States General Accounting Office (GAO) (2002)).

when it has been deemed sufficiently important by the medical community to receive a new code, it is not immediately clear that these technologies are actually still diffusing at the point when the new codes are introduced. Figure 1.2 investigates this question by plotting a diffusion S-curve of the fraction of new codes used in a treatment category by year relative to new code introduction. In particular, each point on the graph is a relative-year coefficient from the following equation estimated on the entire sample:

$$y_{ipdt} = \beta_0 + \sum_t \beta_{1t} RY_{t(p)} + \gamma_p + \tau_t + \varepsilon_{ipdt} \tag{1}$$

where where i indexes patients, d indexes operating physicians, p indexes treatment category, and t indexes calendar years and y is a dummy for whether the new or old code within a treatment category is used. $RY_{t(p)}$ are indicators for year relative to the codification of the new code in treatment category p. The omitted category is relative year 1.[12] It is clear that while diffusion seems relatively slow at this point already, it has not stopped and the innovations introduced by the new codes seem to still be gaining use in their relevant markets.

1.4.2 Empirical framework

In order to analyze the data, I use a simple linear probability model that relates new procedure use to physician age. This specification allows for a simple, transparent, yet flexible way of measuring the effect of interest. In particular, I run empirical specifications of the following form:

$$y_{ipdt} = \beta_0 + \beta_1 A_{dp} + \beta_2 X_{idt} + \gamma_p + \tau_t + \varepsilon_{ipdt} \tag{2}$$

where i indexes patients, d indexes operating physicians, p indexes treatment category, and t indexes calendar years. y is an outcome of interest (most importantly, a dummy for whether the new or old code within a treatment category is used), A_{dp} is operating physician d's age in the year when the new code for treatment category p was introduced, X is a vector of patient observables (dummies for patient 5-year age bins, sex, and race, Medicaid receipt, and 26 chronic conditions lagged by a year),[13] γ_p is a vector of treatment category fixed effects, and τ_t is a vector of calendar

[12] Since relative year 0 is defined as the year in the October of which the code change was implemented, I omit it from the specification due to the partial coverage involved and run the specification on years 1-7.

[13] These include acute myocardial infarction, Alzheimer's disease, atrial fibrillation, cataract, chronic kidney dis-

year fixed effects. As discussed above, the necessary assumption to interpret the coefficient β_1 as causal is selection on observables - conditional on the treatment category, older and younger physicians are similar on all relevant unobservable characteristics that may be correlated with adoption propensity. I explore the validity of this assumption by successively adding patient and doctor observables to my specifications below and showing that the results are not sensitive to doing so.

This specification models the decision to use a new procedure as a linear probability process. Specifically, each year increase in doctor age at codification shifts the *absolute*[14] probability of new procedure use by an equal amount across the age distribution, various treatment categories, and in all calendar years. The inclusion of treatment category and calendar year fixed effects controls flexibly for the baseline level of diffusion in the technological branch and secular changes in adoption across relative years.

The choice of model in the case of binary dependent variables has received much attention in the econometric literature (see, e.g., Angrist and Pischke (2008) for an exposition). The linear probability model has the advantage of offering easy-to-interpret estimates of the marginal effects of interest, while not taking a stance on the "correct" non-linear functional form to use. Nonlinear methods such as probit and logit have the advantage of having predicted probabilities in the unit interval and not inducing model-driven heteroskedasticity in the residuals. I choose linear probability for the baseline specifications due to its simplicity and transparency. However, in Table A3 in the appendix, I perform robustness checks showing the results from estimating the baseline specifications using logit. The resulting marginal effects are very close to those estimated by OLS.

1.5 Results

In this section, I document the negative relationship between age and new technology use. This relationship is robust to the inclusion of a variety of patient observables and stems from extensive-

ease, chronic obstructive pulmonary disorder, heart failure, diabetes, glaucoma, hip/pelvic fracture, ischemic heart dsease, depression, osteoporosis, rheumatoid arthritis/osteoarthritis, stroke/transient ischemic attack, breast cancer, colorectal cancer, prostate cancer, lung cancer, endometrial cancer, anemia, asthma, hyperlipidemia, benign prostatic hyperplasia, hypertension, and acquired hypothyroidism.

[14]One can model the effect of age as being dependent on the baseline rate of diffusion in each treatment category. This is what popular non-linear models such as logit and probit do. However, as discussed in this section, the average marginal effects computed in this way are not much influenced by the choice of model.

margin adoption decisions as opposed to intensive-margin usage choices. The age-based diffusion differences remain stable even after a few years have passed, suggesting that the age penalty is persistent rather than temporary. Finally, in order to interpret the magnitude of this effect, I perform two counterfactual exercises investigating the impact on overall diffusion of, first, changing everyone's adoption propensity to that of younger physicians and, second, of a few decades of physician aging at the rate of aging of the overall workforce.

1.5.1 Main results

Figure 1.3, which is a binscatter of the regression from specification (2) including the full set of controls offers a visual summary of the baseline results in this paper. The slope is -0.0019 implying that a 10-year increase in doctor age at the introduction of a new code is associated with a 1.9 percentage-point decrease in the probability of new code use. The effect is also economically meaningful - since the average probability of new code use in this sample is 16 percent, the marginal effect of a 10-year-increase in age translates to about 12 percent. The pattern in the relationship between age and adoption is is fairly linear. In particular, there is no evidence of an initial learning period where physicians improve on their adoption propensity. The shape of the age-adoption profile will receive further attention in the conceptual framework below. Finally, the effect of interest is present throughout the age distribution and is not driven by outliers in the data.

I proceed by showing the regression specification coefficients from running various versions of specification (2). This allows me to test for statistical significance and observe changes in the coefficients as more observables are added, which will serve as an indicator for selection. Table 1.2 presents the results from this analysis. Column (1) contains the baseline specification regressing the new procedure indicator on doctor age at new code introduction including treatment category and calendar year fixed effects. Standard errors are clustered at the treatment category level. The coefficient is statistically significant at the 5% level and indicates that a 10-year-increase in age at the introduction of a new code is associated with a 1.8 percentage-point decrease in the probability of new code use.

A major reason why older and younger physicians may differ in their adoption probabilities is if their patient populations are different in ways that are relevant for technological adoption. Columns (2) and (3) of Table 1.2 aim to evaluate the likelihood that this is a substantial issue in

the current setting. Column (2) adds indicators for 26 lagged chronic conditions, and column (3) - binary indicators for patient sex, race, 5-year-age-bins, and receipt of Medicaid - to the regression. The coefficients on doctor age remain essentially unchanged. This exercise demonstrates that the observable characteristics of patient populations for older and younger doctors practicing within the same treatment category do not appear to be differentially correlated with adoption propensity. This is, of course, no guarantee that there are no unobservable patient characteristics important for technology adoption that differ across older and younger physicians. However, it is an indicator that suggests that patient selection is unlikely to play an important role in this setting.

Another major set of counfounding factors are potentially unobserved doctor characteristics that are correlated with age and adoption. Most salient among these are the institutions where doctor practice and cohort-based changes in adoption propensity. Column (4) investigates this possibility by including fixed effects for doctor gender, AMA hospital, and medical school. Physician gender is a salient characteristic that has been changing across cohorts as more women have entered surgical specialties. Controlling for it is intended to test for the possibility that cohort effects play a major role by investigating the change in coefficient magnitude due to this addition. Similarly, hospital and medical school fixed effects are intended to control for doctor selection across medical institutions. These are imperfect controls as there may be other cohort characteristics that I do not observe that may be important for adoption and since physician selection across institutions may include a dynamic component that is not captured by controlling for hospitals at a point in time. Nonetheless, the fact that the coefficient remains essentially unchanged due to the addition of these characteristics indicates that, again, unobserved doctor characteristics may not be salient factors driving the age-adoption results.

Finally, column (5) breaks up the independent variable into 10-year bins in order to allow for non-linear effects in age in light of previous results that have found an inverse-U shape to relevant age-outcome profiles. The omitted category is doctors that are 30-39 years old at codification. The results indicate that the observed pattern is linear in nature as opposed to the hump-shaped profile observed in various studies in labor economics. One conjecture as to the source of this difference that I will not be able to test in the current setting is the role of medical school. In particular, while previous studies attribute the hump-shape to the competing effects of experience and general cognitive decline, the presence of structured residency training well into a doctor's practice career

may effectively cut off from observation the initial upward slope in outcomes due to procedural experience.

One way to interpret the magnitude of these results is in terms of a counterfactual. How much would the diffusion of medical procedures in the economy change if older physicians had the adoption propensity of younger doctors? This can be quantified by calculating the change in the use of new procedures for each of the age groups used in the analysis so far based on the estimates from the baseline specification (2). In particular, I calculate the change in the overall average usage of new technologies if each 10-year age group increased its probability of new-code utilization in order to match the usage among the youngest age group of 30-39 year olds. In practice, this involves taking an average of the estimated age group regression coefficients from column (5) in Table 1.2 weighted by the fraction of patients seen by each age group. The majority of patients in the sample (over 130,000) are seen by physicians in the 40-49 and 50-59 age groups, so these estimates have an outsize influence on the final calculation. I find that the absolute probability of new code usage will increase by 3 percentage points, which represents a 19-percent increase in the rate of new-technology usage.

1.5.2 Extensive vs intensive margin

Extensive margin

The nature of the age-based barriers to adoption can be further characterized by determining whether it originates at the point of adoption (extensive margin) or in the secondary stage of using the procedure conditional on adopting. In particular, the outcome in Table 1.2 combines both intensive and extensive margins as it is influenced both by whether physicians are more likely to use the new code at all and, if they do, by how intensively they use it. One way of disentangling the two is to investigate the effect of age on any use of the procedure (adoption) by physicians. I do this by collapsing dataset to doctor-treatment category level and asking the question of whether physicians that are older at codification are less likely to be seen using the procedure at all in a specific period of time after new code introduction. In particular, I run the following specification:

$$y_{pd} = \beta_0 + \sum_b \beta_{1b} Age\,Bin(b)_{pd} + \beta_2 X_d + \gamma_p + \tau_{t(p)} + \varepsilon_{pd} \tag{3}$$

where y_{pd} is a dummy for whether doctor d is observed using the new code in treatment category p in 1,3,5, or 7 years, $Age\,Bin_{pd}$ are 10-year age bins, X_d is a vector of physician observables, including physician sex, 2014 hospital ID, and med school ID dummies, and $\tau_{t(p)}$ are year dummies for the year t when the new code in treatment category p was introduced. One barrier to this approach is the fact that I only observe 20% of the Medicare beneficiaries that a doctor sees, or in other words, a small fraction of all the procedures that a doctor performs. This introduces a large amount of measurement error on the left-hand side, which will decrease the power of this analysis. In order to mitigate this problem, I keep only doctors that have at least 5 procedures of the respective treatment category in sample.

Table 1.3 shows the results from this exercise, where the outcome in columns (1)-(4) is whether the physician uses the new code in 1,3,5, or 7 years respectively. The results display the same linear pattern where older age groups are less likely to use the newer codes at all and indicate that the barriers to adoption seem to occur in the initial learning stage of the procedure. Comparing the results in these columns to column (5) of Table 1.2 indicates that the extensive-margin gradient is, if anything, potentially steeper than the combined one suggesting that while there is a large discrpancy between older and younger physicians in terms of learning a new procedure, older physicians are just as likely or potentially more likely than young physicians to use it conditional on adopting.

Figure 1.4 illustrates these results graphically by plotting the coefficients from the four regressions side by side. The figure makes two things salient. First, the linear pattern in adoption is preserved across different ways of defining the period of extensive margin adoption. Second, most of the technological diffusion occurs within 3 years of codification with only small increases in the spread of these new procedures after that.

Intensive margin

Given the results from the previous subsection, it may be useful to confirm that intensive-margin differences between older and younger physician adopters are less pronounced. I do this by running the baseline specification (2) on the subset of adopters in the years after they have been observed

using the new code for the first time.[15] The results are in Table 1.4 and Figure 1.5 respecitvely. Table 1.4 repeats the regressions in Table 1.2 on the specified subsample and Figure 1.5 is a binscatter of the regression including the full set of controls in column (4). It is easy to see that while there is still a negative relationship between age and new-code usage, the differences are much diminshed in both absolute and relative terms. The binscatter confirms that the slope is nearly flat. This analysis confirms that differences in the intensity of new-code usage conditional on adopting are less important in this context than extensive-margin adoption decisions.

1.5.3 Dynamic patterns

The results in Section 1.5.1 combine the effect of age on new code use over all years after codification within a treatment category into a single estimate. We may be interested in the dynamic patterns of this effect. For instance, if older doctors are less likely to adopt in the first few years after codification, but then rapidly catch up, this pattern will still show up as an age-adoption effect in the baseline estimates, but may merit a different policy response than if older physicians were continually falling behind younger physicians over all years post introduction. One way of investigating the dynamic patterns of adoption is to look into the differential effect of year relative to codification for younger and older physicians by running the following specification:

$$y_{ipdt} = \sum_t \beta_{1t} RY_{t(p)} xDA_{dp} + \sum_t \beta_{2t} RY_{t(p)} + \beta_3 X_{idt} + \gamma_p + \tau_t + \varepsilon_{ipdt} \tag{4}$$

where $RY_{t(p)}$ are relative year indicators for year relative to introduction for treatment category p, DA_{dp} is doctor age at codification for treatment category p and the set of controls is the full set of controls from column (5) in Table 1.2 above (i.e., including both patient and doctor observables, as well as calendar year and treatment category fixed effects). Since relative year 0 is defined as the year in the October of which the code change was implemented, I omit it from the specification due to the partial coverage involved and run the specification on relative years 1-7. The omitted category is relative year 1. Intuitively, this specification estimates the relative-year new-code use profile for different age groups and takes differences across age groups within relative years to

[15] As has been remarked in the past, even if baseline specification (2) satisfies selection on observables, this is not necessarily true in the case of an intensive-margin regression since I am conditioning on an outcome, namely extensive-margin adoption (Angrist and Pischke (2008)). Hence, this part of the analysis should be thought of as an accounting exercise.

evaluate how the effect of age changes as a new code goes through its diffusion cycle.[16]

The results are displayed in Figure 1.6 below where I have plotted the coefficients β_{1t} from the regression above. It is apparent that the age difference arises immediately at codification (which is unsurprising given the advanced stage in technological diffusion that these procedures are likely to be in) and then *increases* as time goes by. The estimates suggest that a 10-year increase in age decreases the probability that a physician will use the new code in relative year 3 by slightly over 4 percentage points while it does so by only 1 percentage point in relative year 1. This indicates that older physicians are not catching up to and are in fact falling behind younger ones as a new code goes through its diffusion process.

1.6 Conceptual framework

In order to interpret the patterns found in the data, I construct a single-agent technology adoption model that incorporates the standard framework on the effects of aging from psychology and economcis research. I begin with a multi-period economy with J doctors indexed by j. At the start of each period doctors face a probability $p(a_j)$ of retiring, where a_j is doctor j's age (normalized to the number of periods a doctor has been present in the sector) and $p(a_j)$ increasing in a_j. Retiring doctors are replaced by young doctors who start at age $a_j = 0$. The economy has a finite set of available surgical procedures S with each doctor performing a subset S_j of S and individual procedures being indexed by s. Each period a doctor performs procedure s on n_j^s patients.

Each period, a new treatment technology available for a randomly chosen procedure s enters the market. Adoption costs c_j^s for new technology in procedure s consist of two components. The first is a "systematic" component σ_j^s that represents aspects of the task that require skills that are are transferrable from the doctor's previous experience with procedures in this surgical category. Examples include using tools and devices specific to this surgical category or knowledge of the relevant anatomic area. This component of adoption costs represents the popular idea that there are returns to experience and that "experiential capital" (in the terminology of Agarwal et al. (2009))

[16]To show the intuition behind this specification, I plot, in Appendix Figure A1 the coefficients from a regression of new code use on relative year using the same controls performed separately for physicians 30-39 at codification and physicians 60-69 at codification. The coefficients of interest β_{1t} in (4) take the difference between the two curves in each realtive year and average this difference across all ages. Even in that figure, one can see the increasing difference between older and younger physicians as time goes by.

has positive effects on performance.

The second component of adoption costs, ι_j^s, consists of "idiosyncratic" skill requirements that do not benefit from previous experience in this surgical category. Instead, these adoption barriers come from novel concepts or methodologies that a surgeon needs to internalize. Examples include the use of new surgical tools or learning how to operate on a new organ within the same anatomic area. This component of adoption costs represents the idea within psychology that one's overall ability to deal with novel concepts and challenges is useful in learning how to perform complex tasks.

These two components of adoption costs showcase the idea that there are benefits and costs to the aging process. As one gets older, one gains experience with the world and the specific tasks she engages with and this experience pays off in terms of performance (crystallized intelligence). At the same time, one's ability or willingness to engage with new concepts and challenges diminishes, leading to losses in one's general analtical capabilities (fluid intelligence). This trade-off implies that the quality of one's decision making follows an inverted-U shape. The sweet-spot in terms of decision-making and ability to deal with challenges has been identified as middle age in previous studies.

I formalize the two adoption cost components in the following way:

$$c_j^s = \sigma_j^s + \iota_j^s \tag{5}$$

$$\sigma_j^s = f(e_j^s) \tag{6}$$

$$\iota_j^s = \alpha(a_j) + \varepsilon_j^s \tag{7}$$

where e_j^s represents surgeon j's previous experience with procedure s, and $\alpha(\cdot)$ is a function representing the impact of aging on surgeon analytic skill or fluid intelligence. Additionally, ε_j^s represents other randomly distributed unobserved shocks to the difficulty of learning procedure s by surgeon j, which includes unobserved procedure difficulty or match quality of the procedure with the surgeon. Note that since every surgeon starts off at age 0 and performs a constant number of procedures per period, then $e_j^s = n_j^s * a_j$, so that older surgeons using a particular procedure

s have more experience with that procedure. Lastly, I assume that $f'(e_j^s) < 0$, $\alpha'(a_j) > 0$, and $\alpha''(a_j) + f''(e_j^s) \geq 0$.

All of the major assumptions have standard interpretations within the fluid/crystal intelligence framework. The additive separability of σ_j^s and ι_j^s is a simplification that models crystal intelligence and fluid intelligence as two different processes that are separately affected by aging. This is a standard approach in the literature and while potential interactions between these two forces could be interesting in their own right, exploring them is beyond the scope of this study. The assumptions that $f'(e_j^s) < 0$ and $\alpha'(a_j) > 0$ simply formalize the opposite effects of age on crystal and fluid intelligence described above: older doctors have more experience, but may have lower analytical capital due to the aging process. The assumption on curvature, $\alpha''(a_j) + f''(e_j^s) \geq 0$, guarantees that there is a point in the age profile after which increases in age necessarily lead to lower adoption propensity.

The overall shape of the cost-age profile is determined by the exact form of $\alpha(a_j)$ and $f(e_j^s)$. For instance, Agarwal et al. (2009) assume that $\alpha(a_j)$ is linear with age and $f(e_j^s)$ is concave due to decreasing returns to experience. This yields a hump-shaped profile where performance initially increases with age until it hits a "sweet spot" in middle age after which performance declines. However, the exact shape and properties of these functions depend on the characteristics of the tasks of interest. For instance, if the novel tasks that are being learned are simply extensions of old tasks, one could imagine a world where idiosyncratic adoption costs do not change with age because analytic capital plays little role in the relevant context. In that case, one would expect a flat or even positive age-adoption profile as the dominant force in the adoption costs is $f(\cdot)$, the to experience. Similarly, if the tasks being learned are completely novel experience may have little impact on adoption costs and the age profile will be determined by $\alpha(\cdot)$, so one may expect to observe no gains to performance in early years and a generally downward-sloping age-adoption-cost profile.

Next, I turn to the benefit-side of the physician decision problem. The relative benefit for patient i treated by the new alternative in s is b_i where b_i follows a distribution $F_j(\cdot)$ with a positive mean (so the procedure is on average beneficial to patients). For simplicity, I assume that each procedure is affected by innovation only once. A physician using s has to decide whether to learn

the new technology affecting the procedure. A doctor's utility if she adopts is:

$$\left\{ \sum_{t=0}^{\infty} [\beta(1-p(a_j+t))]^t \sum_{i \in n_j^s} b_i \right\} - \left\{ c(a_j) \right\} \tag{8}$$

The first term shows the present discounted value of adopting for the doctor's patients with t indexing time periods and i indexing patients. The second term refers to the adoption costs for the technology. The expression above shows that a doctor will adopt if:

$$\varepsilon_j^s \leq \left\{ \sum_{t=0}^{\infty} \phi_t^j(a_j) \theta_i^j(n_j) \right\} - (\alpha(a_j) + f(n_j^s * a_j)) = T(a_j, n_j^s) \tag{9}$$

where $\phi_t^j(a_j) = [\beta(1-p(a_j+t))]^t$ is the discount factor adjusted for a doctor's individual retirement probability and $\theta_i^j(n_j^s) = \sum_{i \in n_j^s} b_i$ is total single-period patient benefit.

The implied age-adoption profile is determined by the exact assumptions on $p(\cdot)$ and $c(\cdot)$. On the benefit side, the current structure of the problem guarantees that $\frac{d}{da_j}\phi_t^j(a_j) < 0$. This comes from the fact that older physicians have less expected practice time, which means that the benefits of learning a new skill accrue to fewer patients. Since the adjusted individual discount factor ϕ_t^j decreases with age due to increased retirement probability, older doctors will also find it less beneficial to adopt a new technology. More subtly, doctors of the same age that are less exposed to innovation due to the fraction of their patients treated by the affected service are also less likely to adopt. $c(a_j)$ may increase or decrease initially depending on the relative importance of experience and analytic skill in adoptoin costs. This in turn means that the presence or absence of an initial increase in the propensity to adopt is an indicator for the nature of the new surgical procedures and the kinds of skill capital required to adopt them.

On the cost side, $c(a_j)$ may increase or decrease initially depending on the exact properties of $\alpha(a_j)$ and $f(n_j^s * a_j)$. This in turn is dependent on the nature of the skills that need to be acquired. Technologies that rely mostly on novel sets of skills that rely on analytical rather than experiential capital will likely have $\alpha(a_j)$ as the dominant force behind the age-adoption profile. This implies little gain to experience and a mostly negative slope to costs. Technologies that do not require much analytical capital or that rely mostly on previous experience to be adopted are likely to exhibit initial decreases in adoption costs due to age and thus an inverted-U shape to the age-

adoption profile, which is familiar from studies in other domains of economics and psychology.[17]

The baseline empirical results suggest that adopting new surgical procedures has high requirements of analytical capital relative to experiential capital. In particular, the patterns in the figures above show a mostly linear relationship between age and adoption with no apparent gains to experience in the younger age ranges. This negative and linear relationship is reminiscent of the negative and linear functional form on analytical capital degradation assumed by Agarwal et al. (2009) and implies that surgical procedure adoption requires the acquisition of novel skills that are not readily available from the surgeon's previous experience in a treatment category.

These findings have important implications for policy. They suggest that barriers to adoption do not arise from frictions in the ability of surgeons to apply skills that they already have to a relevant domain. This would be the case for instance if younger and older surgeons differed only in their knowledge of new procedures entering the market. Instead, the results indicate that the barriers to adoption have to do with a more fundamental mismatch betweem the skill-set of older surgeons and what is required to learn newer technologies. In the following section, I offer additional evidence that is consistent with this idea and other empirical predictions of the model.

1.7 Further empirical tests

In this section, I confront the model with the data in order to test two key empirical predictions - the type of adoption coss associated with a procedure are key for the shape of the age-adoption profile and age affects the present discounted value of patient benefits. First, I show that the age-adoption profile is significantly steeper for a subgroup of procedures that are considered novel and difficult to learn, namely minimally invasive procedures. Second, I repeat the baseline analysis in two healthcare technology adoption contexts where information rather than skill acquisition is likely to be the main barrier to adoption, diagnostic codes and prescription pharmaceuticals, the age-adoption profile is flat. Lastly, I use the conceptual framework to derive a simple test of the first channel that relies on including an interaction of physician age with a measure of her annual

[17]Notice that it is also possible to model the benefits of adoption as being dependent on the skill and experience components of aging. While this structure may generate some interesting patterns through interactions between those components and the probability of retirement, which is also dependent on age, the basic conclusion that experience-heavy technologies should push the age-adoption profile towards positive gradients and that analysis-heavy technologies should push it towards more negative values still stands.

arrival rate in the baseline specification.

1.7.1 Adoption costs

I proceed by testing the idea that the skill acquisition costs involved with technology adoption are likely drivers behind the shape of the age-adoption profile. In particular, the linearity in the documented age-adoption profile for surgical procedures indicates that technology adoption in this context requires mostly analytic capital and does not benefit much from previous experience with surgical procedures within the same treatment category. This suggests that the age-adoption profile should be flat in settings where previous experience is highly applicable or where analytic capital is not required. I test this implication of the model in two ways. First, I show that the age-adoptin profile is steeper for minimally invasive surgeries, which are popularly thought to be quite difficult to learn. Second, I explore two settings where the main technology-adoption barriers are likely to be informational in nature: usage of new diagnostic codes and of new prescription pharmaceuticals.

This distinction is of policy as well as theoretical interest. In particular, prior research has suggested that information can be a key barrier to the adoption of new technologies (Agha and Molitor (2018)). The adoption of new surgical procedures naturally requires both information and skill acquisition. Therefore, a possible policy response to the existence of an age-adoption profile would be to strive to inform surgeons better of the available procedures that are potentially relevant to them. However, taking the theory developed here seriously would suggest that in addressing the age-adoption profile as a policy concern through information interventions would only be beneficial if the informational barriers to acquisition grow harder to surmount with age. I will aim to show that this is not the case. In particular, even if information is important to technology diffusion in general, this importance does not vary systematically with age. Thus, a policy response to the existence of an age-adoption profile will have to focus on helping surgeons support the analytically-intensive skill-set they need to stay at the technological frontier.

Minimally invasive procedures

The theoretical model from the previous section predicts that as the difficulty of learning the new procedures increases in terms of analytic resources required, the age-adoption profile should become steeper. This is a straightforward testable prediction. In practice, however, it is difficult

to find a consistent measure of the difficulty of learning multiple new procedures. A simpler approach is to identify subgroups of procedures which are generally considered difficult to adopt and test whether the slope of the age-adoption profile changes within those subgroups. In the surgical context, multiple authors have identified minimally invasive approaches to common surgeries as a group of procedures which is difficult to learn and substantively different from its traditional open counterparts. In particular, minimally invasive surgeries (MIS), which perform the procedure through small incisions in the patient's body, offer restricted access and information flow to the surgeon as she needs to operate using specialized tools based on a video feed (Edwards and Bailey (2000), Sood et al. (2015), Bansal et al. (2016)). This suggests that a simple test of the model would be to check whether minimally invasive procedures have a steeper age-adoption profile than the average procedure.

I perform this test in two steps. First, I use the procedure's ICD9-CM description to identify code changes that pertain to the addition of minimally invasive approaches to the respective treatment category. In particular, I define these as code changes where the new code has the words "thoracoscopic", "laparoscopic", "endoscopic", or "bronchoscopic" in its description and where the old code in the treatment category does not. For instance, code "07.84 - Thoracoscopic total excision of thymus" was differentiated in 2007 from code "07.82 - Other total excision of thymus" and thus this code change is classified in the minimally invasive category due to the presence of "thoracoscopic" in the description of "07.84". Note that the entire treatment category (observations of both 07.82 and of 07.84) are included in the minimally invasive category. The resulting group has 28 codes in the minimally invasive category. Next, I run a version of the baseline specification that includes an interaction of doctor age with the minimally invasive group indicator:

$$y_{ipdt} = \beta_0 + \beta_1 1(MIS)_p \times A_{dp} + \beta_2 A_{dp} + \beta_e X_{idt} + \gamma_p + \tau_t + \varepsilon_{ipdt} \tag{10}$$

where $1(MIS)_p$ is an indicator for minimally invasive surgery as described above. The model predicts that $\beta_1$should be negative, meaning that the slope of the age adoption profile is steeper for the more difficult to learn MIS. The results from this exercise are presented in Table 1.5. Columns (1)-(4) successively add observables to the group of controls as in the baseline specification. The coefficient on β_1 is negative and significant as predicted by the model. The magnitudes involved

are also quite substantial - while β_2 implies that a 10-year increase in age is associated with a 0.6 percentage-point decrease in the probability of adoption, the magnitude of the gradient increases to 4.6 percentage-points for minimally-invasive surgeries. Therefore, the results from this exercise are consistent with the implications of the theoretical model. In particular, technological adoption that requires higher amounts of fluid intelligence has a steeper age-based gradient.

Diagnostic codes

I continue the tests of the model implications by looking at analogous settings where information rather than skill-acquisition is likely the most important barrier to adoption. A particularly attractive feature of the ICD9-CM diagnosis code setting is that its institutions are analogous to the ones that govern procedure codes. Speicifically, diagnosis codes are revised through biannual committee meetings, where new codes are differentiated from older designations to reflect advances in medical knowledge. Similarly to procedure codes, diagnosis codes are hierarchically organized in increasing specificity with regards to the body system and organ affected as well as the type of condition that is being recorded. Medical conditions branch out of a parental category as more is learned about a specific condition. This means that many of the foundations of the procedure analysis are preserved in the diagnosis code setting as well.

There are two prominent differences, however. First, diagnosis codes are used both in the outpatient and inpatient setting. This means that in principle the analysis would not be restricted to inpatient procedures only. Second, identifying the exact physician responsible for a particular diagnosis may not be straightforward. This is especially true in inpatient settings where multiple doctors are taking care of a particular patient.

In order to adjust for this feature of the environment, I focus on a specific setting where the physician responsible for a diagnosis can be easily identified, namely PCP office visits. In particular, I follow an analogous methodology to the one used for procedure codes in identifying diagnosis codes which come from a single parental code which continued being used even after the new code introduction. I take code introductions between 2001 and 2013 and look at the primary diagnosis on Medicare provider claims in the same period submitted for services performed by PCPs in an office setting. The resulting sample has over 5 million observations with 137 thousand doctors in almost 300 treatment categories. More sample construction details are presented in the appendix.

I run regressions of the following form analogous to the one used in the main analysis above:

$$y_{ipdt} = \beta_0 + \beta_1 A_{dp} + \beta_2 X_{idt} + \gamma_p + \tau_t + \varepsilon_{ipdt} \tag{11}$$

where all relevant variables have definitions identical to the one from above. The outcome of interest y_{idpt} is whether the primary diagnosis on the claim is one of the old or new codes post codification of a new diagnosis. The results are presented in Table 1.6 below. The age-adoption profile for new diagnosis codes is essentially flat and precisely estimated. The table indicates that there are no statistically significant differences between younger and older doctors in their propensity to use the new diagnostic code within a diagnostic category. The addition of various observables does not affect the magnitude of the estimates, which is consistent with the procedure analysis from above. Furthermore, the presented estimates are statistically significantly different from the estimates for medical procedures suggesting that the lack of statistical significance is not due to statistical noise.

The appendix contains a binscatter for this regression (Figure A5). It also presents some of the summary tables and figures that correspond to the diagnosis code sample (Figure A4, Table A6). The most notable feature is that the diffusion of these new codes is much more muted than the one for procedure codes. There is only a slight upward trend, indicating that the entry of new diagnosis codes occurs at the flat portion of the S-curve of the diagnostic technology.

Overall, the results point to two conclusions. First, technologies where the adoption barrier is information costs seem to have a much flatter age-adoption profile. Second, interpreted through the lens of the model, the results imply that there is little gain to experience or penalty to aging with respect to adopting new diagnostic codes. This is consistent with the finding that procedures that require a higher skill investment exhibit a steeper age-adoption profile. I turn next to testing this hypothesis in another setting where information is likely to be important - prescription pharmaceuticals. [18]

[18] One caveat to this analysis is that it is agnostic about the role of relative patient benefits in diagnostic versus procedure codes. In particular, an additional factor that could lead to a flatter age-adoption profile in the case of diagnostic codes is lower relative patient benefits to receiving the correct new diagnosis as opposed to the incorrect old diagnosis. While this is certainly a possibility, the small magnitude and statistical insignificance of the estimates makes it difficult to draw conclusions from an interaction-based analysis similar to the one employed in the case of procedure codes above.

Prescription pharmaceuticals

I continue the investigation of the role of information as an adoption barrier by analyzing age and adoption in prescription pharmaceuticals. Existing research on pharmaceutical markets suggests that the main barriers associated with the adoption of new drugs are largely informational in nature. This may include knowledge that a new drug exists, knowledge of its benefits relative to potential competitors, as well as knowledge of what patients are likely to benefit from its prescription (Dickstein (2018)). Therefore, a test of the effect of age on new drug adoption in this setting will reveal the extent to which age impacts the adoption barriers associated with informational frictions as opposed to skill acquisition.

Pharmaceutical markets are also attractive because drugs are organized in therapautic classes that offer a compelling analogue to the branching structure of ICD9-CM procedure codes. In particular, therapeutic classes are groups of pharmaceuticals that have similar chemical composition and physiological pathways of action. This means that drugs in the same therapeutic class are likely to be used on a similar population of patients by the same physician.[19] However, a major difference would be that in the case of pharmaceutical markets, a new drug gets its own designation immediately upon approval and thus its diffusion process can be monitored from the time it was first marketed, while in the case of procedures, this is infeasible.

I use introductions of new chemical compounds into existing therapeutic classes to investigate the effect of age on pharmaceutical adoption. I use Medicare Part D Event files for 2006-2015 and define new technology as a new compound introduced in an already existing therapeutic class. An important difference with the procedural analysis is that there are multiple chemical compounds per therapeutic class introduced between 2006 and 2015. I choose to run the analysis by defining the use of new technology as the use of any new compound within a therapeutic class. Hence, in order to measure technological adoption propensity, I will be comparing the use of drugs introduced in or after 2006 to drugs introduced before 2006. The risk set of physicians is defined as doctors who were prescribing in the relevant therapeutic class in the year of the new compound's introduction.

The final sample has 55,195,851 observations by 359,965 physicians in 58 therapeutic classes over the 10-year period. Sample construction details are in the appendix. Consistently with the

[19] Nonetheless, just like in the case of procedures, drugs in the same therapeutic class are not perfect substitutes for each other (Johansen and Richardson (2016)).

procedural analysis from above, I run regressions of the following form:

$$y_{icdt} = \beta_0 + \beta_1 A_{dc} + \beta_2 X_{idt} + \gamma_c + \tau_t + \varepsilon_{icdt} \tag{12}$$

where i indexes patients, d indexes operating physicians, c indexes therapeutic classes, and t indexes calendar years. y is a dummy for whether a new (post-2006 approval year) or old (pre-2006 approval year) compound within a therapeutic class is used, A_{dc} is prescribing physician d's age in the year when therapeutic class c first receives a new drug , X is a vector of the same patient observables as above (dummies for patient 5-year age bins, sex, and race, Medicaid receipt, and 26 chronic conditions lagged by a year), γ_p is a vector of treatment category fixed effects, and τ_t is a vector of calendar year fixed effects.

The results are presented in Table 1.7. Column (1) has the baseline specification without any patient observables, column (2) adds lagged chronic condition indicators and column (3) - binary indicators for patient sex, race, 5-year-age-bins, and receipt of Medicaid. Column (4) breaks out the age variable into 10-year bins. The estimates are statistically insignificantly different from 0 and quite precisely estimated in all columns. The result in column (1) sugests that a 10-year incrase in doctor age in the year a therapeutic class receives an innovation is associated with a 0.01 percentage point decrease in the probability of prescribing one of the new drugs that enter the therapeutic class in the sample period. The standard errors, clustered at the therapeutic class level, allow for the rejection at the 5% level of results as high as a 0.05 percentage-point decrease. It is easy to see that the effect of age on adoption propensity in this context is more than 100 times smaller in absolute terms than its effect in the context of ICD9-CM procedures. Of course the mean level of new-drug use in this context is much lower at 1.7% partly because each therapeutic class can have a large number of chemical compounds in it. However, even taking this base level into account the relative magnitude of the age effect is more than 10 times lower than in the case of procedures. Finally, the results are consistent across columns, again suggesting that patient selection is not important in this setting.[20]

Overall the results from this exercise suggest that the effect of age on adoption is mediated through the high fixed costs of skill acquisition in the context of medical procedures. Older physi-

[20] Since all of these drugs are follow-on drugs in their respective therapeutic classes, these results are also consistent with the findings of Glass and Rosenthal (2004), who find no effect of age on the adoption of follow-on drugs.

cians are not at disadvantage in obtaining new information about existing treatments. However, they may be less willing to invest in the skills necessary to learn new procedures. This is an important consideration from a policy perspective because it indicates that simply spreading awareness about new technological options is unlikely to have much of an effect on diffusion. Policy-makers interested in boosting the adoption of new procedures have to take into account the associated skill-acquisition costs faced by physicians in this space.

1.7.2 Present-discounted value of patient benefits

The empirical tests so far explored variation in the adoption costs side as a test of the theoretical framework. However, the model also posits that age can affect adoption through its impact on the benefit side of the physician's decision. In particular, increases in age lead physicians to adopt less because of decreases in the present-discounted value (PDV) of the benefits that will accrue to their patients due to the new technology. The model shows that this channel is affected by the physician's arrival rate - the number of patients she expects to see in her practice over the lifetime of her medical activity. As the number of patients increases, the physician is more likely to take on the fixed costs associated with technology adoption. This can be seen in equation (9) above:

$$\frac{d}{dn_j^s}T(a_j,n_j^s) = \frac{d}{dn_j^s}\theta_i^j(n_j)\left\{\sum_{t=0}^{\infty}\phi_t^j(a_j)\right\} - a_j f'(n_j^s * a_j) > 0 \tag{13}$$

where the sign comes from the fact that $\frac{d}{dn_j^s}\theta_i^j(n_j) = \frac{d}{dn_j^s}\sum_{i\in n_j^s} b_i > 0$. More subtly, however, changes in the arrival rate n_j^s also change the marginal return (or penalty) to an additional year of age in the model. This can be most clearly seen in the cross-partial derivative:

$$\frac{\partial^2}{\partial n_j^s \partial a_j}T(a_j,n_j^s) = \frac{\partial}{\partial n_j^s}\theta_i^j(n_j)\frac{\partial}{\partial a_j}\left\{\sum_{t=0}^{\infty}\phi_t^j(a_j)\right\} - f'(n_j^s * a_j) - a_j^2 f''(n_j^s * a_j) \tag{14}$$

The sign of this expression is ambiguous. In particular, the effect on the marginal return to an additinoal year of age is negative since $\frac{d}{dn_j^s}\theta_i^j(n_j) > 0$ and

$$\frac{\partial}{\partial a_j}\left\{\sum_{t=0}^{\infty}\phi_t^j(a_j)\right\} = \frac{d}{da_j}\left\{\sum_{t=0}^{\infty}[\beta(1-p(a_j+t))]^t\right\} =$$

$$= \left\{ \sum_{t=0}^{\infty} [\beta t(1-p(a_j+t))]^{t-1} [-\frac{d}{da_j} p(a_j+t)] \right\} \tag{15}$$

which is negative since $\frac{d}{da_j} p(a_j+t) > 0, \forall t$.

However, $f'(n_j^s * a_j) < 0$ and $f''(n_j^s * a_j) \geq 0$, so that $[-f'(n_j^s * a_j) - a_j^2 f''(n_j^s * a_j)]$ may be positive or negative depending on the relative size of these effects. The results from the baseline regression indicate that the age-adoption profile is roughly linear, so that $f''(n_j^s * a_j)$ is likely to be small in magnitude. Therefore, equation (14) can be approximated by:

$$\frac{\partial^2}{\partial n_j^s \partial a_j} T(a_j, n_j^s) = \frac{\partial}{\partial n_j^s} \theta_i^j(n_j) \frac{\partial}{\partial a_j} \left\{ \sum_{t=0}^{\infty} \phi_t^j(a_j) \right\} - f'(n_j^s * a_j) \tag{16}$$

That is, increases in the arrival rate increase the penalty of age through the present-discounted value of benefits channel and decrease it through the returns to experience channel. Intuitively, for physicians with many patients, holding all else equal, the return to having an additional expected year in practice is higher since more patients will benefit from this opportunity to use the new technology. This in turn means that younger physicians should be much more likely to invest in the fixed costs required for technology adoption. At the same time, however, if the returns to experience are high, it is possible that a higher arrival rate leads to a shallower age-adoption profile due to the gains from performing more surgeries.

This prediction of the model offers an opportunity for a simple empirical test. In particular, including an interaction of a measure of a doctor's arrival rate with age at codification in the baseline specifications can tell us whether the benefits channel is likely to be important in this setting. Specifically, if the coefficient on the interaction is non-negative or non-significant, then the PDV channel is overwhelmed by the experience channel and thus the negative slope of the age-adoption profile is driven by the adoption-cost sie of the model. However, if the coefficient on the interaction is negative and significant, then the PDV channel is strong enough to show through even the decreases in the age-penalty due to increased experience. While this may not allow for an exact measurement of the magnitude of the benefits channel, it is an indication of its relative importance.

The main challenge for this exercise is choosing a measure of the arrival rate. In what follows, I use the average number of annual procedures that a doctor is observed performing in the sample

period. This is not the only way to measure the arrival rate, but is attractive for two main reasons. First, it is simple, transparent, and easy to interpret. Second, the sample I'm working with is a 20% sample of Medicare beneficiaries. This means that I observe only a small fraction of the actual procedures performed by any single physician. This means that using a more granular measure of arrival rate (e.g., number of patients in a set number of periods before or after adoption or within a particular treatment category) may be exceedingly noisy due to the relatively small number of per-physician observations in the sample. To illustrate, 10th, 25th, 50th, 75th, and 90th percentiles of this variable are 1.33, 2.14, 3.71, 6.63, and 10.62 respectively.

I re-run the baseline specifications from above in the following updated form:

$$y_{ipdt} = \beta_0 + \beta_1 A_{dp} * N_d + \beta_2 A_{dp} + \beta_3 N_d + \beta_2 X_{idt} + \gamma_p + \tau_t + \varepsilon_{ipdt} \tag{17}$$

where N_d is the average number of annual procedures that a doctor is observed performing in the sample period re-centered so that N_d is mean-0 (i.e., to form N_d I subtract from average number of annual procedures the overall mean for each doctor). This is done so that the coefficient β_2 can be interpreted as the age gradient at the mean of the arrival rate distribution as opposed to an arrival rate of 0 (which is not even in the sample). I compute the average by excluding from the total number of physician procedures, the number of new procedures performed by the surgeon in order to avoid a mechanical relationship between the dependent variable and the interaction regressor of interest. The results are shown in Table 1.8 below.

Columns (1)-(4) consecutively add observables as done in the baseline case above. Column (5) interacts the age bins with N_d instead of a linear age term. The coefficient on age at codification is of similar magnitude and significance to what was found in the baseline case. The coefficient on N_d in column (4) implies that an additional procedure performed by a doctor on an annual basis increases the probability that a physician will adopt a new procedure by 1 percentage point. The coefficient on the interaction of interest β_1 is negative as predicted and significant at the 10% level in all specifications. The magnitude implies that an additional 10 annual procedures increases the doctor age gradient by 0.1 percentage point, which is almost a 60% increase in the size of the baseline coefficient.

Despite the significant role for doctor benefits in the age-adoption gradient, the results from this

estimation indicate that PDV can only account for part of the observed effect of age on adoption. In particular, the baseline estimate of the size of the age-adoption gradient is 0.17 percentage points decrease in adoption propensity for each year increase in doctor age. If I assume age has no effect on adoption costs and that the effect is fully mediated through this channel, doctors would need to perform 17 in-sample procedures on average in order for the age gradient to be fully accounted for by this channel. This number of procedures is past the 95th percentile in the analysis sample, indicating that it is unlikely that the benefit-side of the model can account fully for the age-adoption gradient. Nevertheless, this analysis has shown that patient benefits do seem to be important and that the predictions of the model regarding their influence do seem to be borne out in the data. Finally, these findings are in line with the linearity of the age-adoption profile, which, as noted above, suggests that experience gains are not important in this environment.

1.8 Robustness

In this section, I consider robustness of the baseline results to various aspects of the empirical research design. I begin by testing the cohesiveness and technological similarity of the procedures included in the treatment category definition. I then proceed by investigating reasonable changes in the analysis sample. First, I consider changing the set of technologies considered based on the number of procedures observed. Second, I offer two related but different definitions of the risk set and show that the results are robust to these changes.

1.8.1 Treatment category cohesiveness

I begin by testing the idea that medical procedures that are in the same treatment category constitute technological substitutes. In particular, my analysis is based on the assumption that procedures that used to be designated by the same code are an appropriate technological space and the physicians that practice within a treatment category are "at risk" for adoption. This assumption can be false if treatment categories contain disparate procedures which are lumped together without regard to their medical similarity or technological profile. I test this idea in three ways.

The first test of this idea is a simple descriptive exercise. The basic idea is to use other available medical information in a patient's record in order to evaluate whether the patients falling in

the same treatment category tend to be medically similar to each other. If patients falling within the same treatment category have similar medical profiles, this would indicate that the likely procedures performed under that treatment category are also medically similar. A parsimonious way of accomplishing this is to use information on the DRGs for inpatient stays. As discussed above, DRGs are used to categorize patients into reimbursement-relevant groups whose members share similar medical conditions and require similar resource intensity. While the specific surgical procedure performed on the patient plays a role in determining her DRG, the classification relies heavily on patient diagnoses and other medical information such as the presence of relevant comorbidities and is therefore an independent source of information on a patient's medical condition.

In Appendix Figure A2 below (construction details in the appendix), I show that the majority of patients in the same treatment category fall in the same DRG indicating that patients in the same treatment category are indeed similar in their medical conditions and that treatment categories do not tend to contain procedures that treat vastly different medical conditions. Additionally, as explained in the appendix, I find that the vast majority of patients treated using new procedure codes fall within DRGs that are observed with the old procedure codes. This indicates that on average hospital reimbursement tends to stay the same for new and old procedures removing this particular channel as a possible mechanism of action behind the age-adoption profile documented in this study.

The second test of the treatment category as a cohesive group of procedures is use the procedure description to exclude from the analysis sample treatment categories that appear like they may contain disparate groups of procedures. In particular, I exclude all treatment categories which mention the word "other" in their description. Examples are "open and other cecectomy" and "other operations on cul-de-sac". The idea is that treatment categories with non-specific descriptions may be more likely to include unrelated procedures. The results from running the baseline specification with a full set of controls are in column (1) of Table 1.9 below. The doctor-age estimate is negative, significant, and close in magnitude, both in absolute and relative terms, to the baseline estimates. This indicates that the results are not driven by treatment categories that may contain disparate procedures, at least based on their description.

The last test of treatment category as a cohesive group of procedures is based on the hierarchical structure of the ICD9-CM and the tendency of the system to group medically similar procedures

close to each other. Specifically, I restrict the analysis to treatment categories where the new procedure code is in the same two-digit category as the old code. This includes for instance procedures such as "34.51 - Decortication Of Lung" and its corresponding new code "34.52 - Thoracoscopic Decortication Of Lung" because they both fall within two-digit category "34 - Operations On Chest Wall, Pleura, Mediastinum, And Diaphragm." This procedure excludes cases where the new code falls outside of the parent procedure's two-digit category. An example is "88.75 - Diagnostic Ultrasound Of Urinary System" and its corresponding new code "00.25 - Intravascular Imaging Of Renal Vessels" because the former falls within category "88 - Other Diagnostic Radiology And Related Techniques" and the latter - within "00 - Procedures And Interventions, Not Elsewhere Classified." The idea is that procedures which receive new designations far from the original's position within the hierarchical system are more likely to be unrelated to the other procedures in the treatment category.

The results are listed in column (2) of Table 1.9 below. The doctor-age estimate is still negative, significant, and close to the baseline results. This indicates that the results are not driven by code-changes where the new code is in a different two-digit category than the original. Together with the results from column (1) this analysis indicates that treatment categories are likely to be cohesive groups of procedures.

1.8.2 Sample definition

This subsection tests the influence of the partial sample of Medicare patients observed on the baseline results. In particular, since I observe only a 20% sample of Medicare beneficiaries, I see only a fraction of the procedures performed within the treatment categories of interest. If the only effect of this feature of the underlying data is to induce classical measurement error in the dependent variable of interest (i.e., I observe the true adoption rate of physicians within a treatment category with a linearly additive independent error term), then this has no bearing on the bias or consistency of the estimates in my specifications. However, one may be worried for instance that the probability of a new procedure being observed is correlated with the age of the physician (e.g., older physicians are worse at coding new procedures using the appropriate code). This is a general problem that would be exacerbated in an environment where some treatment categories have only a few observations (e.g., less than 20 procedures in the entire sample) since it may create "outlier"

treatment categories that might drive the results.

One way of addressing this issue is to restrict the sample to treatment categories which have a relatively high number of observations. The idea is that larger treatment categories are less likely to experience the measurement error I outlined above and are also less likely to have that measurement error be correlated with the doctor's age since these categories may be financially more important to the hospitals in terms of Medicare reimbursement. In column (3) of Table 1.9, I restrict the sample to treatment categories that have at least 500 procedures in the sample frame. The results are statistically significant and close to those from the baseline specifications indicating that small treatment categories are not driving the results.

Second, I test for robustness to the risk set definition. As outlined above, the baseline sample defines the risk set for each treatment category as the set of doctors who are observed using the treatment category in the year of codification. This definition could be problematic due to the fact that codification occurs in October of each year, so that this risk set includes some physicians who were using the parent code in the months leading up to October and some who were using it after October possibly as a result of the availability of the new code within that category. I therefore check how the results change if I change the definition to doctors who are observed using the treatment category one or two years prior to codification. This significantly changes the sample for two reasons. First, physicians may exit the labor force, the treatment category, or treatment of Medicare patients in the years leading up to codification, so that the set of physicians observed using the treatment category farther back in time is increasingly different from the set of physicians observed using the treatment category at and immediately after codification. Second, the 20% sampling issue means that only a fraction of the procedures that a doctor uses are observed in any given year, so that the doctors who are observed using a given treatment category in 2007 and 2009 for instance is different from the set of physicians observed using the same treatment category in 2006 and 2009.

The results are in columns (4) and (5) in Table 1.9. Column (4) changes the risk set definition to those physicians observed using the procedure in the year prior to codification and column (5) - to those physicians observed using the procedure two years prior to codification. Both results are significant and close in magnitude to those of the baseline specification. As outlined above, the sample decreases substantially with these changes leading to column (5) being significant only at

the 10% level. Still, the results from those columns indicate that the results are not sensitive to the definition of the risk set.

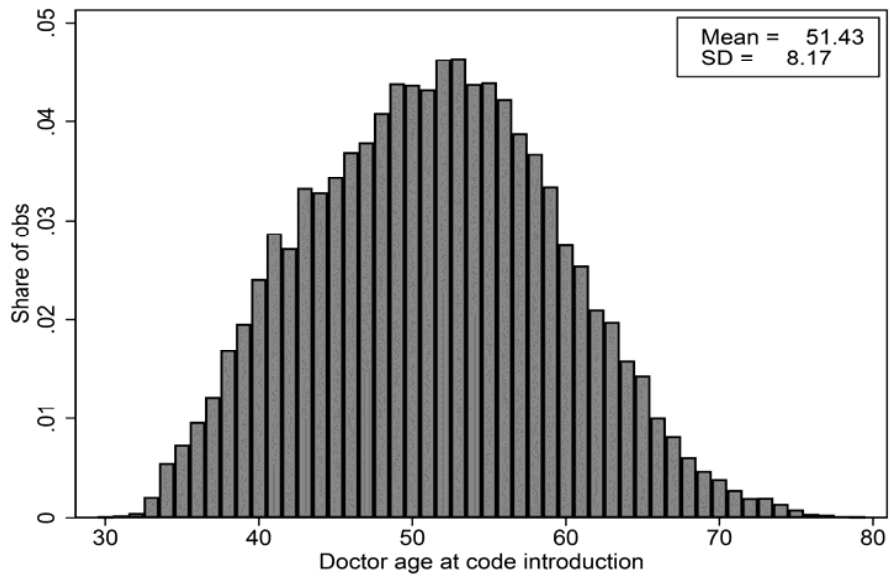
1.9 Conclusion

This paper considers the effect of physician age on the propensity to adopt new technology in the form of medical proceures. Relative to previous literature on the subject, it proposes a method of analyzing a large number of technologies undergoing diffusion with a well-defined risk-set of doctors who may consider adopting the technology and who are likely to be similar on salient unobservable characteristics that may bias estimation. The results indicate that younger doctors are more likely to use new procedures, that this difference is largely explained by extensive-margin decisions whether to learn the procedure in the first place, and that the difference only increases as the diffusion process plays out. Incorporating the arrival rate of new procedures in the analysis implies that the present discounted value of patient benefits is an important channel that explains some, but not all of this effect. An analogous analysis of diagnostic codes and pharmaceuticals suggests that the age-adoption profile is likely highly influenced by the fixed costs of skill acquisition required for the adoption of new medical procedures and that informational barriers are likely to play a negligible part in this setting.

It is important to stress that these findings do not take a stand on welfare. As discussed above, older physicians may face real higher costs and lower benefits of adoption, which should induce lower adoption. To the extent that physicians are fully internalizing the expected patient benefits that arise from the use of new technology and weigh that against their own real costs of adoption, the results here are not indicative of an inefficiency in the market. However, to the extent that a policy maker believes that new medical technology has patient benefits that may not be internalized by the physician (or that the market has multiple equilibria due to technology spillovers), this paper suggests that the most effective interventions intended to accelerate the diffusion process are interventions that target the skill-acquisition costs associated with learning new procedures such as structured training classes and increased financial support for learning procedures.

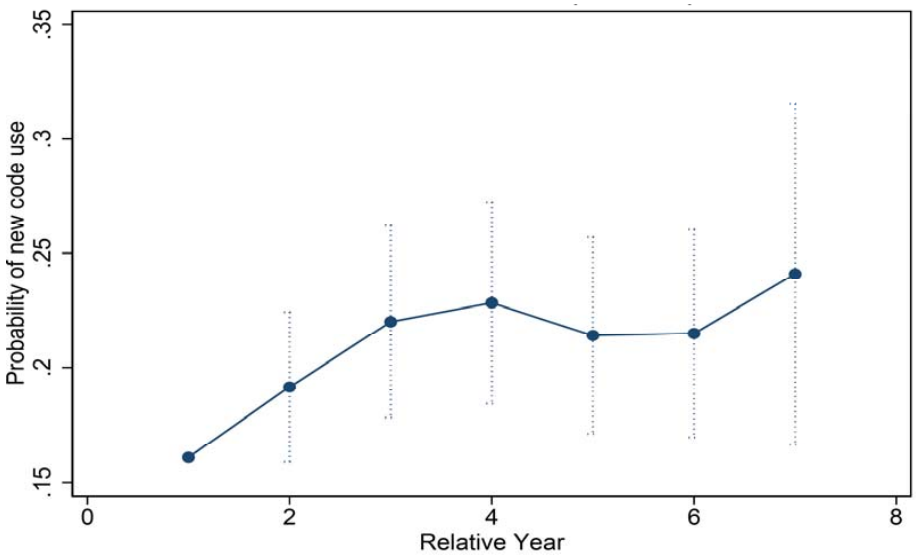
1.10 Figures and Tables

Figure 1.1: Distribution of Doctor Age at Codification



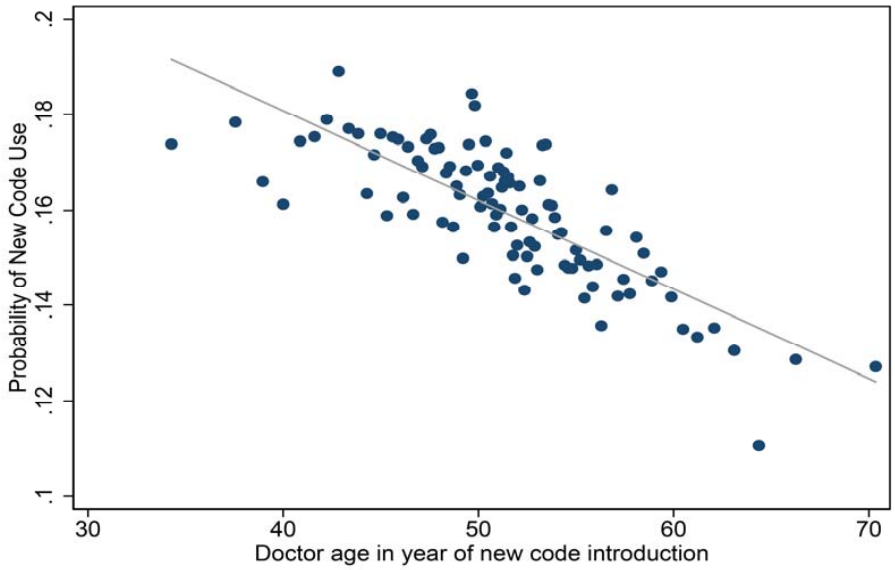
Notes: The histogram displays the distribution of doctor age at new code introduction for all treatment categories. Each observation represents a procedure on an inpatient claim.

Figure 1.2: Diffusion S-Curve



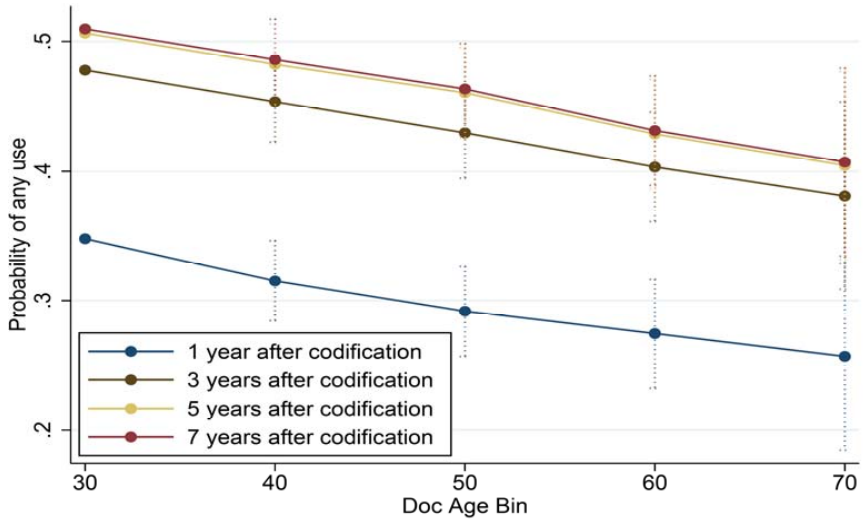
Notes: The figure plots the coefficient from a regression of new code dummy on year relative to new code introduction with treatment category fixed effects and controls for calendar year, patient age, race, sex, chronic conditions, and Medicaid receipt, and doctor controls for AMA hospital ID, med school ID, and gender.

Figure 1.3: Baseline Regression Binscatter



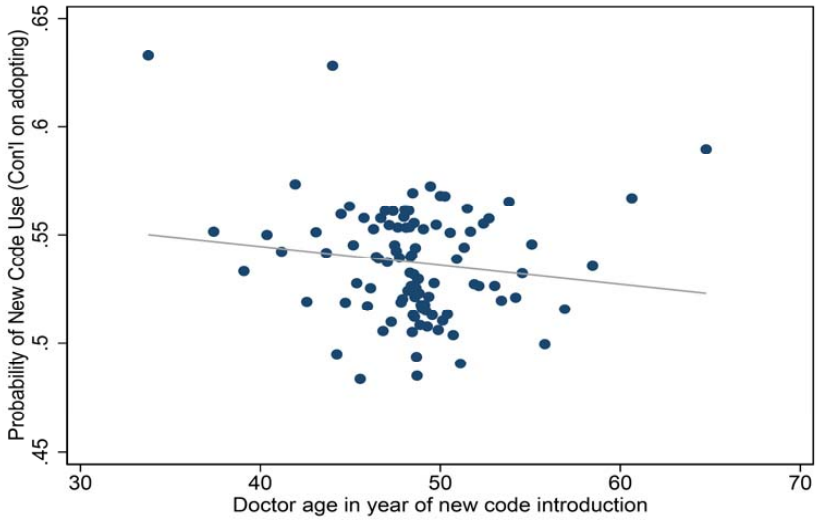
Notes: The figure is a binscatter of the baseline specification (4). The figure is constructed by estimating the residuals from separate regressions of the new code dummy and doctor age on the full set of patient and physician controls, dividing them into quantiles, adding the means of the dependent and independent variable respectively, and plotting hte results. The grey line is a line of best fit whose slope equals the slope in the main specification.

Figure 1.4: Extensive Margin Results



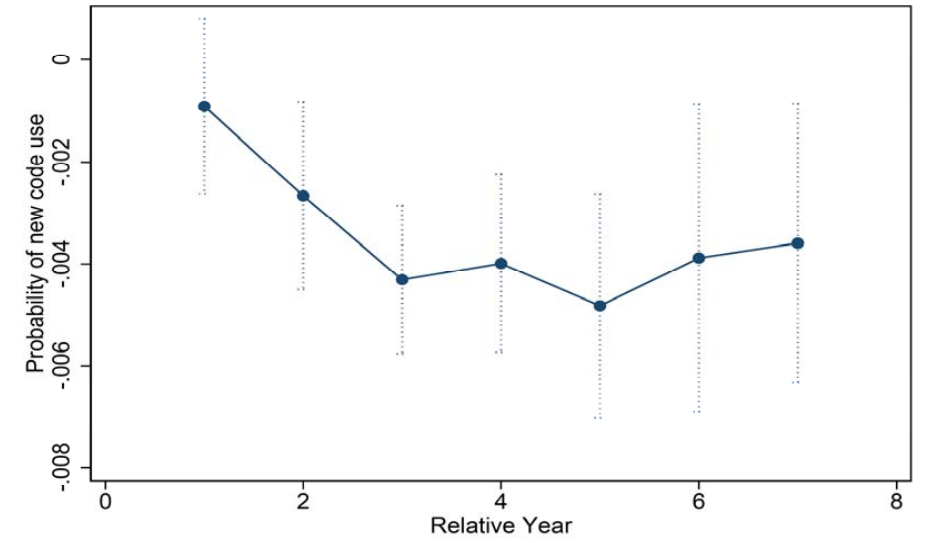
Notes: Figure shows the coefficients estimated from running specification (5) on different measures of extensive margin adoption. The bands are 95-percent confidence intervals based on standard errors clustered at the treatment category level.

Figure 1.5: Intensive Margin Reults



Notes: The figure plots the coefficient from a regression of new code dummy on year relative to new code introduction with treatment category fixed effects and controls for calendar year, patient age, race, sex, chronic conditions, and Medicaid receipt, and doctor controls for AMA hospital ID, med school ID, and gender. The sample is all operations by surgeons who have adopted the new code in the treatment categor after they are first observed using the new code.

Figure 1.6: Dynamic Patterns of Age-Based Adoption Differences



Notes: Figure shows the coefficients estimated from running specification (6) using the full set of patient and doctor controls. The bands are 95-percent confidence intervals based on standard errors clustered at the treatment category level.

Table 1.1: Sample Summary Statistics

Summary Statistics

Panel A: Patient Level

	Mean	S.D
Share of patients with any chronic condition	0.09	0.29
Share of white patients	0.89	0.31
Share of male patients	0.53	0.50
Share on medicaid	0.16	0.37
Patient age	73.50	10.26
Share treated with new code	0.16	0.37
Number of patients in sample	167,908	
Number of observations in sample	178,699	

Panel B: Doctor-Year Level

	Mean	S.D
Age	53.02	8.45
Share male	0.96	0.20
Total years in sample	3.53	2.18
Number of treatment categories	1.55	0.90
Probability of using new code	0.14	0.31
Number of doctors in sample	27,476	

Notes: Table shows summary statistics for the sample. Panel A shows means calculated at the patient procedure level, while panel B does so at the doctor-year level.

Table 1.2: New Code Use Baseline Regressions

	(1)	(2)	(3)	(4)	(5)
VARIABLES	1(New code)	1(New code)	1(New code)	1(New code)	1(New code)
Doc Age at RY 0	-0.00177***	-0.00178***	-0.00177***	-0.00188***	
	(0.000643)	(0.000638)	(0.000635)	(0.000543)	
Doc Age 40-49					-0.0201***
					(0.00640)
Doc Age 50-59					-0.0348***
					(0.00895)
Doc Age 60-69					-0.0529***
					(0.0143)
Doc Age 70-79					-0.0564***
					(0.0177)
Observations	178,699	178,699	178,699	178,699	178,699
R-squared	0.308	0.310	0.311	0.353	0.353
Dep Var Mean	0.159	0.159	0.159	0.159	0.159
Treatment Cat F.E.	x	x	x	x	x
Chronic Conditions		x	x	x	x
Demographics			x	x	x
Phys Observables				x	x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running specification (2) on the main analysis sample including successively more controls to the specification. Standard errors are clustered at the treatment category level.

Table 1.3: Extensive Margin Analysis

VARIABLES	(1) 1(Adopt in 1Yr)	(2) 1(Adopt in 3Yrs)	(3) 1(Adopt in 5Yrs)	(4) 1(Adopt in 7Yrs)
Doc Age 40-49	-0.0327**	-0.0242	-0.0241	-0.0237*
	(0.0159)	(0.0168)	(0.0147)	(0.0137)
Doc Age 50-59	-0.0564***	-0.0483**	-0.0458**	-0.0461**
	(0.0178)	(0.0195)	(0.0192)	(0.0184)
Doc Age 60-69	-0.0737***	-0.0745***	-0.0777***	-0.0781***
	(0.0215)	(0.0205)	(0.0267)	(0.0255)
Doc Age 70-79	-0.0911**	-0.0967**	-0.102**	-0.102**
	(0.0369)	(0.0420)	(0.0499)	(0.0494)
Observations	8,763	8,763	8,763	8,763
R-squared	0.598	0.729	0.767	0.770
Dep Var Mean	0.247	0.345	0.365	0.367
Treatment Cat F.E.	x	x	x	x
Phys Observables	x	x	x	x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the coefficients estimated from running specification (3) on different measures of extensive margin adoption. Standard errors are clustered at the treatment category level.

Table 1.4: New Code Use (Intensive Margin)

VARIABLES	(1) 1(New code)	(2) 1(New code)	(3) 1(New code)	(4) 1(New code)	(5) 1(New code)
Doc Age at RY 0	-0.000974**	-0.000962**	-0.00101**	-0.000865*	
	(0.000431)	(0.000422)	(0.000421)	(0.000500)	
Doc Age 40-49					-0.0453***
					(0.0104)
Doc Age 50-59					-0.0303**
					(0.0129)
Doc Age 60-69					-0.0586***
					(0.0159)
Doc Age 70-79					0.0271
					(0.0436)
Observations	52,462	52,462	52,462	52,462	52,462
R-squared	0.134	0.144	0.145	0.220	0.220
Dep Var Mean	0.159	0.159	0.159	0.159	0.159
Treatment Cat F.E.	x	x	x	x	x
Chronic Conditions		x	x	x	x
Demographics			x	x	x
Phys Observables				x	x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running specification (4) on the sample of adopters after they are first seen using the new code including successively more controls to the specification. Standard errors are clustered at the treatment category level.

Table 1.5: Minimally Invasive Surgery Analysis

VARIABLES	(1) 1(New code)	(2) 1(New code)	(3) 1(New code)	(4) 1(New code)
Doc Age at RY 0	-0.00457***	-0.00456***	-0.00456***	-0.00443***
x MIS Flag	(0.000620)	(0.000612)	(0.000610)	(0.000563)
Doc Age at RY 0	-0.000649*	-0.000659*	-0.000656*	-0.000772***
	(0.000360)	(0.000359)	(0.000358)	(0.000271)
Observations	178,713	178,713	178,713	178,713
R-squared	0.310	0.312	0.312	0.354
Dep Var Mean	0.159	0.159	0.159	0.159
Treatment Cat F.E.	x	x	x	x
Chronic Conditions		x	x	x
Demographics			x	x
Phys Observables				x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running specification (10) on the baseline analysis sample including successively more controls to the specification. Standard errors are clustered at the treatment category level.

Table 1.6: New Diagnostic Code Use Analysis

VARIABLES	(1) 1(New code)	(2) 1(New code)	(3) 1(New code)	(4) 1(New code)	(5) 1(New code)
Doc Age at RY 0	0.000544	-5.44e-05	-5.49e-05	-5.92e-05	
	(0.000447)	(7.90e-05)	(7.91e-05)	(7.88e-05)	
Doc Age 40-49					-0.00169
					(0.00233)
Doc Age 50-59					-0.000852
					(0.00173)
Doc Age 60-69					-0.00535
					(0.00407)
Doc Age 70-79					-0.00257
					(0.00265)
Observations	5,319,067	5,319,067	5,319,067	5,319,067	5,319,067
R-squared	0.098	0.668	0.668	0.668	0.668
Dep Var Mean	0.133	0.133	0.133	0.133	0.133
Treatment Cat F.E.	x	x	x	x	x
Chronic Conditions		x	x	x	x
Demographics			x	x	x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running specification (11) on the diagnostic code use analysis sample including successively more controls to the specification. Standard errors are clustered at the treatment category level.

Table 1.7: New Drug Use Analysis

VARIABLES	(1) 1(New drug)	(2) 1(New drug)	(3) 1(New drug)	(4) 1(New drug)
Doc Age at RY 0	-9.29e-06	-1.28e-05	-3.81e-06	
	(1.88e-05)	(1.84e-05)	(1.64e-05)	
Doc Age 40-49				0.00111*
				(0.000571)
Doc Age 50-59				0.000649
				(0.000462)
Doc Age 60-69				0.000581
				(0.000590)
Doc Age 70-79				-0.000492
				(0.000728)
Observations	55,195,851	55,195,851	55,195,851	55,195,851
R-squared	0.163	0.163	0.164	0.164
Dep Var Mean	0.017	0.017	0.017	0.017
Therap. Class F.E.	x	x	x	x
Chronic Conditions		x	x	x
Demographics			x	x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running specification (12) on the pharmaceutical analysis sample including successively more controls to the specification. Standard errors are clustered at the therapeutic class level.

Table 1.8: Arrival Rate Analysis

VARIABLES	(1) 1(New code)	(2) 1(New code)	(3) 1(New code)	(4) 1(New code)
Doc Age at RY 0	-9.93e-05*	-9.91e-05*	-9.90e-05*	-0.000135*
x Arrival Rate	(5.92e-05)	(5.87e-05)	(5.82e-05)	(6.85e-05)
Doc Age at RY 0	-0.00178***	-0.00178***	-0.00178***	-0.00199***
	(0.000631)	(0.000625)	(0.000621)	(0.000510)
Arrival Rate	0.00804***	0.00797***	0.00796***	0.0103***
	(0.00282)	(0.00277)	(0.00275)	(0.00294)
Observations	178,713	178,713	178,713	178,713
R-squared	0.310	0.312	0.312	0.354
Dep Var Mean	0.159	0.159	0.159	0.159
Treatment Cat F.E.	x	x	x	x
Chronic Conditions		x	x	x
Demographics			x	x
Phys Observables				x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running specification (17) on the baseline analysis sample including successively more controls to the specification. Standard errors are clustered at the treatment category level.

Table 1.9: Robustness

	(1)	(2)	(3)	(4)	(5)
VARIABLES	1(New code)	1(New code)	1(New code)	1(New code)	1(New code)
Doc Age at RY 0	-0.00214***	-0.000609**	-0.00188***		
	(0.000667)	(0.000281)	(0.000573)		
Doc Age at RY -1				-0.00105***	
				(0.000381)	
Doc Age at RY -2					-0.000553*
					(0.000294)
Observations	43,856	115,734	172,450	133,216	100,898
R-squared	0.373	0.381	0.355	0.415	0.508
Dep Var Mean	0.183	0.096	0.159	0.154	0.103
Treatment Cat F.E.	x	x	x	x	x
Chronic Conditions	x	x	x	x	x
Demographics	x	x	x	x	x
Phys Observables	x	x	x	x	x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running various robustness checks on the baseline specification. Standard errors are clustered at the treatment category level.

2 Chapter 2
Surgical Innovation and the Challenge of Changing Demographics

2.1 Introduction

The question of whether technological innovation is driven by scientific opportunity or demand-pull incentives has long been of substantial interest to academic economics (Acemoglu and Linn (2004)). From a theoretical perspective, standard economic models yield strong predictions about the behavior of potential inventors when faced with changes in expected demand. Testing them can therefore further our knowledge of economic behavior. From a policy perspective, innovation is frequently linked to important outcomes such as productivity growth and thus understanding its drivers is an important objective. This question is especially salient in the context of healthcare, where innovation has been credited with achieving substantial improvements in patient outcomes, as well as inducing increases in costs, over the past century (Cutler and Kadiyala (2003), Nordhaus (2003), Cutler (2005), Murphy and Topel (2006), Jayachandran et al. (2010), Chandra and Staiger

(2007), Chandra and Skinner (2012)). Perhaps unsurprisingly much of the existing empirical research on the impact of demand-side changes on innovation comes from this part of the economy. The core of this evidence is formed by a successful line of research that has investigated the impact of various shocks to potential market size on innovation in the pharmaceutical sector (Acemoglu and Linn (2004), Finkelstein (2004), Dubois et al. (2015)).[21]

A somewhat overlooked aspect of this line of inquiry is that the answer to this question are inextricably linked with the market institutions under which innovation takes place. Much of the theory on this question assumes that innovation is driven by profit-maximizing firms in a setting where firms are incentivized to innovate through government-sanctioned monopolies due to intellectual property protection. This structure matches the market institutions of the pharmaceutical sector quite well. However, this type of market organization is far from ubiquitous and is, by some metrics, the exception rather than the norm. For example, a survey of 1,478 R&D labs by Cohen et al. (2003) finds that there is a variety of mechanisms other than patents that firms use to protect intellectual property, including secrecy, lead time, use of complementary marketing and manufacturing capabilities. In fact, patents are one of the less popular methods among a majority of firms and industries with the salient exception of pharmaceuticals.

A radically different form of market organization can be observed in another important domain of medical innovation, surgeries. Surgical innovation has played a crucial role in cost and benefit changes over the past half a century. For instance, a third of the huge decrease in cardiovascular mortality that occurred in the middle of the twentieth century comes from the development of surgical interventions such as coronary artery bypass graft, angioplasty, and catheterization (Cutler and Kadiyala (2003)). Surgical innovation is organized around a very different set of institutions

21

One reason for interest here is that the rapidly changing demographics of Western societies induce changes in market size due to demand (Acemoglu and Linn (2004)). This gives rise to the question of whether current market structures in the sector are well-equipped to respond to such these changes. A separate, but related reason is that many popular government policies in the health sector (e.g., patent term, generic-entry regulation, excise taxes, health insurance coverage expansions) have a direct effect on the market size for various medical products. This means that innovation effects due to changes in those policies are a first-order concern (Finkelstein (2004)). These effects have always held a prominent place in public policy discussions. One recent example is the ongoing debate about pharmaceutical price controls in the US, a policy which will directly impact the potential market size for pharmaceutical firms by limiting the potential surplus that can be extracted from the firm's monopoly. Perhaps unsurprisingly, the ability of industry to deliver ongoing innovative products under these proposed policies has been a central point in these debates (see, e.g., Wayne Winegarden (2020)).

than pharmaceuticals. Novel surgical development is organized around the independent activity of multiple decentralized surgeons and academic medical centers (Chang and Luft (1991)). Intellectual property protection is weak and government regulation is very light Meier (2015)). Hence, innovating surgeons can likely recoup only a fraction of the value of their investment (Chang and Luft (1991)). The development of a new surgery is a highly incremental process that relies on the adoption and improvement activity of multiple practitioners as well as communication through professional meetings and publications as opposed to the directed and structured development and marketing activity of a pharmaceutical company (Chang and Luft (1991), Scott-Morton and Kyle (2011)).

At the same time, formal evaluation of new methods and approaches is rare and fraught with methodological difficulties due to reasonable ethical constraints (Savulescu et al. (2016)). This means that only the major procedures that show vast differences in (non-randomized) patient outcomes are likely to survive to the end of the development stage (Chang and Luft (1991)). This is a process which is likely to direct innovative efforts quite differently from pharmaceuticals, where RCTs with sometimes small relative benefits can help generate profitable inventions (Scott-Morton and Kyle (2011)). In light of the radically different market institutions and the deliberate blunting of financial incentives, the traditional mechanism said to underlie the impact of potential market size changes, the question of whether surgical innovation responds to changes in potential market size at all looms large.[22] This project aims to provide an answer.

I do so by overcoming two of the main challenges faced by empirical researchers in this context: measuring surgical innovation and finding credible ways of studying its causes. First, the weak IP protection around surgeries means that surgeons are less likely to patent their inventions (with some special exceptions addressed below) (Meier (2015)). This means that a traditionally popular method of measuring innovative activity, namely patent counts, is not reliable in this setting as patents for new surgical techniques do not track well all new surgical techniques.[23] Additionally, the lack of government regulation in this space means that exogenous quasi-experimental variation

[22]From a theoretical perspective, the answer to this question can help us refine traditional economic models of innovation and explore the applicability of alternative mechanisms to encourage research and devleopment. For a policy-maker, an answer can help with understanding the drivers of innovation in the important setting of surgical innovation, and, more broadly, in contexts where intellectual property protection has only a limited scope for incentivizing innovators.

[23]While individual novel surgeries may be identified through, e.g., claims data, doing so at scale has been difficult.

is hard to find. The present study thus aims to address both of these challenges by offering a novel measure of surgical innovation and by identifying a source of exogenous variation that can be used to study this important economic area. Using these tools, I explore two questions. First, does surgical innovation look different from pharmaceutical innovation? Second, does surgical innovation respond to changes in potential market size.

The paper's first contribution towards this goal is to use a novel measure of surgical innovation. This measure allows me to offer the first, to my knowledge, systematic description of surgical inventions over the past three decades. Specifically, I use the annual revisions to the ICD9-CM system of coding medical procedures to identify novel surgical approaches. This method yields the universe of important surgical advancements over the past 20 years. Importantly for my empirical approach, these surgical procedure codes are organized in a hierarhichcal system based on the part of the human anatomy impacted by the procedure. This organization naturally lends itself to assignment to a well-defined market for surgical procedures.[24]

This measure allows me to document that at least in the aggregate pharmaceutical and surgical innovation markets seem to follow different trends in accordance with their different institutional settings. In particular, I juxtapose trends in overall counts of novel pharmaceutical and surgical innovation in a comparable set of therapeutic markets. The results from this investigation show that the patterns in surgical invention are strikingly different from those in pharmaceutical invention. While the trends in some markets for surgical innovation match those for pharmaceuticals quite closely, other markets show no or even negative correlation between the two. These differences lend additional weight to the claim that surgical innovation markets are different from markets for pharmaceutical innovation and thus to the question of whether surgical innovation responds to changes in potential market size at all.

I then move onto the main part of the analysis and investigate the impact of potential market size on surgical innovation. I solve the well-known problem of reverse causality between potential market size and innovation by using demographic shifts over three decades as a quasi-exogenous

[24]There is one other study that I am aware of in the economics literature which uses the idea that procedure codes may track innovation. Lichtenberg (2014) studies the impact of medical innovation on cancer mortality and uses, as one of his measures of medical innovation, novel procedures as measured by cancer-related radiation and surgical CPT codes that were introduced after 1995. While the author uses a different coding system, with a different definition of innovation based on that coding system, for the purposes of investigating a different question, the core insight remains close to the one used in this study.

source of variation in the independent variable of interest. Inspired by the work of Acemoglu and Linn (2004), this approach is particularly well-suited in the case of surgical innovation because patient age predicts strongly the incidence of various surgical procedures.

I estimate a positive and significant elasticity of surgical innovation with respect to potential market size. My baseline estimate of 1.57 is lower in magnitude than those reported in Acemoglu and Linn (2004) and Dubois et al. (2015) for pharmaceuticals, but still implies a substantial response of surgical innovations to changes in potential market size. On the extensive-margin, a doubling of potential market size increases the probability that a given surgical market receives an innovation by 85%. Finally, I find that innovation responds more strongly to lags than leads in market size. These results suggest that surgical innovation will respond to increased demand for novel procedures from older patients, but they leave the question of how this happens unanswered.

These results have important implications for economic policy. A long-running argument in economics posits that one of the most prominent benefits of intellectual-property protection and financial incentives is the ability of markets to deliver welfare-enhancing innovation (Goldman and Lakdawalla (2011)). This paper offers evidence that while this may be sufficient in some cases, it is by no means necessary. The presence of significant levels of surgical innovation that respond to changes in market size indicates that non-standard market institutions may offer a complement, if not an alternative, to traditional profit-maximizing research and development. This possibility is especially salient in the current climate of concerns about the productivity and direction of corporate pharmaceutical research (e.g., Pammolli et al. (2010), Oostrom (2020)). Naturally, the feasibility and desirability of such arrangements depends.on the exact mechanism of action in this setting, but this paper lays the foundations for future exploration in this direction.

2.2 Institutional background

2.2.1 Surgical innovation

The drivers of surgical innovation are of substantive economic interest due to, among other factors, the unique organization of markets in this domain. Unlike markets for pharmaceutical and device innovation, these markets are not centered around the activity of profit-maximizing private companies granted limited monopolies and regulated by public agencies. Instead, they differ along three

major dimensions: development, regulation, and diffusion. I explore each of these dimensions in detail below.

The development of novel surgeries depends on the character of the innovation under consideration. Minor alterations to existing techniques occur in the course of the surgeon's normal practice if the improvement is considered more efficient or beneficial for the patient's specific situation. For instance, vagotomy is a treatment for peptic ulcer disese involving the severance of the vagus nerve. The procedure can be performed by ligating different sections of the nerve. More selective approaches can be considered variations to the overall procedure. Over the course of the surgeon's practice and through communication via peer reviewed publications and professional conferences, the novel alteration is perfected and diffused (Chang and Luft (1991)).

Major novel surgeries follow a somewhat more formal process that usually occurs at academic medical centers. After conceiving of a novel surgical approach, the innovating surgeon usually tries it on animal models to confirm physiological feasibility and compatibility with life. After experimentation with animal models has been considered sufficiently successful, the new procedure is attempted on a suitable patient. The medical code of ethics plays an important role in this step of the process. Typically only patients without a viable medical alternative are considered eligible for major novel procedures. If the procedure is considered to have a suitably successful outcome, it is shared with the wider medical community and improved upon through professional journals and meetings (Chang and Luft (1991)).

An example may help clarify this process. The first coronary artery bypass surgery was performed in 1960 by Robert Goetz in Albert Einstein College of Medicine-Bronx Municipal Hospital Center in New York. Prior to performing the procedure on a patient, the surgical team had developed the concept and trained their skills on dogs. The graft was successful and when the patient eventually died 13 months later, an autopsy revealed that the graft was still functional. Subsequently, the procedure was performed in Johns Hopkins in 1962 and in Houston in 1964. Spread and further development of the surgery was somewhat sporadic due to the high mortality among eligible patients and the need to confirm procedure graft patency, which could take years if patients survived for longer. The procedural approach which would eventually become the gold standard in CABG was not performed until 1968 in Saint Luke's Hospital in NYC (Melly et al. (2018)).[25]

[25]It is also worth pointing out that various ideas and incremental surgical advancements which would make CABG

Still, the procedure was considered so revolutionary and beneficial that it had achieved widespread acceptance by 1974 (Bunker et al. (1978)).

The specific features of this process mean that it is not nearly as regulated as the corresponding processes in pharmaceutical and device innovation. The reason is that the natural regulatory framework to be applied in this case is that of human subjects research, as outlined by the famous Belmont Report. The report distinguishes between the use of novel methods for thereapeutic and research purposes. Research is performed primarily for the purpose of obtaining generalizable knowledge with any benefits accruing to its subjects taking a secondary role in the objective function. As such, research is governed by a host of federal regulations collectively named the Common Rule. [26] However, therepautic use of a novel method is performed primarily for the benefit of the patient and is not governed by the former regulations. Surgical innovation strikes an uneasy balance between these two categories. Many surgical innovators would see their work as primarily benefiting the patient and potentially creating generalizable knowledge as a secondary boon from the procedure. However, such activity could easily cross the line into territory more akin to human-subject research[27] and this line has not been defined with any reasonable degree of specificity either by the Belmont Report or by subsequent federal regulation. In practice, it is up to individual surgeons to decide whether their activity constitutes research and thus whether it should fall under the relevant established rules and guidelines. In such cases, internal review boards at the surgeon's institution will review the research protocol and permit or forbid the use of a procedure and the broader research design (Laakmann (2015)).

The large amount of physician discretion in this setting and lack of formal guidelines is rooted in a host of difficult issues faced by doctors and patients. The first is a recognition that there is a vast amount of heterogeneity among patients in terms of medical needs, anatomical features, and personal preferences when it comes to surgical procedures.[28] Our current intellectual framework

possible date back to 1910 when the concept of operating on the coronary circulation was first described (Melly et al. (2018).

[26] One example is the requirement for all research protocols and the system of obtaining informed consent to be screened by an internal review board (Laakmann (2015)).

[27] For instance, in the *Ancheff v. Hartford Hospital* trial, the plaintiff asserted that his treatment with a dosage of the drug gentamicin that was double the conventional FDA-aproved dose constituted a part of an undisclosed clinical trail. Ultimately, the Connecticut Supreme Court decided that the jury has the prerogative in determining whether such a program constitutes research as the Belmont Report offers insufficient guidance (Laakmann (2015)).

[28] An example of this is the treatment of breast cancer via mastectomy (complete removal of the breast) or lumpectomy (local excision of the tumor). While the two approaches have statistically similar outcomes, there is a vast

therefore recognizes that a physician should be able to adapt to this heterogeneity in her approach. The second is that medicine is subject to an intense amount of technological change and knowledge accumulation, which means that the frontier of medical solutions and knowledge is constantly shifting and a physician should be allowed to shift with it. The third is that there is still a large number of difficult medical conditions for which there is no standard or widely accepted treatment (or even any treatment at all). Binding a physician with a regulatory standard may take away her ability to serve patients who may have little other recourse for alleviation of their issues (Laakmann (2015)).

An example may help clarify some of the difficult issues that surround these choices. the first successful cardiac transplant in the US, which was only the fourth attempted human cardiac transplant in the world, did not undergo institutional review as the operating surgeon, Dr Norman Shumway of Stanford, claimed he was trying to save the patient's life, and thus the urgency of the patient's condition was considered reason enough to forego research approval (Chang and Luft (1991), Golf (2017)). However, the possibility of cardiac transplantation had been an area of active interest and arguably a broader research agenda that spanned surgeons from multiple countries and generations, each of whom incrementally contributed to the anatomic knowledge necessary to perform such a difficult procedure. The feasibility of such a procedure in humans was an area of active academic debate where, naturally, empirical evidence from actual attempted operations held much sway (Golf (2017)). Cast in this light, Dr Shumway's life-saving procedure had many of the hallmarks of a research project that would have been subject to federal regulation.

In practice, one of the primary factors constraining the use of novel procedures in the US is a surgeon's liability under the court system. Patients who suffer injury or bad outcome during a procedure may bring a federal malpractice lawsuit against the performing physician.[29] In such cases, the plaintiff has to prove four things: that the surgeon had a duty to the patient, that she breached the standard of care, that the patient suffered compensable injury as a result, and that the defendant's actions caused the injury. Surgical innovation in of itself is not grounds for lit-

difference between them in terms of other characteristics. Women electing the former need to consider the cost of disfigurement and possibility of cosmetic surgery, while those choosing the latter may have to undergo chemotherapy and radioation therapy as well as risk local recurrence. The choice in this setting cannot be ethically imposed by a standardized system of guidelines with our current state of knowledge (Laakmann (2015)).

[29]While the pathway for a novel procedure which was performed as part of a formal research study is slightly different, the main parameters of the court's decision process remain unchanged.

igation. The defendant has to prove that such innovation breached the standard of care and the court's decision in such cases is often governed by the surgeon's reputation, the existence of other surgeons who believe the attempted procedure to be feasible (as established through expert testimony, papers, profesisonal meetings and conferences, etc), and the existence of a widely accepted non-innovative treatment for the patient's condition. In general, courts have been cognizant of the surgeon's need to adapt to a patient's needs as well as the changing technological landscape of medicine and thus innovation in of itself has not been a ground for successful litigation. However, even here innovation is generally seen as acceptable when the primary beneficiary is the patient as opposed to the accumulation of knowledge for society (Mastroianni (2006)).

Differences between surgical and pharmaceutical innovation extend to the diffusion process as well. In the case of pharmaceuticals and devices, the FDA requires companies to conduct large-scale randomized controlled trials to verify the safety and efficacy of a drug. Therefore, the evaluation of the relative benefits of a novel treatment largely happens prior to its mass diffusion. Once FDA approval is granted, the inventing firm spreads knowledge about the drug through promotional and informational campaigns directed at physicians (Scott-Morton and Kyle (2011)). In contrast, due to the lack of a formal regulatory evaluation and approval processes in the case of novel surgeries, the decision about what new surgeries work well and thus should be adopted, is largely left up to the professional surgical community. This means that the surgeries that are likely to spread are those that have a large difference in outcomes relative to the existing treatment (if any) and those that treat untreatable conditions where the alternative is extremely low quality of life or death. This evaluation is done informally through the use of case-studies or small-sample analyses of treated patients. RCTs are rarely used and if they are conducted at all, this usually happens after a procedure has gained somewhat wide acceptance (Chang and Luft (1991)).[30][31] Perhaps unsurprisingly, the lack of standardized procedures to determine the relative usefulness of

[30] An example of this process is the spread of laparoscopic cholecystectomy in the 1980s and the 1990s. The procedure was not evaluated by an IRB since it was used in treating individual patients. Neither surgeons, nor insurers viewed it as experimental since it was not part of a formal research study and the fundamental objective of the procedure remained the same as that of open cholecystectomy. Eventually, RCTs were performed to evaluate the procedure, but only after it had spread widely and become standard accepted practice in the community (Mastroianni (2006), Laakmann (2015)).

[31] Some of the reasons for this are practical in nature: the difficulty in actually implementing double-blinding in randomization as well as the need to account for differences in skill level and experience of surgeons with each procedure. A different set of constrants has to do with the fact that randomizing a patient into a surgical procedure will almost certainly violate traditional medical ethics (Savulescu et al. (2016)).

new prcoedures has been an active area of interest in the medical literature on surgical innovation (see, e.g., Savulescu et al. (2016)).

These differences in the organization of markets for surgical and pharmaceutical innovation have led to differences in the incentives behind innovation. The literature on pharmaceutical innovation has argued that the benefits to a company considering whether to engage in the process are the expected profits from having a patent-induced monopoly on the market for its novel molecule for a preset period of time post-invention. The costs include the often substantial fixed costs of innovation which have to do with the discovery of a new chemical compound and the costs of guiding this product through the multiple stages of testing and regulatory approval (Scott-Morton and Kyle (2011)).[32]

In surgical innovation, on the other hand, intellectual property plays very little role. The reason is that patent law does not protect novel surgeries, unless they contain a novel device.[33] This means that the innovating surgeon is generally not able to recoup the costs of her innovative effort through a patent-enforced monopoly. Therefore, the financial returns to innovative activity are more limited than in the case of pharmaceuticals (Chang and Luft (1991)).

The sources and quantity of funding for surgical innovation are consistent with the lack of promise of financial returns for this activity. Surgical innovation relies largely on non-private sources of funding. These include grants from the National Institutes of Health (NIH), National Science Foundation (NSF), and Veterans' Administration (VA), as well as funding from an academic center's School of Medicine or Department of Surgery. While industry and private foundations do provide funding for some projects, this funding tends to be a minority of the overall sources of finance for innovative surgical activity. For example, in a survey of Association of Academic Surgery members, Rikkers et al. (1985) find that only 24% of financing for surgical innovation comes from the latter two sources with the rest coming from federal or academic funds.

[32] DiMasi et al. (2003) estimate that the cost of innovation per new molecule is over $800 million in 2000 dollars.

[33] The 1997 Omnibus Consolidated Appropriations Act created a new section in the patent law's provision of damages limitations. The new law "deprives patentees of remedies for infringement by a medical practitioner's performance of a medical activity, absent exceptions" (Meier (2015)). The major exception under the law is when patented use of a composition of matter contributes to the novelty of the method under considerations. This is commonly taken to refer to medical drugs and devices. The law is widely interpreted to mean that while surgical procedures that do not involve novel devices may be patented, patent infringement litigation is not available to patentees when the patent is violated in the process of providing medical care, which would cover the vast majority of uses of the patented invention. However, in cases where novel devices are part of the surgical method being patented, such infringement litigation is still available (Meier (2015)).

Additionally, funding for surgical innovation is traditionally considered to be relatively low and decreasing in priority for federal agencies providing funds for biomedical research. For instance, Moses et al. (2015) finds that health services research receives only 20% of total healthcare research funds. Hu et al. (2015) find that NIH grants to academic surgeons have declined by 20% since 2004 and Adler and Chen (2010) find that only 6% of American Association of Endocrine Surgeons (AAES) were funded for endocrine surgery research in the 90s and this number further dropped to 3% in the 2000s. These patterns are consistent both with the lack of financial returns to innovation and with the high amount of uncertainty that needs to be resolved before the value of any given innovation is determined.

Since innovative activity in this space largely takes place in academic centers, various researchers have proposed non-financial incentives for innovation. For instance, some authors have suggested that innovating surgeons are largely incentivized through increased academic prestige and higher chances of tenure due to their innovative work (Chang and Luft (1991)). Others have posited that user-driven innovation may play an important role in settings such as medicine (DeMonaco et al. (2006)). The underlying idea is that product manufacturers may be unable to capture the entire consumer surplus of additional product variety due to information asymmetries and transaction costs. However, product users can capture the entire surplus from their own innovative activity even if they can capture none of the benefits accruing to other users. Thus, if individual surplus is high enough and fixed costs of innovation low enough, user-driven innovation can drive product variety in a market. This result requires two conditions: that users derive utility from their own invention and that information asymmetries and transaction costs prevent firms from capturing all of the consumer surplus (Henkel and von Hippel (2004)). Medical innovation due to physicians fits these conditions quite closely. Since medical ethics puts patient well-being in a primary position in a surgeon's objective function this renders improvements to said well-being through innovation self-rewarding. Similarly, physicians dealing with a difficult medical problem have access to the myriad important and often difficult to communicate issues that concern patients with a particular condition. Such "sticky" information makes it difficult for firms to offer the optimal amount of product variety (Hippel (1994)). [34]

[34]Funding for this activity comes from federal agencies (NIH), industry, private funds, and clinical research funds. However, the costs of this innovative process tend to be lower both because of the lower regulatory burden and because of the decreased ability of actors to recoup their investment after an innovation has been proven to be effective.

Therefore, one may expect higher market size to incentivize higher innovative activity in both surgical and pharmaceutical markets. However, this effect is likely to act through different channels. In the case of pharmaceuticals, the likely channel is increased expected profits due to higher quantity consumed. For surgeries, however, a larger potential market means a more important problem to be solved. This can mean either higher prestige awarded to those that manage to offer feasible solutions or increased self-reward to those who are primarily concerned with their patients' well-being. In either case, one should expect an increase in innovative activity as the need for a particular surgery increases.

I formalize these forces through a framework in the spirit of Benabou and Tirole (2003) model of intrinsic and extrinsic motivation in the Appendix. The framework features an agent who decides whether to complete a research project that has both a private and a public benefit, but which has an a priori unknown cost of completion. The social planner who has private information on the project costs faces a choice of how much of an external reward to offer to the researcher. Under standard distributional assumptions on the cost signal, the social planner's problem has a unique solution with higher project costs inducing higher extrinsic rewards, but with higher extrinsic rewards causing agents to revise their expectations of project costs upwards.

The model clarifies several of the patterns to be expected from the empirical analysis and of the institutional setup discussed above. First, all else equal, higher intrinsic motivation leads to higher researcher involvement and research output. To the extent that larger market size implies a more important or prestigious problem to solve, one should then expect that increases in market size will lead to more research output. Second, the costs and research production process drive the choice of extrinsic rewards. For projects that are very expensive, but which have relatively low private benefits, as is pharmaceutical innovation, the power of intrinsic motivation to motivate research may be limited, leading to proportionately higher extrinsic rewards (e.g., IP-induced market protection). On the other hand, processes which consist of multiple steps with correlated research costs (such as the case with surgical innovation) may induce the social planner to reduce the extrinsic motivation even more as the costs of the agent's expectation updating increases. Finally, since offering extrinsic rewards is socially costly (e.g., consumer deadweight loss from monopoly pricing), intrinsically motivated innovation is preferable from a social welfare perspective.

2.2.2 ICD9 Codes and Revisions

To make progress on answering the question whether surgical innovation does indeed respond to changes in market size, I use revisions in the International Classification of Diseases 9th Edition, Clinical Modification (ICD9 CM) system of coding medical procedures to measure surgical innovation. ICD9 CM is a complete system of classification for medical procedures (both surgical and non-surgical) performed on patients in the US. It is mandatory for use on inpatient claims submitted by hospitals for reimbursement by Medicare. It is jointly maintained by the Center for Medicare and Medicaid Services (CMS) and National Center for Health Statistics (NCHS) (United States General Accounting Office (GAO) (2002)). The system is maintained and updated through biannual meetings of the ICD-9-CM Coordination and Maintenance Committee. These meetings focus on proposed revisions to the system, usually to add codes for new and distinct medical procedures and are attended by representatives from relevant federal agencies, private industry, the American Hospital Association, and the American Medical Association among others. Summaries from the proceedings (and, more recently, audio and video recordings), as well as final changes to the system, including new procedure codes, are posted on the relevant agencies' websites (Centers for Medicare & Medicaid Services (CMS) (2020)).

In deciding whether to add a new code for a given medical procedure, the CMS and NHCS consider whether it is sufficiently different from similar procedures that already have codes, whether it is safe and effective, and whether it is widely accepted by the medical community. While there are no hard rules about when a procedure is sufficiently differentiated and accepted to receive a new code, the meetting attendees discuss various sources of evidence such as empirical studies on patient outcomes and technical details of the procedure implementation to make a determination. The goal is to maintain an expressive, yet succinct system of coding medical procedures. The chief constraint is that for a variety of reasons the system allows for a limited number of codes to be created,[35] so that minor modifications to existing procedures are unlikely to be judged

[35]The system is hierarchichically organized based on human physiciology and surgical approach. This means that similar procedures are grouped together. However, each procedure code uses up to four digits with similar procedures sharing the first two to three digits in their designation. This means that there is only a finite number of novel procedures that fall into a given category (For instance, "Operations on the Cardiovascular System" are typically expressed by codes with first two digits in the range 35-39. Within this range, operations that have to do with insertion and removal of pacemaker and pacemaker components reside in the range 3770-3789). Space constraints have been an issue for the CMS previously with the agency having to create novel subcategories in the "incorrect" categories in order to keep the system up to date (Continuing with the pacemaker example, in 2002, CMS decided that six new

sufficiently important to receive their own designation (United States General Accounting Office (GAO) (2002), Centers for Medicare & Medicaid Services (CMS) (2020)).

An example of this process is given by the addition of code 81.88, "Reverse total shoulder replacement" to the system in 2010. Conventional shoulder replacement, marked by code 81.80 up until that point, involves replacing the shoulder joint by placing an artificial ball and stem at the top of the patient's humerus (a long bone forming part of the upper arm) and socket in her glenoid cavity (a part of the shoulder forming the joint connecting it with the arm). The procedure is used to treat conditions such as osteoarthritis and restore functionality to the patient's upper arm. However, in patients with serious conditions such as rotator cuff arthropathy (a type of degenerative joint disease), this procedure may not be indicated. In such patients, shoulder replacement may be performed by swapping the positions of the ball and socket implants. While the procedure does not restore full functionality, it tends to alleviate pain and improve quality of life. After examining evidence for the different surgical approach and patient population for this procedure relative to traditional shoulder replacement, the CMS judged that reverse shoulder replacement is sufficiently different that it requires its own code. The agency revised the system so that code 81.88 was added for this new procedure, while code 81.80, which up until that point was used to denote both kinds of shoulder replacement was revised to denote only traditional shoulder replacement (Centers for Medicare & Medicaid Services (CMS) (2009)).

This example illustrates some of the strengths in using revisions to the ICD-9-CM system as a measure of surgical innovation. The first is almost universal coverage. Since ICD-9-CM is designed to offer a complete system of describing procedures (not least because it is legally required of hospitals to use), the agencies in charge are incentivized to keep track of all changes in medical practice in order to keep the system functional and up-to date. The second is an objective (from an economics researcher's perspective) measure of what constitutes an innovation. Specifically, as highlighted in the previous section, surgical innovation has an incremental character that is closely tied to its diffusion and acceptance among professionals. Therefore, it can be challenging to determine at what point in its development process, a new surgical approach can be deemed a

procedures that include pacemakers with defibrillation capabilities warrant their own codes. However, at that point, the pacemaker code section was already full and there was no space for new codes. Consequently, the CMS assigned the relevant codes to codes starting with 00, "Procedures and Inverventions, Not Elsewhere Classified" (United States General Accounting Office (GAO) (2002)).

fully-fledged innovation.[36] The determination of the CMS in this case, based on input by various professionals in the area, offers one way of doing so without introducing the need for a possibly arbitrary judgment call from the researcher.

Naturally, the example also highlights some of the weaknesses of this measure. The first and foremost is that there may be a sometimes substantial lag between the time a procedure is first performed and the time it is finally codified For instance, the earliest records of laparoscopic hernia repair in the US date back to the 1980s, but the ICD9-CM code for this approach did not come about until 2008 (LeBlanc and Ger (2001), Centers for Medicare & Medicaid Services (CMS) (2007)). Given the incremental nature of surgical innovation, it is not immediately obvious that the first time a procedure is performed should be considered the time of its "invention", but it does highlight the issue of how to determine when exactly an invention occurred. In this paper, I follow precedent in the literature and deal with this issue by aggregating invention and potential market size measures into multi-year windows (Acemoglu and Linn (2004)).

The second weakness is that the code system my present some ambiguity in how to count innovations. This stems from the structural limits to the expressiveness of the ICD9 system. Specifically, as highlighted above, the system's hierarchichal structure combined with limited spots for new digit imposes a hard limit to to the number of codes that can be incorporated. This means that in some cases, closely related, but distinct medical procedures may be grouped under the same code. For instance, there is a single code for multiple-vessel percutaneous angioplasty (3605), while the procedure may be importantly differentiated by the number of vessels involved and the type of device used (stent, catheter, or laser) (United States General Accounting Office (GAO) (2002)). Therefore, it is not a priori clear whether a single new code always means the same thing when applied to different systems and procedures. This can be a problem if the degree to which this happens is correlated with the underlying innovative potential of a particular CCS category. One prominent way in which this could happen is due to differential saturation of the various subsections of the code system is.

In this paper, I choose to be agnostic about this issue and treat a new code as a self-contained piece of innovation. This is equivalent to assuming that there is an underlying true measure of novel

[36]This is much less challenging in the case of pharmaceutical innovation for instance, where a new molecule that is approved by the FDA offers an easy to measure relatively uncontroversial measure of innovative output.

procedures and the former specificity issues amount to classic measurement error on this measure. One justification for this assumption is that even if different sections of the system are variously constrained by space, the CMS has introduced "miscellaneous" sections, i.e., codes beginning with "00" and "17", where codes from full subsections may be placed. This decreases the incentive to aggregate different procedures into single codes. At the same time, the CCS classifier correctly puts procedures from the "misc" and regular sections into the relevant CCS group (United States General Accounting Office (GAO) (2002)).

The final weaknesses have to do with what I cannot observe using this measure. Since it places an implicit threshold on the level of development of a procedure before it can be measured, this also means that I cannot observe innovative activity that hasn't reached this threshold yet. This includes for instance novel procedures that are still in the process of incremental improvement and discussion by the medical community. Lastly, I do not observe who is primarily responsible for the innovation, be it surgeon or medical center. This means while I can conduct the analysis in cases where aggregate causal factors are of interest (such as market size), I am constrained in my ability to talk about specific mechanisms.

2.3 Analysis Sample

2.3.1 Surgical markets

I use the Agency for Healthcare Research and Quality (AHRQ)'s Clinical Classification Software (CCS) to define surgical markets. This is a three-level grouping of procedures into medically meaningful categories performed with the help of medical and coding professionals. At the highest level, there's 16 possible categories corresponding to groups of procdures performed on body systems - "Operations on the cardiovascular system", "Operations on the nervous system", etc. At the second level, we have procedures for specific approaches or organs: "Heart valve procedures", "Coronary artery bypass graft", etc. At the lowest level, there's further specification of the kind of procedure performed: "Bypass of one coronary artery", "Bypass of two coronary arteries", etc. There are 206 second-level categories and 345 third-level categories. I employ the second level of calssification in my analysis. Categories at this level have the benefit of being broad enough to include multiple procedures directed towards similar counditions while being specific enough for the

included procedures to be similarly affected by purported demographic changes. Since the lowest level includes mostly variations of a single procedure it is too specific to be a market, while the highest level may include groups of procedures which are differentially affected by demographic changes.[37]

In calculating relevant market sizes and innovation counts, I focus on a broad definition of surgical procedures developed by AHRQ. This definition includes invasive procedures ("involving incision, excision, manipulation, or suturing of tissue that penetrates or breaks the skin or enters a body cavity through an existing orifice; typically requires use of an operating room; and also typically requires regional anesthesia, general anesthesia, or sedation to control pain") as well as "a broader group of diagnostic and less invasive therapeutic surgeries...that may not fit the more strict definition of surgery applied for the narrow flag, but are often performed in surgical settings" (Agency for Healthcare Research and Quality, Rockville, MD (2020b)). Examples include C-sections, biopsies, joint replacements, catheterization. Example of procedures that do not fit definition are blood transfusion, vaccination, and radiosurgery. Since the focus of this study is surgical procedures and since non-surgical procedures are likely to have different market dynamics, I restrict both innovation counts and market size counts to ICD9 CM codes that fulfill this definition.

2.3.2 Potential market size

One of the advantages of the CCS taxonomy is that it allows a relatively straightfoward construction of the independent variable of interest, potential market size. My empirical strategy relies on exploiting the impact of broad demographic shifts due to the aging US population on the size of surgical markets. In order to implement this, I estimate the age composition of the patients in a given surgical market at the start of the sample period and then project how market size would

[37] An example may help clarify this point. Appendix Figures B1 and B2 show the age profiles of CCS Level 2 categories 10.3 (nephrectomy) and 10.4 (kidney transplant). The figures plot the share of individuals in the population in a given age bin that get the procedure in question. Thus Figure B1 shows that the incidence of nephrectomy rises steadily over an individual's middle age until it peaks in the 60s and 70s. Figure B2 shows that the peak of kidney transplant incidence occurs substantially earlier, in one's 40s and 50s. This means that the same population shift towards old age may impact these two procedures differently - while the effect may be unambiguously to increase the number of nephrectomy patients, it may increase or decrease the number of kidney transplant patients depending on the exact shape of the demographic shift. Nonetheless, both of these procedures fall under CCS Level 1 category 10 - "Operations on the urinary system." Therefore, calculating a single age profile for Level 1 procedures as an exposure to subsequent demographic shifts may mask important heterogeneity among the Level 2 procedures that make the group up and thus bias the empirical results.

change based on demographic shifts alone. The aim is to purge variation in the independent variable coming from surgical innovation itself - any shifts in market size due to increased quantity and/or quality of the surgeries in that market.

In order to estimate the time-invariant age-composition of each market, I use two sources of data - the National Inpatient Sample (NIS) and the National Survey of Ambulatory Surgery (NSAS). The National Inpatient Sample is a representative sample of all hospital discharges in the US prepared by the AHRQ as part of its Health Cost Utilization Project (HCUP). The NSAS is a nationally representative survey of patients performed in outpatient settings. Since 1994 is the earliest year when the NSAS is available and since I need data on both inpatient and outpatient surgeries in order to construct accurate measures of the age composition of each surgery's patients, I take 1994 as the earliest year in the analysis. Both the NSAS and NIS contain information on patient characteristics and the medical procedures that the patients underwent during their stay. For each CCS Level 2 market, I use the surveys to estimate the total number of surgeries performed in that category in 1994 by 10-year age bins.

I then use the CPS to estimate the total population in each age bin in each year between 1994 and 2013. I then use this market information to project the total market size forward from 1994. Specifically, potential market size is defined as:

$$M_{ct} = \sum_{a} u_{ca} i_{at} \tag{18}$$

where M_{ct} denotes the projected number of surgeries needed in CCS category c in period t, u_{ca} represents the number of surgeries from CCS category c per person in 10-year-age-group a (calculated using NIS and NSAS in 1994), and i_{at} represents the number of people in age-group a in period t (calculated using CPS). This method is analogous to that used by Acemoglu and Linn (2004).

The age profiles of the CCS Level 2 procedures indicate that demographic shifts are likely to be a powerful influence on projected market size. Figure 2.1 displays the variation in the age profile across categories. In order to provide a summary visual measure of this variation, I divide the CCS Level 2 categories into three age groups (0-29 year-old, 30-59 year-old, and 60 year-old and older) based on the age group that is responsible for most of the patients for the respective

category. The figure shows the number of CCS Level 2 categories in each of these age bins. The figure shows that there is a large number of categories in each of the coarse age bins. This means that broad demographic shifts such as the overall aging of the US population are likely to affect different CCS Level 2 categories quite differently and thus that empirical tests based on changes in projected market size are likely to have power to detect significant changes in the number of surgical innovations for a particular market.

The existence of multiple waves of the NSAS and NIS allows me to examine the stability of the age-profiles of these procedures. To do so, I use the methodology from above to construct age profiles for all years when both datasets are available (1994, 1995, 1996, and 2006). I then examine the correlation between the age profiles in these various waves of the sample. The results are presented in Table 2.1. The first row computes raw correlations between the age-bin-surgery counts in each pair of years shown in the columns. The second row computes these correlations within each level-2 CCS surgical category and then averages the correlations over all categories. Both rows show a high degree of persistence of the age profiles for these surgical categories. This suggests that CCS is a valid classification of surgical procedures and a reasonable candidate for market definition.

2.3.3 Pharmaceutical classes

Another advantage of the CCS system is that it allows for a relatively straightforward comparison with well-established classifications of pharmaceutical products. Specifically, the anatomy-based organization of CCS corresponds closely to the higher levels of the Anatomic Therapeutic Chemical (ATC) pharmaceutical classification system. ATC classifies pharmaceuticals based on the anatomic system they act upon and the drug's pharmacological properties. The system has five levels. The top level classifies drugs based on the system they act upon. The second level uses the drug's therapeutic area and each successive level adds more pharmacological detail to the classification system. For example, the anticoagulant dabigatran etexilate used in the prevention of blood clots and strokes and approved by the FDA in 2010 has ATC Level 5 code B01AE07. Its Level 1 through Level 4 categorizations are respectively "B - Blood and blood-forming organs", "B01 - Antithrombotic agents", "B01A - Antithrombotic agents", and "B01AE-Direct thrombin

inhibitors" (World Health Organization (2020)). [38]

ATC is a popular system for use in research on pharmaceutical innovation (e.g., Agha et al. (2020), Dubois et al. (2015)). One of its advantages in the present context is the fact that the higher levels of the system are based on anatomic systems that match closely the systems used in CCS. This allows me to define markets (anatomic systems) where pharmaceutical and surgical innovation are comparable. I can therefore compare patterns in surgical innovation to those in pharmaceutical innovation, which has been much more extensively studied in the literature.

I begin by identifying CCS and ATC categories where there is a clear match based on the anatomic system where the procedure or pharmaceutical acts. The crosswalk is shown in Table 2.2 below. For most of the systems, the match is fairly clear, but there are some categories where there is no obvious match. These include for instance "Obstetric procedures" on the CCS side and "Antiparasitic products, insecticides, and repellents" on the ATC side. I exclude these categories and focus on the systems where there is a clear match. This leaves me with ten systems: alimentary tract, hemic/lymphatic system, cardiovascular system, dermatologicals, genito-urinary system, endocrine system, musculo-skeletal system, nervous system, respiratory system, and sensory organs.

I measure pharmaceutical innovation in each category as the number of new moleculary entity (NME) approvals by the FDA. A new molecular entity is a drug that contains at least one active moiety that has not been approved by the FDA before. There are multiple ways of measuring pharmaceutical innovation (e.g., total FDA approvals). However, the number of NMEs is typically employed as a measure of novel and impactful inventions since it requires that the new products considered contain a chemical component that has not been previously approved by the authorities (see, e.g., Dubois et al. (2015)). Since the relevant comparison is with surgical procedures which did not exist or were not recognized as separate procedures previously, this seems like the most natural comparison group.

I obtain the list of NME approvals between 1994 and 2013 from the FDA. Using three large web databases on biochemical products (BioPortal, Drugbank, the National Library of Medicine's PubChem), I find the ATC classification for each of these products and assign them to one of the

[38] Since pharmaceuticals can act through multiple channels and be used for different conditions, it is possible for a drug to have multiple ATC designations. However, of the 640 new molecular entities in my sample, 557 have a single ATC classification.

ten systems used for this part of the analysis.[39] I also use the ICD9 revisions to count the number of new surgeries in each of these markets over the sample period. I also aggregate to four-year intervals as I do in the main analysis.

2.4 Results

2.4.1 Comparison with pharmaceutical markets

I begin by displaying the results from the descriptive analysis comparing surgical and pharmaceutical innovation. Figure 2.2 shows the total innovation counts in each of these markets ordered by pharmaceutical innovation counts. There are two take-aways from the figure. First, the most innovative markets as measured by NMEs and new surgical codes conform with popular ideas of what areas of the body are traditionally treated by pharmaceuticals and surgeries. For instance, the cardiovascular system is commonly perceived as an area of intense technological development over the past few decades and ranks highly both in terms of surgical and pharmaceutical innovation. The nervous system is an area of the body where until recently few surgical remedies have been available, but where drugs play an important role (e.g., through the recent rise in anti-depressant medication (Cutler (2005))). Accordingly, this system ranks highly in pharmaceutical innovation counts, but not as highly in surgical innovation counts. Finally, we have systems such as the musculo-skeletal system where recent advancements in joint replacement and minimally invasive surgery have been quite important, but where pharmaceutical remedies have been less salient (see, e.g., Hansen and Bozic (2009)).

Second, the figure presents more evidence that surgical and pharmaceutical innovation markets are quite different. For instance, while some classes (e.g., nervous system) rank quite highly in terms of pharmaceutical innovation and somewhat middling in terms of surgical innovation, others (e.g., alimentary system) rank quite highly in terms of innovation for both. The overall correlation in rank between the two sets of innovation is 0.39, indicating innovative efforts are positively, but imperfectly correlated.

[39] If a new molecular entity has multiple ATC designations, I assign it to all of the systems under which it falls. The underlying assumption is that if an innovation affects two separate systems, it can be counted as two separate innovations. However, since the vast majority of NMEs in the sample has a single ATC match, this is not an important restriction.

Figure 2.3 then takes this evidence one step further by showing the trends in these markets over time. Specifically, it depicts the total innovation counts in each of the body system markets identified in the previous section in four-year intervals over the sample period. The graph shows that the dynamics of surgical innovation are quite different from the dynamics of pharmaceuticals. While some markets seem to show parallel trends (e.g., sensory organs), others show no or even negative correlation between the two (e.g., cardiovascular system). These differences are consistent with the different institutional organization of the two kinds of markets highlighted above. Most of the pharmaceutical markets show a stable or smoothly decreasing trend over time. This is consistent with concerns voiced by some researchers of decreasing industry output of innovative drugs (Pammolli et al. (2010)). It is also consistent with a research process that is stable, directed, and driven by rational design. The surgical innovation counts on the other hand exhibit quite disparate trends and higher variance around the trend. This pattern is consistent with the decentralized nature of the process and the fact that innovative activity is driven by incremental technological breakthroughs achieved by individual surgeons.

Finally, Figure 2.4 shows a simple scatterplot of the two innovation measures. Specifically, it shows the raw number of new procedures versus new pharmaceuticals in each ATC - Level 1 bin and 4-year time interval. The scatterplot shows that the two measures are essentially uncorrelated. In particular, the slope of the line of best fit is 0.02. This is yet another indication that at least in the aggregate, the two sets of markets follow different trends and processes.

2.4.2 Impact of potential market size

Visual evidence

The previous section has documented evidence consistent with the possibility that the different institutional setup of surgical and pharmaceutical markets has resulted in different market dynamics. These institutions such as market competition and profit incentives have great theoretical importance in driving the responsiveness of innovation to potential market size. These two observations give rise to the concern that markets that need more surgical innovation may not receive it. Accordingly, this section investigates whether increases in market size do indeed lead to higher levels of surgical innovation.

I begin with a visual representation of the variation underlying my results. Figure 2.5 shows the trends in the share of population due to individuals in three broad age groups - 0-29,30-59, and 60+. The figure shows that the share of individuals in the youngest age group is consistently declining, while that of individuals in the oldest age group is consistently increasing over the sample period. The share of individuals in the middle age group displays a flat inverse-U shaped trend with an initial increase and a slow decline after the early 2000s.

Figure 2.6 then scales these population trends by the probability that an individual in these age groups will receive surgery using the 1994 NIS and NSAS sample. This transformation gives the projected share of surgeries in each age group in the sample period where the trend movements are coming from the population movements displayed in Figure 2.5. Since older individuals are significantly more likely to receive surgery, the figure shows high share of surgeries due to the oldest and middle age group. The increase in projected surgeries due to the oldest group is much more pronounced as is the decrease due to the youngest group. The inverse-U shape due to the middle age group is also much more pronounced here.

Lastly, Figure 2.7 then connects these trends with the trends in innovation counts. In particular, I begin by taking all CCS categories and assigning them to an age group the group where the most procedures from the respective category were performed (e.g., for "Procedures of the Eye", the group 60 and over had the highest number of procedures in 1994 and thus it was classified as a "60 and over" group of procedures). Then, for each age group, I plot the share of total surgical innovations that occurred in CCS categories in that group in four-year intervals starting in 1994. The trends displayed in Figure 2.7 mirror those from Figure 2.6. In particular, there is a general increase in the share of new procedures for the oldest age group and a general decline in procedures for the other two age groups that is more pronounced for the middle age group after the early 2000s when the share of population due to the middle-aged group starts steeply declining. This is precisely the type of variation that I will be exploiting in the analytical specifications. These specifications use this variation much more efficiently, however, since they use more detailed age profiles for each CCS category that exploits variation coming from demographic shifts at a finer level.

Analytic specifications

The first econometric challenge in this setting comes from the likely non-linearity of the functional dependence between innovation and potential market size and the fact that this is a setting where the dependent variable takes on the value of zero in a non-trivial fraction of observations. In particular, I take as a starting point and target for estimation the result of traditional models of innovation (e.g., Acemoglu and Linn (2004)) that new product entry is an exponential function of the independent variables:

$$E[N_{ct}|\zeta_{ct}] = \exp(\alpha \log M_{ct} + \zeta_c + \mu_t) \tag{19}$$

where N_{ct} represents the number of new procedure codes in category c in period t, M_{ct} is market size as above, ζ_c is CCS-category fixed effects and μ_t denotes period-fixed effects. The parameter of interest α is the elasticity of innovation with respect to potential market size.

The distribution of the dependent variable is typically assumed to be Poisson. While a Poisson regression will yield consistent and efficient parameter estimates if the distributional assumptions are correct, these estimates may be both inconsistent and inefficient if these assumptions turn out to be wrong (Blackburn (2007), Correia et al. (2013)). Thus, concerns about functional form assumptions usually steer researchers towards using a linear model. The benefit of doing so is that under standard assumptions, OLS will yield the best linear approximation to the true non-linear conditional expectation function regardless of the latter's exact form. in a minimum mean-squared error sense (Angrist and Pischke (2008)).

This leads to the second challenge, which is that the preferred log-log form of the linear specification cannot be effectively implemented in a setting where the dependent variable takes on values of 0 (in the current case, this happens because there are periods of no new surgeries for some markets). The traditional approach has been to set the log of the number of innovations equal to 0 in those cases and include a dummy for "0 on the left-hand side" as an independent variable, which yields biased estimates (Acemoglu and Linn (2004)). An additional issue in this kind of specification is that even if all the dependent variable values are strictly positive and the log-log functional form is correct, heteroskedasticity in the error term may bias the coefficients of interest badly (see, e.g., Blackburn (2007)).

There is no single widely accepted solution to the combination of all of these issues, so I take an approach that is standard in the literature and estimate both non-linear and linear models that are relatively better at solving different sides of these problems. For my non-linear specification, I use a quasi-maximum likelihood (QML) Poisson count model. The QML has the benefit of requiring only that the conditional mean function be specified correctly for the consistency of the parameter estiamtes. In other words, even if the distribution of potential innovation is not Poisson, the QML procedure will still yield consistent estimates of α. The cost of this approach is that if the distribution is not Poisson, then the parameter estimates may be inefficient (Blackburn (2007), Correia et al. (2013)).

To implement a linear specification, I use the inverse hyperbolic sine (IHS) transformation of the number of new codes for a CCS category in a given period. The IHS transformation allows me to approximate a log-log specification while allowing for markets and periods of zero-innovation on the left-hand side:

$$IHS(N_{ct}) = \beta \log M_{ct} + \zeta_c + \mu_t + \varepsilon_{ct} \tag{20}$$

Here β also has an interpretation of elasticity of innovation with respect to market size. The benefit of this approach is that it is simple, transparent, and yields an unbiased estimate of β if the errors are homoskedastic conditionally independent of $\log M_{ct}$. However, the estimate of β may differ markedly from the estimate of α due to three reasons. First, the left-hand side is an approximation to, and not an exact match of the functional form in equation (19). Second, as mentioned previously, in the presence of heteroskedasticity in the error term, the estimates of β may be biased. Third, estimates of equation (20) implicitly use the linear functional form to extrapolate to cells with zero innovation in all periods. The Poisson QML on the other hand relies on a subsample of markets which change the number of innovations from period to period. This implicitly drops all markets that have no innovation whatsoever during the sample period.

The results are presented in Table 2.3 below. The first three columns show estimates from the Poisson model, while the last two columns show estimates from the linear model. Colum (1) shows the main Poisson specification from equation (19) above. Columns (2) and (3) use instead the lag or lead value of that variable respectively. Column (4) shows the baseline linear specification in

(2), while column (5) uses a 0-1 indicator for any innovation in a given CCS category on the LHS.

The results in all columns are positive and significant. The interpretation of the estimate in column (4) is that a 1% increase in expected market size is associated with a 1.5% increase in surgical innovation in that market. Similarly, Column (1) shows that a 1% increase in expected market size is associated with a 20% increase in surgical innovation. Lag maket size is more highly correlated with innovations than lead market size suggesting that expectations of market size changes do not play a big role in this setting. The estimate in column (5) indicates that at least some of the market size effect comes from the extensive margin.

As highlighted above, there are a variety of reasons why the estimates from the linear and QML estimations may differ. The observation counts in columns (1) and (4) are quite different, indicating that the effective set of markets which inform the estimate changes substantially between the Poisson QML and linear regression. As suggested by the discussion above, both methods of estimation have relative strengths and weaknesses. While QML has the benefit of consistency even under misspecification, it may not deal well with a large fraction of zero-innovation markets. The linear estimation on the other hand may yield biased estimates in the presence of heteroskedasticity and relies at least partially on functional form to deal with zero-innovation markets. In what follows, I will use the linear estimation results as the preferred set of parameter estimates because the large fraction of zero-innovation markets seem to play an important role in this setting and because it is more in line with other estimates of the role of market size in driving innovation in the healthcare context.

To investigate the possibility that the presented here are driven by outliers, I create a binscatter of the regression in equation (20). Specifically, I residualize the inverse hyperbolic sine of innovations on the left-hand side from year and market fixed effects and plot them against the similarly residualized values of log expected market size. The results are presented in Figure 2.8. The figure clearly shows that while the relationship between the two variables of interest is somewhat noisy, the positive slope is not driven by outliers.

Given the importance of no-innovation markets in the current setting, I prefer the estimates in column (4) as my baseline specification because they are in some sense agnostic towards the presence of such markets. These estimates are lower than the comparable estimates of around 3% found in Acemoglu and Linn (2004) and Dubois et al. (2015). It is notable as well that the high end

of the estimates in these two papers (12% in the former and 32% in the latter) are also comparable to the high-end estimates here. These results suggest that while novel surgeries are responsive to market size changes, they are less responsive than pharmaceuticals. This is again consistent with the different institutional setup of the surgical market. supply-side constraints on the role of profits due to incrased expected market size in driving innovation.

2.5 Discussion

The results from this investigation indicate that even though surgical markets lack the institutions prescribed by standard theory as instrumental for innovation, they are likely to respond to the changes in demand for procedures due to the aging of the US population. This result is policy-relevant for a variety of reasons. First, perhaps obviously, it suggests that regulatory intervention to boost innovation in order to satisfy increased demand among the elderly is not required. Calls for private-sector involvement of various aspects of healthcare in order to improve outcomes have been common throughout the past few decades (as visible in, e.g., the design of Medicare Part D (Megellas (2006))). Some surgeons have even attempted (unsuccessfully) to license procedures they invent (Judge (1997)). This paper suggests that government reform towards strengthening intellectual property rights is not called for in this setting.

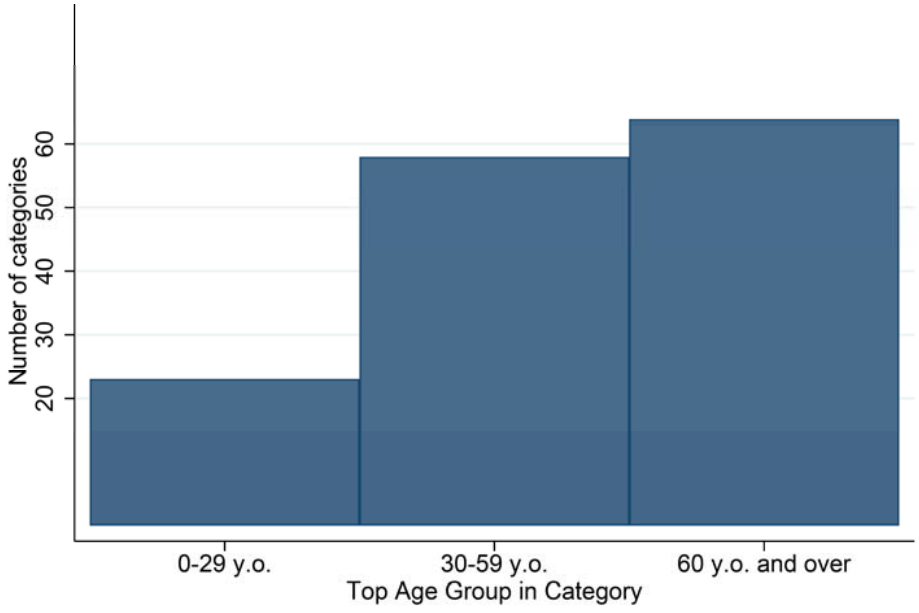
Second, the results raise some interesting possibilities for optimizing medical innovation. The model of physician-driven innovation has proven to be a robust alternative to firm-based research and development in this context. One cannot help but wonder whether it hides a possible solution (or part of a solution) for some of the issues with pharmaceutical innovation that have gained salience in recent years. First, a number of economists have raised worries about a possible decline in innovative productivity.among pharmaceutical firms (Pammolli et al. (2010)). Second, recent research has shown that firm profit incentives may be introducing a bias in the size and significance of measured pharmaceutical effectiveness, a result which raises the worry that some not insignificant fraction of pharmaceuticals entering the market are ineffective or less effective than widely believed (Oostrom (2020)). Physician driven innovation follows a wholly different development process and is likely less affected by financial bias.

The extent to which this can be helpful depends of course on the questions of “what kind of

innovation can we achieve" and "what exactly are the mechanisms at play". As highlighted in the institutional background section above, innovations that can be developed using this model are likely to be less capital-intensive than what can be achieved using profit maximization. Additionally, whether the mechanism is user-driven innovation or academic prestige or even happenstance innovation can matter a great deal for the possible implementation of this model in other settings or its improvement in the current setting. These are questions that will be left for further study.

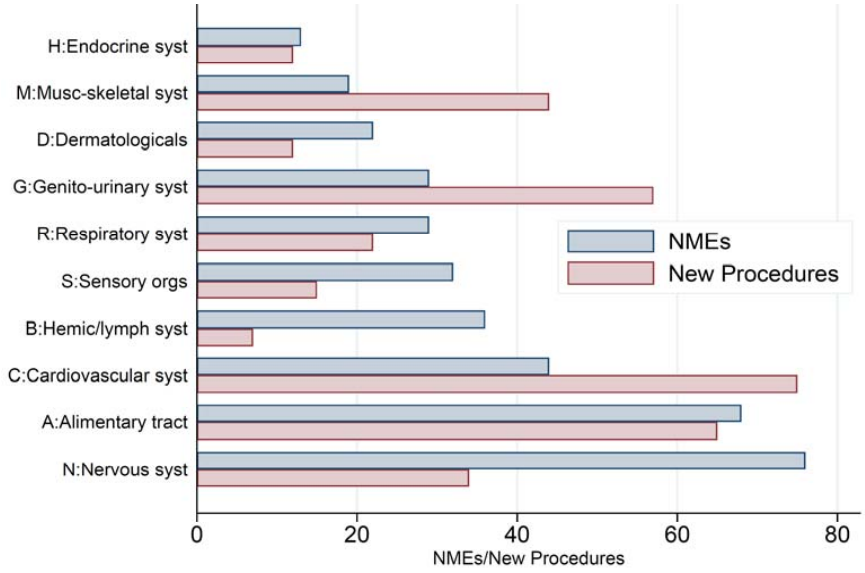
2.6 Figures and Tables

Figure 2.1: Age Profile Variation



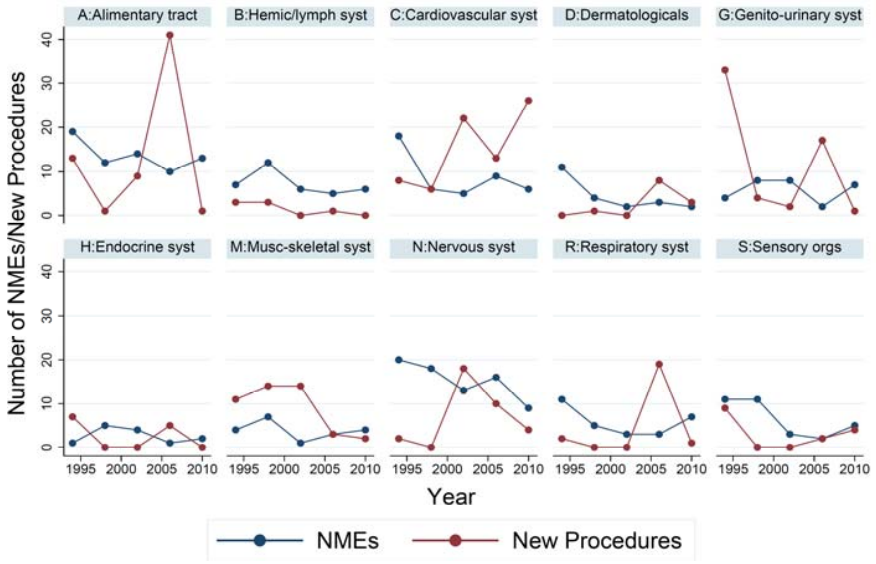
Notes: The histogram displays the distribution of doctor age at new code introduction for all treatment categories. Each observation represents a procedure on an inpatient claim.

Figure 2.2: Total Innovation Counts Ordered by Number of NMEs



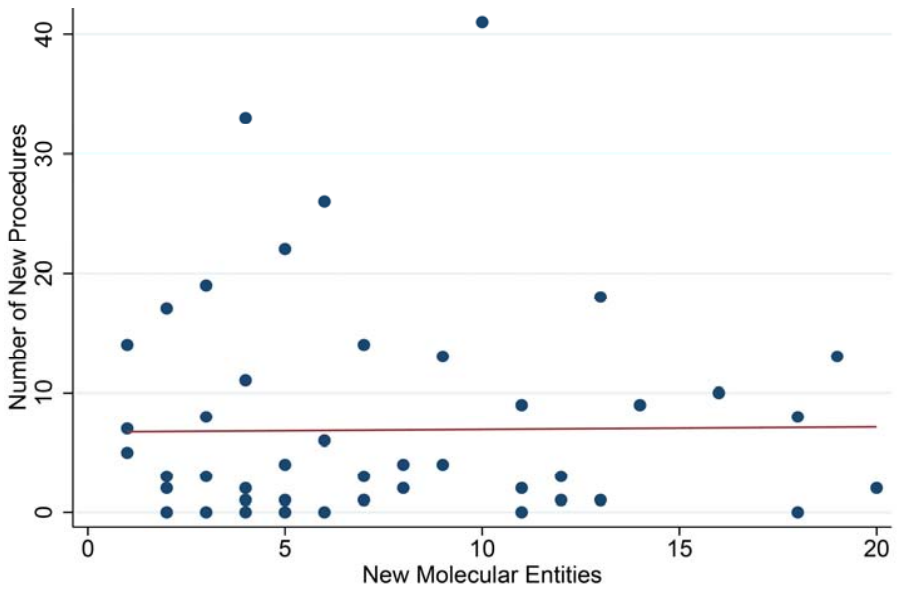
Notes: Figure shows the total number of pharmaceutical and surgical innovations in ten groups defined so as to match relevant ATC Level 1 categories. The categories are ordered by the total number of pharmaceutical innovations. Pharmaceutical innovations are defined as New Molecular Entities. Surgical innovations are defined as new ICD9 procedure codes.

Figure 2.3: Innovation Trends by ATC Class



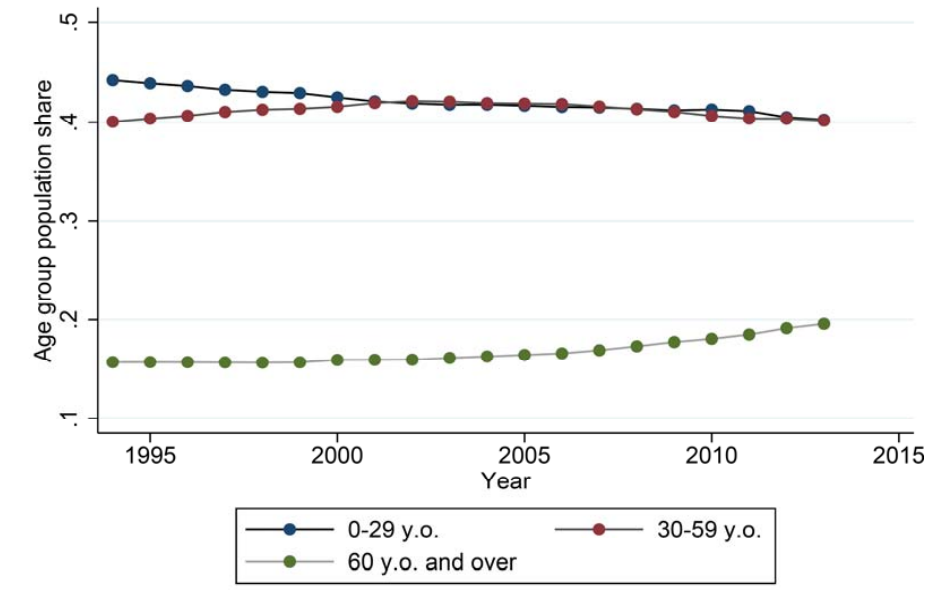
Notes: Figure shows the trends in the raw number of pharmaceutical and surgical innovations in ten groups defined so as to match relevant ATC Level 1 categories. Each dot represents the total number of surgical or pharmaceutical innovations over a four-year interval starting in the year defined by the dot's x-coordinate. Pharmaceutical innovations are defined as New Molecular Entities. Surgical innovations are defined as new ICD9 procedure codes.

Figure 2.4: Scatterplot of Surgical vs Pharmaceutical Innovation Count



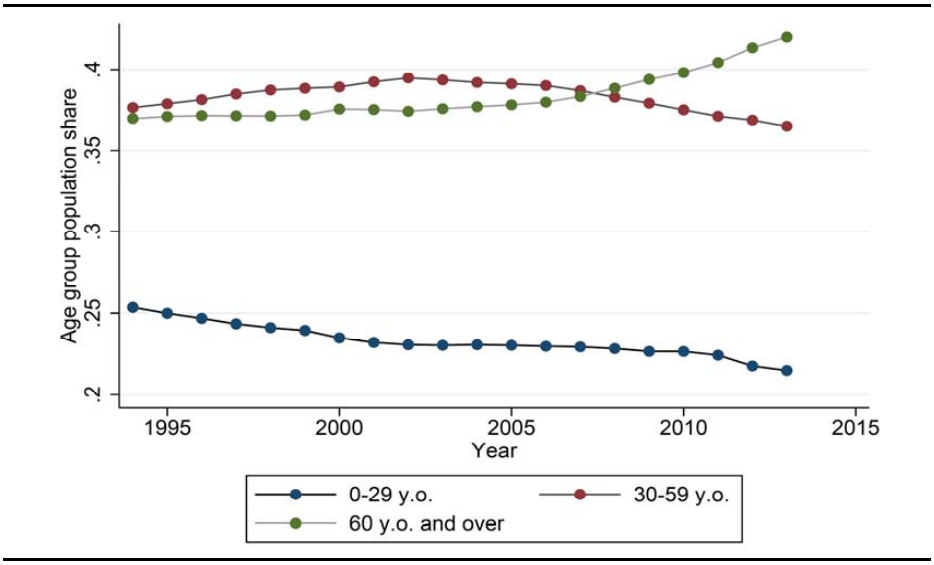
Notes: Figure shows a scatterplot of the raw number of surgical and pharmaceutical innovations in ten groups defined so as to match relevant ATC Level 1 categories. Each dot represents the total number of surgical or pharmaceutical innovations over one of the sample's four-year intervals. The red line represents the line of best fit between these two variables (slope of 0.02). Pharmaceutical innovations are defined as New Molecular Entities. Surgical innovations are defined as new ICD9 procedure codes.

Figure 2.5: Share of Population in Different Age Groups



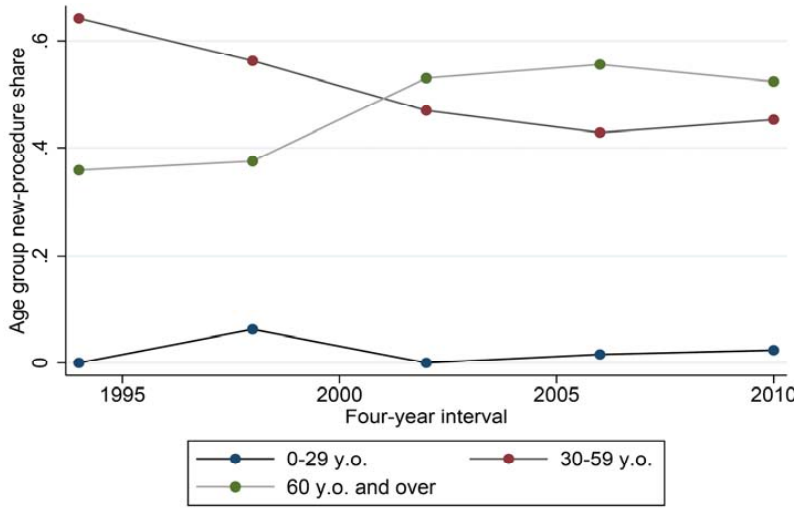
Notes: The figure shows trends in the share of total US population in each of three age bins over the sample period 1994-2013: 0-29 year-old, 30-59 year-old, and 60 year-old and older. The poulation count data comes from the American Community Survey.

Figure 2.6: Projected Share of Surgical Procedures for Different Age Groups



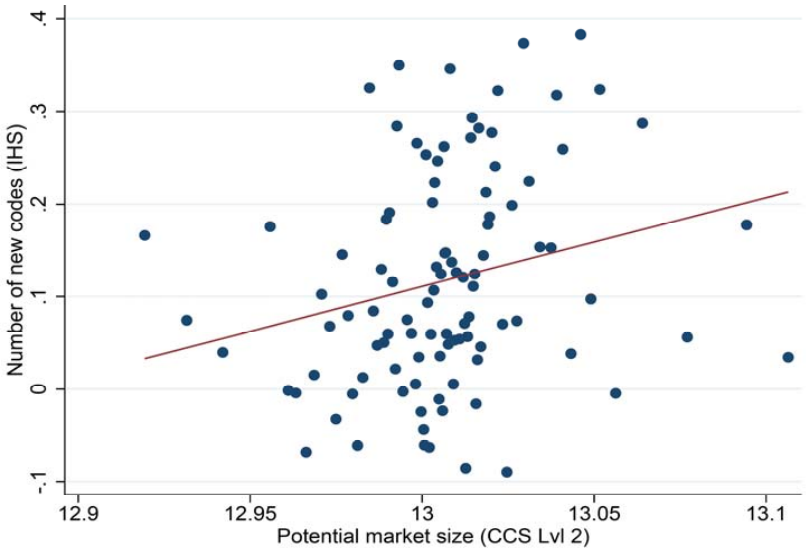
Notes: The figure shows trends in the share of total projected US surgeries in each of three age bins over the sample period 1994-2013: 0-29 year-old, 30-59 year-old, and 60 year-old and older. I begin by calculating the share of total surgeries in each age group in 1994 using the National Inpatient Sample and National Survey of Ambulatory Surgery. Then I project the number of surgeries forward using only demographic changes to the number of people in each group calculated using the American Community Survey.

Figure 2.7: Share of New Procedures Belonging to Different Age Groups



Notes: The figure shows trends in the share of total surgical innovation in each of three age bins over the sample period 1994-2013: 0-29 year-old, 30-59 year-old, and 60 year-old and older. The surgical innovaiton counts come from introductions of novel ICD9 CM surgical procedure codes. Each novel surgery is assigned to the age group that has the highest share of patients receiving surgeries in the procedures CCS-Level 2 category. These counts are then aggregated to four-year intervals the first year of which is indicated by each dot's x-coordinate.

Figure 2.8: Baseline Regression Binscatter



Notes: Figure shows a binscatter of the baseline specification (2) regressing the inverse hyperbolic sine of surgical innovation on log projected market size including year and market fixed effects.

Table 2.1: Stability of Age Profiles

	1994/1995	**1995/1996**	**1996/2006**
Correlation	0.996	0.997	0.954
Mean correlation by category	0.906	0.911	0.795

Notes: Table shows the correlation between the estimated age profiles of CCS Level 2 categories across different waves of the National Inpatient Sample/National Survey of Ambulatory Surgery. The age profile of each category is the share of surgeries performed on patients from each 10-year-age bin. The top row shows the raw correlation between the shares across all age bins and categories between each pair of years. The bottom row first calculates this correlation across age bins within a category and then averages these correlations for each pair of years.

Table 2.2: CCS-ATC Designation Correspondence

CCS Level 1 Class	ATC Class
1: Operations on the nervous system	N: Nervous System
2: Operation son the endocrine system	H: Systemic hormonal preparations, excluding sex hormones and insulins
3: Operations on the eye	S: Sensory organs
4: Operations on the ear	S: Sensory organs
5: Operations on the nose, mouth, and pharynx	S: Sensory organs
6: Operations on the respiratory system	R: Respiratory system
7: Operations on the cardiovascular system	C: Cardiovascular system
8: Operations on the hemic and lymphatic system	B: Blood and blood forming organs
9: Operations on the digestive system	A: Alimentary tract and metabolism
10: Operations on the urinary system	G: Genito-urinary system and sex hormones
11: Operations on the male genital organs	G: Genito-urinary system and sex hormones
12: Operations on the female genital organs	G: Genito-urinary system and sex hormones
13: Obstetrical procedures	No match
14: Operations on the musculoskeletal system	M: Musculo-skeletal system
15: Operations on the integumentary system	D: Dermatologicals
16: Miscellaneous diagnostic and therapeutic procedures	V: Various
No match	I: Anti-infective for systemic use
No match	L: Antineoplastic and immunomodulating agents
No match	P: Antiparasitic products, insecticides, and repellents

Notes: Table shows the correspondence between CCS Level 1 and ATC Level 1 categories used for the comparison between surgical and pharmaceutical innovation in the text.

Table 2.3: Effect of Potential Market Size on Surgical Innovation

	(1)	(2)	(3)	(4)	(5)
VARIABLES	New code ct	New code ct	New code ct	IHS of new code ct	1(Any new code)
			CCS Level 2 (141 categories)		
Market Size	19.71***			1.569**	0.853**
	(4.849)			(0.702)	(0.396)
Lead Market Size		19.13***			
		(5.158)			
Lag Market Size			22.88**		
			(10.09)		
Observations	305	228	204	705	705
R-squared				0.368	0.393
Dep Var Mean	1.11	1.25	1.31	0.35	0.12
CCS & Year Interval F.E.	x	x	x	x	x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running specifications (1) and (2) on the main analysis sample. Columns (1)-(3) show results from using a Poisson specification using current, lag, or lead projected market size respectively. Standard errors are clustered at the CCS Level 2 category level.

3 Chapter 3
Geographic Variation in Healthcare Utilization: The Role of Physicians

The most expensive piece of medical equipment, as the saying goes, is a doctor's pen.

Atul Gawande, "The Cost Conundrum"

3.1 Introduction

The intensity of per-capita healthcare utilization varies widely across the U.S, but its causes are not well understood. A leading example is the substantial geographic variation among over-65 Medicare enrollees. Existing evidence suggests supply-side factors are key drivers of these disparities (Skinner, 2011; Chandra et al., 2012), explaining about half of the observed geographic variation (Finkelstein et al., 2016). This supply side encompasses both the practice styles of local

physicians[40] and other aspects of the practice environment such as hospital capacity, physical capital, and organizational, financial, and legal incentives. Recent work has documented substantial variation in practice patterns across physicians (Currie and MacLeod, 2020; Chan et al., 2020), but little is known about the quantitative importance of this variation relative to area-level differences in practice environments.[41] Decomposing the role of supply-side factors along these lines is essential for understanding potential impacts of policies aiming to reduce health care expenditures or improve health outcomes.

This paper quantifies the role of physicians in explaining the geographic variation in Medicare utilization. One widely-held view is that physicians play a large role; for example, Cutler et al. (2019) link patient and physician responses to treatment vignettes with area Medicare expenditures and argue that physician beliefs about treatment are the key driver of regional variation. On the other hand, Molitor (2018) examines changes in physicians' treatment intensity for Medicare heart attack patients as they move across regions and concludes that cardiologist practice styles explain relatively little variation in treatment intensity; other aspects of the practice environment instead appear to be the main driver.

Our empirical approach analyzes variation in patient-physician encounter networks in a nationally-representative sample of over-65 Traditional Medicare beneficiaries to fit parsimonious models for two different outcomes: average utilization ("treatment intensity") by a patient-physician pair, and the number of physicians seen by each patient in each year. Both models are identified by quasi-experimental variation in patient and physician migration, with the utilization model also allowing flexible within-region matching of physicians and patients conditional on observables. We follow prior literature in focusing on variation across hospital referral regions (HRRs).

To isolate variation in physicians' average treatment intensity, we estimate a three-way fixed effects linear regression of utilization at the level of patient-physician pairs, allowing for fixed unobserved heterogeneity in patient demand, physician practice style, and other regional supply factors. We estimate substantial variation across physicians in average treatment intensity. The

[40] In keeping with previous literature, we use practice style to signify the preferences, training, and experience of individual physicians. This defnition is embodied in time-invariant physician factors that affect spending as well as physician mover relative year effects in our empirical analysis.

[41] For example, Reinhardt (2019) writes: "It takes a hard-core economist, beholden to faith-based theories, to believe that these [geographic] variations reflect demand by patients, forged by their knowledge about the appropriate care for their medical conditions. Much more probably these variations reflect the belief among physicians of what is the appropriate care or the financial incentives physicians face to favor one treatment over another."

majority of this variation is within region, although there are also pronounced differences in average physician effects across HRRs.

For the analysis of the number of physicians seen by each patient in a year, we estimate a two-way fixed effects Poisson regression that decomposes the number of physicians seen into patient demand and regional supply factors. We find a high amount of variation in both patient and regional factors. For instance, moving the average patient to an HRR that has a one standard deviation higher place effect increases her expected number of encounters by 0.26. We also find that patients and places contribute roughly equally to the observed geographic variation in number of encounters.

We then combine estimates from the two models to construct a sequential decomposition of regional variation in annual per capita healthcare utilization, quantifying the roles of patients, physicians, and residual (non-physician) area supply-side factors. Consistent with Finkelstein et al. (2016), we find that around half of regional utilization variation arises from the non-random distribution of patients with different demand for care. We further find that over 90 percent of this patient component arises from differences in demand for the number of physicians seen in a year, a channel potentially related to healthcare fragmentation. A much smaller share of the overall patient component is attributable to differential demand for per-physician utilization, with virtually no role for sorting of patients with different demand-side preferences across physicians with different practice styles. Understanding heterogeneity in the demand for physician quantity thus appears central for characterizing the role of patients in regional utilization gaps.

The next step in our decomposition is to investigate the drivers of the remaining supply-side variation. We find that around three-fifths of the supply-side component of variation is due to differences in physician practice style. The remaining two-fifths of the supply-side component reflects residual (non-physician) place-based factors. We again find an important role of places on the number of physicians seen, consistent with earlier findings on the importance of healthcare fragmentation on utilization (Agha et al., 2019; Frandsen et al., 2015; Baicker and Chandra, 2004). Sorting of patients with different demand-side preferences across physicians of different practice styles has almost no impact.

The bottom line is that variation in physician practice style explains around 30 percent of the overall geographic variation in healthcare utilization. This is a significant share that still leaves

much room for other place-related supply side factors to play a role. We find a more limited role of physicians when focusing on specific physician specialties, such as cardiologists and primary care physicians, which have been the focus on prior literature. The more limited role for physicians in explaining geographic variation arises in part because of a strong negative correlation between where high intensity physicians are located and regions' residual supply-side factors.

Our analysis builds on a growing empirical literature that uses quasi-experimental changes in location or matching to separately identify individual heterogeneity from the systematic effects of geography or institutions, typically using linear fixed effect models. Prominent examples in the study of healthcare spending include Song et al. (2010), Molitor (2018), and Finkelstein et al. (2016). A similar approach is used outside healthcare to study determinants of wage variation (Abowd et al., 1999), wage inequality (Card et al., 2013; Card et al., 2016), neighborhood effects (Aaronson, 1998; Chetty and Hendren, 2018), cultural assimilation (Fernandez and Fogli, 2006), workplace shirking (Ichino and Maggi, 2000), brand preferences (Bronnenberg et al., 2012), and tax reporting (Chetty et al., 2013), among other topics. While most of this literature decomposes outcome variation in terms of two sources (e.g. patients and places, or workers and firms), our setting requires modeling a third dimension (physicians) in addition to patients and other supply-side factors.

The remainder of the paper is organized as follows. Section 3.2 presents the data and summary statistics. Section 3.3 describes our model and empirical strategy. Section 3.4 presents our findings. Section 3.5 concludes.

3.2 Data and Summary Statistics

3.2.1 Data and Variable Definitions

Our analysis draws from a 20 percent random sample of Medicare beneficiaries, covering encounters between patients and physicians from 1998 to 2013. We define an encounter by aggregating all unique patient-physician interactions in each year, including claims for inpatient and outpatient care as well as physician services. For each interaction we observe the type and quantity of care provided, as well as the primary diagnosis. We also observe patient demographic information, including age, sex, race, and zip code of residence, as well as physician clinical specialty. Appendix

C.1 provides more detail on our sample and variable construction.

Our geographic unit of analysis is a hospital referral region (HRR), as defined by the 1998 Dartmouth Atlas. An HRR is a collection of zip codes aggregated to approximate a tertiary hospital market. Following Finkelstein et al. (2016), we use the patient's zip code of residence to define location, regardless of where the claim is incurred. We categorize a patient as a "non-mover" if her HRR of residence does not change during our sample period, and as a "mover" if her HRR of residence changes exactly once. We drop patients whose HRR of residence changes more than once.[42] To match the timing at which we observe each patient's residence, we define all outcomes for year t to be totals of claims submitted between April 1 of year t and March 31 of year $t+1$. As in Finkelstein et al. (2016), we measure our primary outcome of interest (healthcare utilization) through an index of the use of medical resources that adjusts for regional variation in administratively set prices.

Following Molitor (2018), we use the claims data to define mover and non-mover physicians. For each physician practicing in a given HRR, we define a "practice episode" as the period between the first and last encounter date with patients residing in that HRR. We then define the physician's primary HRR as the one whose practice episode contains the most patients and her secondary HRR as the highest-patient-count HRR whose practice episode does not overlap with that of the primary HRR. A physician mover is defined as a physician who has both a primary and secondary HRR, and physician non-movers are defined as those with only a primary HRR. [43][44]

As a supplementary analysis, we investigate the determinants of healthcare spending variation in three specialty settings (primary care physicians, cardiologists, and dermatologists) that have been the focus of previous literature and where doctor practice style may be expected to play an important role. We form these subsamples by restricting the baseline sample to encounters with

[42] Following Finkelstein et al. (2016), we also exclude a small number of patient movers for whom the location of observed claims does not clearly shift from the origin location to the destination location. Specifically, we drop any patient mover for whom the share of claims in the destination HRR as a share of claims in either origin or destination HRRs does not increase by at least 0.75 in the years after the move.

[43] Analogously to the case of patients described in the previous footnote, some physicians in our sample are observed treating patients from multiple HRRs without satisfying the definition of a move. Unlike the case of patients, however, there is a relatively high number of such physicians, making us wary of excluding all of them. Since the drivers of such cross-HRR practice are poorly understood, we prefer to rely on the more transparent quasi-exogenous patient moves for our identification. We therefore code physicians separately for each HRR in which they practice within a given year.

[44] Some observations have a missing physician ID. We include a separate fixed effect for each HRR that is assigned to the observations in that HRR with missing physician IDs.

physicians from the respective specialty. In all cases, we define a doctor's specialty as the specialty listed on the plurality of her claims submitted in her capacity as a performing physician. For cardiologists and dermatologists, we use "cardiology" and "dermatology" as the relevant specialty. For primary care physicians (PCPs), we use "internal medicine", "family medicine", and "general practice". Finally, for the PCP subsample, following previous literature (e.g., Fadlon and Parys, 2019), we assign each patient in each year a single "primary" PCP based on the plurality of PCP-related claims submitted on account of that patient and assign all PCP spending for that patient and year to the primary PCP.

3.2.2 Sample Restrictions and Summary Statistics

We impose several restrictions to arrive at our main analysis sample. From the 16.7 million original patients, we drop a 75% random sample of non-movers in order to simplify computation. We further exclude all patient-years where the patients are younger than 65 or older than 99, where the patients are enrolled in Medicare Advantage, and where patients are not subscribed to Medicare Part A and B for all months in a year. Our baseline analysis sample contains 144.1 million encounters between 3.1 million patients and 2.1 million physicians, corresponding to 23.7 million patient-years and 11.5 million physician-years. The estimation sample is restricted to the largest connected set of physician, places, and patients, which includes the vast majority (99%) of encounters in the full analysis sample.[45]

Panel A of Table 3.1 summarizes our sample of patient movers and non-movers. The two groups are broadly similar, with movers being slightly older and more likely to be female, white, and living initially in the South and West. Both non-movers and movers have on average around $7,800 in utilization a year and see between five and six physicians annually.

Panel B of Table 3.1 likewise summarizes our sample of physician movers and non-movers. These two groups have a similar geographic distribution, though non-movers have a lower annual average utilization for Medicare patients of around $48,000, compared with $130,000 for movers. Physicians in the former group see around 41 Medicare patients annually on average, while those in the latter group see 106 Medicare patients annually.

[45] Specifically, the largest connected set has 142.7 million out of 144.1 million encounters, 2.9 out of 3.1 million patients, and virtually all physicians (excluding 300 of 2.1 million).

Figure 3.1 summarizes the geographic distribution of patient utilization across HRRs. Average annual HRR-level utilization is $6,975 with a standard deviation of $776. The map of this figure illustrates the high degree of geographic variation in utilization, with the South and Midwest outpacing lower-spending areas in the West and Northeast.[46]

As detailed in the next section, our empirical framework leverages the quasi-exogenous migration of both patient and physician movers in order to characterize the drivers of geographic variation in Figure 3.1. There is significant variation in the types of moves made by both physicians and patients. To illustrate this, Figure 3.2 shows the gap in average log utilization between the origin and destination HRRs of patient movers (Panel A) and physician movers (Panel B). Both distributions appear symmetric, indicating no systematic imbalance in moves from high- to low-utilization areas. The standard deviations of these distributions are also substantial, at 0.25 for patients and 0.39 for physicians, with a substantial share of moves in both cases involving changes in region-average log spending of more than 0.5.

A natural question is of course *why* patients in this sample are choosing to move. Data from the Health and Retirement Survey (Finkelstein et al., 2016) and the Longitudinal Survey of Aging (Choi, 1996), which ask individuals for their reason to move, lead to similar conclusions. The most frequently reported reason for moves in this age group is to be near/with children or other kin, followed by health reasons, financial reasons, or other amenities..

3.3 Model and Empirical Framework

3.3.1 Model

Our primary analysis is based on two models of healthcare utilization: one characterizing the average annual utilization per physician-patient pair, and one characterizing the number of physicians seen by each patient in each year. We show below how these models can be combined to decompose the observed geographic variation in utilization across regions into components attributable to patient demand, physician practice style, and other regional supply-side factors. We consider a set of physicians d, patients i, years t, and HRRs j.

[46]The geographic distribution of spending remains fairly stable across years of our sample. For example, the rank correlation between a HRR's utilization in the first and second half is 0.9.

Encounter Utilization

We refer to each physician-patient pair observed in a given year (i,d,t) as an *encounter*. Both patients and physicians may move across HRRs. We let $j(i,t)$ denote the HRR of patient i in year t, and we let $j(d,t)$ denote the HRR of physician d in year t. We assume that patients and physicians both differ in their average taste for more intensive treatment, as captured by respective parameters α_i and δ_d, and that other area supply factors shift utilization via a parameter γ_j. Differences in patient tastes for intensity (α_i) may reflect preferences or unobserved health; differences in physicians' tastes for intensity (δ_d) may reflect differences in their beliefs and in their training. Finally, differences in other area supply factors (γ_j) might include other aspects of the practice environment such as hospital capacity, physical capital, and organizational, financial, and legal incentives.

Patient i's utility from a utilization level of y from physician d in year t is assumed to take the form $u_{idt}(y) = \alpha_i y - \frac{1}{2}(y - h_{idt})^2$, where higher values of h_{idt} represent worse latent patient health. Physicians choose utilization to maximize $u_{idt}(y) + \delta_d y$ net of costs $c_{idt}(y) = (\gamma_{j(i,t)} + g_{idt})y$, where higher values of g_{idt} denote higher marginal costs. The resulting utilization in encounter (i,d,t) is given by:

$$\begin{aligned} y_{idt} &= \arg\max u_{idt}(y) + \delta_d y - c_{idt}(y) \\ &= \alpha_i + \delta_d + \gamma_{j(i,t)} + h_{idt} + g_{idt}. \end{aligned} \tag{21}$$

Equation (21) specifies a three-way fixed effects model for encounter utilization y_{idt} in terms of an individual effect α_i, a physician effect δ_d, and a residual place effect γ_j capturing non-physician area supply factors. Utilization y_{idt} is defined as the log spending of patient i with physician d in year t. This is only observed (or defined) for the subset of patient-physician matches that actually take place that year. Our specification assumes that α_i, δ_d, and γ_j are additively separable in the equation for log utilization; this has the intuitive implication that patient, physician, and other area supply factors affect the *level* of utilization multiplicatively. Thus, for example, the (level) utilization of patients who are sicker or prefer more intensive care (i.e., have higher α_i) will vary more across physicians than that of patients who are healthy or rarely seek care (i.e., have low α_i).

To bring this model to data, we assume the residual variation in patient health (h_{idt}) and in

cost of care (g_{idt}) can be forecasted by a time effect (τ_t) and sets of time-varying patient and physician observables x_{it} and w_{dt} given encounter location locations j: i.e. that $E[h_{idt} + g_{idt} \mid x, w, \text{j}] = \tau_t + x_{it}'\beta + w_{dt}'\theta$. This yields a linear regression model of

$$y_{idt} = \alpha_i + \delta_d + \tau_t + \gamma_{\text{j}(i,d,t)} + x_{it}'\beta + w_{dt}'\theta + \varepsilon_{idt}, \tag{22}$$

satisfying $E[\varepsilon_{idt} \mid x, w, \text{j}] = 0$. The residual ε_{idt} captures the unforecastable component of utilization, which is idiosyncratic conditional on the identity of the physician and patient, the year and location of encounters j, and the observables (x, w).

Encounter Quantity

We model the number of physicians seen by each patient in each year, a quantity we denote by N_{it}, as a Poisson random variable with mean λ_{it}. Thus, for each $k = 0, 1, 2, \ldots$

$$Pr(N_{it} = k) = \frac{\exp(-\lambda_{it})\lambda_{it}^k}{k!}. \tag{23}$$

A Poisson approximation for N_{it} is appropriate for counting matches between patient i and a large number of potential physicians d, an event indicated by $\mathbf{1}[\pi_{idt} > \nu_{idt}]$ for uniform ν_{idt}, given a small average match probability π_{idt} and sufficiently uncorrelated ν_{idt} (Walsh, 1955). We model the expected match rates λ_{it} by

$$\ln \lambda_{it} = \alpha_i^n + \tau_t^n + \gamma_{\text{j}(i,t)}^n + x_{it}'\beta^n, \tag{24}$$

allowing for persistent unobserved heterogeneity across both individuals and regions with the fixed effects α_i^n and γ_j^n. This model would follow if, for example, the ν_{idt} were independent across physicians d and the predicted match probabilities were multiplicative in patient and place effects and the controls.

HRR Utilization

The encounter log spending model (21) implies a specification for log patient-year spending y_{it} and, in turn, the average log spending $\bar{y}_j$ in each region j. We next show how estimates of this

specification can be used, along with estimates of the encounter quantity model (23)-(24), to decompose regional utilization differences into factors attributable to patients, physicians, and places.

Let D_{it} denote the set of physicians d which patient i sees in year t, with $N_{it} = |D_{it}|$. Aggregating over this set, we obtain log patient-year spending as

$$\begin{aligned} y_{it} &= \ln\left(\frac{1}{N_{it}}\sum_{d\in D_{it}} \exp y_{idt}\right) + \ln N_{it} \\ &= \alpha_i + \tau_t + \gamma_{\mathrm{j}(i,t)} + x_{it}'\beta + \varepsilon_{it}, \end{aligned} \tag{25}$$

where we normalize y_{it} to zero when $N_{it} = 0$. Here $\varepsilon_{it} = \ln N_{it} + \ln\left(\frac{1}{N_{it}}\sum_{d\in D_{it}} \exp\left(\delta_d + w_{dt}'\theta + \varepsilon_{idt}\right)\right)$ captures the potential contribution of physicians to patient i's log spending in year t, either through the number of encounters N_{it} or through per-encounter utilization.

To better characterize the role of physicians in per-encounter utilization, we further decompose ε_{it}. Let $D^*(n,j,t)$ denote a random set of physicians of size n practicing in HRR j in year t and define

$$\overline{\delta}_{it} = E\left[\ln\left(\frac{1}{N_{it}}\sum_{d\in D^*(N_{it},\mathrm{j}(i,t),t)} \exp\left(\delta_d + w_{dt}'\theta + \varepsilon_{idt}\right)\right) \mid N_{it}\right] \tag{26}$$

as the typical contribution of physicians to patient i's utilization in time t if she were to select N_{it} physicians at random from her HRR $\mathrm{j}(i,t)$. The expectation in $\overline{\delta}_{it}$ is taken both with respect to the random sets of physicians $D^*(N_{it},\mathrm{j}(i,t),t)$ and the unforecastable contribution of utilization ε_{idt}. Thus $\overline{\delta}_{it}$ captures the typical utilization due to the regional availability of physicians with different practice styles, removing differences in how patients select different physicians from an HRR. To capture the importance of such physician selection, we further define

$$\sigma_{it} = E\left[\ln\left(\frac{1}{N_{it}}\sum_{d\in D_{it}} \exp\left(\delta_d + w_{dt}'\theta + \varepsilon_{idt}\right)\right) \mid N_{it}\right] - \overline{\delta}_{it} \tag{27}$$

as the patient's expected difference in physician-driven utilization given her actual chosen set of physicians D_{it} and a random set of the same size. We then have $\varepsilon_{it} = \overline{\delta}_{it} + \sigma_{it} + \nu_{it}$ where $\nu_{it} = \varepsilon_{it} - E\left[\ln\left(\frac{1}{N_{it}}\sum_{d\in D_{it}} \exp\left(\delta_d + w_{dt}'\theta + \varepsilon_{idt}\right)\right) \mid N_{it}\right]$ gives an idiosyncratic component of pa-

tient utilization. Substituting this into Equation (25) yields

$$y_{it} = \alpha_i + \tau_t + \gamma_{\mathrm{j}(i,t)} + x'_{it}\beta + \ln N_{it} + \overline{\delta}_{it} + \sigma_{it} + \nu_{it}. \tag{28}$$

Aggregating Equation (28) across patients and years, we obtain a model of the average utilization $\bar{y}_j \equiv E[y_{it} \mid \mathrm{j}(i,t) = j]$ in HRR j:

$$\bar{y}_j = p_j(\overline{\alpha}_j + \gamma_j + \overline{N}_j + \overline{\delta}_j + \overline{\sigma}_j) \tag{29}$$

where $p_j = Pr(N_{it} > 0 \mid \mathrm{j}(i,t) = j)$ denotes the probability of positive utilization among patient-years in HRR j, $\overline{\alpha}_j$ is the average patient-year component $\alpha_i + \tau_t + x'_{it}\beta$ among those with positive utilization in HRR j, $\overline{N}_j$ is the average number of log physician encounters among those with positive utilization in HRR j, $\overline{\delta}_j$ is the average physician component $\overline{\delta}_{it}$ for patients in HRR j, and $\overline{\sigma}_j$ is the average selection component for patients in HRR j.

We use estimates of Equations (29) and the Poisson model (23)-(24) to compute counterfactual HRR utilization in six incremental steps. First, we set $\overline{\sigma}_j = 0$ to simulate a counterfactual in which there is no systematic matching of higher-utilization patients to higher or lower-utilization physicians across HRRs. Second, we equalize $\overline{\delta}_j$ across regions to simulate a counterfactual in which there is additionally no systematic sorting of physicians with different practice styles across HRRs. Third, we equalize $\overline{\alpha}_j$ to simulate a counterfactual in which there is additionally no systematic sorting of patients with different per-encounter utilization demand across HRRs. Fourth, we use the Poisson model to equalize patient effects on the average number of physicians seen, affecting both p_j and $\overline{N}_j$. This step simulates a counterfactual in which there is additionally no systematic sorting of patients with different demand for physician quantity. The fifth and sixth steps similarly equalize place effects on per-encounter utilization and the number of physicians seen, respectively. These remaining steps quantify the remaining variation in non-physician supply-side factors across regions. Each counterfactual step is defined formally in Appendix C.2.[47]

It is worth emphasizing that each step in our counterfactual analysis, while quantifying the relative importance of physicians, patients, and places in geographic utilization variation, represents a partial-equilibrium analysis. In practice, policies which affect patient or physician mobility, or

[47] Although our decompositions are sequential, we obtain consistent results with alternative sequencing.

which equalize residual supply-side factors across regions, are likely to have a variety of general equilibrium effects that are outside of our model and scope of this analysis.

3.3.2 Estimation and Identification

In our baseline estimation of the utilization and encounter models, Equations (22) and (24), the patient observables x_{it} consist of year effects, dummies for five-year age bins, and relative-year fixed effects $\rho_{r(i,t)}$ for patients who move between HRRs, where for a mover who moves during year t_i^*, the relative year is $r(i,t) = t - t_i^*$ (we normalize $\rho_{r(i,t)}$ to zero for non-movers). The physician observables w_{it} include similar relative-year fixed effects for physician movers, again normalized to zero for non-movers.

We estimate the parameters of the utilization model (22) by a three-way fixed effects linear regression on the full set of physician-patient matches that occur each year. We use these estimates to form simulation-based estimates of average physician utilization and selection terms $\overline{\delta}_{it}$ and σ_{it}.[48] In calculating all HRR- and physician-year-averages, we upweight patient non-mover encounters by a factor of four to take into account our patient sampling procedure. Finally, we estimate the parameters of the encounter quantity model (24) by a two-way fixed effect Poisson regression in the full sample of patient-years, both those with and without physician encounters, using the same vector of patient observables x_{it} as in the utilization model.[49]

Identification of the utilization model from equation (22) leverages the variation from patient and physician moves across regions, as well as the within-region variation in patient-physician matches. Absent migration, relative patient and physician utilization effects within each HRR may be identified by conditionally idiosyncratic matching between physicians and patients (Abowd et al., 1999). Quasi-experimental movement across HRRs identifies place effects, and thus the average patient and physician effect in each HRR.

To build intuition for the roles that movers play in identification, consider a special case with no time effects or time-varying controls and where a group of patients in each region see a representative non-moving physician $d(j)$. Identification of the variation in combined physician-place

[48] Specifically, for each patient and year, we take a random draw of physicians from her HRR with the number of physicians equalling her actual number of encounters for this patient and year. We use these simulated encounters averaged over 100 random draws to form estimates of $\overline{\delta}_{it}$ and σ_{it}.

[49] Hausman et al. (1984) establish the consistency of conditional maximum likelihood estimation of such models, which we implement with two high-dimensions using the algorithm of Guimaraes (2014).

effects $\overline{\gamma}_j \equiv \delta_{d(j)} + \gamma_j$ is then given by a "common trends" assumption on the utilization of patient movers: that the unobserved trends in patient health and cost of care ε_{idt} for movers between different HRRs are typical in the population. Formally, for the individuals who move from HRR k to HRR j between time $t-1$ and t the observed utilization trend is

$$\begin{aligned} &E[y_{idt} \mid \mathrm{j}(i,d,t) = j, \mathrm{j}(i,d,t-1) = k] - E[y_{id,t-1} \mid \mathrm{j}(i,d,t) = j, \mathrm{j}(i,d,t-1) = k] \\ &= \overline{\gamma}_j - \overline{\gamma}_k + E[\varepsilon_{idt} - \varepsilon_{id,t-1} \mid \mathrm{j}(i,d,t) = j, \mathrm{j}(i,d,t-1) = k], \end{aligned} \tag{30}$$

identifying relative place effects $\overline{\gamma}_j - \overline{\gamma}_k$ when $E[\varepsilon_{idt} - \varepsilon_{id,t-1} \mid \mathrm{j}(i,d,t) = j, \mathrm{j}(i,d,t-1) = k] = 0$. This common trends assumption would be violated if (for example) patients move in response to an expected change in their healthcare needs.

Just as patient migration can separate the contribution of patient utilization effects from other factors, physician migration can separate the contribution of physicians. To again see this simply, consider a group of physicians who move from HRR k to HRR j between time $t-1$ and t and treat a representative group of non-moving patients in each period. By the same logic as before, when these movers are representative in terms of their unobserved trends in ε_{idt} a comparison of average physician utilization before and after the move identifies the difference in $\overline{\alpha}_j + \gamma_j$, where $\overline{\alpha}_j$ denotes the average α_i of non-moving patients in HRR j. Combining these differences with the identified difference in α_i and $\overline{\gamma}_j$ from the patient mover quasi-experiment allows for a full separation of the variation in α_i, δ_d, and γ_j.

In practice, identification of Equation (22) is also assisted by the inclusion of time-varying patient- and physician-level controls x_{it} and w_{dt}. Including time and patient age effects weakens the key common trends assumptions to allow movers and matching to vary across these dimensions. Similarly, including relative year effects for movers allow for arbitrary differences in utilization before and after a move, while imposing the restriction that these changes are the same regardless of the origin and destination of the move. While allowing for persistent unobserved patient, physician, and region heterogeneity, the assumption of common trends in per-encounter utilization imposes several important restrictions on the data. Most notably, it requires there to be no sudden changes in utilization demand which coincide with the timing of moves or the systematic matching to particular physicians. This assumption would be violated if, for example, patients systematically

respond to a negative shock to their health by moving to high-utilization areas or seeking out more intensive physicians. We show some supportive evidence for common trends in the next section.

Identification of the nonlinear model of encounter quantity (equation (24)) similarly follows from exogenous patient migration. Specifically, as we show in Appendix C.2, conditional maximum likelihood estimates of the place effects γ_j can be obtained from contrasts in log growth rates of encounter quantity across movers with different origin and destination pairs. In this case the nonlinear model imposes a common trends assumption on potential log encounter growth rates (instead of potential level changes, as in the linear utilization model). We again build support for this assumption in the next section.

3.4 Results

3.4.1 Preliminary descriptive results

Before estimating our main models, we present some descriptive evidence of the average changes in utilization when patients and physicians move across different regions. This analysis illustrates a key source of the variation we use to estimate the models, and provides some support for the key identifying assumptions. Our approach builds on the earlier analysis of Finkelstein et al. (2016) and Molitor (2018) who study utilization changes for patient and cardiologist movers, respectively.

To motivate this event study analysis, we consider a restricted version of our patient-year utilization model (25) in which the contribution of physicians is additively separable in a patient's location and year: $\varepsilon_{it} = \bar{\varepsilon}_{\mathrm{j}(i,t)} + \tilde{\varepsilon}_i + \check{\varepsilon}_t + \eta_{it}$. Here $\bar{\varepsilon}_j$ captures the overall regional contribution to utilization via physician availability and matching, while $\tilde{\varepsilon}_i + \check{\varepsilon}_t$ captures residual physician-driven utilization specific to the patient and year and η_{it} is an idiosyncratic error. For each patient mover i, write $d(i)$ and $o(i)$ as the indices of her destination and origin HRRs, respectively, and write $\Delta_i = \bar{y}_{d(i)} - \bar{y}_{o(i)}$ as the average difference in observed utilization between her destination and origin HRR. Further write $S_i = \left(\gamma_{d(i)} + \bar{\varepsilon}_{d(i)} - (\gamma_{o(i)} - \bar{\varepsilon}_{o(i)})\right)/\Delta_i$ as the share of this observed utilization difference that is due to the causal effect of places, either through physicians ($\bar{\varepsilon}_{d(i)} - \bar{\varepsilon}_{o(i)}$) or other regional supply-side factors ($\gamma_{d(i)} - \gamma_{o(i)}$). We can then rewrite the utilization (25) for

movers as

$$y_{it} = \tilde{\alpha}_i + \tilde{\tau}_t + (\mathbf{1}[r(i,t) > 0] S_i)\,\Delta_i + x_{it}'\beta + \eta_{it}, \tag{31}$$

where $\tilde{\alpha}_i = \alpha_i + \tilde{\varepsilon}_i + \gamma_{o(i)} - \bar{\varepsilon}_{o(i)}$ and $\tilde{\tau}_t = \tau_t + \check{\varepsilon}_t$.

Equation (31) motivates a patient event study regression of the form

$$y_{it} = \tilde{\alpha}_i + \tilde{\tau}_t + \theta_{r(i,t)}\Delta_i + x_{it}'\tilde{\beta} + \eta_{it}, \tag{32}$$

where the relative-year coefficients θ_r capture how average annual utilization changes in the years preceding and following a move across HRRs, as a share of the average observed difference in observed between the destination and origin HRRs Δ_i. Given the model, we expect these coefficients to be near zero for the years preceding a move ($r < 0$) and to reflect the average share S_i of utilization differences which is due to place-based factors (both due to physicians or other area supply-side conditions) for the years following a move ($r > 0$).

While we derived Equation (32) for patient movers, an analogous derivation follows for average physician utilization and physician movers. In this case the change in relative-year coefficients θ_r following a move captures the average share of observed physician utilization differences across regions which is due to place-based factors, either due to patients or other area supply-side factors. We describe this derivation in more detail in Appendix C.2, where we also show how both derivations can be extended to relax the assumption of additively-separable ε_{it}, allowing for time-varying differences in the contribution of physicians via the available physician stock $\overline{\delta}_{it}$, patient-physician matching σ_{it}, and the number of physicians seen N_{it}.

Panel A of Figure 3.3 plots estimates of the relative-year coefficients in Equation (32). We estimate this regression on the full sample of patient movers and non-movers, normalizing Δ_i to zero for the latter group. The outcome y_{it} is again the log annual spending of patient i in year t, normalized to zero when no spending takes place. The control vector x_{it} again contains indicators for five-year age bins and the relative-year main effects $\rho_{r(i,t)}$.

The patient event study shows that the average utilization of movers is stable in the years preceding a move, conditional on the controls, while in the years following a move average patient utilization changes sharply in the direction of the observed difference in average HRR utiliza-

tion. The lack of pronounced pre-trends is consistent with the identifying assumption of quasi-experimental patient migration, while the sustained post-move jump indicates an important role of regional supply-side factors in annual patient utilization.[50] Quantitatively, we estimate an event study jump of around 0.5. This is consistent with the earlier finding of Finkelstein et al. (2016) and suggestive that supply-side factors (both due to physicians and other regional drivers) account for around half of the observed geographic variation in annual patient utilization with the remaining 50% due to differences in patient demand.

Panel B of Figure 3.3 presents an analogous event study analysis of utilization changes for physician movers. We now plot estimates of relative-year coefficients from a regression of

$$y_{dt} = \tilde{\alpha}_d + \tilde{\tau}_t + \theta_{r(d,t)}\Delta_d + w'_{dt}\tilde{\beta} + \eta_{dt}, \tag{33}$$

where now y_{dt} denotes the log spending of physician d in year t, the control vector w_{dt} includes relative-year fixed effects, and Δ_d gives the difference in average log utilization between a moving physician's origin and destination HRR (again normalized to zero for non-movers). The relative-year coefficients θ_r thus now capture the change in log utilization for physician movers in each relative year r as a share of Δ_d.

As with the patient event study, the physician event study shows that the average utilization of movers is stable prior to a move (consistent with our identifying assumption of quasi-experimental physician migration). Following a move, physician utilization changes sharply by around 70-80% of the observed difference in average utilization between her destination and origin HRR. This is broadly consistent with the earlier finding of Molitor (2018) in showing that regional factors play an important role in physician behavior, while also suggesting an important role of persistent physician practice styles. Here an event study jump of 0.7 suggests that regional factors (both due the local demand of patients and other factors) account for around 70% of the observed geographic variation in annual physician utilization, with the remaining 30% due to differences in physician

[50] One possible source of bias in this exercise may come from endogenous moves caused by patients seeking better care due to worsening health status. While we cannot fully rule out such sources of bias, the patterns we observe in the data suggest that they are likely to be small. Gradual worsening of health status that would lead to eventual relocation would tend to show up as pre-trends in our motivating event studies. Although sudden negative health shocks that cause immediate movement to a more intensive area might occur without causing pre-trends, such changes might lead to a spike in utilization immediately following a move. However we also find relatively flat post-trends in our event study analysis.

practice intensity.

Lastly, to illustrate the source of identifying variation underlying our Poisson model of number of encounters, we present an event study analysis from a regression of the form

$$\ln N_{it} = \tilde{\alpha}_i^n + \tilde{\tau}_t^n + \theta_{r(i,t)}^n \Delta_i^n + x_{it}' \tilde{\beta}^n + \eta_{it}^n, \tag{34}$$

where now the outcome is the log number of encounters of patient i in year t (normalized to zero for patient-years with no encounters), the control vector x_{it} and relative-year coefficients $\theta_{r(i,t)}$ are defined as in Equation (32) above, and Δ_i^n is the average observed difference in number of log encounters between destination and origin HRRs for patient movers (normalized to zero for non-movers). Equation (34) can be derived from a linear approximation to the Poisson model (23)-(24).[51] Just as with the patient- and physician-utilization event studies presented in Figure 3.3, the shape of the pre- and post-trends and the size of the event study jump illustrate the patient mover variation that underly our encounter model estimates.

The results from this analysis are presented in Figure 3.4. As with the utilization event studies, both the pre- and post-trends are stable, which is consistent with the assumption of quasi-exogenous patient migration. The event study jump of around 0.6 indicates that place-based factors account for 60% of the observed geographic variation in encounter quantity, with differences in patient demand accounting for the remaining 40%.

Taken together, the event studies in Figures 3.3 and 3.4 suggest a non-trivial role for patient, physician, and place heterogeneity in affecting geographic distribution of healthcare utilization. At the same time, this descriptive analysis cannot by itself yield a complete accounting of the importance of patients and physicians in driving spending differences. To do so, we estimate the two models described in Section 3.3.

3.4.2 Model Estimates

We summarize our estimates of the encounter quantity model (23)-(24) and per-encounter utilization model (22) graphically in Figures 3.5 and 3.6. Given our interest in geographic variation we

[51] Formally, using Equation (24), $E[\ln N_{it} \mid x, \mathrm{j}] \approx \ln E[N_{it} \mid x, \mathrm{j}] = \tilde{\alpha}_i^n + \tau_t^n + (\mathbf{1}[r(i,t) > 0] S_i^n) \Delta_i^n + x_{it}' \beta^n$ for movers, for $\tilde{\alpha}_i^n = \alpha_i^n + \gamma_{o(i)}^n$ and where $S_i^n = (\gamma_{d(i)} - \gamma_{o(i)})/\Delta_i^n$ denotes the share of the observed difference in average log encounters attributable to the place effects γ_j^n.

plot histograms of the average estimated patient and physician effects across HRRs along with the estimated region effects for each model. Broadly, these estimates confirm the findings of the motivating event studies that each of these factors plays an important role in driving geographic variation in healthcare utilization.

Figure 3.5 shows results for the number of encounters. The two panels indicate that patients and places exhibit a high dispersion, but contribute equally to overall differences in the HRR-average number of encounters. Panel A is a histogram of the average expected number of encounters in each HRR if each patient's effect was replaced with the sample's average patient effect, The figure thus displays the geographic variation induced by places alone if differences in patient population across HRRs were eliminated. The results show that moving the average patient one standard deviation up in the distribution of HRR effects increases the expected number of encounters by about 0.26. Panel B repeats this exercise for patients by plotting a histogram of the average expected number of encounters in each HRR if each patient's place effect was replaced with the sample's average place effect. Again, the figure shows geographic variation induced by differences in patient population alone if differences in place effects were eliminated. The standard deviation of these averages is almost exactly equal to that in Panel A at 0.25. It indicates that a one standard deviation increase in average patient population intensity is associated with an expected increase of 0.25 in demanded encounters.

Panel A of Table 3.3, which reports the variance-covariance matrix for the HRR-averages of the patient and place components of the encounter quantity model, shows that they are positively correlated. The variance of both components is around 0.06-0.07 consistent with the results shown in Figure 3.5. The covariance is 0.02, or about a third of the individual variance of each of these components. This shows that patients that demand more encounters also sort to places which induce higher encounter numbers, thereby contributing to overall geographic dispersion in spending. As we will show in our formal counterfactual analysis, the factors inducing higher encounter numbers play a crucial role in driving cross-HRR variation in utilization.

Figure 3.6 shows results for per-encounter utilization. The three panels plot histograms of the HRR-average patient, physician, and residual area supply side effects.[52] The figure shows that

[52]The non-linearity of the Poisson model requires us to transform the estimated effects into predicted values in order to visualize their variation on a meanintful scale. However, the linear per-encounter utilization model allows for the effects to be plotted directly, with the variation displayed on the histograms providing an accurate sense of the

while all three factors exhibit a high degree of geographic variation, average place and physician dispersion (with a standard deviation of 0.13) surpasses average patient dispersion (with a standard deviation of 0.02).[53] 3.2 shows furthermore that the 0.13 standard deviation in HRR-average physician effects is a small share of the overall variation in practice style differentiation: the overall standard deviation of physician fixed effects is 0.88, with an average standard deviation of 0.87 within HRRs. In other words, a physician with one standard deviation more intensive practice style is estimated to have an 88% higher level of spending per patient-year encounter. Given this estimate, we find that moving a physician's treatment intensity from the bottom decile to the top decile would almost triple the amount of healthcare spending among Medicare beneficiaries.[54]

Table 3.3, Panel B, reports the variance-covariance matrix for the individual components of the utilization model: patient effects α_i, effects δ_d, and residual are supply-side effects γ_j, averaged to the HRR level. As shown in Figure 3.6 the HRR-average physician effects have a standard deviation of around 0.13, similar to the standard deviation in residual-supply-side effects. Average patient effects have a smaller standard deviation of 0.03, and are positively correlated with average physician practice intensity and with residual supply-side factors. There is a strong negative correlation between HRR-average physician practice intensity and residual area supply side factors. Using the estimated covariance and variance in Table 3.3 we estimate this correlation as -0.85. This suggests that while physician practice intensity varies widely across regions, a patient who moves from a low-intensity to a high-intensity HRR need not see a large increase in her utilization, as the change in available physicians may be offset by other supply-side factors.[55] We now investigate this possibility directly.

3.4.3 Geographic Variation Counterfactuals

To decompose the geographic variation in utilization into its constituent components, we conduct the counterfactual analysis described in Section 3.3.1. These decompositions augment the analysis

magnitude of the estimated variation.

[53]The relatively low variation of the HRR-average patient effects may be surprising given the fact that patients explain 50% of overall geographic variaiton in spending. As we show in our counterfactual results below, much of this contribution comes through patient demand for physician encounters.

[54]Appendix Figure C3 visualizes the underlying geographic variation in physician effects. Areas with most-intensive physicians tend to be in the South, Mid-West and South-West of the country, while low-intensity areas tend to be in the West and North-East.

[55]One caveat is that these tables use the raw estimates and do not correct for sampling variation.

in Section 3.4.2 by incorporating both the variance and covariance terms for the individual effects in the utilization model, and by incorporating estimates from the encounter quantity model.

Table 3.4 presents the decomposition results between specific groups of high- and low-utilization HRRs. Specifically, we show the decomposition for HRRs above and below the median level of average patient spending (columns 1-3), the top and bottom quartile (columns 4-6), and the top and bottom decile (columns 7-9). In each case, the first column shows how the the absolute difference in average utilization changes as we restrict different parts of the model, the second column shows the percentage change, and the third column shows the cumulative percentage change in utilization. Appendix C.2 provides more details on these counterfactual calculations.

Since the results are qualitatively similar across the comparisons, we focus our attention on the above / below median HRR decomposition. The difference in average log patient utilization between HRRs above/below the median is 0.272. This difference drops by 5%, to 0.257, once we eliminate patient-physician selection. This small effect is consistent with the low variance in average selection shown in the last row of Table 3.3, Panel B. The difference decreases by a further 31% of the original difference, to 0.173, once average differences in physician practice styles are removed. In other words, differences in physicians practice style explain about 30 percent of the difference in utilization between high and low utilization HRRs. This substantial but not overwhelming drop is consistent with the finding in Table 3.2, that while there is a high level of variation in physician effects, much of this variation occurs within healthcare markets.

Eliminating patient differences in utilization per encounter and number of encounters further reduces the above/below median average LnRBU difference by 3% and 50% of the original difference. The small magnitude of the impact of patient-driven differences in utilization per encounter is consistent with the small variance of HRR-average patient effects from the utilization model, documented in Table 3.3, Panel B. The quite large role for patient-driven differences in the number of encounters is consistent with the large overall patient demand share in Figure 3.3. Variation in residual place-based factors account for the remaining 11% of cross-geographic variation in spending, consistent with Panel A of Figure C2. Interestingly, eliminating residual place effects on utilization per encounter actually increases spending differences by 16%. This stems from the strong negative correlation of residual place effects with average physician practice styles and selection shown in Table 3.3, Panel B.

Taken together, this decomposition of regional variation into patient, physician, and other supply side factors yields several striking findings. First, the single biggest driver of cross-HRR variation is patient demand. Consistent with Finkelstein et al. (2016), we find that around half of regional utilization variation arises from the non-random distribution of patients with different demand for care. Our analyses here further indicate that almost all (over 90 percent) of the patient component reflects differences in patient demand for the number of different physicians seen in a year, rather than differential patient demand for per-physician utilization, or sorting of patients with different demand-side preferences across physicians with different practice styles. These results are consistent with existing literature pointing to the importance of practice style variation in individual specialties and of fragmentation in the delivery of healthcare (Agha et al., 2019; Frandsen et al., 2015; Baicker and Chandra, 2004). Somewhat more surprising is the fact that patient-physician selection seems to play no role in these differences; pairing of more expensive patients with more expensive physicians does not seem to vary systematically across different HRRs.

Second, differences in physician practice styles across regions explain about 30 percent of geographic variation in utilization. Physicians display substantial differences in practice-styles: a 90th percentile physician will spend about 3 times more per patient encounter than a 10th percentile physician. This large variation in physician practice style is consistent with recent findings similarly documenting substantial variation in practice patterns across physicians in prescribing anti-depressants (Currie and MacLeod, 2020) and in interpreting chest x-rays (Chan et al., 2020). However, our estimates suggest that most of this variation is within rather than across regions, which mitigates the contribution of physician heterogeneity to regional variations.

Finally, we find a relatively small role for residual supply-side area factors in contributing to regional variation. These factors, which may include provider capacity and organizational forms as well as regional norms, are only about half as important as physician practice style in contributing to regional variations.

3.4.4 Specialty Case Studies

Taken together, our primary analysis finds that physicians do not play the overwhelming role in driving geographic variation. This is broadly consistent with the finding of Molitor (2018) for cardiologist treatment decisions, though it is at odds with some conventional wisdom and earlier

suggestive findings for other specialties (e.g. Cutler et al., 2019). Notably, our model estimates do not suggest that variation in physician practice style is limited but that this variation is mostly found within regions and is negatively correlated with other place-based drivers of utilization across regions.

To confirm the robustness of these findings, we conduct a series of subsample analyses that decompose the geographic utilization of three physician specialities: primary care physicians (PCPs), cardiologists, and dermatologists. These are specialties where one may expect a high role of individual practice styles, with PCPs and cardiologists being of particular focus in the earlier literature. We first identify physicians of each specialty and estimate versions of our two models, (22) and (23)-(24), for their utilization and encounter quantity. We then decompose differences in specialty-specific average log spending across HRRs, as in Table (3.4), using the sequence of counterfactuals derived in Section (3.3.1).

Table 3.5 shows that differences in physician practice style play even less of a role in the observed variation in specialty-specific utilization across regions than in our main sample analysis. The table displays our geographic counterfactual decomposition of the spending difference between above- and below-median HRRs for log annual spending for the three specialties.[56] Focusing on columns (1)-(3), we see that the overall difference between above and below median-HRRs for log annual patient PCP spending is 0.422. This difference drops down by 2% once the selection channel is eliminated and remains essentially constant (or even increases slightly) once physician practice style differences are eliminated. Patients account for a relatively smaller share of overall spending differences (7% total, of which 4% is attributed to patient demand for utilization per encounter and 3% to patient demand for number of encounters. The remaining 93% can be attributed to residual place-based supply-side components, of which 26% are due to effects on utilization per encounter and 67% are due to effects on number of encounters. Just as in our baseline results, physicians do not explain a majority of the geographic variation in patient spending. We observe a similar pattern in the cardiologist and dermatologist decompositions in columns (4)-(6) and (7)-(9). Physicians explain between 3% and 10% of overall variation. While the relative role of patients and places shifts across specialties, it does so in patterns that are consistent with previous studies

[56] As before, these counterfactual results are very stable across different ways of defining the high- and low-spending HRRs based on percentiles of the HRR spending distribution.

(e.g., Molitor (2018) on cardiologists).[57]

However, while physicians explain little of the overall geographic spending differences, variation in physician practice styles as a whole is quite significant and mostly found within regions. Table 3.6 illustrates this point by presenting the within- and between-HRR components of physician practice style variation for each of the specialty samples. In all three cases, the majority of variation in practice styles occurs within rather than across medical markets. This echoes the results from our baseline analysis and underlines once more the fact that our findings are consistent with an important role for physicians in determining spending levels overall.

Finally, we note that, just as in the baseline sample, part of the reason why physicians do not explain more of the geographic variation in spending is because average practice styles are negatively correlated with other place-based drivers of utilization. Table 3.7 reports the variance-covariance matrix for the individual components of the encounter quantity model (Panel A) and utilization model (Panel B) averaged at the HRR-level for each of the specialty samples. A strong negative correlation between average practice styles and other place-based factors is easily observable in all three cases. As in the baseline sample, a patient moving from a low-spending area to a high-spending area may not witness a large increase in utilization because average physician practice styles and residual place-based factors partially cancel each other's influence.[58]

3.5 Conclusion

While patient demand and regional supply-side factors have both been shown to play important roles in driving geographic variation in U.S. healthcare spending, the role of physicians is less well-understood. We fill this gap by leveraging the quasi-experimental migration and matching of Traditional Medicare patients and physicians, estimating a model of encounter quantity and per-encounter utilization that allows for variation in patient demand, physician practice intensity, and other regional supply side differences. Consistent with past work, we find that around half

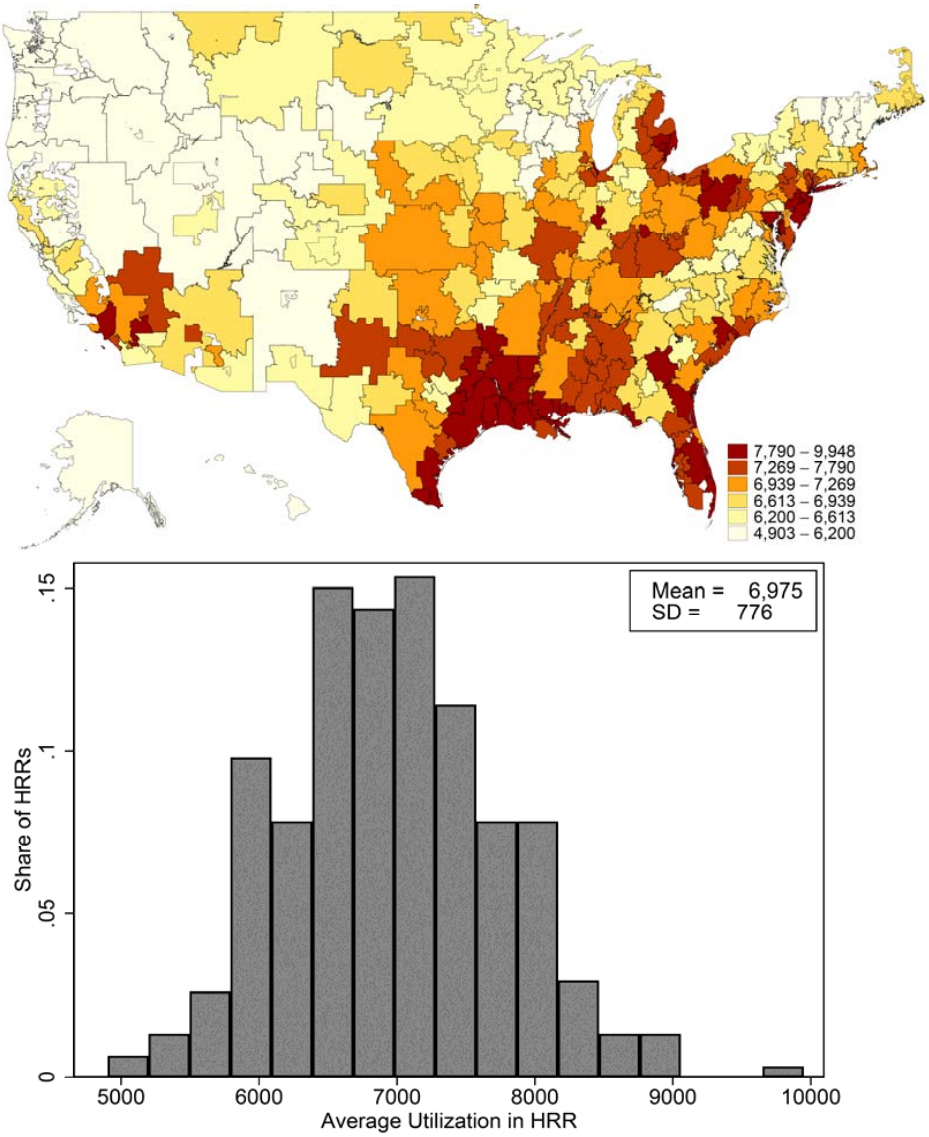
[57] Appendix Figures C4, C5, and C6 confirm the outsize role of non-physician place effects in driving geographic differences in physician utilization for PCPs, cardiologists, and dermatologists via an event study analysis paralleling that of Figure 3.3.

[58] Interestingly, a high negative correlation between patients and places is also visible in the case of PCPs in Panel A, which indicates that negative sorting of patients who demand more encounters to places that are less likely to supply them may also be decreasing the role of place-based factors in explaining geographic variation in spending. In the other two cases, the two components of encounter quantity are only weakly correlated with each other.

of regional utilization variation arises from the non-random distribution of patients with different demand for care. On the supply side, we find that three-fifths of place-based drivers of utilization are due to differences in physician practice style. The number of different physicians a patient sees each year, a measure which could be interpreted as health care fragmentation, plays an important role in the supply side and drives the majority of the demand side. Non-physician factors account for two-fifths of the supply-side drivers and one-tenth of the overall variation. We find that the modest role of physician practice styles in explaining geographic variation arises not because of limited variation in physician practice intensity, but because of a strong negative correlation between where high intensity physicians are located and average regional per capita spending; both within and between HRRs there is substantial variation.

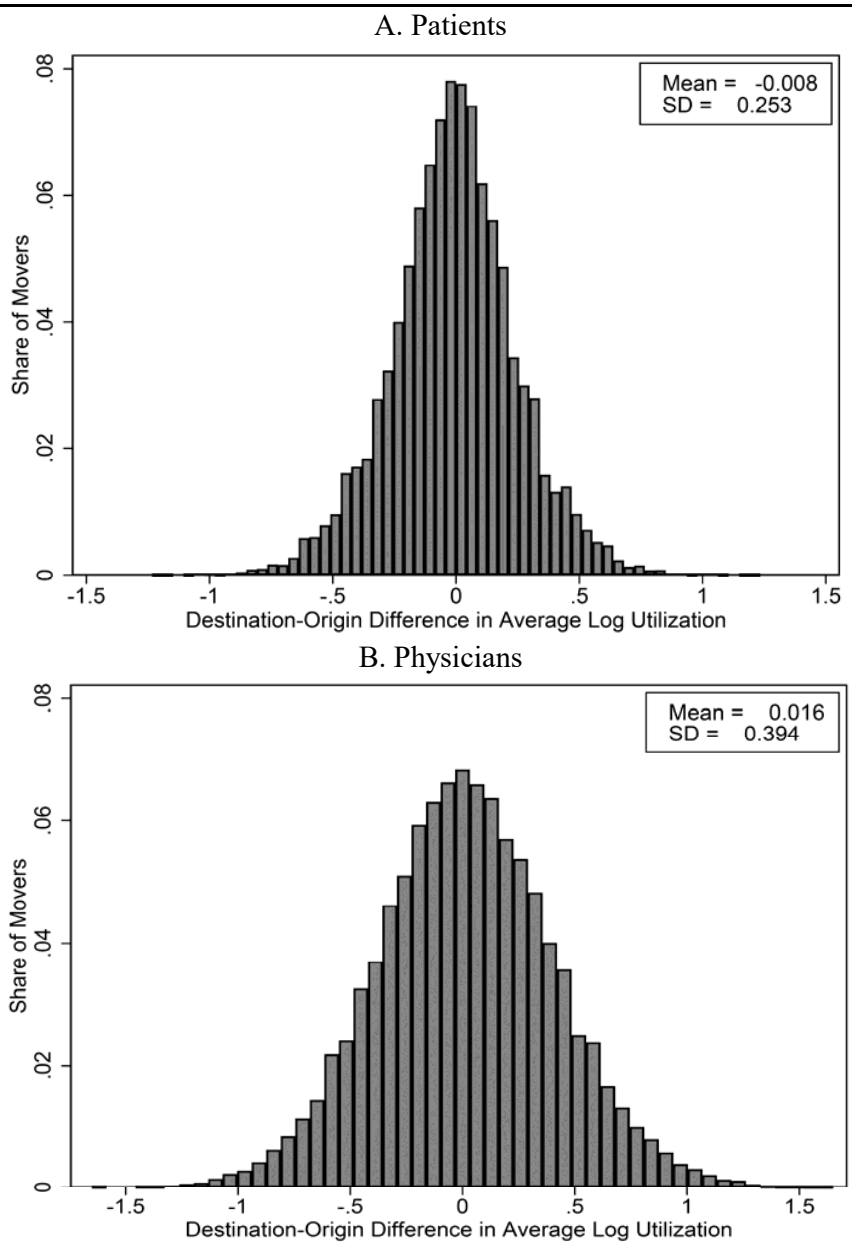
3.6 Figures and Tables

Figure 3.1: Distribution of Annual Patient Utilization Across HRRs



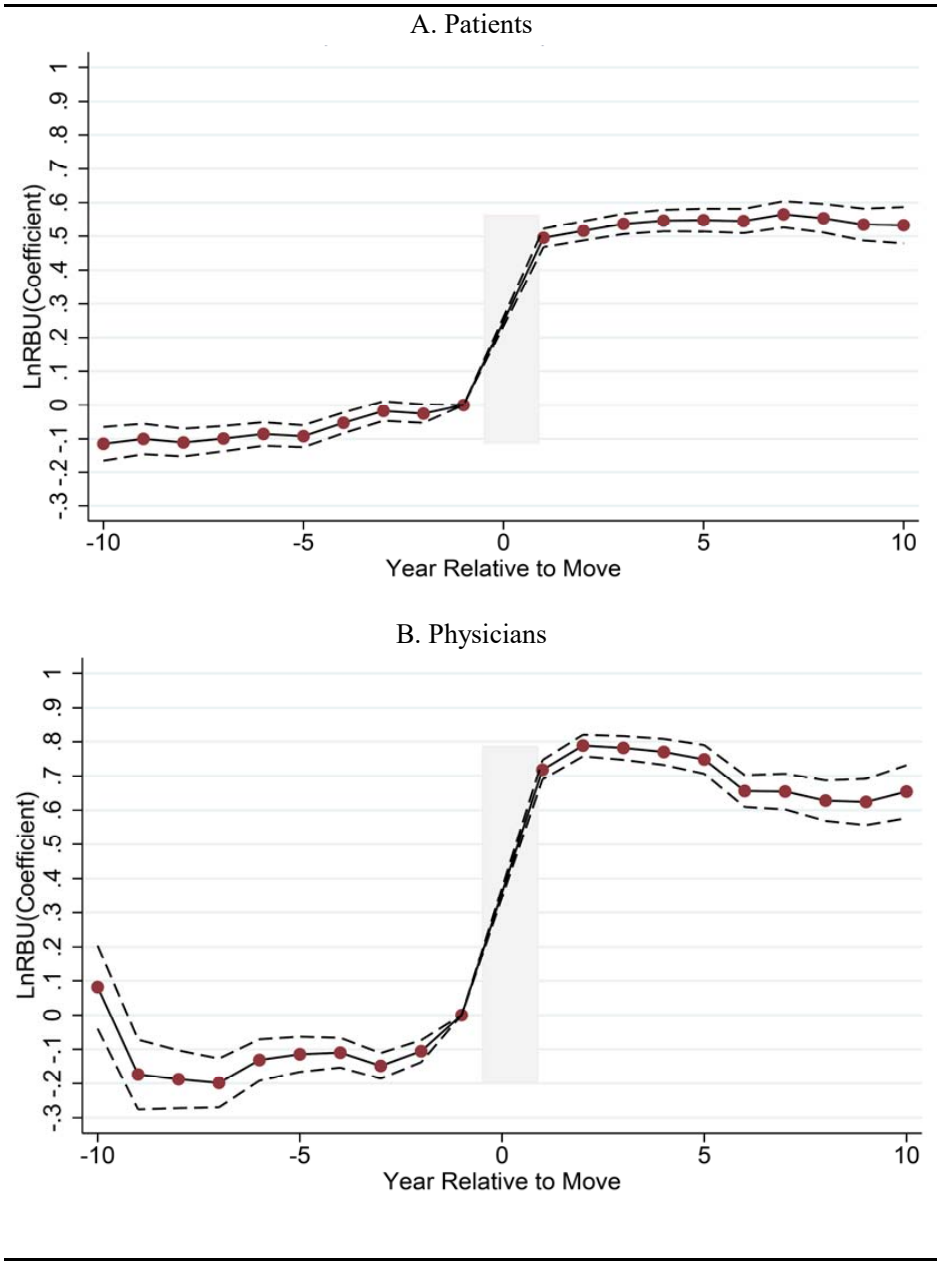
Notes: The map shows the distribution of average annual patient utilization by HRR, in quintiles defined in the legend. The histogram displays the marginal distribution of average annual patient utilization by HRR. The sample is the baseline sample of all patient-years for movers and non-movers (N = 23,678,685 patient-years).

Figure 3.2: Destination-Origin Gaps in Average Patient Utilization



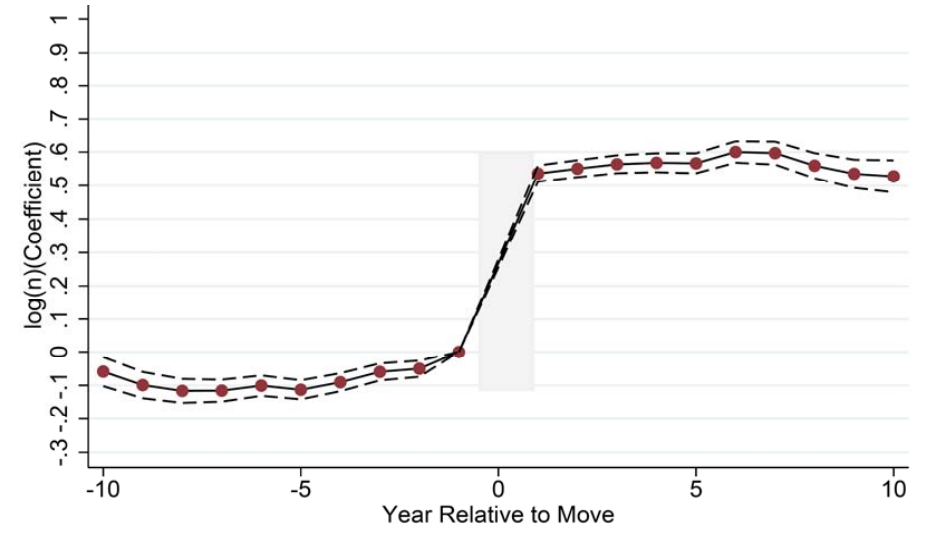
Notes: This figure shows the difference in HRR-average annual utilization between the origin and destination HRRs of patient (Panel A) and physician (Panel B) movers. The samples are all patient movers (Panel A, N = 6,011,147 patient-years) and physician movers (Panel B, N=1,069,099 physician-years).

Figure 3.3: Mover Event Study for Per-Encounter Utilization



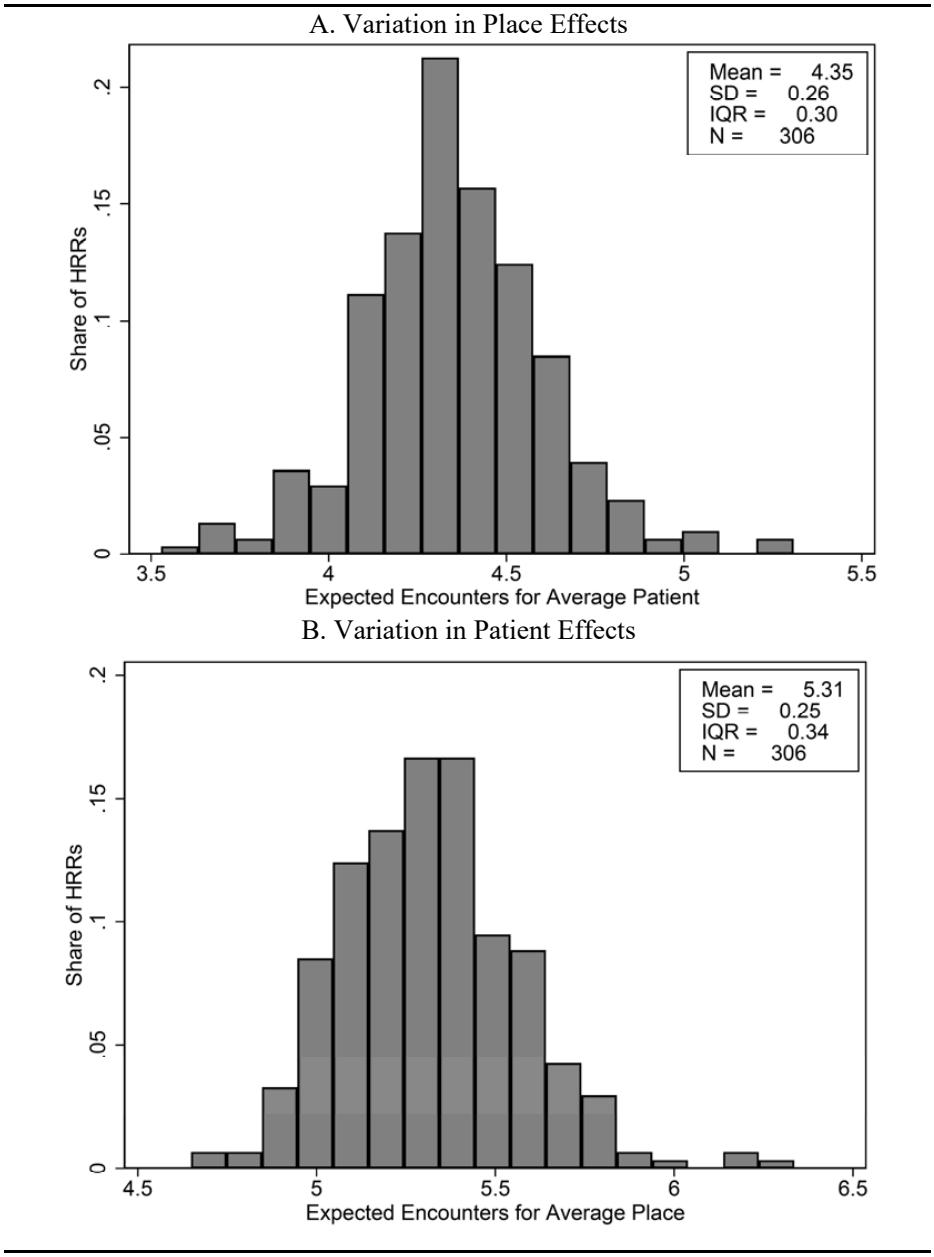
Notes: This figure shows the estimated θ_r coefficients in Equation (32) for patient movers (Panel A) and in Equation (33) for physician movers (Panel B). The coefficient for relative year -1 is normalized to 0. The dependent variable in Panel A is normalized log annual patient spending and the control vector includes indicator variables for five-year age bins and relative-year effects for movers. The dependent variable in Panel B is normalized log annual physician spending and the control vector includes relative-year effects for movers. Dashed lines indicate upper and lower bounds of the 95 percent confidence interval for each θ_r estimate. The sample is all patient-years (Panel A, N = 23,678,685 patient-years) or physician-years (Panel B, N = 11,472,923 physician-years).

Figure 3.4: Mover Event Study for Encounter Quantity



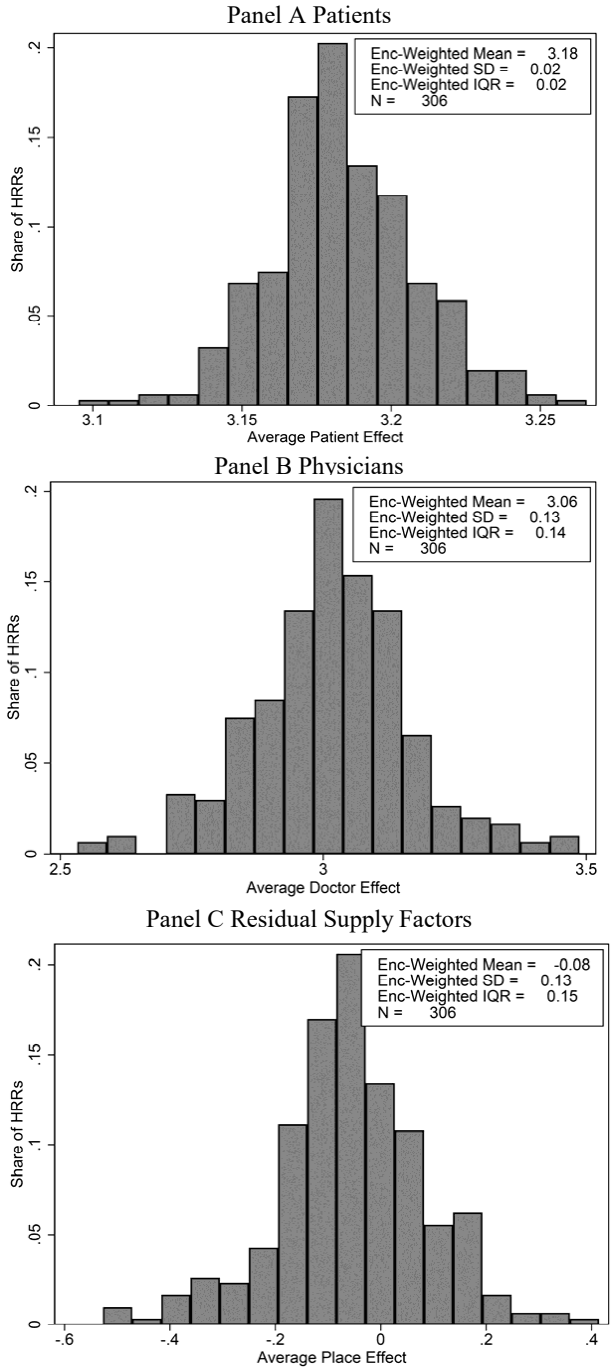
Notes: This figure shows the estimated θ_r coefficients in Equation (34) for patient movers. The coefficient for relative year -1 is normalized to 0. The dependent variable is normalized log annual number of encounters, set equal to zero for patients with zero encounters, spending and the control vector includes indicator variables for five-year age bins and relative-year effects for movers. Dashed lines indicate upper and lower bounds of the 95 percent confidence interval for each θ_r estimate. The sample is all patient-years (N = 23,678,685 patient-years).

Figure 3.5: HRR-Average Patient and Place Components: Encounter Quantity



Notes: Histograms display the geographic distribution of the patient- and place-components of the estimates from our Poisson encounter quantity model (23)-(24). The top panel replaces each patient's own patient component (patient, year, relative-year, and age-bin fixed effects) with the average for the entire sample, then simulates the projected number of encounters for each patient using these parameters and averages down to the HRR level. The bottom panel does the same, but replaces the place effect with the average for the entire sample instead. The sample is the baseline sample of all patients (N = 23,678,685 patient-years).

Figure 3.6: HRR-Average Patient, Physician, and Regional Fixed Effects: Per-Encounter Utilization



Notes: Histograms displays the distribution of HRR-average components of utilization estimated from our encounter-level model (22). Panel A shows the distribution of HRR-averages of patient fixed effects from the specification. Panel B shows the distribution of HRR-average physician effects. Panel C shows the distribution of the estimated residual supply factors. In all panels, we list the mean, standard deviation, and interquartile range of the respective component weighted by the number of encounters that occur in that HRR. The sample is the baseline sample of all encounters (N = 144 million encounters).

Table 3.1: Sample Summary Statistics

	(1) Non-movers	(2) Movers
A. Patients		
Share female	0.56	0.60
Share white	0.85	0.88
Mean age first observed	71.02	72.51
Share first observed residence:		
Northest	0.20	0.17
South	0.39	0.42
Midwest	0.25	0.20
West	0.16	0.22
Annual utilization:		
Mean	$7,811	$7,765
S.D.	$12,138	$9,687
Number of chronic conditions:		
Mean	2.98	3.32
S.D.	2.15	2.03
Annual number of encounters:		
Mean	5.14	5.64
S.D.	3.67	3.40
Number of patients	2,443,020	650,264
B. Physicians		
Share first observed residence:		
Northest	0.20	0.19
South	0.38	0.36
Midwest	0.24	0.25
West	0.18	0.20
Annual utilization:		
Mean	$48,454	$130,082
S.D.	$904,700	$239,895
Annual number of encounters:		
Mean	41.42	106.38
S.D.	526.38	105.77
Number of doctors	7,874,704	134,045

Notes: In Panel A, rows for female, white, age first observed, and first observed residence report the shares of patients with the given characteristics among movers and non-movers. The sample is the baseline sample of all patient-years for movers and non-movers (N = 23,678,685 patient-years). Panel B has the analogous statistics for the sample of all physician movers and non-movers (N = 11,472,923 physician-years)

Table 3.2: Within- and Between-HRR Components of Physician Practice Style Variation

Component	Standard Deviation
Overall	0.88
Within-HRR	0.87
Between-HRR	0.13

Notes: This table is based on estimation of Equation (22) and displays the overall (row 1) standard deviation in physician fixed effects as well as within- (row 2) and between- (row 3) HRR components. Row 2 displays the standard deviation of physician fixed effects after the average of physician effects in each physician's HRR is subtracted from her fixed effect estimate. Row 3 shows the standard deviation of these HRR-average estimates.

Table 3.3: Covariance Matrices of Per-encounter Utilization Components

A. Encounter Quantity Model				
Residual place-based factors	0.066			
Patients	0.022	0.062		
B. Encounter Utilization Model				
Residual place-based factors	0.020			
Patients	0.000	0.001		
Physicians	-0.017	0.001	0.020	
Patient-physician selection	-0.005	0.000	0.003	0.005

Notes: This table is based on estimation of from our Poisson encounter quantity model, equations (23)-(24). (Panel A), and encounter utilization model, equation (22) (Panel B), in the baseline sample and lists the variance-covariance matrix for the individual components in the two models averaged to the HRR level.The patient component in both panels includes patient effect, relative years, calendar years, as well as 5-year age bins. The physician component in Pael B includes physician effects and relative years. The place and patient components in Panel A are formed analogously to those in Figure 3.5. The place component replaces each patient's own patient component (patient, year, relative-year, and age-bin fixed effects) with the average for the entire sample, then simulates the projected number of encounters for each patient using these parameters and averages down to the HRR level. The patient component does the same, but replaces the place effect with the average for the entire sample instead.

Table 3.4: Geographic Variation Counterfactuals

	Above/below median			Top/bottom 25%			Top/bottom 10%		
	Level	% decline (increment)	% decline (cumulative)	Level	% decline (increment)	% decline (cumulative)	Level	% decline (increment)	% decline (cumulative)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Observed	0.272			0.451			0.658		
Patient-physician selection	0.257	-5%	-5%	0.425	-6%	-6%	0.631	-4%	-4%
Physicians	0.173	-31%	-36%	0.281	-32%	-38%	0.426	-31%	-35%
Patients	0.031	-53%	-89%	0.066	-48%	-85%	0.127	-45%	-81%
Of which: patient effects in utilization per encounter	*0.165*	*-3%*	*-39%*	*0.271*	*-2%*	*-40%*	*0.415*	*-2%*	*-37%*
Of which: patient effects in # encounters	*0.031*	*-50%*	*-89%*	*0.066*	*-45%*	*-85%*	*0.127*	*-44%*	*-81%*
Residual place-based factors	0.000	-11%	-100%	0.000	-15%	-100%	0.000	-19%	-100%
Of which: residual place-based effects in utilization per encounter	*0.075*	*16%*	*-72%*	*0.130*	*14%*	*-71%*	*0.232*	*16%*	*-65%*
Of which: residual place-based effects in # encounters	*0.000*	*-28%*	*-100%*	*0.000*	*-29%*	*-100%*	*0.000*	*-35%*	*-100%*

Notes: This table is based on estimation of Equation (22), Equation (23)-(24), and the counterfactuals described in Section 3.4.3. Each pair of columns partitions HRRs into two groups based on percentiles of average log patient utilization. The first row in each panel reports the observed difference in average log patient utilization between the two areas. Each successive row reports this difference under a particular counterfactual, along with the incremental and cumulative percentage change relative to the first row baseline. The second row reports the counterfactual difference if there were no differential physician selection within regions. The third row reports the difference if there additionally there were no variation in physician intensity in healthcare within an encounter across region. The fourth row reports the difference if additionally there were no differential sorting of patients' demand for healthcare within an encounter across regions. The fifth row reports the difference if additionally there were no differential sorting of patients' demand for healthcare encounters across regions. The sixth row reports the difference if additionally there were no variation in place effects on healthcare utilization within an encounter. The last row reports the difference if additionally there were no place effect on number of encounters across regions. The sample is all encounters (144 million encounters of 3 million patients with 2 million physicians).

Table 3.5: Geographic Variation Counterfactuals (Specialties)

	Above/below median (PCPs)			Above/below median (Cardiologists)			Above/below median (Dermatologists)		
	Level	% decline (increment)	% decline (cumulative)	Level	% decline (increment)	% decline (cumulative)	Level	% decline (increment)	% decline (cumulative)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Observed	0.422			0.511			0.472		
Patient-physician selection	0.412	-2%	-2%	0.482	-6%	-6%	0.443	-6%	-6%
Physicians	0.413	0%	-2%	0.465	-3%	-9%	0.460	4%	-3%
Patients	0.393	-5%	-7%	0.393	-14%	-23%	0.259	-43%	-45%
Of which: patient effects in utilization per encounter	*0.458*	*-8%*	*-10%*	*0.458*	*-2%*	*-10%*	*0.458*	*-8%*	*-10%*
Of which: patient effects in # encounters	*0.393*	*4%*	*-7%*	*0.393*	*-13%*	*-23%*	*0.259*	*-35%*	*-45%*
Residual place-based factors	0.000	-93%	-100%	0.000	-77%	-100%	0.000	-55%	-100%
Of which: residual place-based effects in utilization per encounter	*0.313*	*-19%*	*-26%*	*0.367*	*-5%*	*-28%*	*0.225*	*-7%*	*-52%*
Of which: residual place-based effects in # encounters	*0.000*	*-74%*	*-100%*	*0.000*	*-72%*	*-100%*	*0.000*	*-48%*	*-100%*

Notes: This table is based on estimation of Equation (22), Equation (23)-(24), and the counterfactuals described in Section 3.4.3. Each group of columns partitions HRRs into two groups based on the median of average log patient utilization for each specialty sample. The first row in each panel reports the observed difference in average log patient utilization between the two areas. Each successive row reports this difference under a particular counterfactual, along with the incremental and cumulative percentage change relative to the first row baseline. The second row reports the counterfactual difference if there were no differential physician selection within regions. The third row reports the difference if there additionally there were no variation in physician intensity in healthcare within an encounter across region. The fourth row reports the difference if additionally there were no differential sorting of patients' demand for healthcare across regions. The fifth and sixth row break this change into two separate sequential steps eliminating patient effects on the demand ofr care within an encounter and for healthcare encounters respectively. The seventh row reports the difference if additionally there were no variation in place effects on healthcare utilization. The last two rows break this change into two separate sequential steps eliminating place effects on care within an encounter and number of encounters across regions respectively. The samples are the three specialty samples as listed in the columns.

Table 3.6: Within- and Between-HRR Components of Physician Practice Style Variation (Specialties)

Component	Standard Deviation (PCPs)	Standard Deviation (Cardiologists)	Standard Deviation (Dermatologists)
Overall	0.62	0.93	0.58
Within-HRR	0.61	0.88	0.48
Between-HRR	0.14	0.32	0.32

Notes: This table is based on estimation of Equation (22) in each of the specialty samples listed in the columns and displays the overall (row 1) standard deviation in physician fixed effects as well as within- (row 2) and between- (row 3) HRR components. Row 2 displays the standard deviation of physician fixed effects after the average of physician effects in each physician's HRR is subtracted from her fixed effect estimate. Row 3 shows the standard deviation of these HRR-average estimates.

Table 3.7: Covariance Matrices of Per-Encounter Utilization Components (Specialties)

A. Encounter Quantity Model

	PCPs		Cardiologists		Dermatologists	
Residual place-based factors	0.1139		0.0104		0.0021	
Patients	-0.0719	0.1041	0.0002	0.0035	-0.0008	0.0026

B. Encounter Utilization Model

	PCPs				Cardiologists				Dermatologists			
Residual place-based factors	0.055				0.265				0.242			
Patients	0.000	0.002			0.004	0.004			-0.004	0.007		
Doctors	-0.033	0.001	0.037		-0.231	-0.004	0.220		-0.237	0.000	0.247	
Patient-doctor selection	0.002	0.001	-0.002	0.003	0.051	0.005	-0.047	0.062	0.014	-0.001	-0.015	0.017

Notes: This table is based on estimation of from our Poisson encounter quantity model, equations (23)-(24). (Panel A), and encounter utilization model, equation (22) (Panel B), in each of the three specialty samples and lists the variance-covariance matrix for the individual components in the two models averaged to the HRR level.The patient component in both panels includes patient effect, relative years, calendar years, as well as 5-year age bins. The physician component in Pael B includes physician effects and relative years. The place and patient components in Panel A are formed analogously to those in Figure 3.5. The place component replaces each patient's own patient component (patient, year, relative-year, and age-bin fixed effects) with the average for the entire sample, then simulates the projected number of encounters for each patient using these parameters and averages down to the HRR level. The patient component does the same, but replaces the place effect with the average for the entire sample instead.

A Appendix for Chapter 1

A.1 Additional sample characteristics and an evaluation of the cohesiveness of treatment categories

Apart from spanning a wide variety of systems and conditions, the codes in this analysis are representative of the lion's share of the kinds of code changes that the CMS tends to implement in terms of the number of parent and descendant codes involved. Table A2 in the appendix is an attempt at such a classification. Of the 636 coding changes that have occurred over the past few decades, the vast majority, 440, have involved a single parent code and a single descendant code like the treatment categories in this analysis. A relativley small number, 52, have involved the introduction of a completely new code without a parental code. 44 are cases of the complete splitting off of a parental code into multiple descendant codes with the subsequent deletion of the parental code. Finally, 100 changes have involved multiple parental codes contributing to a new descendant code, sometimes over multiple years and code change cycles. The table shows that most of the code revisions that the CMS has implemented involve the splitting off of one or multiple procedures from a single parental code that continues its existence afterwards. Cases of "pure innovation" where a new code is added without an antecedent or of "mixed innovation" where multiple parental codes contribute to a given descendant code are relatively rare. Overall, the treatment categories selected represent what may be considered usual types of codification events and span a wide variety of body systems indicating that the results presented are not driven by idiosyncracies in the kinds of operations that are being considered.

Given the fact that new code introductions frequently involve the splitting off of procedures from the parental code and that one cannot observe all the different kinds of procedures that are performed within a treatment category (since not all of them have codes), one concern for this analysis is whether treatment categories could be including disparate procedures that are otherwise unrelated to one another. If that is the case, then doctors working within the same treatment category are not a well-defined risk set for the new procedures that are split off from that category and the older procedures that remain within the parent code may not represent technologies that are closely related to the procedure represented by the new code.

In order to evaluate the medical similarity of patients within a treatment category, I collect information on the DRGs for the inpatient stay for the procedures in the main sample. Currently, there are over 700 active DRGs, but this number has fluctuated over the years. In order to simplify the analysis, I aggregate the DRGs up one level into DRG "Groups" by classifying together clusters of DRGs that refer to the same condition and procedure taken, but differentiate reimbursement by the presence of complicating conditions. For instance, DRGs 001 and 002 refer to "heart transplant" with and without major complicating conditions. While the presence of such conditions affects the resource intensity of care, the underlying medical condition and surgical approach are the same in the two cases and thus I aggregate them and other similar DRG clusters into a single group.

Note that even with this level of aggregation, each treatment category has patients from a variety of DRGs. This is because the determination of a DRG comes from a complex combination of patient diagnoses and surgical procedures and has a hierarchical component where the most resource intensive condition for a hospital stay takes precedence over others. This means that even if a particular surgical procedure of interest was performed on the patient, it may not be treating

the medical condition that ends up determining the patient's DRG.

For each treatment category, I calculate the share of observations from that treatment category that fall within the top DRG group, the second most used DRG group, etc. I then average these across treatment categories, weighting by the number of observations in each treatment category, and plot these average shares for DRG group ranks up to 10. The results are shown in Figure A2. The graph shows that even with the complex nature of DRG determination, the amount of clustering of patients within a treatment category is remarkably high. The most popular DRG group receives, on average, 60 percent of the observations in a treatment category and the top 2 - over 80 percent. This implies that the patients treated within a treatment category tend to be medically similar and is an indication that treatment categories do not include disparate and unrelated procedures. Additionally, almost 98% of observations associated with newly introduced codes fall within one of the DRGs previously used for the old treatment category indicating that new codes are used on patients that are medically similar to those that are treated in the parent category. This is consistent with the hierarchichal nature of the ICD9-CM, which tends to place similar procedures in close proximity with each other in the numbering system.

A.2 Impact of using components of technology rather than broad technology

In this section, I evaluate the potential impacts of using a narrow definition of surgical categories in my analysis. As outlined above, some authors and practitioners may find it more natural to define surgeries as groups of ICD9-CM codes as opposed to individual codes. This poses two types of challenges to my analysis. The first is econometric in nature and concerns the potential bias or inconsistency of the estimators introduced by a potential mismatch between the definition of technology used in this paper and that used by practitioners. The second is an external validity challenge - if there is a mismatch between the technology space defined in this study and in individual healthcare studies, how useful are these estimates for drawing broader conclusions about technology diffusion in this space?

To answer the first question, I use a simple econometric framework that incorporates the model I've relied on in my analysis and adds the feature of treatment category groups. In particular, imagine a world with a single surgical operation X (e.g., repair of inguinal hernia) that has two subcategories, A and B, indexed by p (e.g., repair of direct inguinal hernia and repair of indirect inguinal hernia) where each of the subcategories corresponds to a treatment category in my analysis. Imagine further that treatment of patient i with a new procedure within a treatment category p by doctor j is determined through a simple threshold rule:

$$1\{New\}_{ijp} = (\beta_0 + \beta_1 Age_j - \psi_i \geq \theta_p)$$

where $1\{New\}_{ijp}$ is a new-procedure indicator, Age_j is doctor j's age in years, ψ_i is an unobserved patient appropriateness factor, and θ_p is treatment-category-specific threshold. If ψ_i is uniformly distributed,[59] this implies that

$$Pr\{New\}_{ijp} = \beta_0 + \beta_1 Age_j + \theta_p$$

[59]This is a simplified version of the random utility model proposed by Heckman and Snyder (1977) where unobservable shocks across choice alternatives are independently, but not *identically* distributed.

which then leads to the baseline regression specification:

$$1\{New\}_{ijp} = \beta_0 + \beta_1 Age_j + \gamma_p + \varepsilon_{ijp}$$

The estimate of the age effect is simply:

$$\hat{\beta}_1 = \frac{Cov(Age_j, 1\{New\}_{ijp})}{Var(Age_j)}$$

This estimate would be unbiased and consistent under the normal assumptions of conditional independence of the error term from doctor age conditional on observables.

How can the presence of a broader technological definition affect the validity of this assumption? First, it is easy to see that if patients are randomly distributed across the two treatment categories A and B, then the presence of X as a grouping of A and B has no bearing on the estimator. However, if the presence of a higher grouping of procedures induces doctors to select patients into one category or the other non-randomly, this may lead to bias in the estimator. To see a simple case of this, suppose that a doctor uses a simple threshold rule to determine which sub-category to assign the patient to:

$$1\{A\}_{ijp} = \eta_i \geq \tau_p$$

This can correspond either to the resolution of uncertainty in the patient's underlying medical condition (e.g., does she require a direct or indirect hernia repair) or to the documentation of the procedure afterwards in the presence of ambiguity (was the procedure performed closer to direct or indirect hernia repair). In that case, we have (assuming a similar distribution of doctor age across A and B):

$$\hat{\beta}_1 = \frac{Cov(Age_j, 1\{New\}_{ijp})}{Var(Age_j)} =$$

$$= \frac{Cov(Age_j, 1\{New\}_{ijp} | \eta_i \geq \tau_p)}{Var(Age_j)} \times Pr(\eta_i \geq \tau_p) + \frac{Cov(Age_j, 1\{New\}_{ijp} | \eta_i < \tau_p)}{Var(Age_j)} \times Pr(\eta_i < \tau_p)$$

If η_i, the patient appropriateness for treatment category A is independent from her appropriateness for treatment with new technology, then the above reduces to the unbiased case. However, if patient treatment category appropriateness is correlated with the likelihood of receiving the new surgical procedure, the coefficient above can be biased. The bias can be positive if large treatment categories within a surgical procedure have most of the patients that are more likely to get the new technology or negative if large treatment categories have fewer of the new-technology appropriate patients. Additionally, some of the treatment categories within a surgical procedure may be dropped out of my analysis due to sample restrictions. The overall direction of the bias in the sample will depend on the balance of these factors across broader surgical technologies.

Naturally, evaluating the sensitivity of the results to these potential issues can be difficult without a proper definition of broader surgical technology in all cases. However, the fact that the estimates do not change substantially due to resonable perturbations in the set of treatment categories used as described in the Robustness section above is one indication that this kind of bias

may not be very important in this setting.

This framework can also help me answer the second challenge listed above - about the relevance of the estimator to other definitions of technological space when it comes to surgical procedures. One potential interpretation of this question is about whether the best way to define the age-adoption parameter $\beta_1$is at the treatment category level or broader surgical procedure level. To the extent that adopting a particular broader surgical procedure requires familiarity with multiple subcategories, the procedure-level parameter will be a weighted average of the treatment-category-level parameters. Since my sample restrictions work at the treatment-category level, it is possible that the set of treatment categories considered cannot be neatly partitioned into a set of broader surgical procedures, all of whose components are in the sample. In that sense, it is possible that the overall parameter estimate from my analysis does not match that which may be obtained if one were to perform a similar analysis on surgical procedures more broadly defined. While my definition of the technological space is driven in many ways by the issue of measurement, the fact that the estimates are very robust to changes in the set of treatment categories considered indicates that this kind of cross-treatment category/within-surgical-procedure variation in β_1 unlikely to be important.

A.3 Impact of delay in codification

In this section, I consider the likely effect of the delay between innovation and codification on my estimates. In particular, as outlined in the institutional background section above, surgical innovation is a gradual and incremental process and there could be substantial delay between the time a procedure is first used (or even first widely established) and the time it receives its own ICD9-CM code. This delay means that I'm able to measure the age-adoption gradient only after the innovation has had some time to diffuse in the relevant market. This feature of my setup may have an effect on the estimated age-adoption gradient.

I use the model I developed in the theoretical section to identify two channels through which this may occur. The first is entry into the market of younger physicians. In particular, since codification could occur more than a decade after a procedure is first performed, it is possible and likely that medical programs may begin incorporating it into their curricula by the time I first observe the new procedure. This means that at least some of the physicians in the sample are likely to have seen the procedure during their education and thus to have benefitted from formal instruction in mastering it. This will push the age-adoption gradient to be steeper than if it were measured when an innovation first entered the market.

The second channel is increased exposure. In particular, as a technology becomes more widely used, there could be adoption spillovers as new users have a firmer foundation and support network to draw upon in learning the new procedure. This means that as time goes by, the parameters of the adoption process change as the market and its institutions may adapt to the presence of new technology. The extent to which this benefits younger or older physicians is ambiguous, however. On the one hand, it is possible that adoption in the early stages of a technology requires a lot of fluid intelligence and ability and willingness to deal with unfamiliar concepts, which will tend to steepen the age-adoption profile as older physicians will be less likely to adopt. However, it is also possible that only individuals with a lot of experience and clout with the institutions involved will be able to make the most out of adoption. In that case, older physicians may be more likely to experiment with a new technology, which would flatten the age-adoption profile. The overall

direction of these changes is ambiguous.

A.4 Impact of medical school and residency

The impact of medical education is a natural candidate as a potential mechanism for the documented age-adoption profile effect. Physicians who go to highly qualified programs may be more prone to adopt new technology either because of underlying skill type or because of the education they obtain. In order to test this idea, I used medical school rankings produced by US News & World Report to create a dummy for a doctor's medical school belonging to the top 15 in the nation (Abigail Hess (2019)). I then ran the same specification as in equation (2) above, but adding an interaction between this dummy and doctor age. The results are presented in Table A4. Attending a top medical school does not seem to have a statistically significant influence on the age effect on propensity to adopt.

Another related question is whether the results are driven by the fact that the analysis sample may contain some doctors who have not completed their residency and some who have. In particular, it is possible that the effect is driven by the completion of residency whereby current residents are learning all the new technologies due to their program while physicians who have completed their training and who tend to be older are not doing so. In order to investigate this possibility, I re-run the baseline specification (2) on the subsample of observations due to physicians who have completed their residency.[60] The results are presented in Table A5 and are essentially identical to the main results in the paper. This is perhaps unsurprising given that less than 98 out of 178,000 observations in the sample are due to residents.[61]

Nonetheless, a slightly different form of the same question is whether the results documented here are due to the possibility that getting to the frontier is easier when one is close to the frontier. In particular, doctors are brought to the skill frontier at the time they graduate. If there are large fixed costs to skill investment that increase with the number of new skills that have not been acquired in the past, one would expect that the rest of a physician's career is marked by a slow decay of skills relative to the medical frontier. This idea would imply that physicians who are close to the moment they graduated would be much more likely to acquire new skills relative to those of the same age who are farther from graduation. A simple way of testing this possibility is to re-do the baseline regression and see if observations due to residents fall above or below the regression line. I perform this exercise in Figure A3 below. Specifically, I define as "residents/recent graduates" individuals who are either residents or are within 5 years of their graduation date and were 40 at the time they graduated. This gives me about 6,000 observations in this category, which I split into 4 quantiles based on age at code introduction. I then split the remaining 172,000 observations in 96 quantiles, so that each quantile in the two subsamples has roughly an equal number of observations. I plot these points in a binscatter having partialled out all relevant controls from the regressions above from the dependent and independent variables and having added back the overall mean of each subsample to each point. The regression line I fit is the regression line using the overall sample.

As is apparent from Figure A3, recent graduates lie *below* the regression line. This indicates that recent graduates are less likely to use the new code in a treatment category relative to more experienced physicians at their age. While this is somewhat counterintuitive, it is consistent with

[60] I drop physicians with missing values on year of training from this exercise.

[61] The rest of the sample differences are due to physicians with missing values on year of training.

previous literature in age-decision profiles that indicates that initial experience is beneficial to decision-making. In particular, one possibility is that individual doctors need some experience "on the job" before gaining the confidence and skills necessary to adopt new technology successfully.

A.5 Diagnostic code sample construction

I take all physicians, identified by UPIN-NPI pair,[62] who submit claims between 2001 and 2013 and identify PCPs as those submitting the majority of their claims under a specialty identified as "general practice", "family practice", or "internal medicine". I limit to claims by PCPs where the place of service is identified as "office" in order to remove observations where PCPs performed services in inpatient or other institutional settings where the ultimate agent behind the patient diagnosis and medical service provided is unclear. I identify doctor age using the AMA Masterfile and various patient observables using the denominator file as in the basline sample. I further limit to claims whose primary diagnosis code satisfies the diagnostic category restrictions. In particular, I look at diagnostic categories which were revised exactly once between 2001 and 2013 where there is a single parental code and where both the parental and descendent codes are in use up to and including 2013. Table A6 below includes some summary statistics for the sample, while Figure A4 shows the overall diffusion pattern (S-curve) for the codes adjusting for patient mix and parental diagnostic category. As noted above, the diffusion curve is somewhat flatter than that in the procedure code analysis, perhaps reflecting the possibility that diagnostic knowledge diffuses more quickly than medical procedures. Figure A5 shows a binscatter of the regression in column (4) of Table 1.6. The flat age-adoption profile in the case of diagnostic codes is evident with no outliers seemingly driving the presented results.

A.6 Prescription pharmaceuticals sample construction

I use Medicare Part D files for 2006-2015. These are prescription-level datasets that contain all prescriptions filled by 20% of Medicare beneficiaries with data on the exact drug and package used (NDC11 codes), quantity filled, date and location where the prescription was filled as well as the NPI of the prescribing physician. I combine all prescriptions of a single compound by a given physician to a given beneficiary in a single year into a single observation. I use the FDA Orange Book to identify the approval date for each chemical compound and define a drug's therapeutic class using the American Hospital Formulary Service® (AHFS) 8-level classification of 256 therapeutic classes as used by Einav et al. (2018). Drugs in a class have similar mechanisms of action or chemical structure and are used to treat similar or related diseases. I use the set of NDCs considered by Einav et al. (2018) and focus on therapeutic classes that received a new chemical compound between 2006 and 2015 based on FDA approval date. I use the AMA Masterfile to identify the age of the prescribing physician and the Medicare Denominator file to add patient characteristics.

I begin with the sample of all baseline NDCs used in Einav et al. (2018). The list, generously provided by the authors, has information on the corresponding chemical entity (non-proprietary name), related brands used to market the drug (proprietary name), the therapeutic class designations of each drug and the application number used in the FDA application. I use the application

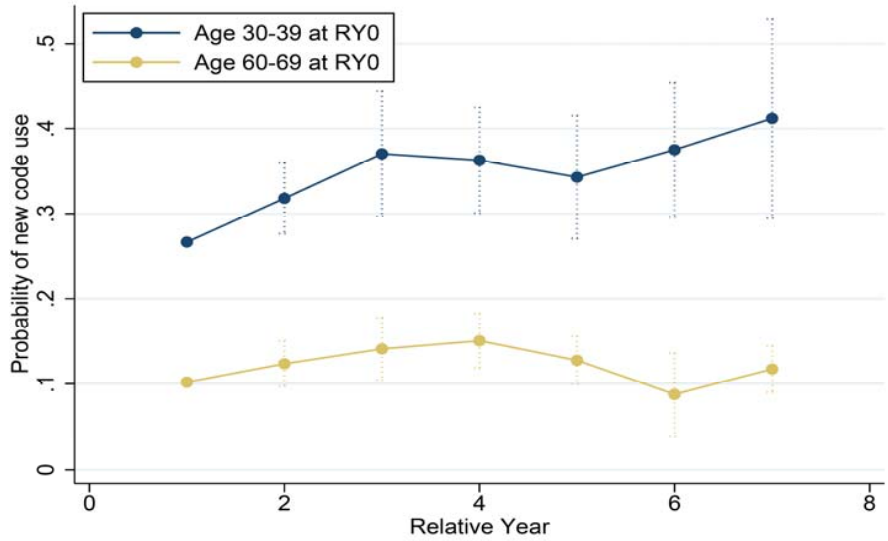
[62] I use the NBER's UPIN-NPI crosswalk to fill-in any missing values

number for each chemical compound to merge approval year information from the FDA Orange Book. Chemical compounds are identified as "Ingredients" in the Orange Book. For each ingredient, I find the earliest listed approval year (separate drugs using the same ingredient such as generics get different approval years) and designate the earliest year as the approval year of the chemical compound (i.e., the first year the chemical compound entered the market in any form). There are 88 out of 2,537 application numbers that have multiple ingredients associated with them and I drop these from the analysis. Additionally, while typically a single ingredient may correspond to multiple non-proprietary names (due to variations in spelling), there are 44 (out of 987) non-proprietary names that are associated with multiple ingredients. I drop these from the analysis as well.

I define a therapeutic class as "treated" if any of its drugs received its earliest approval between 2006 and 2015. Note that a single drug may be part of multiple therapeutic classes as it may be used to treat disparate medical conditions. In those cases, I define all therapeutic classes that include a newly approved drug as treated. I take all treated classes and merge them on to the Part D event files. I collapse down to the beneficiary-physician-drug-year level, so that all prescriptions for a given chemical compound (ingredient) by a single doctor for a single patient in a single year count as a single prescription event, or observation, in the sample. If a prescription drug has multiple therapeutic classes associated with it, I include all prescription events for that drug as separate observations for each of the drug's therapeutic classes. The resulting dataset is at the prescription event level where each prescription event corresponds to a single therapeutic class used for the respective drug.

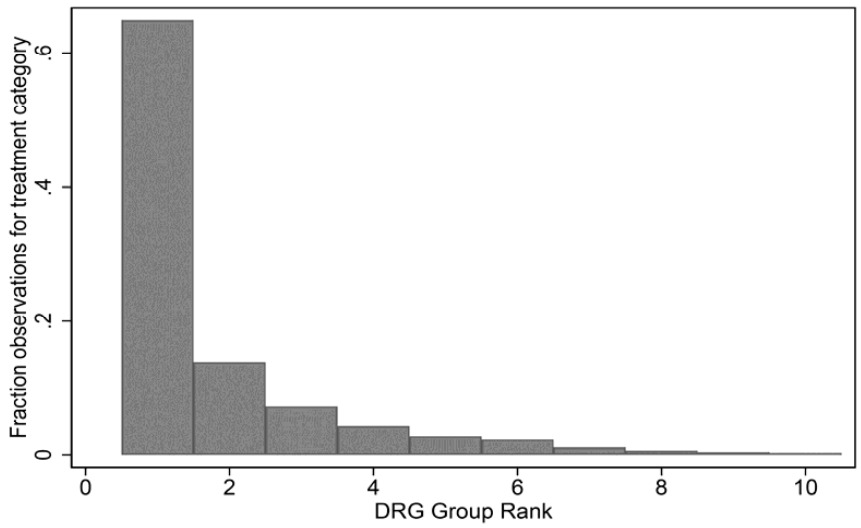
A.7 Appendix Tables and Figures

Figure A1: Relative Year Effects for Two Age Groups



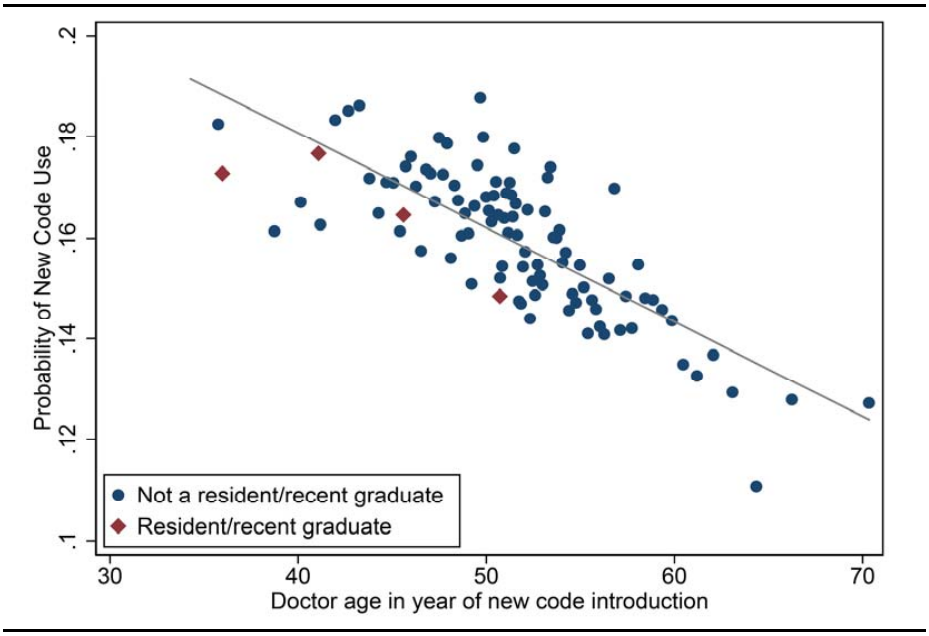
Notes: The figure plots the coefficients from a regression of new code use on relative year using the same controls performed separately for physicians 30-39 at codification and physicians 60-69 at codification. The dotted lines represent 95-percent confidence intervals.

Figure A2: Share of Observations by DRG Group Rank



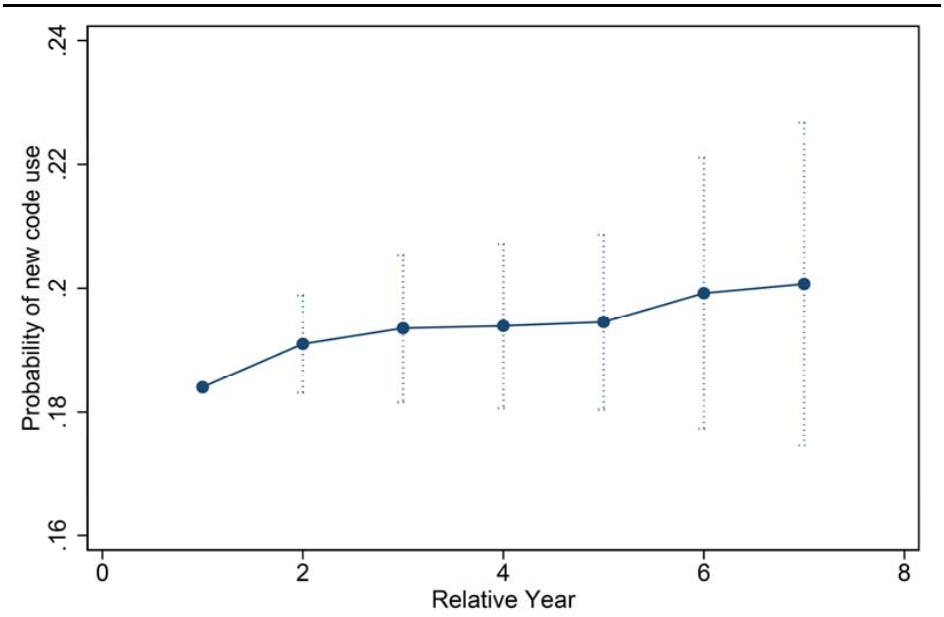
Notes: The figure shows the share of observations in each treatment category that falls within the first, second, third, etc. most populated DRG group for that treatment category. Shares are averages across treatment categories weighted by the number of observations within that treatment category.

Figure A3: Binscatter of Regression Showing Residents and Non-Residents



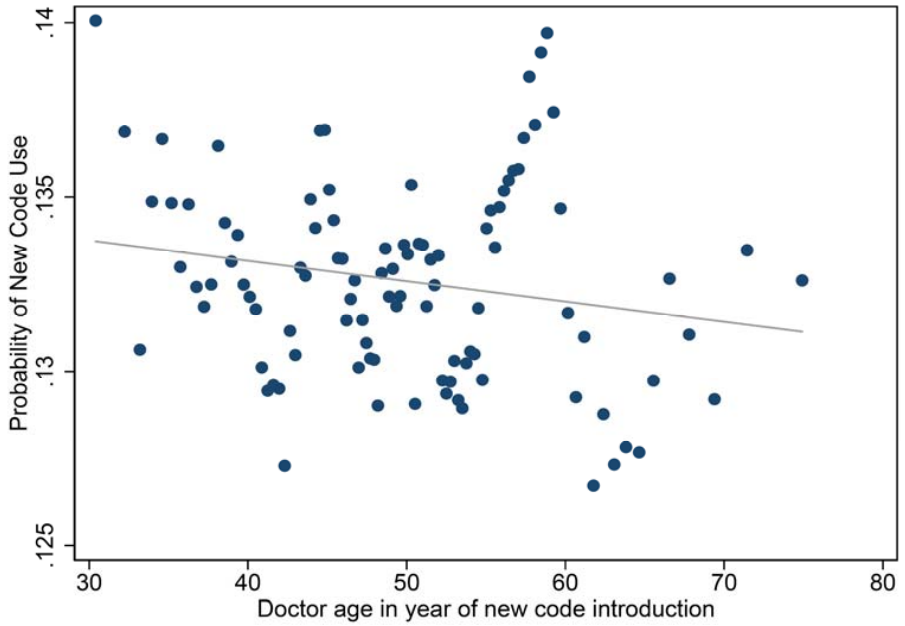
Notes: The figure shows a binscatter of the baseline regression splitting observations into residents/recent graduates (doctors within 2 years of completing residency that are 40 or less at the time of completion) and more experienced physicians (not fulfilling those conditions).

Figure A4: New Diagnosis Code Diffusion S-Curve



Notes: The figure plots the coefficient from a regression of new diagnostic code dummy on year relative to new code introduction with treatment category fixed effects and controls for calendar year, patient age, race, sex, chronic conditions, and Medicaid receipt, and doctor controls for AMA hospital ID, med school ID, and gender.

Figure A5: New Diagnosis Code Baseline Regression Binscatter



Notes: The figure is a binscatter of the baseline specification (9). The figure is constructed by estimating the residuals from separate regressions of the new code dummy and doctor age on the full set of patient and physician controls, dividing them into quantiles, adding the means of the dependent and independent variable respectively, and plotting hte results. The grey line

Table A1: Code Changes by Anatomic System

	Number of Code Changes	% of Total Code Changes	Number of Observations	% of Total Observations
Operations On The Digestive System	26	31%	35,518	20%
Miscellaneous Diagnostic And Therapeutic Procedures	15	18%	7,854	4%
Operations On The Cardiovascular System	14	16%	40,509	23%
Operations On The Musculoskeletal System	7	8%	39,724	22%
Operations On The Female Genital Organs	6	7%	2,272	1%
Operations On The Respiratory System	4	5%	2,527	1%
Procedures And Interventions , Not Elsewhere Classified	3	4%	46,460	26%
Operations On The Integumentary System	3	4%	1,488	1%
Operations On The Nervous System	3	4%	2,090	1%
Operations On The Endocrine System	2	2%	19	0%
Operations On The Urinary System	1	1%	163	0%
Operations On The Nose, Mouth, And Pharynx	1	1%	75	0%

Notes: Table shows the distribution of of ICD9-CM procedure code changes in sample by body system they represent.

Table A2: Code Changes

Change Type	Number of instances
Multiple previous codes	100
No previous code	52
One previous code, which was subsequently deleted	44
One previous code, which was kept after the change	440

Notes: Table shows the number of ICD9-CM procedure code changes by type.

Table A3: Baseline Analysis Using Logit

VARIABLES	(1) 1(New code)	(2) 1(New code)	(3) 1(New code)
Doc Age at RY 0	-0.0212***	-0.0213***	-0.0214***
	(0.00414)	(0.00415)	(0.00410)
Observations	173,804	173,801	173,801
Avg marginal effect	-0.0020	-0.0020	-0.0020
Treatment Cat F.E.	x	x	x
Chronic Conditions		x	x
Demographics			x
Phys Observables			

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running the baseline specification using a logit model. Standard errors are clustered at the treatment category level.

Table A4: Impact of Medical School Rank

	(1)
VARIABLES	1(New code)
Doc Age at RY 0 x	-0.000322
Top 15 Flag	(0.000686)
Doc Age at RY 0	-0.00184***
	(0.000550)
Observations	178,699
R-squared	0.353
Dep Var Mean	0.244
Treatment Cat & Yr. F.E.	x
Chronic Conditions	x
Demographics	x
Physician Observables	x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running the main specification with an interaction with an indicator for having attended a top-15 medical school. Standard errors are clustered at the treatment category level.

Table A5: Impact of Removing Residents

VARIABLES	(1) 1(New code)	(2) 1(New code)	(3) 1(New code)	(4) 1(New code)	(5) 1(New code)
Doc Age at RY 0	-0.00177***	-0.00177***	-0.00177***	-0.00187***	
	(0.000645)	(0.000641)	(0.000637)	(0.000536)	
Doc Age 40-49					-0.0196***
					(0.00671)
Doc Age 50-59					-0.0343***
					(0.00903)
Doc Age 60-69					-0.0526***
					(0.0143)
Doc Age 70-79					-0.0567***
					(0.0180)
Observations	174,563	174,563	174,563	174,563	174,563
R-squared	0.307	0.309	0.309	0.352	0.352
Dep Var Mean	0.159	0.159	0.159	0.159	0.159
Treatment Cat F.E.	x	x	x	x	x
Chronic Conditions		x	x	x	x
Demographics			x	x	x
Phys Observables				x	x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table shows the results from running the baseline specification (9) on a subsample of non-residents. Standard errors are clustered at the treatment category level.

Table A6: Diagnosis Code Sample Summary Statistics

Panel A: Patient Level

	Mean	S.D
Share of patients with any chronic condition	0.08	0.27
Share of white patients	0.89	0.32
Share of male patients	0.50	0.50
Share on medicaid	0.18	0.38
Patient age	73.83	10.45
Share treated with new code	0.13	0.34
Number of patients in sample	650,100	
Number of observations in sample	5,319,067	

Panel B: Doctor-Year Level

	Mean	S.D
Age	50.41	9.90
Share male	0.80	0.40
Total years in sample	5.13	2.80
Number of treatment categories	2.58	1.71
Probability of using new code	0.14	0.28
Number of doctors in sample	132,006	

B Appendix for Chapter 2

B.1 A Toy Model of User Innovation

I outline a simple toy model showing how the extent of external rewards may be driven by the nature of the innovative activity. I follow closely Benabou and Tirole (2003)'s framework of extrinsic and intrinsic motivation. Suppose the social planner has a research project which costs $c \in [\underline{c},\bar{c}]$ to complete and has a social benefit B if successfully finished. In this context c may represent the various costs associated with inventing a new drug or researching a new procedure - capital expenditures, opportunity cost of labor (or surgeon time), costs of obtaining financing for the project, regulatory costs of clearing the project's output, etc. The project can be completed by a research agent, who has a private benefit from successfully doing so of b. b can represent for instance any private altruistic benefit from contributing to social welfare through one's work, academic or social prestige due to the accomplishment, the sense of achievement from success in a difficult and important task. The agent does not observe the exact cost of the project before committing to it, but knows that it has a distribution with cdf $F(c)$ and observes a signal σ with cdf $G(\sigma)$ and pdf $g(\sigma)$ of the project's difficulty. I assume that the signal's distribution has the monotone likelihood ratio property (MLRP): for $\sigma_1 > \sigma_2$, $\frac{g(\sigma_1|c)}{g(\sigma_2|c)}$ is decreasing in c. This property guarantees that as the cost increases, high realizations of the signal σ become less likely. In other words, a high realization of σ indicates a lower likely value of c.

The project has a success rate of θ. The social planner may choose to offer an additional reward of r to the researcher contingent upon the successful completion of the project. Such rewards can take many forms - examples are explicit research awards or the opportunity to profit from a temporary government-sanctioned monopoly enforced through intellectual property rights. However, the reward comes at a cost $f(r)$ with $f'(r) > 0$ to the planner. Since this is a social planner, the reward itself is simply a transfer from one party to another. But the planner may suffer economic costs such as the deadweight loss associated with monopolies or with the taxation required to obtain the funds for r. The cost of the reward is assumed to increase with the size of the reward as would be the case, for instance, if a pharmaceutical's monopoly period were to be extended for a longer period of time inducing a higher deadweight loss to society.

The game is played in the following stages: the social planner observes the project costs and decides on the reward r, then the researcher observes r as well as σ and decides whether to commit to the task with the decision marked by $i \in \{0,1\}$.

The agent's expectation of the project costs is $\hat{c}(\sigma,r) = E[c|\sigma,r]$ and thus she will commit to the project if $\theta(b+r) \geq \hat{c}(\sigma,r)$. The MLRP guarantees that there exists a threshold $\sigma^*(r)$ such that the agent commits to the project if and only if $\sigma \geq \sigma^*(r)$. The social planner then maximizes the following objective function with respect to r:

$$\theta[1 - G(\sigma^*(r)|c)][B - f(r)]$$

As in Benabou and Tirole (2003), in the non-degenerate equilibrium, higher project costs are associated with higher rewards, but the higher rewards also induce agents to revise their expectations of the project costs upwards.

The model allows an interpretation of some of the patterns observed in the data and institutions of surgical innovation. First, a higher degree of intrinsic motivation b will induce, ceteris

paribus, the social planner to offer lower extrinsic rewards r. Thus, surgeons' medical ethics and the promise of professional prestige will be associated with lower explicit rewards from the social planner. Thus, the high intrinsic motivation of the surgeon motivators means that even in the absence of intellectual property rights, innovation may still occur. Second, higher intrinsic motivation and lower rewards are associated with a lower degree of social welfare loss $f(r)$. Thus, surgical innovation, which does not require temporary monopolies is relatively more beneficial to society for a project of the same overall value B. At the same time, the degree to which the social planner can rely on intrinsic motivation here is limited by the magnitude project costs c. If the project costs significantly more than any reasonable realization of b, then the social planner has no choice but to resort to extrinsic rewards to encourage innovation. This is consistent with a world where pharmaceutical innovation has significant financial costs that likely have to be reimbursed through significant external rewards.

The intrinsic motivation effect may be reinforced by the technological profile of the mode of innovation. The model so far has assumed a single project where the social planner has to decide on a reward scheme. However, one may imagine that the innovation project may proceed in multiple separate stages where the agent has to agree to commit to each stage separately. This is especially applicable to the case of surgical innovation where a single novel surgical approach is developed through an incremental process where multiple patients have to undergo the procedure and the procedure's parameters are refined as it is used more and more. In pharmaceutical innovation, on the other hand, the research project is completed in one go and the returns to the agent transpire at that point. If the social planner has to decide on the reimbursement scheme for each stage separately and in advance, the fragmentation of the research process may serve to further lower the social planner's use of an external reward.

Imagine, for instance, that the game was modified in the following way: there are now two sub-projects, with rewards B_1, B_2, the social planner observes the cost c, which is the same for both stages, and offers separate rewards $r_1$and r_2,. Imagine that the agent still observes a signal σ, observes only the reward r_1 before committing to the first project and then, after her commitment decision, she observes the reward r_2 for the second project and decides whether to commit to that one as well. In this case, the social planner's objective function takes the following modified form:

$$\theta[1-G(\sigma_1^*(r_1)|c_1)][B_1-f(r_1)]+\theta[1-G(\sigma_2^*(r_1,r_2)|c_2)][B_2-f(r_2)]$$

Relative to a world where only the first stage exists, the social planner's incentives to offer a high external reward for this first project r_1 are diminished by the fact that the agent will update her expectations of the difficulty of the project based on this reward and thus require a potentially higher reward in the second stage to exert effort. Formally, while $\frac{d}{dr_1}\sigma_1^*(r_1) > 0$ as before, $\frac{\partial}{\partial r_1}\sigma_2^*(r_1,r_2) < 0$, which exerts a downward pressure on the optimal choice of r_1. Hence, the incremental nature of surgical innovation induces the social planner to offer even lower external rewards.

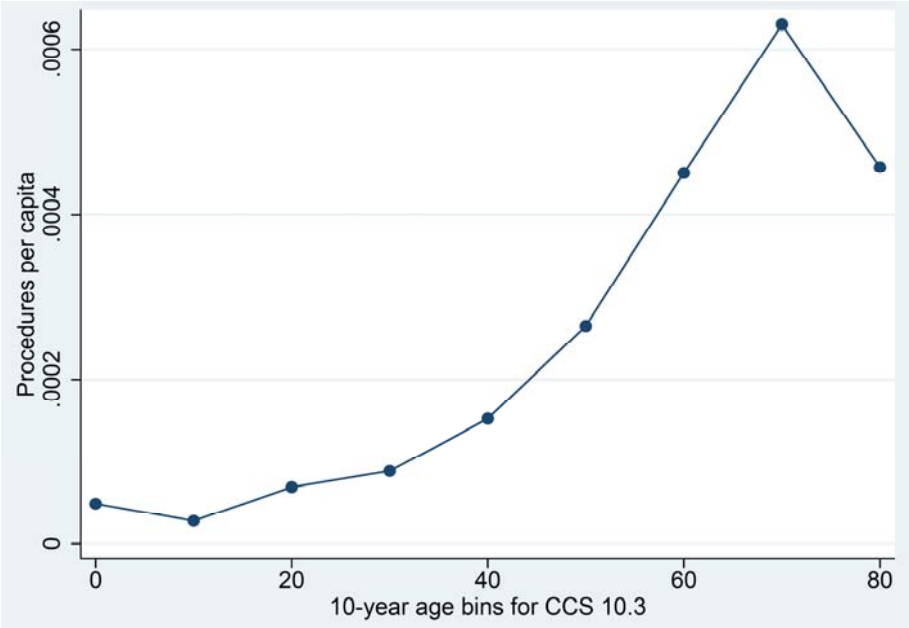
A separate, but related point may be made about the choice of method of offering external rewards. In the case of pharmaceuticals, intellectual property monopolies are the method of choice - the social planner ties the size of the reward to the overall social benefit from the innovative effort as revealed through market demand. The cost of said monopoly is the deadweight loss due to monopoly pricing and decreased patient access to drugs that already exist. However, in a world of fragmented innovation monopoly power may also be associated with a high amount of transaction

costs - developing the next stage of the surgical procedure may require licensing from a number of other agents who have developed methods relevant to the procedure, and this may increase the social costs of offering the external reward substantially. It is therefore perhaps unsurprising that the main method of rewarding innovation financially in this sector may come through one-off grants from various agencies that do a poor job of revealing the benefit of the innovation, but avoid the high social costs that will likely result from the enforcement of intellectual property.

The question of method of rewarding invention is also relevant for the impact of market size on innovation. In the case of pharmaceuticals, sicne the reward for invention is closely tied to overall consumer benefit of the product (through monopoly pricing), market size enters indirectly through the r term. However, in the case of surgeons, we have seen that this reward plays a rather minimal role. The model therefore suggests that market size enters through the surgeon's own intrinsic motivator b. This can occur for instance if surgeons feel more internally motivated to work on problems that will benefit more people and are considered more important for society in gneral. It can also occur if there is more academic prestige associated with solving a bigger problem. Both of these propositions are salient features in doctors' own conceptualization of their work and in the system of academic prestige that underlies much of the work in academic medical centers.

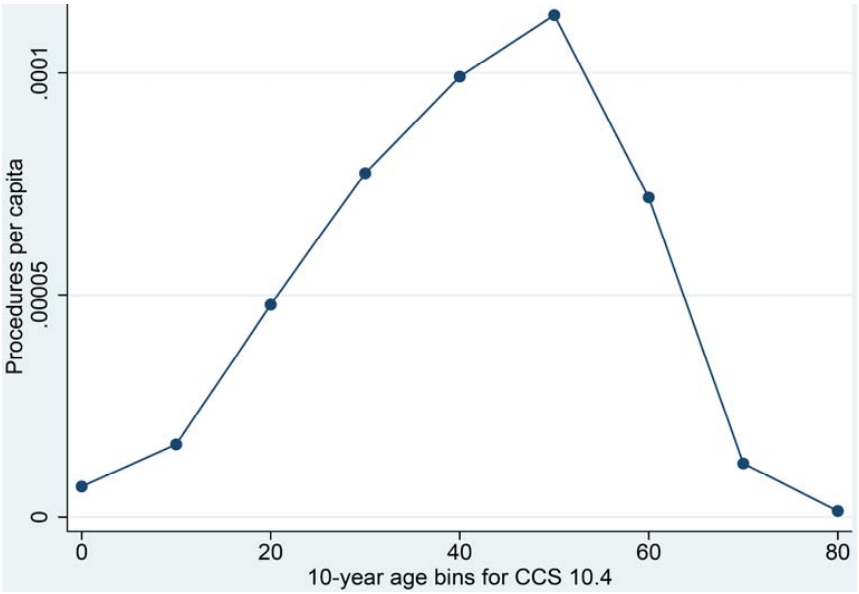
B.2 Appendix Figures and Tables

Figure B1: Nephrectomy Procedures Per Capita



Notes: The figure plots the fraction of individuals in each age bin in the population that get a procedure in CCS category 10.3 (nephrectomy).

Figure B2: Kidney Transplant Procedures Per Capita



Notes: The figure plots the fraction of individuals in each age bin in the population that get a procedure in CCS category 10.4 ((kidney transplant)).

C Appendix for Chapter 3

C.1 Data Appendix

In order to construct our baseline set of patient-physician encounters, we identify the physician in the claims data who is most likely to be responsible for the patient's treatment decision. In the inpatient and outpatient claims data, this is the attending physician, and in carrier files this is the referring physician. In particular, federal law requires that attending physicians attest to each patient's diagnoses and procedures. Medicare also requires that the attending physician be listed on any claim that contains services other than nonscheduled transportation. Similarly, Medicare requires that any claims that contain services and items that are the results of a physician's order or referral list the referring physician.[63] A physician-patient encounter therefore consists of all services provided to that patient as a result of the decisions of that physician in her capacity as an attending or referring physician. This procedure ensures that we pick out the relevant decision maker and that coverage of this decision maker in the data is maximized due to the relevant federal regulations.

The raw sample contains the interactions of approximately 13 million patients and 1.7 million physicians. We identify physicians through a combination of their National Provider Number (NPI) and Universal Provider Identification Number (UPIN). An individual physician is a unique combination of UPIN-NPI. We use a crosswalk prepared by the National Bureau of Economic Research (NBER) which matches UPINs and NPIs that appear together in the claims data. If a single NPI appears with multiple UPINs and vice versa, each unique combination is considered an different physician. An exception is made when a UPIN only appears in combination with a single NPI or a missing value, in which case missing values are set to the unique NPI that corresponds to said UPIN. We follow a similar procedure when there is a single UPIN or a missing value corresponding to a given NPI.

18 million out of the 144 million encounters in our baseline sample have a missing physician identifier. The vast majority of these encounters have a missing physician identifier because there is no UPIN or NPI listed on a subset of a patient's medical claims in a given year. Additionally, there are instances of claims in the raw data with erroneous values for the relevant physician's UPIN and NPI. An erroneous value is a value that does not follow the correct format of these two identifiers: a letter followed by five digits for UPINs and 10 digits for NPIs. We set these values to missing as well. An exception is made for surrogate UPINs such as "INT000" and "RES000", which are sometimes listed on claims when the responsible physician is yet to be issued a UPIN (e.g., if she is a resident) (Solutions, 2017). These UPIN values are kept intact and not set to missing.

Patients can exit our sample for three primary reasons: death, switching to Medicare Advantage, and exiting the 65-99 age window. About a third of patients die in our sample window, and their mortality is similar for movers and non-movers. About a fifth exits from switching to Medicare Advantage. We observe the average non-mover for 10.5 years and the average mover for 11.6 years, with part of the difference being mechanical since the mover label is contingent on observing a patient for at least 2 years. Correspondingly, we observe the average non-mover physician for 10.6 years and the average mover physician for 8.1 years.

We exclude patients who moved more than once and whose share of claims in the destination

[63] 42 CFR 412.46 - Medical review requirements, Medicare Processing Manual Chapters 25 and 26

HRR did not exceed that in the origin HRR by at least 0.75 after the move. For analytic tractability our baseline analysis excludes patients whose HRR of residence changes multiple times.

To construct our physician mover definition, we follow a methodlogy analogous to the one developed by Molitor, 2018. We begin by calculating the number of patients that each physician sees in each HRR we observe her. We keep physicians who are observed in more than one HRR and HRRs where each of these physicians sees more than one patient. For each of these physician-HRR combinations, we form treatment episodes as the period of time between the first and last encounter of a physician with a patient in that HRR. We define a physician's primary HRR as the one in which she sees the most patients. A physician's secondary HRR as the highest-patient-count HRR whose treatment episode does not overlap with that of the primary HRR (we take the one closest in time to the primary HRR in case of a patient-count tie). A physician's origin HRR could be the primary or secondary HRR depending on which one occurs earlier in time. For each move, relative year -1 is the final year of the origin HRR treatment episode and relative year +1 is the first year of the destination HRR treatment episode.

We assign a separate physician ID to each practice episode for physician non-movers. We do the same for physician movers for any non-primary and non-secondary practice episode. In practice, this means treating physicians who see patients from multiple HRRs as multiple independent physicians. We do so to avoid using cross-HRR identifying variation coming from an individual physician's practice patterns outside of a move. We prefer to rely on variation coming from quasi-exogenous cross-HRR physician moves, which is much better understood and more easily verified as satisfying our identifying assumptions.

We construct our specialty visit subsamples using the following methodology.[64] For each specialty, we begin by identifying all carrier claims for patients in the baseline sample where the performing physician specialty matches the desired specialty. For the PCP subsample, this includes "general practice", "family medicine", or "internal medicine". For the cardiology and dermatology subsamples, this includes "cardiology" and "dermatology" respectively. We further narrow these claims down to physicians whose specialty in the Carrier file matches the specialty of interest (i.e., we require both that the claims be "PCP", "cardiology", or "dermatology" claims based on the performing physician and that the performing physician be a "PCP", "cardiologist", or "dermatologist" based on her claims in the Carrier file). We define a physician's Carrier file specialty in a two-step procedure. First, we determine whether the physician is a PCP or a non-PCP based on whether the majority of her submitted claims as performing physician list one of the three PCP specialties ("general practice", "family medicine", or "internal medicine"). If the physician is not a PCP, we define her specialty as the specialty that has the plurality of the physician's submitted claims as a performing physician in a given year. For the PCP subsample, we further drop any claims whose place of service is a hospital's inpatient department or the ER .

C.2 Econometric Appendix

Mover Identification of the Poisson Model

This appendix shows how the Poisson fixed effects model of Hausman et al. (1984) identifies causal effects under a "common growth rates" assumption, similar to the common trends assumption

[64] In defining the PCP sample, we draw inspiration from Fadlon and Parys (2019).

identifying effects in linear regression models. Suppose

$$y_{it} \sim Poisson(\lambda_{it}) \tag{35}$$

where $\ln \lambda_{it} = x_{it}'\beta$. The log-likelihood first-order condition for this model is

$$0 = \sum_i c_i \sum_t x_{it} \left(\frac{y_{it}}{c_i} - \frac{exp(x_{it}'\beta^*)}{\sum_s exp(x_{is}'\beta^*)} \right) \tag{36}$$

for $c_i = \sum_t y_{it}$. Suppose $t \in \{0,1\}$ and $x_{it}'\beta = \alpha_i + \tau T_{it} + \gamma D_{it}$, for binary D_{it} and where $T_{it} = \mathbf{1}\{t = 1\}$. Then we can rewrite (36) as

$$\begin{aligned}
0 &= \sum_i c_i \sum_t \begin{bmatrix} T_{it} \\ D_{it} \end{bmatrix} \left(\frac{y_{it}}{c_i} - \frac{\exp(\tau^* T_{it} + \gamma^* D_{it})}{\sum_s \exp(\tau^* T_{is} + \gamma^* D_{is})} \right) \\
&= \sum_{c=1}^{\infty} cN_c \left(\frac{1}{N_c} \sum_{i:c_i=c} \sum_t \begin{bmatrix} T_{it} \\ D_{it} \end{bmatrix} \left(\frac{y_{it}}{\sum_t y_{it}} - \frac{\exp(\tau^* T_{it} + \gamma^* D_{it})}{\sum_s \exp(\tau^* T_{is} + \gamma^* D_{is})} \right) \right) \\
&= \sum_{c=1}^{\infty} cN_c \begin{bmatrix} \frac{N_{c,0\to1}}{N_c} \left(S^1_{c,0\to1} - \frac{\exp(\tau^*)}{1+\exp(\tau^*+\gamma^*)} \right) + \frac{N_{c,1\to0}}{N_c} \left(S^0_{c,1\to0} - \frac{\exp(\gamma^*)}{\exp(\tau^*)+\exp(\gamma^*)} \right) \\ \frac{N_{c,0\to0}}{N_c} \left(S^1_{c,0\to0} - \frac{\exp(\tau^*)}{1+\exp(\tau^*)} \right) + \frac{N_{c,0\to1}}{N_c} \left(S^1_{c,0\to1} - \frac{\exp(\tau^*+\gamma^*)}{1+\exp(\tau^*+\gamma^*)} \right) \end{bmatrix} \\
&\quad + \sum_{c=1}^{\infty} cN_c \begin{bmatrix} \frac{N_{c,1\to1}}{N_c} \left(S^0_{c,1\to1} - \frac{\exp(\gamma^*)}{\exp(\gamma^*)+\exp(\tau^*+\gamma^*)} \right) + \frac{N_{c,1\to1}}{N_c} \left(S^1_{c,1\to1} - \frac{\exp(\tau^*+\gamma^*)}{\exp(\gamma^*)+\exp(\tau^*+\gamma^*)} \right) \\ \frac{N_{c,1\to0}}{N_c} \left(S^1_{c,1\to0} - \frac{\exp(\tau^*)}{\exp(\tau^*)+\exp(\gamma^*)} \right) + \frac{N_{c,1\to1}}{N_c} \left(S^1_{c,1\to1} - \frac{\exp(\tau^*+\gamma^*)}{\exp(\beta^*)+\exp(\tau^*+\gamma^*)} \right) \end{bmatrix},
\end{aligned} \tag{37}$$

where $N_c = |\{c_i : c_i = c\}|$ is the number of individuals with $\sum_t y_{it} = c$, $N_{c,j\to k}$ is the same for individuals who switch from $D_{i0} = j$ to $D_{i1} = k$, and

$$S^t_{c,j\to k} = \frac{1}{N_{c,j\to k}} \sum_{\substack{i:c_i=c \\ (j\to k)}} \frac{y_{it}}{c_i} \tag{38}$$

is the average $y_{it}/\sum_t y_{it}$ among such individuals.

Suppose, as in a canonical difference-in-difference setting, that there are two types of individuals i: those who remain untreated in both periods ($D_{i0} = D_{i1} = 1$) and those who switch into treatment ($D_{i0} = 0$, $D_{i1} = 1$). Then (37) becomes

$$0 = \sum_{c=1}^{\infty} cN_c \begin{bmatrix} \frac{N_{c,0\to1}}{N_c} \left(S^1_{c,0\to1} - \frac{\exp(\tau^*+\gamma^*)}{1+\exp(\tau^*+\gamma^*)} \right) \\ \frac{N_{c,0\to0}}{N_c} \left(S^1_{c,0\to0} - \frac{\exp(\tau^*)}{1+\exp(\tau^*)} \right) + \frac{N_{c,0\to1}}{N_c} \left(S^1_{c,0\to1} - \frac{\exp(\tau^*+\gamma^*)}{1+\exp(\tau^*+\gamma^*)} \right) \end{bmatrix}, \tag{39}$$

the solution to which can be written

$$\tau^* = \ln \left(\frac{\sum_{i:0\to0} y_{i1}}{\sum_{i:0\to0} y_{i0}} \right) \tag{40}$$

$$\beta^* = \ln \left(\left(\frac{\sum_{i:0\to1} y_{i1}}{\sum_{i:0\to1} y_{i0}} \right) / \left(\frac{\sum_{i:0\to0} y_{i1}}{\sum_{i:0\to0} y_{i0}} \right) \right). \tag{41}$$

The Poisson treatment coefficient thus thus asymptotically sets

$$\beta = \ln\left(\frac{E[y_{i1} \mid D_{i1}=1]}{E[y_{i0} \mid D_{i1}=1]}\right) - \ln\left(\frac{E[y_{i1} \mid D_{i1}=0]}{E[y_{i0} \mid D_{i1}=0]}\right),$$

the difference in log growth rates among those treated and untreated in period 1.

Letting $y_{it}(0)$ and $y_{it}(1)$ denote untreated and treated potential outcomes of individual i in time t, we have under an assumption of common log growth rates,

$$\frac{E[y_{i1}(0) \mid D_{i1}=1]}{E[y_{i0}(0) \mid D_{i1}=1]} = \frac{E[y_{i1}(0) \mid D_{i1}=0]}{E[y_{i0}(0) \mid D_{i1}=0]}, \tag{42}$$

that

$$\beta = \ln\left(\frac{E[y_{i1}(1) \mid D_{i1}=1]}{E[y_{i1}(0) \mid D_{i1}=1]}\right). \tag{43}$$

This shows that in the simple difference-in-difference setting the Poisson fixed effect regression identifies the log percentage effect of treatment on the treated.

A similar result holds for the simplest mover design, in which individuals either switch out of or into treatment in $t \in \{0,1\}$. Specifically, it can be shown that the Poisson treatment coefficient asymptotically sets

$$\beta = \frac{1}{2}\left(\ln\left(\frac{E[y_{i1} \mid D_{i1}=1]}{E[y_{i0} \mid D_{i1}=1]}\right) - \ln\left(\frac{E[y_{i1} \mid D_{i0}=1]}{E[y_{i0} \mid D_{i0}=1]}\right)\right). \tag{44}$$

Thus, under the same common log growth rate assumption

$$\begin{aligned}\beta &= \frac{1}{2}\left(\ln\left(\frac{E[y_{i1}(1) \mid D_{i1}=1]}{E[y_{i0}(0) \mid D_{i1}=1]}\right) - \ln\left(\frac{E[y_{i1}(0) \mid D_{i0}=1]}{E[y_{i0}(1) \mid D_{i0}=1]}\right)\right) \\ &= \frac{1}{2}\left(\ln\left(\frac{E[y_{i1}(1) \mid D_{i1}=1]}{E[y_{i1}(0) \mid D_{i1}=1]}\right) + \ln\left(\frac{E[y_{i0}(1) \mid D_{i0}=1]}{E[y_{i0}(0) \mid D_{i0}=1]}\right)\right),\end{aligned} \tag{45}$$

the average log percentage treatment-on-the-treated effect across the two time periods.

Event Study Decomposition

This appendix shows how the estimates from our encounter utilization model (22) can be represented graphically in an event study framework. Following the logic in Section 3.4.1, note that we can rewrite the patient-level utilization model (28) for patient movers as

$$y_{it} = \tilde{\alpha}_i + \tau_t + (\mathbf{1}[r(i,t) > 0]S_i)\Delta_i + x_{it}'\beta + \ln N_{it} + \overline{\delta}_{it} + \sigma_{it} + \nu_{it}, \tag{46}$$

where $\tilde{\alpha}_i = \alpha_i + \gamma_{o(i)}$ and here $S_i = (\gamma_{d(i)} - \gamma_{o(i)})/\Delta_i$ denotes the share of the observed difference in utilization between a mover's destination and origin HRRs due to the place effects γ_j. This S_i differs from the share defined in Section 3.4.1 by isolating the other (non-physician) regional supply-side drivers of utilization. We do not assume that the physician component $\varepsilon_{it} = \ln N_{it} + \overline{\delta}_{it} + \sigma_{it} + \nu_{it}$ is additively separable in place and patient effects, as in Section 3.4.1.

Equation (46) suggests an enriched event study decomposition of the different drivers of geographic variation in healthcare utilization. Note that $\ln N_{it}$ is directly observed and that $\overline{\delta}_{it}$ and σ_{it} may be estimated by first-step estimates of the physician-level parameters δ_d and θ and an assumption on the distribution of residual utilization variation ε_{idt} (for example, that ε_{idt} is *iid* given (x, w, j). Then, by subtracting estimates of $\ln N_{it} + \overline{\delta}_{it} + \sigma_{it}$ from observed utilization y_{it}, we obtain a model like Equation (32). This suggests that a patient-level event study using this adjusted $\tilde{y}_{it} \equiv y_{it} - (\ln N_{it} + \overline{\delta}_{it} + \sigma_{it})$ as an outcome may capture a weighted average of non-physician place effect shares S_i, by taking into account the predicted change in physician availability, sorting, and encounter quantity in ε_{it}. The difference between this adjusted event study of $\tilde{y}_{it}$ and the motivating patient-level event study in Figure 3.3 can furthermore be evaluated by performing event studies on each of the subtracted components $\ln N_{it}$, $\overline{\delta}_{it}$, and σ_{it}. These auxiliary event studies replace y_{it} in Equation (32) with each of these components, with the sum of event study coefficients θ_r across specifications equaling, by construction, the difference between the adjusted and original event study jumps. A large event study jump in a regression of $\overline{\delta}_{it}$ would, for example, suggest that a sizable proportion of the aggregate place effect in Figure 3.3 is due to differences in the availability of physicians with different utilization effects across different HRRs. Similarly, event study jumps in $\ln N_{it}$ or σ_{it} would suggest that some of the aggregate place effect in Figure 3.3 arises from systematic differences in the number of encounters or matching of patients to physicians across HRRs.

We use the parameters from our encounter-level estimation to perform this decomposition. We first form simulation-based estimates of average physician utilization and selection terms $\overline{\delta}_{it}$ and σ_{it}.[65] We then use these estimates as components in estimating the enriched event study decomposition (46).[66]

The results from this event study decomposition are presented in Appendix Figure C2. In Panel A, we reproduce the patient-level event study using adjusted annual spending $y_{it} - (\ln N_{it} + \overline{\delta}_{it} + \sigma_{it})$ as the outcome. As discussed above, the height of the resulting event study jump estimates a weighted average of the share of place effects in average log utilization differences, net of any potential contribution of physicians to geographic spending differences. The figure shows that netting out the contribution of physicians to annual spending differences alters the event study substantially. While the unadjusted event study from regression (32) shows that the place share including physician utilization differences is around 50%, subtracting those differences decreases the estimated place share to around 10%. The figure indicates that non-physician factors account for around 20% of total supply-side effects of places on patient spending, or 10% of overall differences in geographic variation.

Panels B-C of Appendix Figure C2 decompose the difference between the simple patient-level event study in Figure 3.3 and the enriched analysis in Panel A of Appendix Figure C2. Around half of the remaining overall "place effect" is due to an effect on the increased number of physicians seen (Panel B), with half due to increased practice intensity of the available physician stock (Panel C). The event study on the residual selection term σ_{it} is flat, suggesting none of the variation in utilization across HRRs is due to the differential sorting of patients to physicians with different

[65] Specifically, for each patient and year, we take a random draw of physicians from her HRR with the number of physicians equalling her actual number of encounters for this patient and year. We use these simulated encounters averaged over 100 random draws to form estimates of $\overline{\delta}_{it}$ and σ_{it}.

[66] For the 6% of patient-years that have no utilization, we set $\overline{\delta}_{it}$ and σ_{it} equal to 0 since there is no actual patient-physician encounter.

demand and practice styles. Overall, this analysis suggests around 20% of the geographic variation in utilization is due to differences in physician practice style, around 20% is due to differences in the number of physicians seen, and around 10% is due to differences in residual supply-side factors. The remaining 50% appears due to differences in patient demand, as first found by Finkelstein et al. (2016).

It is also worth emphasizing how our enriched analytic framework can reconcile the earlier patient- and physician-level event studies. While Panel A of Figure 3.3 indicates that the share of places in geographic utilization is important, a superficial read of Panel B may cause one to believe that physician practice styles, which account for only 20% of geographic utilization differences for *physicians,* may not be an important driver of the place share for patients. As discussed above, there are many reasons why drawing such a conclusion may be unwarranted, and Panel A of Figure C2 shows precisely and strikingly that such a conclusion is in fact misleading.

Finally, we note that the shape of the pre- and post-trends in all of the figures are consistent with our identifying assumptions and our interpretation of the driving forces behind the observed effects. Our model suggests that utilization should be flat before and after a move and that it should jump discontinuously upon a move. It is clear that in both Figure 3.3 and Figure C2, this is indeed the case. The flat pre-trends everywhere are consistent with our assumption of the exogeneity of moves with respect to unobserved determinants of spending. Additionally, as in Finkelstein et al. (2016), the flat post-trends and the symmetry between positive and negative moves of the same size documented in Figure C1 speak against patient habit-formation as the source of some of the cross-regional differences we observe (Becker and Murphy, 1988).[67]

Spending Variation Counterfactuals

This appendix formally defines our counterfactual analysis of how differences in average log annual patient utilization in each HRR, as represented in equation (29), change as we equalize the various underlying sources of spending differences. This analysis proceeds in five incremental steps: first by shutting down patient-physician selection and then exploring sequentially the effect of eliminating variation due to physicians, patients, and places.

Our first counterfactual sets $\sigma_j = 0$:

$$\bar{y}_j^{(1)} = p_j(\bar{\alpha}_j + \gamma_j + \overline{N}_j + \bar{\delta}_j) \tag{47}$$

The regional distribution of this quantity captures how the geographic distribution of healthcare utilization would change if there were no systematic differences in the allocation of patient encounters to physicians with different practice styles, holding fixed the number of physicians each patient sees and the set of patients and physicians in each region. We then remove regional variation in average physician intensity by setting $\bar{\delta}_j$ to its average value across HRRs $\bar{\delta}$:

$$\bar{y}_j^{(2)} = p_j(\bar{\alpha}_j + \gamma_j + \overline{N}_j + \bar{\delta}) \tag{48}$$

[67] First, Figure C1 shows that for any given $|\hat{\delta}_i|$, the change in log utilization looks symmetric for moves up and down, which would be inconsistent with a story where aggressive treatments started aggressive areas continue after a move since this would introduce an asymmetry between high- and low-utilization areas. Second, habit formation would predict a gradual convergence to the region average after a move. In our setting, the change in utilization after a move is sharp and discontinuous and utilization remains relatively flat for many years after the move was completed.

The regional distribution of this quantity captures how the geographic distribution of healthcare utilization would change if there were no systematic differences in the average practice styles of physicians across regions.

The next two counterfactuals eliminate regional variation due to patients:

$$\bar{y}_j^{(3)} = p_j(\bar{\alpha} + \gamma_j + \overline{N}_j + \bar{\delta}) \tag{49}$$

$$\bar{y}_j^{(4)} = \tilde{p}_j(\bar{\alpha} + \gamma_j + \tilde{N}_j + \bar{\delta}). \tag{50}$$

Here $\bar{y}^{(3)}$ sets $\bar{\alpha}_j$ to its average $\bar{\alpha}$, thus eliminating regional variation due to patient effects on utilization per encounter. $\bar{y}_j^{(4)}$ leverages the encounter quantity model (23) to eliminate regional variation coming from patient effects on the number of encounters. Specifically we define, in contrast to p_j,

$$\tilde{p}_j = 1 - E[\exp(-\alpha_i^n - \gamma_j^n - x_{it}'\beta^n)], \tag{51}$$

the share of individuals with any healthcare utilization in region j, given a random geographic reallocation of patients. This is a known function of the extensive margin model parameters α_i^n, γ_j^n, and β^n. We similarly define $\tilde{N}_j$ as the average log number of physicians seen in region j, when non-zero, under random patient reallocation.

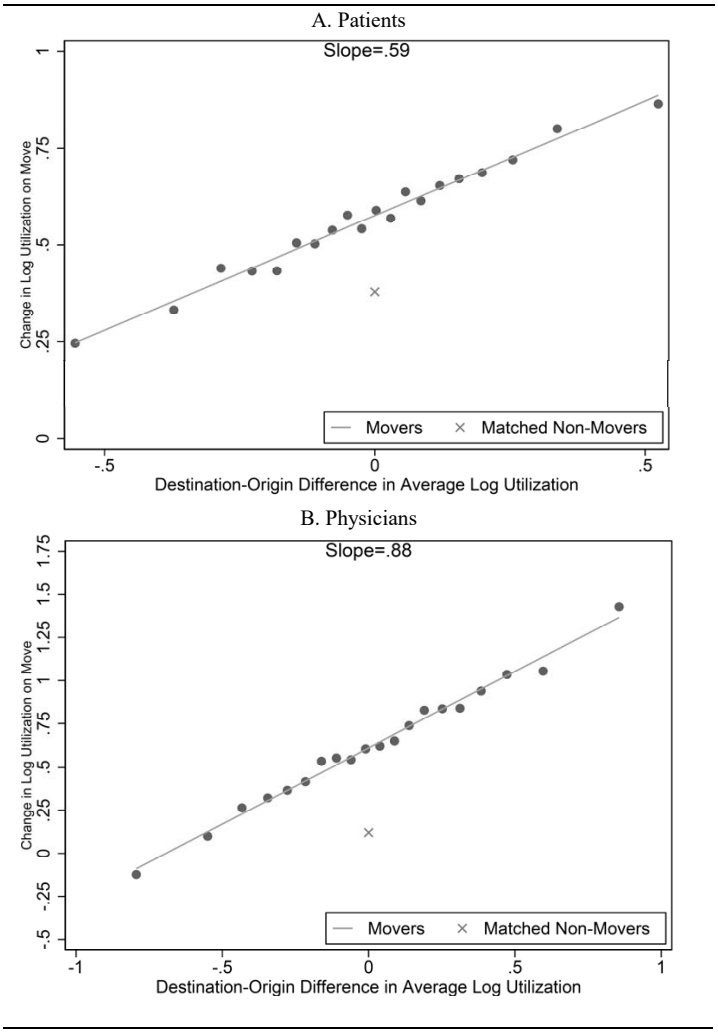
Following these four counterfactuals, the only regional variation left is that due to place effects on utilization per encounter and number of encounters. We separate these two factors through our final counterfactual which sets γ_j to its regional average $\bar{\gamma}$:

$$\bar{y}_j^{(5)} = \tilde{p}_j(\bar{\alpha} + \bar{\gamma} + \tilde{N}_j + \bar{\delta})$$

The only regional variation in $\bar{y}_j^{(5)}$ is due to place effects on the number of physicians seen, $\tilde{p}_j$ and $\tilde{N}_j$. Taken together, Y_j and $Y_j^{(1)} - Y_j^{(5)}$ thus provide a full accounting of the partial equilibrium role that each of the primary factors of interest (physicians, places, and patients) play in the geographic variation in utilization.

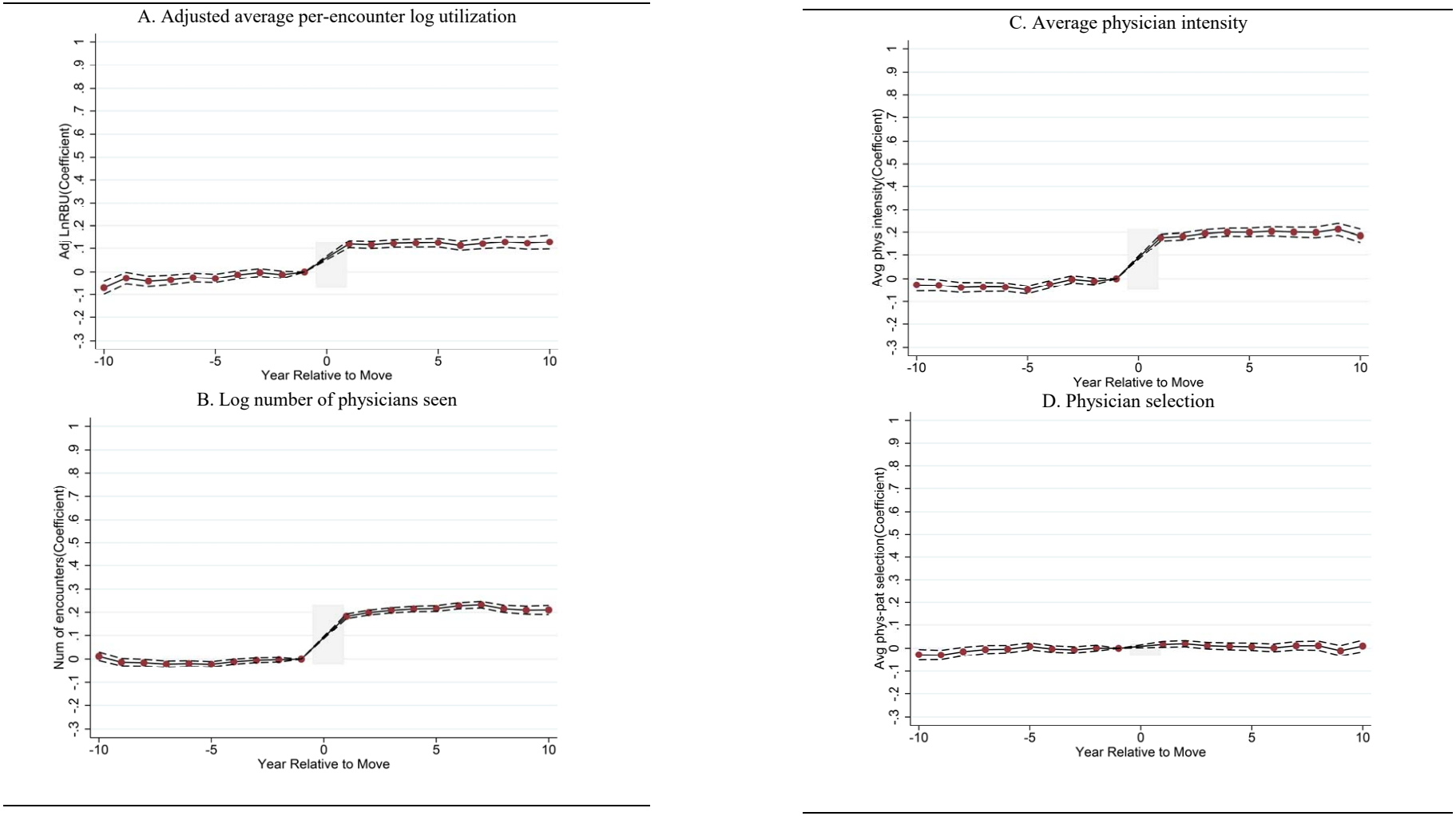
C.3 Appendix Figures and Tables

Figure C1: Change in Log Utilization by Size of Move



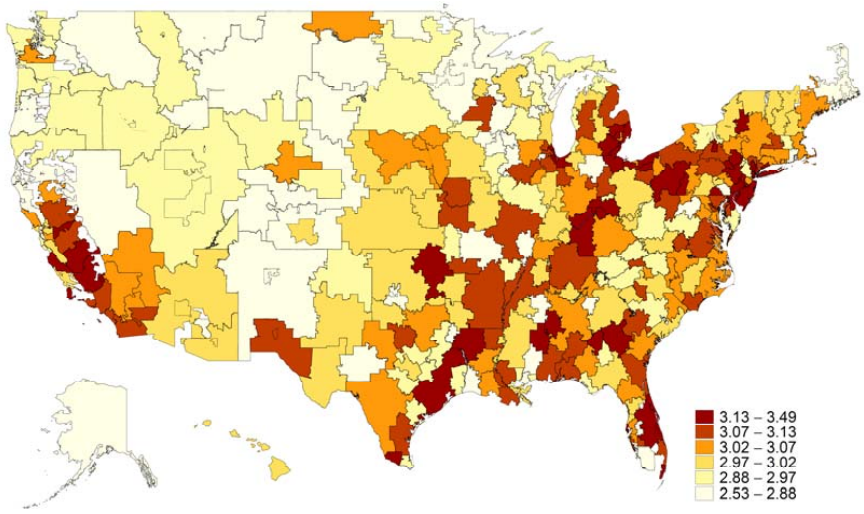
Notes: This figure shows the change in log utilization before and after move for patient (Panel A) and physician (Panel B) movers. For each mover, we calculate the difference in average log utilization between their origin and destination HRRs, then group these differences into ventiles. The x-axis displays the mean of difference for movers in each ventile. The y-axis shows, for each ventile, average log utilization two to five years post-move minus average log utilization two to five years pre-move. The line of best fit is obtained from simple OLS regression using the 20 data points corresponding to movers, and its slope is reported on the graph. The sample is all mover years between two and five years pre-move and between two and five years post-move (N = 2,704,487 patient-years in Panel A and N = 479,057 physician-years in Panel B). For comparison, we also compute the average change in log utilization for a sample of matched non-movers, which we show with the "x" marker on each graph. Specifically, for each patient or physician mover in our data in each calendar year, we randomly draw a non-mover in the same year in the mover's origin HRR (for patients, we also require that the non-mover share the mover's gender, race, and five-year age bin; for physicians, we draw a random non-mover physician in the mover's origin HRR); the union of the selected non-mover patient-years (Panel A) or physician-years (Panel B) forms the matched sample.

Figure C2: Mover Event Study Decomposition



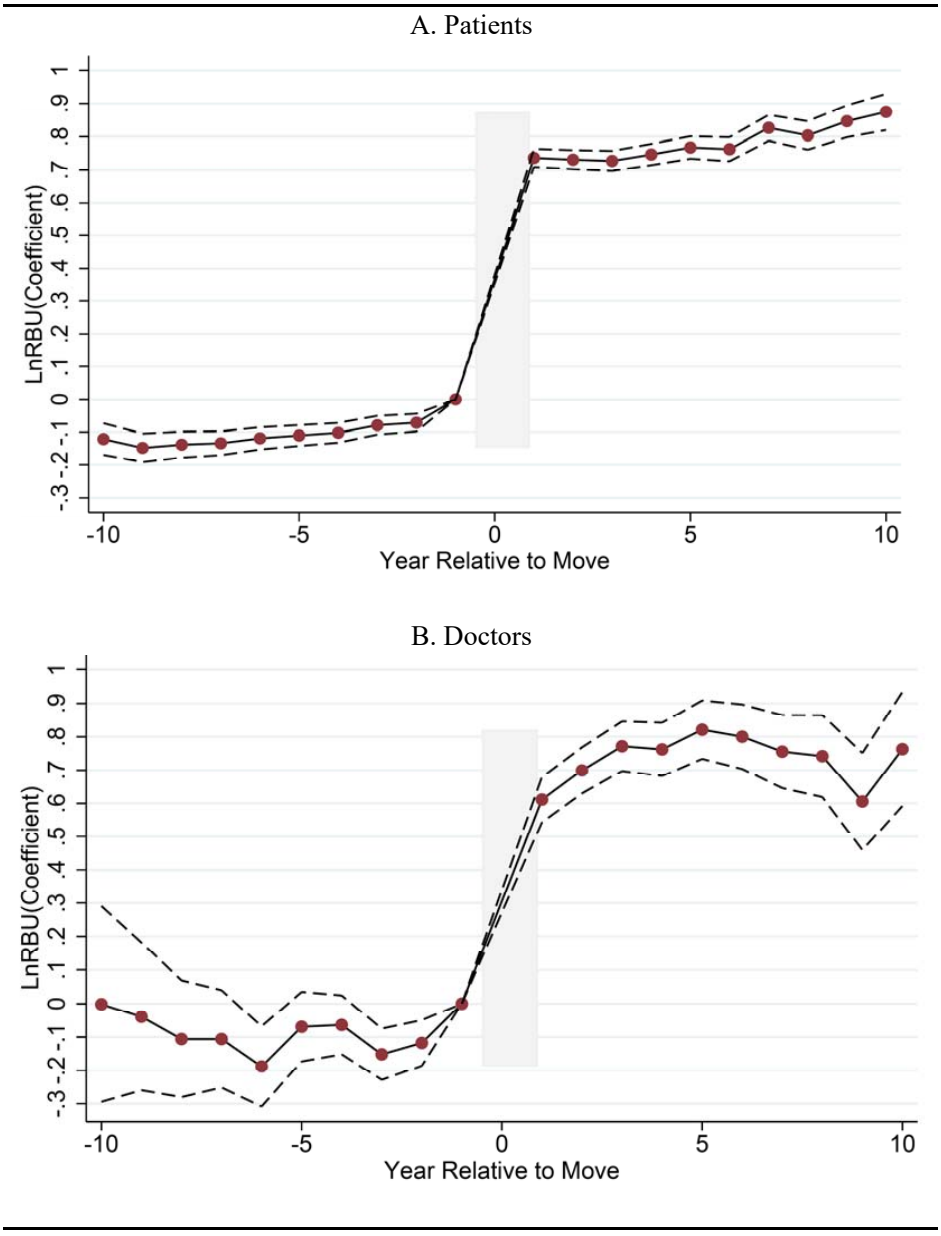
Notes: This figure shows the coefficients θ_r estimated from equation (46) for patient movers. The coefficient for relative year -1 is normalized to 0. The dependent variable is adjusted annual spending $y_{it} - (\ln N_{it} + \overline{\delta}_{it} + \sigma_{it})$ in Panel A, number of physicians seen $\log n_{it}$ in Panel B, average physician intensity $\overline{\delta}_{it}$ in Panel C, and average physician-patient selection σ_{it} in Panel D; xit consists of indicator variables for five-year age bins (Panel A) and relative-year effects (Panel A and Panel B). The dashed lines are upper and lower bounds of the 95 percent confidence interval. The sample is all patient-years (N = 23,678,685 patient-years).

Figure C3: Distribution of Physician Fixed Effects Across HRRs



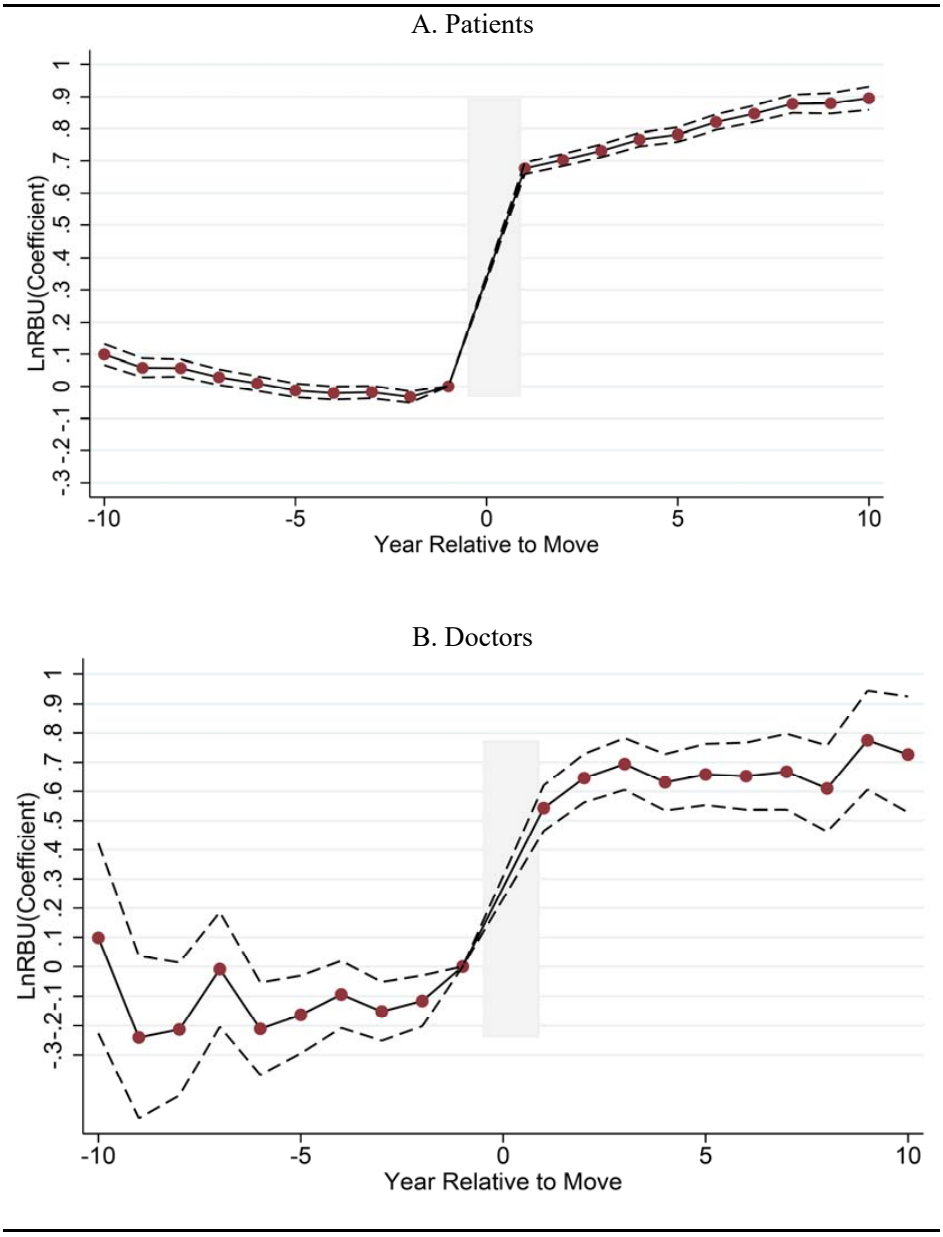
Notes: The map shows the distribution of average physician fixed effects by HRR, in quintiles defined in the legend. Physician effects are estimated from our encounter-level model. The sample is the baseline sample of all encounters (N = 144 million encounters).

Figure C4: Mover Event Study (PCPs)



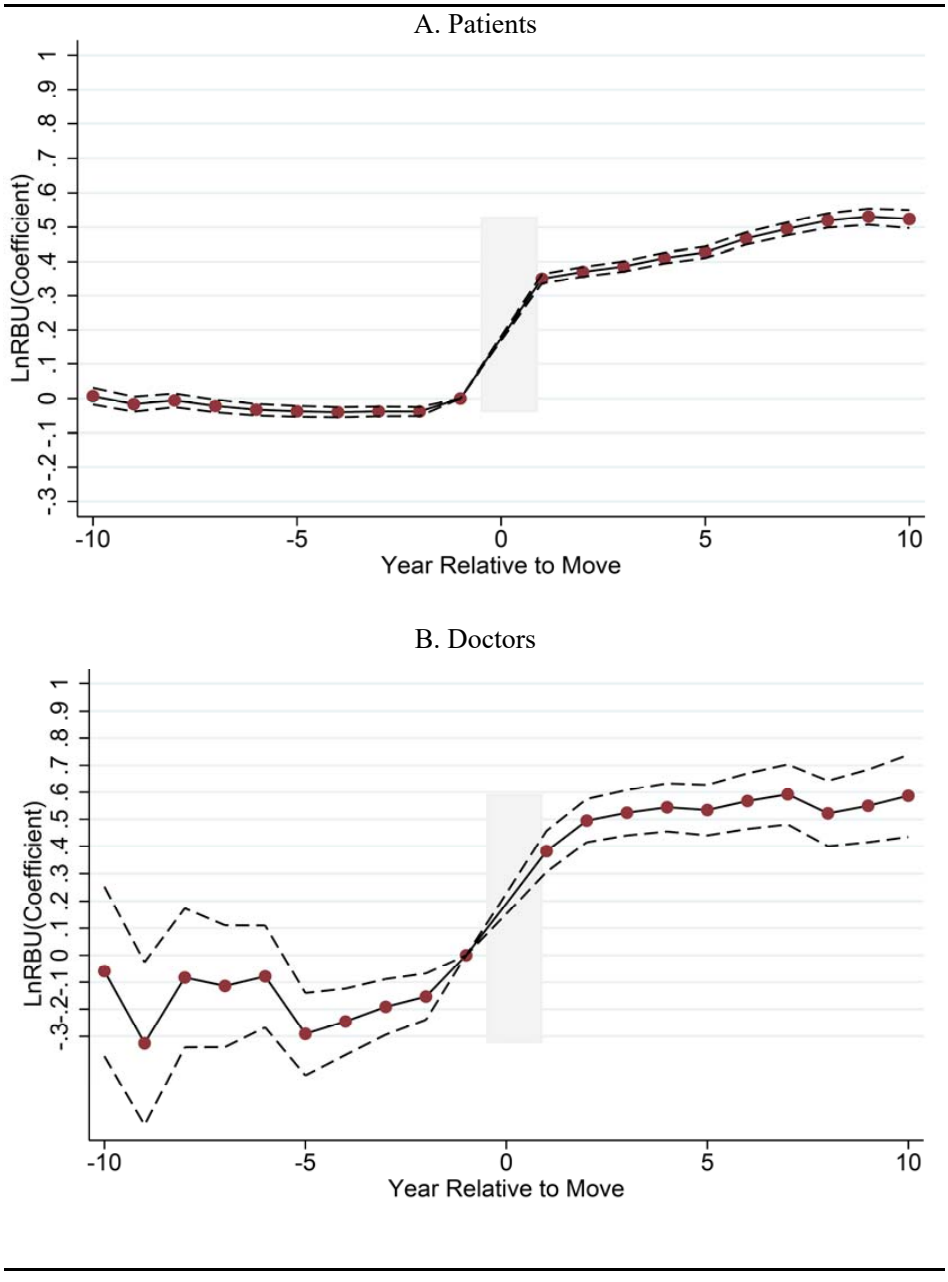
Notes: This figure shows the estimated θ_r coefficients in Equation (32) for patient movers (Panel A) and in Equation (33) for physician movers (Panel B). The coefficient for relative year -1 is normalized to 0. The dependent variable in Panel A is normalized log annual patient spending and the control vector includes indicator variables for five-year age bins and relative-year effects for movers. The dependent variable in Panel B is normalized log annual physician spending and the control vector includes relative-year effects for movers. Dashed lines indicate upper and lower bounds of the 95 percent confidence interval for each θ_r estimate. The sample is the PCP specialty subsample.

Figure C5: Mover Event Study (Cardiologists)



Notes: This figure shows the estimated θ_r coefficients in Equation (32) for patient movers (Panel A) and in Equation (33) for physician movers (Panel B). The coefficient for relative year -1 is normalized to 0. The dependent variable in Panel A is normalized log annual patient spending and the control vector includes indicator variables for five-year age bins and relative-year effects for movers. The dependent variable in Panel B is normalized log annual physician spending and the control vector includes relative-year effects for movers. Dashed lines indicate upper and lower bounds of the 95 percent confidence interval for each θ_r estimate. The sample is the cardiology specialty subsample.

Figure C6: Mover Event Study (Dermatologists)



Notes: This figure shows the estimated θ_r coefficients in Equation (32) for patient movers (Panel A) and in Equation (33) for physician movers (Panel B). The coefficient for relative year -1 is normalized to 0. The dependent variable in Panel A is normalized log annual patient spending and the control vector includes indicator variables for five-year age bins and relative-year effects for movers. The dependent variable in Panel B is normalized log annual physician spending and the control vector includes relative-year effects for movers. Dashed lines indicate upper and lower bounds of the 95 percent confidence interval for each θ_r estimate. The sample is the dermatology specialty subsample.

D Bibliography

References

Chapter 25 - completing and processing the formcms-1450 data set. *Medicare Claims Processing Manual*.

Chapter 26 - completing and processingform cms-1500 data set. *Medicare Claims Processing Manual*.

(2014). Health and retirement study. Public Use Datasets Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).

Aaronson, D. (1998). Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources*, 33(4):915–946.

Abigail Hess (2019). The 15 best medical schools in the us, according to us news & world report. CNBC, `https://www.cnbc.com/2019/03/12/us-news--world-report-the-15-best-medical-schools-in-the-us.html`.

Abowd, J. M., Creecy, R. H., and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, U.S. Census Bureau Center for Economic Studies.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Acemoglu, D. and Finkelstein, A. (2008). Input and technology choices in regulated industries: Evidence from the health care sector. *Journal of Political Economy*, 116(5):837–880.

Acemoglu, D. and Linn, J. (2004). Market size in innovation: Theory and evidence from the pharmaceutical industry. *The Quarterly Journal of Economics, 119(3): 1049-1090*.

Adler, J. and Chen, H. (2010). Is endocrine surgery research dying? *American Journal of Translational Research, 2(2):129-134*.

Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2009). The age of reason: Financial decisions over the life-cycle with implications for regulation. *Brookings Papers on Economic Activity 2: 51-117*.

Agency for Healthcare Research and Quality, Rockville, MD (1994-2011). Healthcare cost and utilization project (hcup) nationwide inpatient sample (nis). `www.hcup-us.ahrq.gov/nisoverview.jsp`.

Agency for Healthcare Research and Quality, Rockville, MD (2009-2014). Healthcare cost and utilization project (hcup) state inpatient databases (sid). www.hcup-us.ahrq.gov/sidoverview.jsp.

Agency for Healthcare Research and Quality, Rockville, MD (2012-2013). Healthcare cost and utilization project (hcup) national inpatient sample (nis). www.hcup-us.ahrq.gov/nisoverview.jsp.

Agency for Healthcare Research and Quality, Rockville, MD (2020a). Healthcare cost and utilization project (hcup) clinical classifications software (ccs) for icd-9-cm. https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.

Agency for Healthcare Research and Quality, Rockville, MD (2020b). Healthcare cost and utilization project (hcup) surgery flags software for services and procedures. https://www.hcup-us.ahrq.gov/toolssoftware/surgflags/surgeryflags.jsp.

Agha, L., Ericson, K. M. M., Geissler, K., and Rebitzer, J. (2018). Team formation and performance: Evidence from healthcare referral networks. *NBER Working Paper No. w24338*.

Agha, L., Frandsen, B., and Rebitzer, J. B. (2019). Fragmented division of labor and healthcare costs: Evidence from moves across regions. *Journal of Public Economics*, 169:144–159.

Agha, L., Kim, S., and Li, D. (2020). Insurance design and pharmaceutical innovation. *NBER Working Paper 27563*.

Agha, L. and Molitor, D. (2018). The local influence of pioneer investigators on technology adoption: Evidence from new cancer drugs. *Review of Economics and Statistics, 100(1): 29-44*.

American College of Surgeons (2012). Ventral hernia repair. www.facs.org/patienteducation.

American Medical Association (2020). Ama physician masterfile. https://www.ama-assn.org/practice-management/masterfile/ama-physician-masterfile.

Angrist, J. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics*. Princeton University Press.

Anthony, D. L., Herndon, M. B., Gallagher, P. M., Barnato, A. E., Bynum, J. P. W., Gottlieb, D. J., Fisher, E. S., and Skinner, J. S. (2009). How much do patients' preferences contribute to resource use? *Health Affairs*, 28(3):864–873.

Arora, S., Panaich, S. S., Patel, N., Patel, N. J., Savani, C., Patel, S. V., Thakkar, B., Sonani, R., Jhamnani, S., Singh, V., Lahewala, S., Patel, A., Bhatt, P., Shah, H., Jaiswal, R., Gupta, V., Deshmukh, A., Kondur, A., Schreiber, T., Badheka, A. O., and Grines, C. (2016). Coronary atherectomy in the united states (from a nationwide inpatient sample). *The American Journal of Cardiology 117(4): 555-562*.

Artis, L., Burkhart, T., Johnson, T., and Matuszewski, K. (2006). Physician factors as an indicator of technological device adoption. *J Med Syst, 30:177-186*.

Ashton, C. M., Peterson, N. J., Souchek, J., Menke, T. J., Yu, H.-J., Pietz, K., Eigenbrodt, M. L., Barbour, G., Kizer, K., and Wray, N. P. (1999). Geographic variations in utilization rates in veterans affairs hospitals and clinics. *New England Journal of Medicine*, 340(1):32–39.

Azoulay, P., Jones, B. F., Kim, J. D., and Miranda, J. (forthcoming). Age and high-growth entrepreneurship. *American Economic Review: Insights*.

Baicker, K. and Chandra, A. (2004). The productivity of physician specialization: Evidence from the medicare program. *American Economic Review,94 (2): 357-361*.

Baker, L. C., Bundorf, M. K., and Kessler, D. P. (2014). Patients' preferences explain a small but significant share of regional variation in medicare spending. *Health Affairs*, 33(6):957–963.

Baker, L. C., Fisher, E. S., and Wennberg, J. E. (2008). Variations in hospital resource use for medicare and privately insured populations in california. *Health Affairs*, 27(2):w123–134.

Baker, L. C. and Phibbs, C. S. (2002). Managed care, technology adoption, and health care: The adoption of neonatal intensive care. *RAND Journal of Economics, 33 (3): 524-548*.

Bansal, V. K., Krishna, A., Misra, M. C., and Kumar, S. (2016). Learning curve in laparoscopic inguinal hernia repair: Experience at a tertiary care centre. *Indian J Surg. 78(3): 197-202*.

Barnato, A. R., Herndon, M. B., Anthony, D. L., Gallagher, P. M., Skinner, J. S., Bynum, J. P. W., and Fisher, E. S. (2007). Are regional variation in end-of-life care intensity explained by patient preferences? a study of the us medicare population. *Medical Care*, 45(5):386–393.

Beckelis, K., Skinner, J., Gottlieb, D., and Goodney, P. (2017). De-adoption and exnovation in the use of carotid revascularization: retrospective cohort study. *BMJ 2017;359:j4695*.

Becker, G. S. and Murphy, K. M. (1988). A theory of rational addiction. *Journal of Political Economy*, 96(4):675–700.

Benabou, R. and Tirole, J. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies, 70: 489-520*.

Bennett, D., Boyle, P., Buchman, A., Laibson, D., and Yu, L. (2011). Cognitive function is associated with risk aversion in community-based older persons. *BMC Geriatrics 11(53)*.

BioPortal (2020). Anatomical therapeutic chemical classification. `https://bioportal.bioontology.org/ontologies/ATC`.

Blackburn, M. L. (2007). Estimating wage differentials without logarithms. *Labour Economics 14(1): 73-98*.

Blume-Kohout, M. and Sood, N. (2013). Market size and innovation: Effects of medicare part d on pharmaceutical research and development. *Journal of Public Economics, 97: 327-336*.

Bourke, J. and Roper, S. (2012). In with the new: the determinants of prescribing innovation by general practitioners in ireland. *Eur J Health Econ 13:393-407*.

Bronnenberg, B. J., Dube, J.-P. H., and Gentzkow, M. (2012). The evolution of brand preferences: Evidence from consumer migration. *American Economic Review*, 102(6):2472–2508.

Bunker, J., Hinkley, D., and McDermott, W. (1978). Surgical innovation and its evaluation. *Science, 200: 937-941*.

Callison, K., Kaestner, R., and Ward, J. (2018). A test of supply-side explanations of geographic variation in health care utilization. *NBER Working Paper No. 25037*.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*.

Card, D., Heining, J., and Kline, P. M. (2013). Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics*, 128(3):967–1015.

Centers for Disease Control and Prevention (1994-1996,2006). National survey of ambulatory surgery. https://www.cdc.gov/nchs/nsas/index.htm.

Centers for Disease Control and Prevention (2020). Icd-9-cm addenda, conversion table, and guidelines. https://www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm.

Centers for Medicare & Medicaid Services (CMS) (2007). Agenda. ICD-9-CM Coordination and Maintenance Committee. ICD-9-CM Volume 3, Procedures September 27-September 28, 2007.

Centers for Medicare & Medicaid Services (CMS) (2009). Agenda. ICD-9-CM Coordination and Maintenance Committee. ICD-9-CM Volume 3, Procedures September 16-September 17, 2009.

Centers for Medicare & Medicaid Services (CMS) (2020). Icd-10 coordination and maintenance committee. https://www.cms.gov/Medicare/Coding/ICD9ProviderDiagnosticCodes/meetings#:~:text=Coordination%20and%20Maintenance%20Committee%20Meetings,and%20is%20led%20by%20CDC.

Centers for Medicare and Medicaid Services (2014). Medicare claims processing manual. Centers for Medicare and Medicaid Services.

Chan, D., Gentzkow, M., and Yu, C. (2020). Selection with variation in diagnostic skill: Evidence from radiologists. *NBER Working Paper No. 26467*.

Chan, J., Shalhoub, J., Gardiner, M. D., Suleman-Verjee, L., and Nanchahal, J. (2013). Strategies to secure surgical research funding: fellowships and grants. *Journal of the Royal Society of Medicine Open, 5(1): 1-5*.

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

Chandra, A., Sabik, L., and Skinner, J. S. (2009). Cost growth in Medicare: 1992 to 2006. In Wise, D. A., editor, *Explorations in the Economics of Aging*, pages 133–157. University of Chicago Press.

Chandra, A. and Skinner, J. (2012). Technology growth and expenditure growth in health care. *Journal of Economic Literature, 50 (3): 645-680*.

Chandra, A. 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.

Chang, S. W. and Luft, H. S. (1991). *The Changing Economics of Medical Technology*, chapter Reimbursement and the Dynamics of Surgical Procedure Innovation, pages 96–122. Washington, D.C.: National Academy Press.

Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, 40(5):373–383.

Chernew, M., Sabik, L., Chandra, A., Gibson, T., and Newhouse, J. (2010). Geographic correlation between large-firm commercial spending and Medicare spending. *American Journal of Managed Care*, 16(2):131–138.

Chetty, R., Friedman, J. N., Leth-Petersen, S., Nielsen, T., and Olsen, T. (2014a). Active vs. passive decision and crowd-out in retirement savings accounts: Evidence from Denmark. *Quarterly Journal of Economics*, 129(3):1141–1219.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679.

Chetty, R., Friedman, J. N., and Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. *American Economic Review*, 103(7):2683–2721.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *Quarterly Journal of Economics*.

Choi, N. G. (1996). Older persons who move: Reasons and health consequences. *Journal of Applied Gerontology*, 15(3):325–344.

Clemens, J. (2013). The effect of u.s. health insurance expansions on medical innovation. *NBER Working Paper 19761*.

Clemens, J. and Gottlieb, J. (2014). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review*, 104(4):1320–1349.

Clemens, J. and Rogers, P. (2020). Demand shocks, procurement policies, and the nature of medical innovation: Evidence from wartime prosthetic device patents. *NBER Working Paper 26679*.

Cohen, W., Nelson, R., and Walsh, J. (2003). Protecting their intellectual assets: Appropriability conditions and why u.s. manufacturing firms patent (or not). *NBER Working Paper 7552*.

Colavita, P., Tsirline, V., Walters, A., Lincourt, A., Belyansky, I., and Heniford, B. T. (2013). Laparoscopic versus open hernia repair: outcomes and sociodemographic utilization results from the nationwide inpatient sample. *Surgical Endoscopy*, 27(1):109–117.

Congressional Budget Office (2008). Geographic variation in health care spending. Washington: Government Printing Press.

Correia, S., Guimaraes, P., and Zylkin, T. (2013). ppmlhdfe: Fast poisson estimation with high-dimensional fixed effects. *arXiv:1903.01690v3 [econ.EM]*.

Crepon, B., Deniau, N., and Perez-Duarte, S. (2003). Wages, productivity and worker characteristics: a french perspective. *CREST-INSEE Working Paper No. 2003-04*.

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

Cutler, D. (2005). *Your Money or Your Life: Strong Medicine for America's Health Care System*. New York, NY: Oxford University Press.

Cutler, D., Deaton, A., and Lleras-Muney, A. (2006). The determinants of mortality. *Journal of Economic Perspectives, 20 (3): 97-120*.

Cutler, D. and Kadiyala, S. (2003). *Measuring the Gains from Medical Research: An Economic Approach*, chapter The Return to Biomedical Research: Treatment and Behavioral Effects, pages 110–162. London, UK: University of Chicago Press.

Cutler, D., Skinner, J., Stern, A. D., and Wennberg, D. (2019). Physician beliefs and patient preferences: A new look at regional variation in spending. *American Economic Journal: Economic Policy*, 11(1):191–221.

Deming, D. and Noray, K. (forthcoming). Earnings, dynamics, changing job skills, and stem careers. *Quarterly Journal of Economics*.

DeMonaco, H., Ali, A., and von Hippel, E. (2006). The major role of clinicians in the discovery of off-label drug therapies. *Pharmacotherapy, 26(3): 323-332*.

Dickstein, M. J. (2018). Efficient provision of experience goods: Evidence from antidepressant choice. *Working Papers 18-17, New York University, Leonard N. Stern School of Business, Department of Economics*.

DiMasi, J., Hansen, R., and Grabowski, H. (2003). The price of innovation: new estimates of drug development costs. *Journal of Health Economics 22: 151-185*.

Doyle, J. J. (2011). Returns to local-area health care spending: Evidence from health shocks to patients far from home. *American Economic Journal: Applied Economics*, 3(3):221–43.

Drugbank (2020). Drugbank database. `https://www.drugbank.ca/`.

Dubois, P., Mouzon, O. D., Scott-Morton, F., and Seabright, P. (2015). Market size and pharmaceutical innovation. *RAND Journal of Economics, 46(4): 844-871*.

Dunn, A., Shapiro, A. H., and Liebman, E. (2013). Geographic variation in commercial medical-care expenditures: A framework for decomposing price and utilization. *Journal of Health Economics*, 32(6):1153–1165.

Edwards, C. and Bailey, R. (2000). Laparoscopic hernia repair: the learning curve. *Surg Laparosc Endosc Percutan Tech. 10(3):149-53*.

Einav, L., Finkelstein, A., and Polyakova, M. (2018). Private provision of social insurance: Drug-specific price elasticities and cost sharing in medicare part d. *American Economic Journal: Economic Policy, 10(3): 122-153*.

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2013). Selection on moral hazard in health insurance. *American Economic Review*, 103(1):178–219.

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*.

Escarce, J. (1996). Externalities in hospitals and physician adoption of a new surgical technology: An exploratory analysis. *Journal of Health Economics 15 (1996): 715-734*.

Fadlon, I. and Parys, J. V. (2019). Primary care physician practice styles and patient care: Evidence from physician exits in medicare. *NBER WP 26269*.

Farrell, D., Jensen, E., Kocher, B., Lovegrove, N., Melhem, F., Mendonca, L., and Parish, B. (2008). Accounting for the cost of us health care: A new look at why americans spend more. McKinsey Global Institute.

Fernandez, R. and Fogli, A. (2006). Fertility: the role of culture and family experience. *Journal of the European Economic Association*, 4(2-3):552–561.

Feyer, J. D. (2008). *Population Aging, Human Capital Accumulation and Productivity Growth, Population and Development Review. A Supplement to Vol. 34, 2008, Population and Development Review*, chapter Aggregate evidence on the link between demographics and productivity, pages 78–99. New York: Population Council.

Finkelstein, A. (2004). Static and dynamic effects of health policy: Evidence from the vaccine industry. *Quarterly Journal of Economics, 119(2): 527-564*.

Finkelstein, A. (2007). The aggregate effects of health insurance: Evidence from the introduction of medicare. *The Quarterly Journal of Economics*, 122(1):1–37.

Finkelstein, A., Gentzkow, M., and Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *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: the content, quality, and accessibility of care. Part 1. *Annals of Internal Medicine*, 138(4):273–287.

Fisher, E. S., Wennberg, D. E., Stukel, T. A., Gottlieb, D. J., Lucas, F. L., and Pinder, E. L. (2003b). The implications of regional variations in medicare spending: the content, quality, and accessibility of care. Part 2. *Annals of Internal Medicine*, 138(4):288–299.

Fisher, E. S., Wennberg, J. E., Stukel, T. A., Skinner, J. S., Sharp, S. M., Freeman, J. L., and Gittelsohn, A. M. (2000). Associations among hospital capacity, utilization, and mortality of u.s. medicare beneficiaries, controlling for sociodemographic factors. *Health Services Research*, 34(6):1351–1362.

Forte, M., Virnig, B., Eberly, L., Swiontkowski, M., Feldman, R., Bhandari, M., and Kane, R. (2010). Provider factors associated with intramedullary nail use for intertrochanteric hip fractures. *J Bone Joint Surg Am. 92(5):1105-14*.

Frandsen, B., Joynt, K., Rebitzer, J., and Jha, A. (2015). Care fragmentation, quality, and costs among chronically ill patients. *American Journal of Managed Care 21(5):355-362*.

Frandsen, B., Powell, M., and Rebitzer, J. (2019). Sticking points: common-agency problems and contracting in the us healthcare system. *The RAND Journal of Economics 50(2): 251-285*.

Frosch, K. (2011). Workforce age and innovation: A literature survey. *International Journal of Management Reviews 13: 414-430*.

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

Garfield, R., Orgera, K., and Damico, A. (2020). The coverage gap: Uninsured poor adults in states that do not expand medicaid. *Kaiser Family Foundation Issue Brief. https://www.kff.org/medicaid/issue-brief/the-coverage-gap-uninsured-poor-adults-in-states-that-do-not-expand-medicaid/*.

Gawande, A. (2009). The cost conundrum: What a texas town can teach us about health care. *The New Yorker*.

Geruso, M. and Layton, T. (2015). Upcoding: Evidence from medicare on squishy risk adjustment. Working Paper 21222, NBER.

Glass, H. E. and Rosenthal, B. (2004). Demographics, practices, and prescribing characteristics of physicians who are early adopters of new drugs. *P&T, 29(11): 699-708*.

Goldman, D. and Lakdawalla, D. (2011). *Handbook of Health Economics, Volume Two*, chapter Intellectual Property, Information Technology, Biomedical Research, and Marketing of Patented Products, pages 763–823. Waltham, MA: North Holland, pp. 763-823.

Golf, N. (2017). History of heart transplantation: a hard and glorious journey. *Brazilian Journal of Cardiovascular Surgery, 32(5): 423-427*.

Gottlieb, D. J., Zhou, W., Song, Y., Andrews, K. G., Skinner, J. S., and Sutherland, J. M. (2010). Prices don't drive regional medicare spending variations. *Health Affairs*, 29(3):537–543.

Gowrisankaran, G., Ho, V., and Town, R. (2006). Causality, learning and forgetting in surgery. *Federal Trade Comission Draft*.

Gratwohl, A., Schwendener, A., Baldomero, H., Gratwohl, M., Apperley, J., Niederwieser, D., and Frauendorfer, K. (2010). Changes in the use of hematopoietic stem cell transplantation: a model for diffusion of medical technology. *Haematologica 95: 637-643*.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223–255.

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

Guimaraes, P. (2014). Poi2hdfe: Stata module to estimate a poisson regression with two high-dimensional fixed effects. Technical report, Statistical Software Components S457777, Boston College Department of Economics.

Guimaraes, P. and Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal*, 10(4):628–649.

Hansen, E. and Bozic, K. (2009). The impact of disruptive innovations in orthopaedics. *Clinical Orthopedics and Related Research, 467(10): 2512-2520*.

Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-r&d relationship. *Econometrica*, 52(4):909–938.

Heckman, J. and Snyder, J. (1977). Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *RAND Journal of Economics 28, S142-S189*.

Hellerstein, J., Neumark, D., and Troske, K. (1999). Wages, productivity, and worker characteristics: evidence from plant-level production functions and wage equations. *Journal of Labor Economics, 17, pp. 409-446*.

Henkel, J. and von Hippel, E. (2004). Welfare implications of user innovation. *The Journal of Technology Transfer, 30: 73-87*.

Henseke, G. and Tivig, T. (2008). *Labor Markets and Demographic Change*, chapter Demographic change and industry-specific innovation patterns in Germany, pages 122–136. Wiesbaden: VS Research.

Hippel, E. V. (1994). 'sticky information' and the locus of problem solving: Implications for innovation. *Management Science 40(4): 429-439*.

Ho, V. and Aloia, T. (2008). Hospital volume, surgeon volume, and patient costs for cancer surgery. *Medical Care. 46(7): 718-725*.

Hoisl, K. (2007). A closer look at inventive output-the role of age and career paths. *Munich School of Management Discussion Paper No. 2007-12, University of Munich*.

Hu, Y., Edwards, B. L., Brooks, K. D., Newhook, T., and Jr, C. L. S. (2015). Recent trends in national institutes of health funding for surgery: 2003 to 2013. *American Journal of Surgery, 209(6): 1083-1089*.

Huesch, M. (2009). Learning by doing, scale effects, or neither? cardiac surgeons after residency. *Health Services Research, 44(6): 1960-1982*.

Huskamp, H., O'Malley, A. J., Horvitz-Lennon, M., Taub, A. L., Berndt, E., and Donohue, J. (2013). How quickly do physicians adopt new drugs? the case of second-generation antipsychotics. *Psychiatr Serv. 64(4): 324-330*.

Ichino, A. and Maggi, G. (2000). Work environment and individual background: Explaining regional shirking differentials in a large Italian firm. *Quarterly Journal of Economics*, 115(3):1057–1090.

Iezzoni, L. I., Heeren, T., Foley, S. M., Daley, J., Hughes, J., and Coffman, G. A. (1994). Chronic conditions and risk of in-hospital death. *Health Services Research*, 29(4):435–460.

Jayachandran, S., Lleres-Muney, A., and Smith, K. (2010). Modern medicine and the 20th century decline in mortality: Evidence on the impact of sulfa drugs. *American Economic Journal: Applied Economics, 2 (2): 1675-1698*.

Johannesson, M. and Lundin, D. (2002). The impact of physician preferences on the prescription of new drugs. *SSE/EFI Working Paper No. 460*.

Johansen, M. and Richardson, C. (2016). Estimation of potential savings through therapeutic substitution. *JAMA Intern Med. doi:10.1001/jamainternmed.2016.1704*.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Review of Economic Studies 76: 283-317*.

Jones, B. F. (2010). Age and great invention. *The Review of Economics and Statistics 92(1): 1-14*.

Judge, L. (1997). Issues surrounding the patenting of medical procedures. *Santa Clara High Technology Law Journal, 13: 181-215*.

Kaiser Family Foundation (2020a). Medicaid and chip income eligibility limits for children as a percent of the federal poverty level. `https://www.kff.org/health-reform/state-indicator/medicaid-and-chip-income-eligibility-limits-for-children-as-a-percent-of-the-federal-poverty-level/`.

Kaiser Family Foundation (2020b). Medicaid income eligibility limits for adults as a percent of the federal poverty level. `https://www.kff.org/health-reform/state-indicator/medicaid-income-eligibility-limits-for-adults-as-a-percent-of-the-federal-poverty-level/`.

Kaiser Family Foundation (2020c). Medicaid income eligibility limits for parents, 2002-2020. `https://www.kff.org/medicaid/state-indicator/medicaid-income-eligibility-limits-for-parents/`.

Kelley, R. (2009). Where can $700 billion in waste be cut annually from the u.s. healthcare system? Ann Arbor: Thomson Reuters.

Koopman, B. O. (1950). Necessary and sufficient conditions for poisson's distribution. *Proceedings of the American Mathematical Society*, 1:813–823.

Korniotis, G. and Kumar, A. (2009). Do older investors make better investment decisions? *Review of Economics and Statistics*, 93(1):244–265.

Kozyrskyj, A., Raymond, C., and Racher, A. (2007). Characterizing early prescribers of newly marketed drugs in canada: a population-based study. *Eur J Clin Pharmacol 63:597-604*.

Laakmann, A. (2015). When should physicians be liable for innovation? *Cardozo Law Review, 36: 913-968*.

LeBlanc, K. and Ger, R. (2001). *Abdominal Wall Hernias*, chapter Evolution of Laparoscopic Hernia Repair, pages 33–36. Springer, New York, NY.

Lee, T. and Wilde, L. L. (1980). Market structure and innovation: A reformulation. *The Quarterly Journal of Economics, 94 (2): 429-436*.

Libicki, M. and Brahmakulam, I. (2004). The costs and benefits of moving to the icd-10 code sets. *RAND Science and Technology Technical Report*.

Lichtenberg, F. (2014). Has medical innovation reduced cancer mortality? *CESifo Economic Studies (60), 1: 135-177*.

Luttmer, E. F. P. and Singhal, M. (2011). Culture, context, and the taste for redistribution. *American Economic Journal: Economic Policy*, 3(1):157–179.

Mandelblatt, J., Faul, L. A., Luta, G., Makgoeng, S., Isaacs, C., Taylor, K., Sheppard, V., Tallarico, M., Barry, W., and Cohen, H. (2012). Patient and physician decision styles and breast cancer chemotherapy use in older women: Cancer and leukemia group b protocol 369901. *Journal of Clinical Oncology*, 30(21):2609–2614.

Mariani, M. and Romanelli, M. (2007). 'stacking' and 'picking' inventions: the patenting behaviour of european inventors. *Research Policy 36: 1128-1142*.

Mark, T., Dirani, R., Slade, E., and Russo, P. (2002). Access to new medications to treat schizophrenia. *Journal of BehavioralHealth Services & Research, 29(1): 15-29*.

Martin, A., Whittle, L., Heffler, S., Barron, M. C., Sisko, A., and Washington, B. (2007). Health spending by state of residence, 1991-2004. *Health Affairs*, 26(6):w651–663.

Mastroianni, A. (2006). Liability, regulation and policy in surgical innovation: The cutting edge of research and therapy. *Health Matrix (16): 351-442*.

McPherson, K., Strong, P., Epstein, A., and Jones, L. (1981). Regional variations in the use of common surgical procedures: within and between England and Wales, Canada and the United States of America. *Social Science & Medicine. Part A: Medical Sociology*, 15(3 Part 1):273–288.

Medicaid and CHIP Payment and Access Commission (MACPAC) (2020). Overview of the affordable care act and medicaid. https://www.macpac.gov/subtopic/overview-of-the-affordable-care-act-and-medicaid/.

Megellas, M. (2006). Medicare modernization: the new prescription drug benefit and redesigned part b and part c. *Baylor University Medical Center Proceedings, 19(1): 21-23*.

Meier, B. (2015). The new patent infringement liability exception for medical procedures; note. *Journal of Legislation, 23(2): 265-281*.

Melly, L., Torregrossa, G., Lee, T., Jansens, J.-L., and Puskas, J. D. (2018). Fifty years of coronary artery bypass grafting. *Journal of Thoracic Disease, 10(3): 1960-1967*.

Meyer, J. (2009). Workforce age and technology adoption in small and medium-sized service firms. *Small Business Economics*.

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

Morris, C. N. (1983). Parametric empirical bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78(391):47–55.

Moses, H., Matheson, D. H. M., Cairns-Smith, S., George, B. P., Palisch, C., and Dorsey, E. R. (2015). The anatomy of medical research us and international comparisons. *The Journal of the American Medical Association, 313(2):174-189*.

Moses, H., Matheson, D. H. M., Dorsey, E. R., George, B. P., Sadoff, D., and Yoshimura, S. (2013). The anatomy of health care in the United States. *Journal of the American Medical Association*, 310(18):1947–1964.

Murphy, K. and Topel, R. (2006). The value of health and longevity. *Journal of Political Economy, 114 (5), 871-904.*

Murray, C., Kulkami, S., Michaud, C., Tomijima, N., Bulzacchelli, M., Iandiorio, T., and Ezzati, M. (2006). Eight americas: Investigating mortality disparities across races, counties, and race-counties in the united states. *PLoS Medicine*, 3(9):1513–1524.

National Library of Medicine (2020). Pubchem database. https://pubchem.ncbi.nlm.nih.gov/.

Nelson, A. and Gnanapragasam, V. J. (2016). Applying for research funding. part 1-sources of funding. *Journal of Clinical Urology. Vol. 9(3) 201-204*.

Newhouse, J. P. and Garber, A. M. (2013). Geographic variation in Medicare services. *New England Journal of Medicine*, 368(16):1465–1468.

Nordhaus, W. (2003). *Measuring the Gains from Medical Research: An Economic Approach*, chapter The Health of Nations: The Contribution of Improved Health to Living Standards, pages 9–40. Chicago, IL: University of Chicago Press.

Nyweide, D., Anthony, D., Bynum, J., Strawderman, R., Weeks, W., Casalino, L., and Fisher, E. (2013). Continuity of care and the risk of preventable hospitalization in older adults. *JAMA Internal Medicine 173(20): 1879-85*.

Office of the Assistant Secretary for Planning and Evaluation (2020). Hhs poverty guidelines. `https://aspe.hhs.gov/poverty-guidelines`.

Oostrom, T. (2020). Funding of clinical trials and reported drug efficacy. *Job Market Paper*.

Pammolli, F., Riccaboni, M., and Magazzini, L. (2010). The productivity crisis in pharmaceutical r&d. *University of Verona Department of Economics Working Paper 6*.

Panaich, S., Arora, S., Patel, N., Patel, N., Patel, S., Savani, C., Singh, V., Jhamnani, S., Sonani, R., Lahewala, S., Thakkar, B., Patel, A., Dave, A., Shah, H., Bhatt, P., Jaiswal, R., Ghatak, A., Gupta, V., Deshmukh, A., Kondur, A., Schreiber, T., Grines, C., and Badheka, A. (2016). In-hospital outcomes of atherectomy during endovascular lower extremity revascularization. *The American Journal of Cardiology*, 117(4):676–684.

Phelps, C. E. (2000). *Handbook of Health Economics*, chapter Information Diffusion and Best Practice Adoption, pages 223–264. North Holland.

Philipson, T. J., Seabury, S. A., Lockwood, L. M., Goldman, D. P., and Lakdawalla, D. N. (2010). Geographic variation in health care: The role of private markets. *Brookings Papers on Economic Activity*, 41(1):325–361.

Pope, G. C., Kautter, J., Ellis, R. P., Ash, A. S., Ayanian, J. Z., Iezzoni, L. I., Ingber, M. J., Levy, J. M., and Robst, J. (2004). Risk adjustment of medicare capitation payments using the cms-hcc model. *Health Care Financing Review*, 25:119–141.

Rangel, S., Efron, B., and Moss, R. L. (2002). Recent trends in national institutes of health funding of surgical research. *Annals of Surgery (236)3: 277-287*.

Reinhardt, U. (2019). *Priced Out: The Economic and Ethical Costs of American Health Care*. Princeton University Press, Princeton, New Jersey.

Rettenmaier, A. J. and Saving, T. R. (2009). Perspectives on the geographic variation in health care spending. Special Publication National Center for Policy Analysis.

Rikkers, L., Bland, K., Kinder, B., Lawrence, P., Glynch, T., Modlin, I., Schulak, J., and Whitehouse-Jr, W. (1985). Funding of surgical research: The roles of government and industry. *Journal of Surgical Research, 39:209-215*.

Samanez-Larkin, G. R. and Knutson, B. (2015). Decision making in the ageing brain: changes in affective and motivational circuits. *Nature Reviews: Neuroscience 16(5): 278-89*.

Savulescu, J., Wartolowska, K., and Carr, A. (2016). Randomised placebo-controlled trials of surgery: ethical analysis and guidelines. *Journal of Medical Ethics, 42 :776-783*.

Schettino, F., Sterlacchini, A., and Venturini, F. (2008). Inventive productivity and patent quality: evidence from italian inventors. *MPRA Paper No. 7765*.

Scott-Morton, F. and Kyle, M. (2011). *Handbook of Health Economics, Volume Two*, chapter Markets for Pharmaceutical Products, pages 763–823. Waltham, MA: North Holland.

Sheiner, L. (2014). Why the geographic variation in health care spending can't tell us much about the efficiency or quality of our health care system. *Brookings Papers on Economic Activity*.

Shewchuk, J. R. (1994). An introduction to the conjugate gradient method without the agonizing pain. Technical report, Carnegie Mellon University.

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

Skinner, J., Fischer, E., and Wennberg, J. E. (2005). The efficiency of medicare. In *Analyses in the Economics of Aging*, pages 129–160. Chicago: University of Chicago Press.

Skinner, J. and Fisher, E. (2010). Reflections on geographic variations in u.s. health care. Dartmouth Working Paper: `http://www.dartmouthatlas.org/downloads/press/Skinner_Fisher_DA_05_10.pdf`.

Skinner, J. and Staiger, D. (2007). Technology adoption from hybrid corn to beta-blockers. In *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, pages 545–570. Chicago: University of Chicago Press.

Skinner, J. and Staiger, D. (2012). Technology diffusion and productivity growth in health care. *The Review of Economics and Statistics*, 97(5):951–964.

Skinner, J. S., Staiger, D. O., and Fisher, E. S. (2006). Is technological change in medicine always worth it? the case of acute myocardial infarction. *Health Affairs*, 25(2):w34–w47.

Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) (2020). Framework for continuing professional development (cpd) courses for practicing surgeons: The new sages course endorsement system. `sages.org/publications/guidelines/framework-for-post-residency-surgical-education-training`.

Solutions, N. H. (2017). Surrogate upins. *https://med.noridianmedicare.com/web/jddme/article-detail/-/view/2230703/surrogate-upins*.

Sommers, B., Beard, C., D'Amico, A., Kaplan, I., Richie, J., and Zeckhauser, R. (2008). Predictors of patient preferences and treatment choices for localized prostate cancer. *Cancer*, 113(8):2058–2067.

Song, Y., Skinner, J., Bynam, J., Sutherland, J., and Fisher, E. (2010). Regional variations in diagnostic practices. *New England Journal of Medicine*, 363(1):45–53.

Sood, A., Jeong, W., Ahlawat, R., Campbell, L., Aggarwal, S., Menon, M., and Bhandari, M. (2015). Robotic surgical skill acquisition: What one needs to know? *J Minim Access Surg. 11(1): 10-15*.

Steffensen, F., Sorensen, H., and Olesen, F. (1999). Diffusion of new drugs in danish general practice. *Family Practice 16(4): 407-413*.

Stern, A. (2017). Innovation under regulatory uncertainty: Evidence from medical technology. *Journal of Public Economics, 145: 181-200*.

Subramanian, U., Weinberger, M., Eckert, G. J., L'Italien, G. J., Lapuerta, P., and Tierney, W. (2002). Geographic variation in health care utilization and outcomes in veterans with acute myocardial infarction. *Journal of General Internal Medicine*, 17(8):604–611.

The Council of State Governments' Summation Report (2001). Medical innovation and the aging of america. https://www.csg.org/knowledgecenter/docs/Conf0110HPFMedInnovAging.pdf.

Tu, P. (2017). Quasi-experimental evidence of physician effects. *Job Market Paper*.

United States Bureau of the Census (1994-2013). Current population survey (cps), march. Inter-university Consortium for Political and Social Research [distributor], 2009-03-12. https://doi.org/10.3886/ICPSR09039.v3.

United States Food & Drug Administration (2020). Approved drug products with therapeutic equivalence evaluations | orange book. https://www.fda.gov/drugs/drug-approvals-and-databases/approved-drug-products-therapeutic-equivalence-evaluations-orange-book.

United States Food and Drug Administration (1994-2013). Compilation of cder new molecular entity (nme) drug and new biologic approvals. https://www.fda.gov/drugs/drug-approvals-and-databases/compilation-cder-new-molecular-entity-nme-drug-and-new-biologic-approvals.

United States General Accounting Office (GAO) (2002). Dual code sets are acceptable for reporting medical procedures. Report to the Chairman, Subcommittee on Health, Committee on Ways and Means, House of Representatives. GAO-02-796.

Walsh, J. E. (1955). The poisson distribution as a limit for dependent binomial events with unequal probabilities. *Journal of the Operations Research Society of America*, 3(2):198–209.

Wayne Winegarden (2020). Price controls are not the answer to expensive drugs. Forbes, https://www.forbes.com/sites/waynewinegarden/2019/10/18/price-controls-are-not-the-answer-to-expensive-drugs/#e353261715ec.

Weinstein, J. N., Bronner, K. K., Morgan, T. S., and Wennberg, J. E. (2004). Trends and geographic variations in major surgery for degenerative diseases of the hip, knee, and spine. *Health Affairs*, no.(2004):var81–89.

Weisbrod, B. A. (1991). The health care quadrilemma: An essay on technological change, insurance, quality of care, and cost containment. *Journal of Economic Literature*, 29(2):523–552.

World Health Organization (2020). Atc/ddd index 2020. https://www.whocc.no/atc_ddd_index/.

Wrobel, J. S., Mayfield, J. A., and Reiber, G. E. (2001). Geographic variation of lower-extremity major amputation in individuals with and without diabetes in the medicare population. *Diabetes Care*, 24(5):860–864.

Zuckerman, S., Waldman, T., Berenson, R., and Hadley, J. (2010). Clarifying sources of geographic differences in Medicare spending. *New England Journal of Medicine*, 363(1):54–62.