Essays in Health Economics

by

Hannah Bolder

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2022

Doctoral Committee:

Professor Martha Bailey, Co-Chair
Professor Helen Levy, Co-Chair
Professor Charlie Brown
Professor Jim Hines
Professor Sarah Miller

Hannah Bolder
hbolder@umich.edu
ORCID iD: 0000-0001-7148-9198

ACKNOWLEDGMENTS

During the course of my PhD I was very fortunate to receive advice, support and assistance from many people and organizations. I am profoundly grateful to my committee members: Helen Levy, Martha Bailey, Jim Hines, Sarah Miller and Charlie Brown. One of the most valuable courses in my PhD career was a demography course taught my Martha, where we studied a variety of cutting edge identification strategies. Through this course and beyond, her guidance on how to plan, structure and carry out a large research project has been invaluable, and I have benefited greatly from her feedback and advice. I am especially grateful that she agreed to continue to serve on my committee after leaving the University of Michigan to join UCLA. Sarah inspired my interest in health economics and during the more difficult times of the PhD, her fascinating research continued to remind me of what initially drew me to the field. Helen has been a fantastic mentor to me over the years, and I am very grateful for her research advice, career guidance and encouragement throughout this process. Helen invited me to co-author a paper which became the second chapter of this dissertation, and while working with her I learned a lot about classic areas of health economics research and how to produce high quality research. Even before he was on my committee, Charlie attended several seminars where I presented my work where he asked probing questions and offered valuable suggestions for improvement. I am particularly grateful for his thorough reading of my dissertation and the ensuing detailed feedback. His critical eye contributed greatly to improving the clarity of my work and helped me work through some complicated econometrics issues. Jim's involvement in my career predated my admission to the PhD program, and I am delighted that he agreed to be a member of my committee. I thank Jim for sticking with me and continuing to ask thought-provoking questions and offer insightful comments, even as I "switched gears" away from taxation and public finance as primary area of interest. All of my advisors contributed to my development as a researcher and continued to engage with my work even as for several of them, my research grew to areas that had less overlap with their own areas of research. I owe a debt of gratitude to Martha, Helen, Jim, Charlie and Sarah.

I am fortunate to have benefited from interaction with many other faculty members during the course of my PhD. Conversations and coursework with Hoyt Bleakley and Joel Slemrod have contributed to my intellectual and professional development. I am grateful to Dean Yang for the opportunity to serve as research assistant and later project manager on his RCT, through which I gained useful experience conducting field experiments and surveys, as well as a broader understanding of the process of conducting research in development economics. Jeff Smith advised me as I was contemplating pursuing a PhD, and I benefited from his mentorship. His econometrics courses led me to see econometrics as not only an integral part of any empirical economics research project, but also a truly fascinating subject. Jeff also introduced me to Helen, who he rightly anticipated would be an excellent match as a mentor.

I am thankful to Laura Flak, Lauren Pulay, Miriam Rahl, Hiba Baghdadi, and Julie Heintz for their administrative support in the Economics Department and in the Population Studies Center and for their patience in answering all my questions and helping me avoid and resolve potential issues. At the Institute for Health Policy and Innovation (IHPI), Phyllis Wright-Slaughter, Patrick Brady and Kathryn Ashbaugh provided advice and administrative support and technical help as I applied to use Medicare data, the AMA Masterfile, the HCUP SID data, and the AHA Hospital data. Without their assistance, the process would have been difficult and very complicated. Phyllis and Chiang-Hua Chang possess a wealth of knowledge about Medicare data are were kind enough to advise me about some of the more complicated aspects of how the data is structured, as well as its strengths and limitations. I would also like to thank Timothy Ahlgren, Benjamin Butler, Josh Errickson, Mark Montague and Neil Tweedy for their help with the HPC Cluster, and Steven Wolodkin for assistance with the Yottabyte computing environment.

The Economics Department, the Population Studies Center, IHPI and the University of Michigan Medical School supported my work on this dissertation. The Michigan Office for Tax Policy Research (OTPR) supported my travel to the National Tax Association conference. IHPI and the Medical School supported my application to reuse Medicare data and the HCUP SID data, which otherwise would have been extremely expensive on my own. I was fortunate to receive Summer Research Apprenticeship (SRA) funding and the William Haber Graduate Fellowship. A research award from the NBER-IFS Value of Medical Research network (supported by NIH grant R24AG048059 and NIA grant R24AG048059) supported my travel to attend and present at professional conferences. I am very grateful for the opportunity to be a visiting student at NBER and to have been invited to attend the NBER Summer Institute - an educational, inspiring and truly incredible academic experience. My research and educational development was supported by an NIA training grant to the Population Studies Center at the University of Michigan (T32AG000221). This fellowship enabled me to take some time off from teaching to pursue joint work with Helen and explore research areas of interest for my dissertation.

Kyle Logue generously offered to engage in an independent analysis of state medical malpractice standard of care laws from 2000 to 2020. I thank him and his research team for their efforts and hard work in this endeavor. Kate Britt was instrumental to the legal research aspect of my dissertation, meeting with me on multiple occasions to explain how the Shepardization process works and helping me access various legal texts that otherwise would have been very difficult to come by.

The Michigan graduate program has great camaraderie, and I thank my classmates and colleagues for their support, friendship and advice as we pursued PhDs together, especially Ben Glass, Will Boning, Luis Baldomero, Justin Ladner, Daniel Hubbard, Gail Lucasan and Huayu Xu. Adam Stevenson was an excellent boss to work for and it was great to be able to commiserate about parenting young children and life as a parent during Covid. I thank Joelle Abramowitz for her friendship and support throughout my PhD.

For their endless support and efforts on my behalf I thank my family, especially my parents Paula and Arthur Bolder and my grandmother Claire Bolder. Without their encouragement I would not have had the gumption to change career paths and pursue a PhD in Economics. Last but certainly not least, I also thank my husband Jon Eguia for the tireless hours spent watching our son while I worked to complete my dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

ABSTRACT

This dissertation investigates whether institutions and incentives affect healthcare decisions and health behaviors. The first two chapters address situations where healthcare decisions may be affected by the legal system and health insurance coverage, and the third explores to what extent health behaviors change in response to taxes.

Chapter 1 investigates whether and to what extent physician decisions to abandon ineffective treatment practices are affected by one aspect of the legal system: medical malpractice standard of care definitions. How medical malpractice is defined varies by state. In some states the standard of care requires doctors to adhere to national standards (defined by customary practice) and in others physicians are required to follow customary practice of the community in which they practice. Given the well-documented and significant geographic variation in treatment practices within the United States, local practice norms can and often do differ from national standards. Legal scholars have hypothesized that local standards of care reduce the incentive for physicians to keep abreast of medical advances, slowing the adoption of new treatments and the de-adoption of ineffective ones. The first aim of this chapter was to categorize the standard of care definitions for all 50 states as national customs-based standards or local customs-based standards, as well as to document how these standards changed over time. The analysis in this chapter is based on an extensive review of primary source material such as State Supreme Court and Appellate Court decisions, U.S. District Court decisions, and statutes. Next the chapter proceeds to examine the effects of these varying definitions on patient care, focusing on the physician decision to discontinue the use of vertebroplasty. [1] Vertebroplasty is a surgical procedure where medical grade cement is injected into a spinal compression fracture to attempt to stabilize it and alleviate pain. In 2009, two influential studies presented evidence that vertebroplasty was no more effective than a placebo "sham surgery." I find that de-adoption occurred rapidly in all states regardless of the standard of care law in place. However standard of care laws do matter for rural sub-state regions: rural areas reduced vertebroplasty use by 0.18 per thousand less in locality states than they would have had a national standard of care applied.

[1] It is worth noting that treatment decisions reflect both supply and demand. The use of language such as "the physician decision" reflects the fact that medical malpractice incentives operate on physicians and are likely to affect the supply channel. It is less obvious how variation in standard of care definitions would elicit a demand response, but this is theoretically possible and this chapter cannot rule out demand responses.

Chapter 2, which is joint work with Helen Levy, examines the effect of insurance coverage on healthcare utilization for seriously ill patients who are hospitalized after seeking care in an emergency room. The paper focuses on Medicaid insurance for low-income people, which was expanded in many states in 2014 to increase the number of eligible individuals. Two channels exist through which Medicaid coverage may affect healthcare use: on the extensive margin more people may gain Medicaid coverage, and because they are now insured they may be more likely to seek care. On the intensive margin, conditional on seeking care, insurance coverage through Medicaid may affect "treatment intensity," or "how much" patients are treated. The paper measures treatment intensity as the length of stay, number of procedures received and total list charges. As in Chapter 1, treatment intensity effects may be physician-driven (supply) or patient-driven (demand). This study focuses on a subset of admissions that are so serious that care cannot be deferred - admission to the hospital is equally likely whether the patient arrives at the emergency room on a weekend day or on a weekday. This allows us to estimate intensive margin effects without the confounding effects of selection into treatment and changes in patient composition before and after the expansions. We find that the 2014 Medicaid Expansions increased the share of patients covered by Medicaid. This increase was partially offset by a decrease in the share of patients with private insurance coverage. We find no statistically significant effects of Medicaid coverage on treatment intensity (number of procedures, length of stay and total list charges) or on mortality. The coefficients were imprecisely estimated because the analysis does not separately identify effects from the coverage gain channel and the crowd-out channel, and these effects likely operate in opposing directions. Intensive margin effects for the coverage gain channel are likely to be positive; Medicaid compensation is higher than for uncompensated care. Treatment intensity effects due to crowd-out are likely to be negative because Medicaid reimburses less generously than private coverage.

The last chapter explores how consumers respond to taxes on sugar-sweetened beverages, colloquially called "soda taxes." As with many taxes, the policy maker's goal in imposing a soda tax is two-fold: to raise revenue and to discourage a behavior - consumption of unhealthy sugary drinks in the present case. This paper evaluates the effects of soda taxes on consumer behavior, comparing taxes in Berkeley and Philadelphia within the same study and using the same methods so that measured differences can be more easily attributed to local supplier and consumer responses rather than to differences in methodology. I first estimate the effects of the taxes on soda prices. In Berkeley 6% of the tax on regular soda was passed on to consumers. That is, for every 1 cent of tax levied on sugary drinks, the market price increased by 0.06 cents. In Philadelphia, 57% of the tax on regular soda was passed through to consumers and 67% of the tax was passed through for diet soda. (In Berkeley diet soda was not taxed.) In both cities pass-through declines with beverage size. Consumption of regular soda in Berkeley decreased by 7.6%, and in Philadelphia consumption of regular and diet soda decreased by 28% and 33%, respectively. In Philadelphia I find evidence of cross-border shopping: stores in neighboring regions lowered prices and purchases of soda from these stores increased.

CHAPTER 1

Evidence Reversal in Medicine: Do Legal Standards of Care Affect Phase-out of Ineffective Practices?

Abstract

In medicine evidence reversals occur when current clinical practice is deemed ineffective or harmful, and evidence shifts in favor of an older standard of care or no treatment. Anecdotal evidence suggests that after an evidence reversal occurs the phase-out or "de-adoption" process for the procedure in question may take over a decade. Understanding barriers to de-adoption contributes to our knowledge of the determinants of productivity in the healthcare sector. Medical malpractice incentives may foster or discourage the implementation of new medical evidence or innovations that improve productivity. This paper investigates the role of a particular legal institution - definitions of the standard of care – in the de-adoption process for vertebroplasty following an evidence reversal in 2009. The legal literature hypothesizes that locality-based standards of care stifle progress by enabling laggards and punishing physicians who choose to practice at the cutting edge. Overall, de-adoption of verteborplasty occurred rapidly in both locality and national standard states. However leveraging state-level variation in how medical malpractice is legally defined and within state variation in the practical application of these laws, difference-in-differences and event study analyses demonstrate that there are significant differences across rural and urban areas within states. In rural areas, where locality rules should have a greater impact, de-adoption occurs more slowly, whereas de-adoption occurs more slowly in urban areas of locality states. This study partially confirms predictions in the legal literature: while locality rules do matter for rural areas within states, as a whole they do not have a statistically significant effect on de-adoption. Results indicate that for rural areas, the locality rule causes vertebroplasty rates to fall by 0.18 per thousand (48% of the mean) less after evidence reversal than they would have had a national standard of care applied. The size of the locality effect is one and a half times the size of the association between de-adoption and physician gender, and more than 2 times larger than the association between having at least 30 years of experience and de-adoption.

1.1 Introduction

Innovation is widely recognized as an important driver of economic growth, and much time has been devoted to studying factors that may spur or hinder its progress. Dis-innovation, or the process of abandoning old and ineffective technologies, is also an important contributor to productivity and is less well-understood. At first glance, the study of dis-innovation may seem trivial: adopting a new technology usually entails abandoning or scaling back the use of an old one. In these cases, one might expect innovation and dis-innovation responses to be opposite but symmetric. However in the field of medicine, which is characterized by a high degree of uncertainty about the benefits of some treatments, practices are often rendered obsolete without an accompanying innovation. Medicine presents a unique and important setting in which to study dis-innovation: evidence reversals in which current clinical practice is overturned are frequent, and the practices that are implicated in these reversals affect millions of people.

The productivity of the healthcare sector is an active area of research. Many experts believe that wasteful healthcare spending constitutes as much as 5% of US GDP (Doyle et al 2017). Many proposals have focused on the amount of waste that could be eliminated by cutting back on low-value treatments or treatments that are not cost-effective, or by discouraging overuse (for example, of antibiotics). It is important to note that medical practices that are subject to evidence reversal are not merely not cost effective, or low value. These reversed practices lack health benefits and sometimes even generate harm, independent of their cost. As such, encouraging doctors and patients to abandon these practices could in theory lead to a reduction in wasteful spending without generating as much controversy as efforts to label certain procedures and treatments low value or not cost-effective.

Understanding which frictions slow the abandonment of out-dated medical practices would expand our understanding of the relationship between dis-innovation and productivity in the medical sector and has important policy implications. One potential source of such frictions is medical malpractice law, to the extent that it is designed suboptimally.. The medical malpractice system can be viewed as an attempt to correct the distortionary outcomes of a patient - physician principal agent problem. Facing a decision among many or no treatments the physician-agent maximizes their utility, which is a weighted sum of patient benefits, physician profit and other factors. Physicians may face a trade-off between profits and patient welfare, which is exacerbated by asymmetric information. (Ellis and McGuire, 1986; Chandra Cutler and Song 2011) If the parameters of the medical malpractice system are set optimally, malpractice risk could in theory help align physician and patient incentives. Deviations from the optimum, however could have smaller effects or even be harmful. Standard of care definitions in medical practice law could have distortionary effects on physician decision making, and these effects might differ between the context of evidence reversal and the adoption of new treatments.

This study uses difference-in-differences and event study methods to estimate how procedure rates change after evidence reversal as well as how procedure rates are affected by state-level differences in malpractice law, specifically standard of care definitions. Physicians have a legal duty to provide treatment that meets the standard of care; in order to prevail in a malpractice lawsuit the plaintiff must establish the relevant standard of care owed to the patient by the physician and how the defendant physician breached it. How the medical malpractice standard of care is established has varied over time and continues to vary by state, as this study documents (see section 1.3.1). In medical malpractice, deference to physician expertise and judgment influenced the development of a standard of care that is based on adherence to customary practice in the community. How a community is defined varies by state: in states that have a local (national) customs based standard of care, physician defendants are judged not based on their adherence to evidence based clinical practice guidelines or based on whether their course of action seems reasonable to a panel of experts, but based on whether they adhered to the local (national) medical customs concerning how to treat a patient in a similar situation.

In the legal community, both proponents of local customs-based standards and national customs-based standards have argued that local customs give rise to lower, less demanding standards of care for physicians. Historically, a country doctor was not expected to possess the same skills and experience as an urban physician, nor to have access to the same quality of facilities and equipment. In fact, rural-urban inequality was one factor used to justify the continued use of local customs-based standards of care. As medical education has become more standardized and communication and information technology has improved some state Supreme Courts have recognized a national standard of care. In these states, physicians may be legally liable for discrepancies between local customs and national practice.

Given these standard of care definitions, to the extent that local practice differs from new guidelines or scientific findings, local customs-based standards of care may dis-incentivize physicians from de-adopting practices that are found to be ineffective. Indeed, the idea that these standards of care discourage innovation is a prominent theory in the scholarly legal literature and has been cited in medical malpractice court cases in arguments that support a national standard of care. (Greenberg (2009); Laakmann (2015); Parchomovsky and Stein (2008); Monico et al. (2005) and Smothers v. Hanks, 34 Iowa 286, 289 (dissent)). Evidence from this study only partially supports this argument. I find that at the state level locality-based standards of care do not have a statistically significant effect on de-adoption. However in rural areas the effects of locality rules are more pronounced – locality rules slow the speed of de-adoption by 0.18 procedures per thousand, or 48% of the baseline mean procedure rate.

In the context of de-adoption, the Medicare population is an important and interesting group to study because their healthcare utilization is very high, and health spending by the elderly is disproportionately high relative to their share of the population. Medicare spending accounts for about 20% of National Healthcare Expenditure. (NHE Factsheet, CMS.gov) In addition, there are methodological advantages: reimbursement and payment incentives are the same for all Medicare beneficiaries and the vast majority of the over-65 population has Medicare coverage, alleviating concerns of selection according to trends in uninsurance rates. This study leverages data from several sources: a 20% sample of Medicare fee-for-service beneficiaries covering the years 2002-2014 provides information about treatment patterns and diagnoses over time in both inpatient and outpatient settings, and the American Medical Association Masterfile (commonly referred to as the AMA Masterfile) contains demographic information about physicians who practice in the United States. Primary source material from court cases and statutes informed the creation of a novel dataset tracking state variation in the standard of care from 2000 to 2015.

Many evidence reversals directly affect the elderly. In this paper, I focus on a single procedure as a case study: vertebroplasty, a surgical procedure used to treat age-related painful vertebral compression fractures caused by osteoporosis.[1] Osteoporosis affects about 8 million people over 60 years old, and causes about 1.5 million vertebral fractures each year. (Wright et al. 2014, Alexandru and So 2012) Osteoporotic fractures have direct costs of between $12 and $18 billion per year in the United States (in 2002 dollars). (Office of the Surgeon General, 2004) Given the general trend of rising life expectancy, the burden of osteoporosis is expected to increase. Vertebroplasty involves using fluoroscopy or (less frequently) Computed Tomography (CT) guidance to inject medical grade cement into a fracture in order to stabilize it. Vertebroplasty may be performed by more than one kind of medical specialist, and while all these specialties face relatively high malpractice risk, there is still variation in malpractice risk: orthopedic surgeons and neurosurgeons are among the highest-risk specialties, radiology is considered either moderate or high risk depending on the study. (Kane 2010; Jena et al. 2011; Studdert et al 2005)

The paper proceeds as follows: Section 1.2 briefly discusses evidence about adoption and de-adoption and the effects of medical malpractice standard of care laws on healthcare delivery. Section 1.3 presents information about legal standards of care and their significance within malpractice cases, and introduces the setting of vertebroplasy and the evidence reversal. Section 1.4 describes the data sources and empirical methods as well as the qualitative methodology used to construct the legal data set. Section 1.5 and 1.6 present results and robustness checks, and Section 1.7 concludes.

[1] Vertebroplasty has also been used to treat vertebral fractures that are caused by traumatic injury or cancer (Knavel et al. 2009 and Fourney et al. 2003), but these represent fundamentally different patient populations and clinical scenarios and were not included in the RCTs that estimated causal effects of vertebroplasty.

1.2 Background

By examining how medical malpractice standard of care definitions affect de-adoption, this paper contributes to two areas of research: physician responses to evidence reversal and the effects of medical malpractice on physician decision making. There is a growing body of research about technology diffusion in healthcare and adoption of new treatments (see for example, Skinner and Staiger 2015 and Chandra et al 2014). However, the determinants of de-adoption rates for medical practices that are found to be ineffective or harmful remains understudied. Since an Institute of Medicine report estimates that de-adoption may take up to 17 years, it is very important to understand the barriers to de-adoption in order to increase efficiency (Committee on Quality of Healthcare in America, Institute of Medicine 2001). There are many studies that investigate how procedure rates change in response to negative changes in guidelines or RCT evidence. Howard and Adams (2012) show in the first year after the United States Preventive Services Task Force (USPSTF) issued revised mammography guidelines in 2009 screening rates did not change, showing no significant response to the major or minor changes in recommendations. Howard et al (2016) document that after evidence emerged that a breast cancer treatment was ineffective, procedure rates declined by 32.6 percentage points, showing that contrary to common belief, sometimes evidence reversals are implemented quickly. Howard et al (2016) hypothesize that de-adoption may have occurred especially quickly in this context because the procedure did not constitute a large share of revenue for surgical oncologists, patients often take a more active role in breast cancer treatment than for other conditions, and because the clinical trial not only demonstrated that the procedure was ineffective, but also potentially harmful. Niven et al (2015) is one of the few studies to compare adoption trends after initial positive evidence to deadoption rates after a subsequent evidence reversal. They find that initial positive evidence increased the use of tight glycemic control in the Intensive Care Unit (ICU); however subsequent evidence that the practice was harmful had no statistically significant effect on practice.

A related body of work examines the role of various factors in increasing or decreasing de-adoption speed, such as Bekelis et al (2017), Howard et al (2016), Kozhimannil et al (2017) and Wang et al (2015). Bekelis et al (2017) focus on carotid revascularization and finds that more experienced physicians decreased their use of the procedure by more than those with less than 12 years of experience. Physicians for whom the procedure accounts for a large share of revenue had the smallest declines in procedure rates. Bekelis et al (2017) distinguish between physicians who scale back their use of the procedure and those who abandon the procedure completely, finding that two-thirds of the decline could be attributed to scaling back and one third to physicians who abandoned the procedure completely. Howard et al (2016) investigate the role of physician incentives in de-adoption after an RCT questioned the benefits of knee arthroscopy. They find that while physicians generally reduced their use of the knee arthroscopy, reductions were smaller in physician-owned surgery centers. Kozhimannil et al (2017) document that guidelines from the American College of Obstetricians and Gynecologists (ACOG) discouraging routine use of episiotomy were issued after a large number of physicians had already begun reducing their use, implying that the ACOG guidlelines were a “lagging indicator.” While the guidelines did not precipitate a sharp decrease in procedure use overall, they did narrow the gap in practice between urban teaching hospitals, which already had lower episiotomy rates, and urban non-teaching hospitals, which were “lagging adopters” of the new evidence against routing episiotomy. Wang et al (2015) analyze responses of specialty journals to evidence reversal, comparing these responses to responses in the journal where the reversal publication was originally published. They find greater resistance to de-adoption in specialty journals, confirming theories of “specialty bias” against medical reversal. Taken together as a whole, these studies provide evidence that financial incentives, physician experience and specialty, as well as teaching vs non-teaching practice setting are all factors that affect de-adoption rates. However these studies have not addressed whether malpractice laws affect the de-adoption of outmoded technologies.

In the medical malpractice literature, Frakes (2013) examines the impact of adopting national standard of care definitions and finds evidence of significant convergence in procedure rates after the new standards are adopted. Frakes and Jena (2016) extend this analysis and find that national customs-based standards of care improve healthcare quality. While other aspects of medical malpractice law have been studied extensively (notably non-economic damage caps) relatively few studies in medical malpractice focus on standard of care definitions.

This study has two components: an analysis of standard of care definitions across states and over time and an empirical analysis of how standard of care definitions affect physician responses to evidence reversal. During the study period 2000 – 2015 standard of care definitions vary across states and remain constant over time. During this period, there were substantial changes in the evidence about vertebroplasty effectiveness. Even with the benefit of hindsight, determining precisely when a medical procedure becomes obsolete is inherently difficult, especially when there is no clearly superior replacement treatment, such as in this setting. Rather than relying on this distinction, using these procedures allows me to focus on how physicians respond to an information shock when new evidence emerges that should cause a significant update to their prior beliefs about treatment efficacy.

1.3 Setting

1.3.1 Evolution of medical standard of care laws

In order to demonstrate negligence in a medical malpractice lawsuit the plaintiff must demonstrate that the defendant physician owed a duty of care, the physician breached the standard of care, the plaintiff was injured or incurred damages, and lastly that these damages or injuries arose as a direct result of the defendant's breach of the standard of care (a causal link). The applicable standard of care is a crucial and often hotly contested element of malpractice lawsuits, and is the focus of this section.

Negligence law applied to medical professionals has evolved somewhat differently than ordinary negligence law due to several factors, among them deference to physician expertise, cognizance of geographic disparities in access to resources such as specialists and technology, and a lack of unity among various schools of medical practice. These circumstances supported using a customs based approach to define a standard of care specific to the community where a physician practiced medicine. [2] The prudent man rule, under which the standard of care is determined based on what a reasonable person would have done when faced with the same or similar circumstances was not generally applied to medical malpractice cases because a lay jury was seen as lacking the necessary expertise to be able to judge whether a physician's conduct was reasonable given the complexities involved. Thus customs based standards of care were commonplace, essentially allowing the medical profession as a whole to self-regulate (Peters, 2000; Monico et al, 2005).

As education and training in medicine became more unified and medical research and conferences became more easily accessible, courts began to debate the value of deferring to local customs and many states adopted a nationwide standard of care. Other states, while not accepting a nationwide standard of care, expanded the definition of the relevant locality to include same or similar communities, or an entire state. Thus under current medical malpractice law, given existing variation in the geographic scope of standard of care definitions physicians are judged as follows: in states that have a local (national) customs based standard of care, physician defendants are judged not based on their adherence to evidence based clinical practice guidelines or based on whether their course of action seems reasonable to a panel of experts, but based on whether they adhered to the local (national) medical customs concerning how to treat a patient in a similar situation.

Although this study does not use variation in the timing of standard of care reforms to identify

[2] While customs have also played a role in ordinary negligence law, courts generally have much less deference to custom in this area and acknowledge the existence of "customary negligence." "Customs and usages themselves are many and various; some are the result of careful thought and decision, while others arise from the kind of inadvertence, carelessness, indifference, cost-paring and corner-cutting that normally is associated with negligence. . . even an entire industry, by adopting such careless methods to save time, effort or money, cannot be permitted to set its own uncontrolled standard." See Prosser and Keeton Chapter 5 page 194.

locality rule effects, it is important to understand how these reforms arose. If national standard states shared a common underlying process or trends that led to these reforms, then they may systematically differ from locality states in ways that may bias the estimates in this study. Frakes and Jena (2016) conduct an in-depth qualitative analysis of the circumstances surrounding the transitions to national standards of care from 1979 to the early 2000s, and they conclude that there is no evidence that these reforms were driven by trends in healthcare quality, and they are likely plausibly exogenous with respect to many other healthcare variables. In particular, they find that the initial reforms arose through the court system rather than through the legislature.[3] Furthermore, when ruling on the geographic scope of the standard of care, the state Supreme Court was presented evidence exclusively about the very narrow medical context of the case at hand. Frakes and Jena (2016) reviewed court records pertaining to these cases and found that no amicus briefs or other documents representing "third party interests" were filed relating to these issues. The language and arguments used by judges in these cases reflect equity-based motivations rather than concerns about trends in the healthcare sector. A reading of later court cases covered during this study period (2006 through 2014) confirms their analysis: judges focused on the ubiquity of many medical conditions and cited past cases to uphold the standard of care rather than constructing new arguments related to public policy goals (See for example Avivi v. Centro Medico Urgente Medical Center, 159 Cal. App. 4th 463).

Establishing the standard of care lies at the foundation of any medical malpractice case. Historically, standard of care definitions had both substantive and procedural implications. Nearly all medical malpractice cases require the plaintiff to present expert testimony from a physician who will testify about what the standard of care is and how the defendant breached it. [4] In a state where the standard of care is determined by accepted practice in the same community as the defendant physician, physicians from other states or even other communities in the same state were historically often prevented from testifying because they were not deemed qualified (Frakes 2013; see for example Strode v. Lenzi, 116 Idaho 214, 775 P.2d 106). Although these procedural implications for qualifying witnesses have been relaxed in most states, it is common for defending physicians to submit a motion for a summary judgment before a case is heard in court, and failing to present appropriate expert testimony about the standard of care risks having the case dismissed (Avivi v. Centro Medico Urgente Medical Center, 159 Cal. App. 4th 463; Arregui v. Gallegos-Main, 153 Idaho 801, 291 P.3d 1000 (2012); Ramos v. Dixon, 144 Idaho 32, 156 P.3d 533 (2007); North Carolina. Smith v. Whitmer, 159 N.C. App. 192, 582 S.E.2d 669 (2003); Barnes v. Conn. Po-

[3]This study finds that many states have since passed legislation about the standard of care. We show that the results are robust to focusing only on the subset of states with no legislation in Table 1.10.

[4]Res ipsa loquitor cases, where negligence is obvious (for example if a surgical instrument is left inside a patient's body or if surgery is performed on the wrong body part, etc), often do not require expert testimony about the standard of care, but a case must meet stringent requirements to be considered res ipsa loquitor.

diatry Grp., P.C., 195 Conn. App. 212, 224 A.3d 916 (2020)). In addition, although most states currently allow physicians from other regions to testify, states with "same or similar community" rules or statewide customs of care may still require expert witnesses to produce evidence that they are familiar with conditions in the defendant's community and the customs and habits of local physicians. In order to reach a jury the burden is on the plaintiff to first establish a standard of care by tailoring the legal standard of care definitions in statutes and case law to the specific medical circumstances of the patient.

Appendix A uses Michigan as a case study to show how case law and statutes have jointly contributed to the evolution of the legal definition of the standard of care over time. Court cases in Michigan illustrate the importance of locality rules, even in states that have expanded the standard of care to include "similar communities" and are not often singled out as "strict locality" states. [5]

1.3.2 Vertebroplasty: development, context and evidence

Vertebroplasty is a procedure used to treat age-related painful vertebral compression fractures caused by osteoporosis. Osteoporosis affects about 8 million people over 60 years old, and causes about 1.5 million fractures each year. (Wright et al 2014) (Office of the Surgeon General, 2004) One of the most common complications of osteoporos is vertebral fractures, which carry a substantial increased risk in mortality, although whether this link is causal is controversial. (Teng, 2008) Osteoporotic fractures have direct costs of between $12 and $18 billion per year in the United States (in 2002 dollars). (Office of the Surgeon General, 2004) Given the general trend of rising life expectancy and the shifting age-distribution of the US population, the burden of osteoporosis is expected to increase. Among adults over 60 years old, a larger share of women are affected by osteoporosis than men. However vertebral compression fractures remains a significant risk for men as well. At age 50, the lifetime risk of a clinically significant vertebral fracture is about 16% for white women and 5% for white men. (Melton 2000) (Cummings & Melton 2002) The 11 percentage point difference in lifetime risk is a combination of greater baseline/underlying risk in women and longer life expectancy in women. Not all vertebral compression fractures are clinically significant (serious enough to cause concern and bother patients). However many others dramatically reduce mobility and quality of life, and require intensive treatment.

Vertebroplasty was originally developed in the late 1980s to treat aggressive vertebral hemangiomas (benign tumors) and was then later adapted to treat vertebral compression fractures in the elderly, which was its most common use in 2008. (Peh et al, 2008). It is most often performed by radiologists, orthopedic surgeons, and neurosurgeons. (See figure) In order to perform a vertebroproplasty, the physician injects medical grade cement into the fracture to stabilize it. This

[5] See Ginsberg (2013) for an analysis of the role of the locality rule in six states: Idaho, Tennessee, New York, Virginia, Arizona and Washington.

mechanical stabilization process as well as the destruction of nerve endings during the cement injection process was thought to provide improved pain relief over conservative medical management (medication, bed rest, use of a back brace, and occasionally physical therapy). With conservative non-surgical management, most osteoporosis-related vertebral compression fractures heal within 6-8 weeks and pain subsides. (Voormolen et al, 2007)

Vertebroplasty complications may occur if cement leaks out of the vertebral body, causing spinal cord or nerve root compression, or cement emboli in large vessels or in lungs (eg "pulmonary embolism") (Buchbinder, 2015). Estimates of cement leakage frequency vary from 30 and 93% [6] although most leaks are asymptomatic and do not require follow-up care. (Peh, 2008) Other relatively rare complications include rib fractures, a bone infection, fat embolism, anesthesia-related complications, and thecal sac injury (which could lead to cerebrospinal fluid leakage). (Buchbinder, 2015) See Peh (2008) for a detailed discussion of the incidence of these rare but serious complications. According to Leake et al. (2011) between 2004 and 2008 inpatient vertebroplasties had complication rates of .58% for pulmonary embolism, .37% for cardiovascular complications, 1.42% post-op surgical and neurological complications, and a .72% mortality rate.

Some researchers believe that vertebroplasty may increase the risk of subsequent vertebral fractures, although whether these additional fractures are a result of vertebroplasty or part of the natural course of osteoporosis is controversial. The evidence remains inconclusive. See Klazen et al (2010), Nagaraja et al. (2013) and Yang et al (2019) for a discussion.

As vertebroplasty proponents point out, conservative medical management is not without risks. Prolonged bed rest may result in further reductions in bone density, decreases in muscle strength and higher risk of deep vein thrombosis and pulmonary embolism. (Barr 2014) In addition, medical management of vertebroplasty may involve the use of opioids, which could lead to abuse. However pain from osteoporotic vertebral fractures can usually be managed with acetaminophen, supplementing with codeine as needed for "breakthrough pain." (Papaioannou, 2002) According to Papaioannou , the dose of codeine would vary between 30 and 60mg every 6 hours, or 18 – 36 morphine milligram equivalents (MME) per day using the conversion factor in the CDC's Opioid Prescribing Guideline. According to this guideline, risks of opioid dependency increase at 20 MME, but 20-50 MME is still a "relatively low dose." In addition, studies have found that exercise reduces pain from vertebral fractures, and likely has more long-term benefits for patients with osteoporosis. (Papaioannou, 2002).

This paper is not the first to examine verteborplasty trends from the early 2000s until 2015. Lad et al (2009) and Gray et al (2008) examine early trends and present evidence of substantial growth in vertebroplasty rates since the late 1900s and between 2001 and 2005 (respectively).

[6] There is substantial variation between studies due to inter-rater reliability issues, method of detection used (fluoroscopy, standard X-rays or CT scans), and probably physician technique and skill. (Schmidt, 2005)

Vertebroplasty rates continued to increase through the mid-2000s (see Leake (2011)) and by 2009, enthusiasm for vertebroplasty had reached such levels that some researchers began to explore the extension of vertebroplasty to address not only existing fractures but as a prophylactic measure (Kamano, 2011 and Kobayashi, 2009). After the RCT evidence emerged in 2009, vertebroplasty rates began to decline (Laratta et al 2017), and continued to do so through 2017 Lopez et al (2020).

1.3.3 Evidence about vertebroplasty effectiveness and defining the reversal "event"

Twenty years after the introduction of vertebroplasty to treat painful vertebral compression fractures, the first RCT evidence emerged in 2007 ("the VERTOS Study"). This study compared vertebroplasty (n=18) to optimal pain medication (n=16) and found evidence that one day after vertebrolpasty there was a statistically significant improvement in pain. Two weeks post-treatment, the vertebroplasty group continued to have lower pain scores (relative to the medical management group); however, the difference was not statistically significant. Patients in the control group had higher medication use and worse scores on disability and quality of life questionnaires. The follow-up period was extremely short due to a high crossover rate of 88%. (Voormolen et al, 2007)

In 2009 two influential vertebroplasty RCTs were published in the New England Journal of Medicine. Buchbinder et al (2009) was a multicenter double-blind study, where 78 participants were randomly assigned to undergo vertebroplasty or a sham surgery. Due to concerns that the positive effects of vertebroplasty documented in prior observational studies and in VERTOS may have been due to the placebo effect, and because placebo effects may be even more pronounced for surgical procedures (Kaptchuk, 2000; and Meissner 2011), the sham surgery was constructed to mimic vertebroplasty as closely as possible from a patient's perspective, including opening a smelly can of cement during the sham surgery. Patients were assessed at 1 week, 1 month, 3 months, and 6 months post-surgery. At all points both groups have improvements in pain, disability and quality of life score but crucially, there was no statistically significant difference between vertebroplasty and sham surgery. Given that many of these fractures resolve within 6-8 weeks, the study examined patients with symptoms for less than 6 weeks versus more than 6 weeks and found no difference. (Buchbinder et al, 2009) The other RCT, conducted independently from the Buchbinder et al 20009 study, is referred to as the INVEST Trial. This multicenter study randomized 131 patients to either vertebroplasty or a sham surgery, and was also blinded. As in the Buchbinder et al. trial, vertebroplasty cement was mixed during the sham surgery intervention. Pain and disability outcomes improved in both groups after the surgeries and these improvements persisted after 1 month of follow-up. As in the Buchbinder et al trial, the vertebroplasty group did not show statistically significant improvement over the sham surgery group. (Kallmes et al, 2009)

[7]

The publication of these sham-controlled trials sparked a vigorous debate about the value of vertebroplasty. In subsequent years another vertebroplasty RCT was published with a medical management control arm (VERTOS II Klazen et al 2010). The study concluded that vertebroplasty is more effective than medical management and encouraged its continued use. In 2010 the American Academy of Orthopedic Surgeons released guidelines recommending against vertebroplasty for vertebral fractures in patients with osteoporosis. A Cochrane review in 2015 concluded that vertebroplasty provides no clinical benefit for osteoporotic vertebral fractures. (Buchbinder et al., 2015). The guidelines were updated in 2018 to include more recent studies but the conclusion remained the same. (Buchbinder et al, 2018) Despite the subsequent "article tennis match," (Albers and Latchaw, 2013) the 2009 sham-controlled RCTs attracted the most attention and remained highly influential. (McConnell et al 2014 and Albers and Latchaw 2013) In an editorial published along with the 2009 RCTs, Weinstein states that "the results may change vertebroplasty from a procedure that is virtually always considered to be successful to one that is considered no better than placebo," though he acknowledges that these articles may not bring about a paradigm shift. (Weinstein, 2009)

The primacy of the 2009 New England Journal of Medicine sham-controlled RCTs is also reflected in citation counts. To date, the Buchbinder et al RCT has 1,612 Google Scholar citations and the INVEST Trial has 1,594. In contrast, the 2010 VERTOS II Trial in Lancet indicating that vertebroplasty is superior to medical management has 950 citations. The literature documenting the recent evolution of vertebroplasty procedure rates generally recognizes the 2009 RCTs as a turning point in the discussion about vertebroplasty effectiveness. (Laratta et al 2017; Lopez et al 2020). In their studies of medical reversal, Prasad, Cifu and colleagues identify the 2009 RCTs as

[7]Some social scientists may be surprised that studies that involve sham surgery can pass ethics review boards, however these vertebroplasty studies were not the first RCTs to use sham surgery as a control arm (see for example Moseley et al. 2002, who used this method to test the efficacy of knee arthroscopy). According to Miller (2003), sham surgery was rarely used as a control arm before 2000 because of ethics concerns. Discussions about the ethics of sham surgeries had mainly focused on a case in which the treatment involved transplanting tissue from aborted fetuses into the brains of Parkinson's patients. Miller (2003) argues that there should not be a blanket prohibition of sham surgery in medical research, but instead, the merits must be evaluated on a cases by case basis. He argues that past discussions have confused the ethics of medical research with those of clinical treatment: "first do no harm" does not apply to clinical research, because even in studies without sham surgery, trial patients are often exposed to risks without sufficient (or any) potential compensating benefits. Researchers should not aim to reduce participants' risk at all costs, but instead attempt to minimize risk subject to engaging the rigorous methods necessary to answer an important scientific question (Miller, 2003). Miller concludes that in some cases the use of sham surgery in research may be warranted, quoting Beecher (1961): "One may question the moral or ethical right to continue with casual or unplanned new surgical procedures - procedures that may encompass no more than a placebo effect - when those procedures are costly of time and money, and dangerous to health or life." Stock (2003) concurs with Miller but takes his argument further, arguing that rather than trying to avoid sham surgery, perhaps we should be trying to induce placebo effects (except when doing so requires overt deception.) A natural follow-up would then be how to optimally design placebo interventions in order to illicit the maximum benefit for patients. In the past couple decades a literature has developed around this line of inquiry. See Kaptchuk and Miller (2015) for a review.

the reversal point as well. (Prasad & Cifu 2011; Prasad et al 2013)

Given the evolution of the literature in favor of and against vertebroplasty, the publication of the two sham-controlled RCTs in 2009 in the New England Journal of Medicine marks a turning point in the evidence and this study follows the precedent set by the literature in defining that date as the evidence reversal event.

1.4 Data

1.4.1 Healthcare Data

The Medicare 20% files provide data on a random sample of 20% of all beneficiaries enrolled in fee-for-service Medicare. The data includes information about medical procedures and services received and diagnoses rendered, as well as rich demographic data about beneficiaries. Files from 2001-2014 allow tracking of beneficiaries and physicians over time. One of the advantages of using Medicare data relative to other claims data is that Medicare data includes claims from multiple settings: inpatient (hospital), hospital outpatient and non-hospital outpatient settings (doctors' offices, clinics, free-standing ambulatory surgery clinics, etc). This is important given recent trends in consolidation because we can observe procedure volume regardless of changes in ownership.

Beneficiaries were included in the analysis if they lived in the United States, were at least 65 years old but less than 95 years old, had both part A and part B coverage for the whole year, and qualified for Medicare due to age alone – rather than due to disability or End-Stage Renal Disease (ESRD). Records from beneficiaries age 95 and older are not included in the analysis due to well-known concerns that Medicare enrollment data includes many extremely aged beneficiaries that are not actually alive (West et al. 2010). Patients diagnosed with ESRD were excluded from the analysis due to their unique health needs and because their healthcare utilization is not representative of the average Medicare beneficiary. Because the procedure rate calculations come from claims data, it is important to exclude individuals whose claims may be unobservable in the data. Therefore in order to be included, a beneficiary must have had part A and B coverage, and in addition they must not have been covered through a managed care plan. During the study period of 2006 - 2014, managed care coverage represented a small but growing share of Medicare plans: 6.8% of Medicare beneficiaries in 2006, and 15.7% in 2014. (Freed et al., 2021) Table 1.1 shows how these sample restrictions affect our sample in 2008.

Since the population of interest is elderly Medicare beneficiaries who have painful vertebral compression fractures, it may seem strange not to further restrict the sample to include only people with this diagnosis. Doing so would likely increase the precision of the estimates, but this strategy is not pursued due to concerns of potential bias. In order for Medicare to accept a vertebroplasty claim, the MAC that processes the claim must determine that the procedure was medically necessary. Therefore it is standard for vertebroplasty procedure codes to be accompanied by compression fracture diagnoses on these claims. Physicians have an incentive to diagnose a vertebral compression fracture if they would consider performing vertebroplasty, however, there is little incentive to do so if the physician would not consider this treatment option. Alternatives to vertebroplasty include over-the-counter pain medications or low-dose opioids, which would not require a vertebroplasty diagnosis. Indeed, there is some evidence that vertebral compression fractures are underdiagnosed and there is also no universally agreed-upon definition of a vertebral fracture (Cummings and Melton 2002; Lenchik et al. 2004; Bottai et al. 2016). Ultimately, concerns about endogenous diagnosis patterns overshadow the potential gains from more preceisely targeting the population of interest.

Given that vertebroplasty can occur as an inpatient or outpatient procedure, it is crucial to analyze both inpatient claims and outpatient claims, which can appear in one or two of the following files: the Medpar file, the Outpatient file and the Carrier file. Which file contains a claim record depends on whether the claim is for physician services or facility fees, the practice setting and institutional ownership structure, and some obscure billing regulations. For example, due to vertical integration, claims for a given clinic may appear in only the Carrier file for some years, but in both the Carrier file and the OP file in other years. Using data from all three files ensures that claims from physician offices and free-standing ambulatory surgery clinics are included – not only hospital outpatient departments. Data from all three files were used to construct procedure rates, and in order to prevent double-counting, vertrobroplasties for a given patient that occurred within 4 days of each other were treated as a single event. The results are not sensitive to this duplicate-counting procedure: Table 1.14 does not drop procedures with dates within 4 days of each other, and the results are not qualitatively different.

Vertebroplasties were identified using ICD-9 procedure code 8165 in the MedPAR file (inpatient facility claims) and using the following CPT codes in the Carrier and Outpatient files: 22520, 22521, 22522, S2360, and S2361. In 2015 CMS transitioned to the ICD-10 coding system and the CPT codes for vertebroplasty changed. There is no one-to-one match between ICD-9 and ICD-10 codes. Numerous attempts to create crosswalks have encountered difficulties – discontinuities in procedure rates and diagnosis frequencies appear at the transition point, and these discontinuities are thought to be a data artifact rather than reflective of true health or utilization patterns. (Mainor et al 2019) Vertebroplasty has a single code in the ICD-9 system, but may match to as many as 14 ICD-10 codes. In addition, it often takes some time for medical coders to adapt to drastic coding changes such as the transition to the ICD-10 system - incorporating procedure rates from this transition period would likely add more noise. Given these difficulties and the small marginal value of one additional year of data, the analysis was stopped at the end of 2014. Data from before

2001 was not included because no vertebroplasties were found in the claims data in 2000. The ICD-9 and CPT codes for vertebroplasty did not change from 2004 to 2014.

In order to test the theory that locality rules are more important for rural areas, this study must define what constitutes a rural area. If a patient lives in a rural area and travels to a nearby urban area to have a procedure, what would matter for the application of the locality rule is the location of the physician's practice, rather than the location of the patient's home. [8] Therefore this study counts a procedure in the rural or urban rate calculations based on the location where it was performed rather than the location of the beneficiary's residence. In order to define rural vs urban areas, the 2010 Census ZCTA to Metropolitan and Micropolitan Area Relationship file was used. A Zip Code Tabulation Area (ZCTA) was defined as urban if at least 50% of its population belonged to one or more metropolitan statistical areas (MSAs). Non-urban ZCTAs were counted as rural. An MSA is a Core Based Statistical Area (CBSA) whose largest urban center has at least 50,000 residents. CBSAs are include an urban center and the surrounding counties that are socially and economically integrated with the urban center. Commuting ties are used to measure social and economic integration. A ZCTA may be located within a CBSA, or it may overlap with more than one CBSA. This definition is similar to the one used by CMS for their 2019 report on Rural-Urban Disparities in Healthcare in Medicare (CMS, 2019). CMS calculations show that the 21.5% of fee-for-service beneficiaries live in rural areas, which is similar to the fraction of our sample in rural areas (about 23%, see Table 1.2). Some studies exclude micropolitan areas (a CBSA whose urban care has a population of between 10,000 and 50,000 people) from their classification of rural areas. When it is important to accurately track populations who may have difficulty accessing healthcare services this would be a suitable definition. The definition used in this study is less concerned with differentiating between small or medium-sized towns and rural populations that may live on farms or in the countryside, and instead focuses mainly on removing the influence of cities, where a judge may find it more difficult to justify applying locality rules. To the extent that small and medium-sized towns have advanced centers for specialty care where a judge would likely find it reasonable to apply a national standard of care, this study's definition of rural would be overly conservative and would underestimate the effects of locality rules in rural areas.

Data from Florida and Maryland was dropped because it was not possible to classify these states according to whether they adhered to a local or national standard of care, as described below.

The AMA Masterfile covers physicians from a wide range of specialties and records information about their training, age, years of experience, medical specialties and sex. Information about physician race is not available. In addition to the rich information about physician training, another

[8] In the rare instance where a malpractice suit is brought against a doctor who practices in one state by a patient who lives in a different state, this is referred to as a case involving "diversity jurisdiction." The case usually proceeds through the federal court system, and the law that applies is the law of the state where the petition was filed (Erie Railroad Co. v. Tompkins, 1938

advantage of the AMA Masterfile is the ability to distinguish active physicians from retired, semi-retired and inactive physicians. This distinction allows for a more accurate measure of how the number of specialists changes over time in a state. Physician-level analysis requires linking UPIN physician IDs from the early 2000s to NPI IDs, which were implemented in mid-2007. To do this, UPIN numbers for physicians who performed at least one vertebroplasty before 2009 were extracted from the data and compared to UPIN-NPI entries in the CMS National Downloadable file (formerly part of Physician Compare) and to a "claims generated cross-walk" created as part of this project. The claims-generated crosswalk is based on a subset of claims filed during the UPIN to NPI transition period that include both the NPI and the UPIN. This cross-walk is based claims for a variety of medical services, not exclusively vertebroplasty or related procedures. Over 90% of the physician UPINs on vertebroplasty claims were successfully matched to NPI IDs.

1.4.2 Methodology – Constructing the Legal Dataset

There is no central repository of legal standard of care definitions by state. In order to determine how each state's standard of care definitions have evolved since 2000, I searched each state's statutes and case law in NexisUni. As a first step, statutes addressing the standard of care (where available) were analyzed to determine whether the state's standard of care has a locality dimension and whether the standard of care differs for general practitioners and specialists. Past versions of the statutes were inspected to determine whether these aspects of the standard of care have changed over time.

Many states do not have statutes that define the standard of care. In these states the standard of care is defined only in the case law. The next step required searching for medical malpractice court cases in each state from 2000-2015. Even in states where statutes contain standard of care definitions, identifying court cases where the statute was referenced is necessary to form a more complete understanding of how the law is interpreted and applied. Key words such as "standard of care," "physician" and "medical malpractice" were used to identify relevant cases. Depending on the number of cases identified for each state in the relevant time period, search results were sometimes narrowed to include only cases from the state's highest court (referred to as the Supreme Court in most states). Once relevant cases were identified, references within the case to past decisions were investigated, and the relevant cases (and sometimes headnotes) were also Shepardized in order to track how later courts explained, interpreted, affirmed, distinguished or overturned the interpretation of the standard of care. Shepardization is an algorithm in NexisUni that facilitates analysis of how common law evolves over time. Shepardizing a case allows the user to view later decisions that cite it and to more easily identify whether the citing decision treats the case neutrally, positively, or negatively. Shepardization displays cases from the same court, other courts within the state, other states, and federal cases. It can be viewed as a kind of "Google Scholar" for court cases. In states that have annotated standard of care statutes in NexisUni, relevant cases mentioned in the annotations were also considered and often Shepardized.

The third step involved searching federal case law. In the rare instance where a malpractice suit is brought against a doctor who practices in one state by a patient who lives in a different state, this is referred to as a case involving "diversity jurisdiction," and the case typically proceeds through the federal court system. Erie Railroad Co. v. Tompkins (1938) established that the law that applies is the law of the state where the petition was filed, so information about a the filing state's standard of care can often be found in the opinions of these diversity cases. The majority of the state classifications in this study are based on state rather than federal court cases because state court cases are far more common and are more informative about how the state's law should be interpreted; state Supreme Courts are the final arbiters of state law. However it was necessary to consult federal case law for a few states when there were very few relevant cases from the state's Supreme Court and Appellate Courts (see for example the discussion of Hawaii below).

Figure 1 and Figure 2 show how standard of care definitions vary by state for specialists and non-specialists, respectively. Legal research revealed that during the study period no states changed their standard of care from local to national customs-based or vice versa, so in Section 1.4.3 below locality status will vary by state but not over time. Twenty states have some form of the locality rule in their standard of care definition for non-specialists and the remaining states have a national standard of care. While the division is similar for the specialist standard of care, several states that reference local customs in the non-specialist standard of care defer to national standards for specialists: Colorado, Louisiana, Michigan, Minnesota, Montana, Pennsylvania and South Dakota. Vertebroplasty is primarily performed by radiologists, anesthesiologsits, orthopedic surgeons and neurosurgeons, so specialist standards of care apply for this analysis. (While courts have disagreed about precisely what kinds of physicians should be classified as specialists, the disagreements often involve less clear-cut cases such as residents training in a particular specialty who have not yet obtained board certification.)

Unfortunately Florida and Maryland were impossible to classify. In Maryland there is an unresolved conflict between statutory and case law. According to Shilkret v. Annapolis Emergency Hosp. Ass'n, 349 A.2d 245 (Md. 1975) Maryland has a national standard of care. However a 1993 statute – Md. Cts. & Jud. Proc. Code Ann. §3-2A-02(c) – states that Maryland has a "same or similar community" standard. Court cases cite both Shilkret v Annapolis and the 1993 statute without clearly stating which standard of care holds (see for example Dingle v. Belin, 358 Md. 354 (2000)). To date, neither that statute nor the Shilkret case has been overturned. Williams (2012) reviews the Maryland case law in more detail.

In Florida, statute 766.102 (2000) stated that a same or similar community standard of care applies for non-specialists and was silent about the issue for specialists, which could be interpreted as tacitly supporting a national standard of care for specialists. However in 2003 the same or similar community language was struck from the statute altogether, along with a number of other substantive revisions. No language about a national standard of care was added, and some subsequent court cases continued to cite both the statute and the same or similar community standard that was applied in earlier court cases (see for example Perez v. United States, 883 F. Supp. 2d 1257 (2012)). Given the inconsistent case law it was not possible to categorize Florida as a locality state or a national standard state.

Hawaii was also difficult to categorize because no state statute governs the standard of care and the case law is very sparse. In McBride v. United States, 462 F.2d 72 (9th Circuit Court of Appeals, 1972) the court states that they "feel confident the Supreme Court of Hawaii would follow the American Law Institute formulation," and adopt a locality based standard of care. However this did not occur – whether in subsequent state court cases or by statute – so the law remains unclear. An independent analysis by a team of legal researchers classified Hawaii as a national standard state due to the "absence of challenges to experts on locality grounds," so this study follows that approach. [9] (See robustness section 6 for more details.)

1.4.3 Empirical Analysis

Correlates of de-adoption

The first part of the results section explores which factors are correlated with vertebroplasty de-adoption using a series of descriptive regressions and difference-in-differences and event study specifications. As a baseline, we first estimate the following equation to see how much vertebroplasty rates decreased after the evidence reversal:

(1) $VertRate_{ist} = \alpha_0 + Post_t + Quarter_t + \delta_s + \epsilon_{st}$

Where $VertRate_{ist}$ is the vertebroplasty rate (per thousand people) in rural or urban region i within state s at time (year and quarter) t. $Quarter_t$ and $delta_s$ are quarter and state fixed effects, respectively. $Post_t$ measures the average change in vertebroplasty rates after the evidence reversal in all states.

We next turn to a difference-in-differences style specification to asses rural-urban differences in de-adoption:

[9] I am extremely grateful to Professor Kyle Logue and the research librarians at the University of Michigan Law School for carrying out this independent analysis to confirm my classification.

(2) $VertRate_{ist} = \alpha_0 + Rural_i + RuralPost_{it} + \gamma_t + \delta_s + \mathbf{X'_{st}\Gamma} + \epsilon_{st}$

$Rural_i$ is an indicator variable that equals one if region i is rural. $RuralPost_{it}$ equals 1 for rural regions after the evidence reversal, and measures the average difference in vertebroplasty rates between rural and urban areas after the evidence reversal. This coefficient should not be interpreted as the causal effect of the evidence reversal on de-adoption in rural areas (relative to urban areas) because factors other than the evidence reversal may have affected vertebroplasty rates differentially in rural areas, such as hospital consolidation. $\mathbf{X_{st}}$ is a vector of covariates that vary at the state-year level, generally including the percent of Medicare beneficiaries who are male, the percent who are Black, and measures of the age distribution. For now I do not control for the age distribution. γ_t are year and quarter fixed effects and all other variables are as defined above.

The following two specifications explore patient age as a determinant of de-adoption.

(3) $VertRate_{iast} = \alpha_0 + Post_t + AgeGroup_a + Quarter_t + \delta_s + \epsilon_{st}$

(4) $VertRate_{iast} = \alpha_0 + AgeGroup_a + AgeGroupPost_{at} + Quarter_t + \delta_s + \epsilon_{st}$

$VertRate_{iast}$ is the vertebroplasty rate for age group a in rural or urban region i of state s at time t. $AgeGroup_a$ is a set of 5 indicator variables for the 5-year age groups between age 70 and 94 (age 65-69 is the reference group), and $AgeGroupPost_{at}$ represents the set of age group and post interactions. (3) estimates the average "within age-group" decrease in vertebroplasty rates after the reversal, which serves as a baseline for the heterogeneity by age group analysis in (4).

Equation (5) below explores whether de-adoption rates differ among female vs male patients.

(5) $VertRate_{ijst} = \alpha_0 + Female_j + FemalePost_{jt} + \gamma_t + \delta_s + PercentBlack_{st} + \epsilon_{st}$

where i indexes rural vs urban regions, j indexes men or women, s indexes states and t indexes time (quarter of year).

The next set of equations investigate the relationship between de-adoption and various supply-side factors: physician gender, years of experience and physician specialty.

(6) $VertRate_{dt} = \alpha_0 + \alpha_1 Variable_d + \beta VariablePost_{dt} + Year_t + \delta_s + \mathbf{X'_{st}\Gamma} + \epsilon_{dt}$

where $VertRate_{dt}$ is the vertebroplasty rate for doctor d in year y, $Year_t$ are year fixed effects, and δ_s are state fixed effects. Standard errors are clustered at the state level. $Variable_d$ is a physician characteristic: gender, years of experience, an indicator for early career physicians and an indicator for experienced physicians. Doctors were included in these regressions if they performed at least one vertebroplasty before 2009. If a doctor submits no vertebroplasty claims in 2010-2014, it is difficult to determine whether they have retired or have simply ceased performing vertebroplasties. Therefore as a crude indicator of retirement, doctors who turned 65 years old before 2015 were dropped. Years of experience is defined as time elapsed since the physician graduated from medical school (as of 2009). (Note that as defined here, years of experience varies among physicians but not by time.) A specialist was classified as an early career physician if, as of 2009, they had completed medical school within the past 15 years. Radiologists, neurosurgeons, orthopedic surgeons and anesthesiologists undergo several years of further training as residents after completion of medical school, so this group of physicians would have been in training during the period when vertebroplasty was rapidly growing in popularity. The indicator for experienced physicians equals one for specialists who had graduated medical school at least 30 years prior to 2009. The following equation is used to explore whether de-adoption rates differed by physician specialty:

(7) $VertRate_{dt} = \alpha_0 + \alpha_1 Specialty_d + \beta SpecialtyPost_{dt} + Year_t + \delta_s + \mathbf{X'_{st}\Gamma} + \epsilon_{dt}$

Four specialties are common among physicians who perform vertebroplasty: radiology, orthopedic surgery, anesthesiology and neurosurgery. $Specialty_d$ is an set of indicators, each of which equals 1 if the physician belongs to the given specialty. Radiology is the ommitted reference group. $SpecialtyPost_{dt}$ interacts the specialty indicators with $Post_t$, so $SpecialtyPost_{dt}$ equals 1 if a physician belongs to the given specialty during a period after the evidence reversal.

Empirical specifications for main results

This study relies on difference-in-differences and event study methods to assess whether physicians in locality and non-locality states respond differently to evidence reversal. There are three main specifications: the first speaks to whether locality matters at the state level and the other two address to what extent locality-based standards of care have a differential effect in sub-state rural areas. I first estimate the following equation:

(8) $VertRate_{st} = \alpha_0 + \beta LocalityPost_{st} + \gamma_t + \delta_s + \mathbf{X'_{st}\Gamma} + \epsilon_{st}$

Where s indexes states and t indexes time (measured by quarter of year). Quarterly data and quarterly fixed effects are used whenever possible because past work has found that osteoporosis likely exhibits seasonality (Wang et al., 2021). [10] $VertRate_{st}$ is the vertebroplasty rate for state s at time t, measured in number of procedures per 1,000. $LocalityPost_{st}$ is an indicator variable that equals 1 after the evidence reversal for states that have a local standard of care for specialists. The evidence reversal event is defined as the publication date of the 2 RCTs in the third quarter of 2009 which showed that vertebroplasty is no more effective than a placebo surgery. Vertebroplasty rates grew very quickly during the early 2000s and stabilized in about 2006, so most specifications use data beginning in 2006 rather than in 2002. This helps ensure that the estimates are not partially picking up the effects of adoption trends and incentives. Most specifications include year and quarter fixed effects γ_t and state fixed effects δ_s Standard errors are clustered at the state level. $\mathbf{X_{st}}$ is a vector of covariates that vary at the state-year level. For now I control for the percent of Medicare beneficiaries who are male, the percent who are Black, and measures of the age distribution. β measures the differential decline in procedure rates in locality states compared with national standard states. That is, if β is positive, then locality rules slow de-adoption by beta procedures per thousand patients. If β is negative, then locality states de-adopt more quickly.

Given the historical context surrounding the promulgation of locality rules it seems likely that they would have a far greater impact in rural areas. Traditionally, in arguments used to defend the locality rule courts have explicitly cited the less advanced skill of the "country doctor" due to unequal training opportunities and low patient volume, as well as other disadvantages of rural areas. (See for example Small v Howard, 128 Mass. 131, one of the most famous cases involving the locality rule.) In addition, to the extent that locality rules have any weight in urban malpractice proceedings, the procedural constraints imposed by the locality rule would be less likely to bind. If an expert witness's standing or qualifications to testify are challenged because the opposing side claims that they do not have sufficient familiarity with the local practices and standards of care, it would be easier for the witness to research and demonstrate familiarity with practices in an urban center than in a small rural town. Given the larger number of doctors practicing in urban areas, the expert witness may have colleagues who have practiced there, or may have interacted with doctors from the urban area in question at conferences or professional meetings. Prior research on the effect of locality rules also supports this claim: Frakes, Frank and Seabury (2017) find that the effects of locality rules on physician supply are three times larger in rural areas. Therefore in addition to examining the effect of locality rules viewed as a state level treatment that applies to all physicians in the state, I will also investigate whether the effect of the locality rule is larger in rural areas, using the following two specifications:

[10] Wang et al. (2021) compared Google searches for osteoporosis over time in 4 English-speaking northern hemisphere countries and 2 English-speaking southern hemisphere countries and found that searches peak during winter months and fall to lower levels during summer months. Seasonality in osteoporosis searches may be due to seasonality in hip fractures, vertebral fractures or both. The physician level regressions above use yearly data because vertebroplasty becomes very rare at the physician-quarter level.

$$(9) VertRate_{ist} = \alpha_0 + \beta TreatedPost_{ist} + \gamma_t + \delta_s + \mathbf{X'_{st}}\mathbf{\Gamma} + \epsilon_{ist}$$

$$\begin{aligned}(10) VertRate_{ist} = \alpha_0 + \alpha_1 Rural_i + \alpha_2 Locality_s + \alpha_3 RuralLocality_{is} + \alpha_4 LocalityPost_{st} \\ + \alpha_5 RuralPost_{it} + \beta RuralAreaLocalityStatePost_{ist} + \gamma_t \delta_s + \mathbf{X'_{st}}\mathbf{\Gamma} + \epsilon_{ist}\end{aligned}$$

Where i indexes rural or urban regions within state s and t indexes a year and quarter. $VertRate_{ist}$ is the vertebroplasty rate for sub-state region i in state s at time (year and quarter) t, measured in number of procedures per 1,000. (9) is a difference in differences specification and (10) is a triple differences specification. (9) is similar to (8) but I disaggregate so that the unit of observation is a sub-state region i in state s at time t. $TreatedPost_{ist}$ is an indicator variable that equals 1 after the evidence reversal for rural areas in states that have a local standard of care for specialists. Depending on the specification, the control group could be 1)urban areas within locality states, 2) rural areas within national standard states, or 3) both.

Although the terms "difference-in-differences" and "triple differences" are used repeatedly to refer to equations 8-10, these specifications do not technically fit the standard difference-in-differences paradigm. The most natural difference-in-differences framework that would estimate the causal impact of locality rules would be to compare states that decide to abandon locality rules to those that don't, before and after such rules are abandoned (in favor of a national standard of care). Thus changes in the laws governing the standard of care would serve as the identifying variation. However there are practical and theoretical concerns with using this estimation strategy for this research question. As a practical matter, there were no changes in states' standard of care laws during this time period. Of theoretical concern for this context is that Frakes (2013) demonstrated that when states adopt a national standard of care (abandoning a locality rule), their procedure rates converge to national average rates. That is, the legal shock induces locality states to move to a new equilibrium with procedure rates that more closely resemble those in national standard states as physicians adapt to the national standard of care. However this convergence in procedure rates occurs in the absence of any evidence reversal. Following an evidence reversal the healthcare market is already in transition: procedure rates (potentially in all 50 states) are falling from their pre-reversal steady state rates to their new lower post-reversal steady state rates. It is theoretically ambiguous how vertebroplasty rates would evolve immediately after a legal shock given that the healthcare sector would already be in a state of transition: there would be no uniform "national standard of care" to converge to. Of course eventually national standard states will reach a new post-reversal steady-state equilibrium with a new standard of care. But without making strong assumptions such as perfect information and rational expectations that are likely untenable in this context it is unclear how physicians would be able to foresee what the new equilibrium standard of care (proxied for by the level of procedures) would be. [11]

Given that the classic difference-in-differences approach exploiting variation in timing of state standard of care reforms is not viable in the context of evidence reversals, this study turns instead to the perspective of heterogeneous treatment effects and asks the following question: if effects of the evidence reversal differ between local and national standard states, under what assumptions is this heterogeneity likely to reflect a causal effect of the standard of care on de-adoption? The main identification assumptions for the state level specification (8) are that had the evidence reversal not occurred, vertebroplasty rates would have evolved similarly in locality and non-locality states (parallel trends). It must also be the case that no other events or policies occurred at the same time as the evidence reversal that differentially affected the locality and non-locality states and that are correlated with the outcomes of interest. If there is an unobserved variable that is highly correlated with states' standard of care definitions, then the estimates of β that come from using national standard states as a control group would not be able to separately identify the true effect of locality-based standards of care from the effect of this unobserved variable. Assumptions for specification (9) are similar, substituting the appropriate treatment and control group definitions, and are discussed in detail in section 1.5.2. In specification (10) I address the concern that an unobserved variable could be correlated with state standard of care definitions and vertebroplasty rates using a triple differences specification that differences out state level time-varying unobserved variables.

An identification assumption that underlies all the specifications that aim to estimate a causal effect of the locality rule on de-adoption is that the locality rule only affected the evolution of procedure rates over time *after* the evidence reversal. In other words, the treatment is having a locality rule in effect after the evidence reversal; locality states (or, alternatively rural areas in locality states) are not "treated" before the evidence reversal. If the locality rule were already affecting vertebroplasty rate dynamics before the evidence reversal in 2009, there would be no reason to expect that vertebroplasty rates would have evolved similarly in locality areas and national standard areas had the reversal not occurred. A static difference in the rates of vertebroplasty before the reversal would not pose a problem; however differential changes in the rates (that is, a lack of parallel trends) would. Frakes (2013) finds evidence that after states adopt national standards of care, procedure rates for C-Sections and some cardiac procedures converge towards national average rates, and 30-50% of the gap disappears. However during the period of interest for this study, no states had standard of care reforms; in fact, standards of care remained largely unchanged

[11] Some would argue that given the high levels of regional variation observed by past researchers and the limited amount of this variation that can be explained by demand-side factors, the concept of a "national standard of care," is a fiction.

since the mid-late 1990s. Medicare recognized billing codes for vertebroplasties in 2002 and 2004 (for CPT codes and ICD-9 codes, respectively), and the rapid growth in Medicare vertebroplasty rates appears to have plateaued around 2005 or 2006 (Figure 5). Locality rules may have affected adoption dynamics, but it seems unlikely that they would have continued to affect the evolution of vertebroplasty rates once rates stabilized in 2006. If other variables were influencing vertebroplasty rate dynamics in a way that differentially affected treatment and control states (or sub-state rural vs urban areas) before the evidence reversals occurred, this should be visible in the coming event study figures (see Section 1.5.2). One advantage of the triple differences specification is that in order for state level medical malpractice reforms to bias the estimated effect of locality, the reforms would have had to have had effects that differ in rural and urban areas.

A final issue for the empirical analysis is the possibility of correlated standard errors at the state level. Failure to adequately account for the clustering structure can lead to over-rejection in hypothesis testing and spurious results. (Bertrand et al. 2004) Results in this study are already clustered at the state level, however in some simulations the wild cluster bootstrap performs better when the number of clusters is not sufficiently large. Although this is not a severe case - over-rejection is far worse when the number of clusters is less than 20 - there is no universally agreed-upon cut-off for what constitutes “too few” clusters. The wild cluster bootstrap has been proposed as a alternative when the number of clusters is smaller than ideal.(Cameron, Gelbach and Miller 2008) As a robustness check the main specifications from Table 1.6 are re-estimated using the wild bootstrap method. The p-values from the wild bootstrap method are very similar, therefore it does not appear that these results are a product of over-rejection from a failure to account for the clustering structure.

1.5 Results

1.5.1 Descriptive Statistics and correlates of de-adoption

Table 1.2 contains statistics for national vertebroplasty utilization in 2008 and in 2014. In both time periods, the majority of vertebroplasty recipients were white and female. The literature finds a similar gender distribution (Hazard et al 2014 and Mehio et al 2011). Patients age 80-84 comprised the largest share of vertebroplasty patients, and the age distribution remained relatively stable over time. All measures of vertebroplasty utilization decreased significantly between 2008 and 2014. The number of physicians who performed at least one vertebroplasty per year fell from 1,474 to 907 (Figure 3). The share of vertebroplasties performed by vascular and interventional radiologists, neurologists and neurosurgeons increased between 2008 and 214, while the share performed by diagnostic radiologists, anesthesiologists and orthopedic surgeons decreased. (Figure 4).

Locality standard states adopted vertebroplasty more quickly and then reached higher vertebroplasty rates than states with a national standard of care (Figure 5). In the first year that I observe Medicare vertebroplasty claims, vertebroplasty rates were about 0.15 per thousand both in states with a locality standard of care and in states with a national standard. Within 2 years, vertebroplasty rates in locality standard states nearly doubled, whereas vertebroplasty rates in national standard states rose more conservatively, to about 0.2 per thousand. After the 2 RCTs came out in 2009, vertebroplasty rates decreased sharply in both locality standard states and states with a national standard of care. Faster initial growth rates in locality states is not surprising, since the locality rule could provide more flexibility for physicians to depart from standard national practices. From Figure 5, it appears that vertebroplasty rates stabilize and the period of rapid adoption ends around 2006, so most subsequent specifications will examine vertebroplasty rates beginning in 2006.

Next, we disaggregate the state level vertebroplasty rates used in Figure 5 to explore heterogeneity by patient age, sex and by rural vs urban practice setting. We first disaggregate so that an observation is at the state-urban/rural-year quarter level (equation 1). In Column 1 of Table 1.3 we see that on average vertebroplasty rates fell by 0.193 per thousand, from a pre-reversal average rate of 0.38 per thousand. Keeping the same level of observation and using an interaction term to explore rural-urban de-adoption differences (equation 2), in Column 2 we see that on average, rural areas do not appear to be more likely to de-adopt slowly, and the magnitude of the relevant coefficient is also small when compared to the effect that age has on a person's risk of vertebroplasty (columns 3 and 4). Next we disaggregate further so that an observation is at the state by rural/urban by age group by year quarter level. Estimates from equation 3 in Column 3 of Table 1.3 shows that on average, age-specific vertebroplasty rates fell by about 0.261 per thousand after the evidence reversal (from a pre-reversal rate of 0.51 per thousand), however this masks substantial statistically significant heterogeneity by age group. The relative decline was concentrated among younger Medicare beneficiaries, with smaller declines among beneficiaries age 85-94.. The estimates in Column 4 (using equation 4) show that vertebroplasty rates decreased by 0.05 per thousand for Medicare enrollees age 65-69 (54% of their pre-reversal mean rate), 0.15 per thousand for 70-74 year olds (62% of their pre-reversal mean rate), 0.27 per thousand for 75-79 year olds (54% of their pre-reversal mean rate), 0.48 per thousand for 80-84 year olds (62% of their pre-reversal mean rate), 0.30 per thousand for 85-89 year olds (43% of their pre-reversal mean rate) and 0.32 per thousand for 90-94 year olds (42% of their pre-reversal mean rate). Note that beneficiaries ages 90-94 had the highest pre-reversal vertebroplasty rate (76 per thousand) but the smallest percentage decrease in response to the reversal. One theory that could explain the slower de-adoption in the oldest age groups is if physicians are more reluctant to refuse to perform vertebroplasty for the sickest patients. For very elderly populations, physicians may prioritize pain alleviation and quality of life, even if benefits are short-lived and arise through a placebo effect. In addition, if vertebroplasty primarily functions through a placebo effect, some of these patients may have already had vertebroplasty and reported reduced pain, making them likely to have a positive placebo response again.

We now turn to heterogeneity by patient gender (equation 5). Unsurprisingly given that osteoporosis disproportionately affects women, before the evidence reversal women were more likely to undergo vertebroplasty than men (Column 5 Table 1.3). However vertebroplasty rates among women showed a greater response to the evidence reversal than for men – a drop of about 0.10 vertebroplasties per thousand, relative to the relatively flat rates for men. This represents 28% of the pre-reversal mean vertebroplasty rate of 0.36 per thousand (at this level of observation).

Table 1.4 uses equations (6) and (7) to explore the relationship between various supply-side factors and de-adoption: physician gender in Column 1, years of experience in Columns 2-4 , and a vector of indicator variables for physician specialties in Column 5. The corresponding event studies are in Figures 6-11. Before the evidence reversal, female physicians were statistically less likely to perform vertebroplasty than their male counterparts, however they also de-adopted more slowly, performing an average of about 0.005 more vertebroplasties per thousand after the evidence reversal than men. Vertebroplasty rates decreased for both male and female physicians after the evidence reversal, and the coefficients in Figure 6 show the relative increase (slower decline) in vertebroplasty rates for female physicians compared to male physicians. It is worth noting that women are remarkably underrepresented in the specialties that typically perform vertebroplasties, comprising only 4.3% of the 3,098 physicians in the sample. Note that the pre-reversal mean vertebroplasty rate in Table 1.4 is about 0.016 per thousand, so an association magnitude of 0.005 is substantial – an effect size of about 31% of the baseline mean.

Columns 2- 4 of Table 1.4 investigate the role of experience in physician de-adoption behavior. Years of experience, defined as time since graduation from medical school as of 2009, does not have a measurable impact on deadoption: the coefficient is both tiny and not statistically significant. It is possible that while years of experience in general does not matter, physicians who witnessed the rise of vertebroplasty early in their career may respond differently to the evidence reversal. Their beliefs about vertebroplasty effectiveness may be less firmly cemented, so they may be more likely to reduce their use of vertebroplasty. On the other hand, they may believe that the marginal benefits of time spent reviewing recent medical literature and guidelines for new evidence is small given that their training occurred recently so there may be informational barriers to de-adoption. Column 3 focuses on a cohort of early career physicians who would have been in training during the period when vertebroplasty was rapidly growing in popularity. Among these physicians verterboplasty rates fell by 0.002 per thousand less than they fell for more experienced physicians after the evidence reversal (13% of the baseline mean vertebroplasty rate). Unfortunately the event study for early career physicians shows some pre-trends, indicating that this is likely not a reliable estimate of a causal effect. Prior studies have found that experienced physicians respond more to negative evidence and scale back more quickly. (See Bekelis et al, 2017) In the context of vertebroplasty this does not appear to be true – among physicians who had graduated medical school at least 30 years prior to 2009, vertebroplasty rates fell by less than among the general physician population. The effect size of 0.0029 per thousand represents about 19% of the baseline mean vertebroplsaty rate. Estimates from equation (7) show that compared to radiologists (the omitted specialty), anesthesiologists, neurosurgeons and orthopedic surgeons performed fewer vertebroplasties before 2009 and had a diminished response to the evidence reversal. The effects were large and statistically significant for neurosurgeons (0.008 per thousand, 27% of the baseline mean) and orthopedic surgeons (0.007 per thousand, about 50% of the baseline mean).

1.5.2 Effect of locality-based standards of care on de-adoption

State-level difference in differences

As a benchmark, and because standard of care definitions are decided by state legislatures or state-level courts, the first specification in Column 1 of Table 1.5 models locality rules as a state-level treatment and uses a state-year-quarter as the level of observation (specification 8 in empirical methods section 1.4.3). In this specification, both rural and urban areas within a locality state are considered treated so the state level procedure rate includes procedures performed in both areas. Difference in differences comparing locality states and national standard states before and after the reversal fail to detect any statistically significant difference in de-adoption between locality states and states with a national standard of care. The corresponding event study in Figure 12 shows that the lag three periods (quarters) before the evidence reversal is statistically different from zero, and the statistically insignificant difference in differences estimate is due to noisy period-specific coefficient estimates rather than a product of dynamic effects over time. As a robustness check, the data is aggregated from quarterly to yearly state procedure rates and the estimates are similar: no statistically significant difference in de-adoption between locality and national standard states (coefficient -0.15, confidence interval [-0.41, 0.12]).

Columns 2 and 3 of Table 1.5 show how the measured effect of the locality rule changes as the rural-urban distinction is incorporated to various degrees. Column 2 of Table 1.5 disaggregates the state procedure rates into rural and urban rates, but the effect of locality is not allowed to vary between rural and urban areas. To the extent that the effect of locality differs for rural and urban areas, the estimate in Column 2 averages these two effects, giving equal weight to rural and urban areas. Given that a larger share of the population lives in urban areas, the disaggregated specification in Column 2 places more weight on the effects of locality in rural areas compared to the estimates in Column 1. Placing more weight on the effects in rural areas changes the sign of the coefficient: the association between the locality rule and post-reversal procedure rates becomes small and positive. Column 3 of Table 1.5 models locality-based standards of care as a treatment that only affects rural areas of locality states, and there are three control groups: rural areas of national standard states and urban areas of both national standard states and locality states. The estimates indicate that before the reversal, vertebroplasty rates in rural areas of locality states were 0.314 per thousand lower than in rural areas of national standard states. However after the evidence reversal, procedure rates were 0.109 per thousand higher than in the other control group areas, suggesting that de-adoption occurred more slowly in rural areas with a locality rule.

Given the history of the locality rule and the context in which courts apply it today, it seems more likely to carry weight in rural areas of a state than in modern urban areas. In addition, as a matter of procedural law it is much easier and less costly for an expert witness to acquire sufficient information to demonstrate familiarity with the standard of care and the practice environment in an urban center, so the locality rule is less likely to be a constraint that binds in urban areas. The following analysis continues to focus on vertebroplasty rates that are disaggregated to the sub-state level (rural vs urban rates with a given state) in order to investigate whether the locality rule slows de-adoption in rural areas, both as an additional substantive question of interest and as a means of reducing potential noise from including regions where the locality rule is likely to have less effect.

Disaggregating to sub-state (rural vs urban) areas

Table 1.6 turns to a set of difference in differences and triple differences specifications in order to estimate causal effects of locality-based standards of care on de-adoption in rural areas. The difference-in-differences specifications in Columns 1 and 3 correspond to equation 9 and the triple differences estimates in Column 3 correspond to equation 10. The dependent variable in all columns of Table 1.6 is the vertebroplasty rate (per thousand Medicare beneficiaries) in sub-state (rural or urban) area r in state s in year y and quarter q. There are two plausible difference in differences specifications that can estimate the effect of locality based standards of care on rural areas: 1) differences in how rural and urban procedure rates react to evidence reversal within locality states (where states with a national standard of care are excluded from the analysis), and 2) differences between rural areas of locality states and rural areas of national standard states (excluding all vertebroplasties in urban areas from the analysis). Results from the former "within locality states" specification are displayed in Column 1 and the corresponding trends graph is Figure 13. The coefficient on $Rural \times Post$ reflects the causal effect of locality rules on de-adoption under the following assumption: in locality states, if the rural areas had had a national standard of care, procedure rates would have evolved similarly in urban and rural areas. Stated differently, if these states had had a national standard of care, procedure rates would have evolved similarly in rural and urban areas before and after the reversal. This approach views urban areas as "untreated," or unaffected by locality rules so they serve as a control group. The estimates in Column 1 show that locality-based standards of care decrease vertebroplasty de-adoption by about 0.104 vertebroplasties per thousand (28% of the baseline mean). The corresponding event study is displayed in Figure 14 shows that none of the pre-treatment coefficients are statistically different from zero and although they are not precisely estimated, the point estimates are tightly clustered around zero. The Column 1 difference in difference estimate reflects a gap between rural and urban areas that emerged in the second and third quarter of 2010 and continued to widen over time.

The second difference in differences specification introduced above compares rural areas of locality states to rural areas of national standard states, before and after the evidence reversal (Table 1.6 column 2). In order for the coefficient on $Locality \times Post$ to represent a causal effect, one must assume that had the rural areas in locality states had a national standard of care, procedure rates would have evolved similarly to rates in rural areas in national standard states after the reversal. Column 2 shows that locality rules decrease de-adoption by 0.153 procedures per thousand compared to rural areas in national standard states, although this effect is not statistically significant. The trends graph corresponding to this cross-state difference in differences specification is Figure 16. The event study in Figure 15 shows that none of the pre-reversal coefficients are statistically different from zero, and there doesn't appear to be a clear pre-trend. While the point estimates on 4 of the lead coefficients are substantively different from zero and lower than the 4 preceding coefficients, the confidence intervals on these estimates are much wider than the other leads (about twice as large) and still do include zero. The event study shows that to the extent that rural areas of locality states may have de-adopted more slowly than rural areas of national standard states (mid-2010 through early 2013) this gap was short-lived. As a robustness check, Table 1.9 reproduces these estimates where the level of observation is age group a in rural or urban area r in state s in year y and quarter q, and the results are similar.

There are several reasons to be cautious in assigning a causal interpretation to either of these difference in difference estimates. The first specification compares rural and urban areas within locality states, however de-adoption behavior in rural and urban areas might differ for reasons other than differences in the strength of the locality rule. Since the early 2000s the rural healthcare sector has been shrinking. In particular, a growing number of rural hospitals have closed: 25 between 2005 and 2010 and 138 since 2010. (Sheps Center for Health Services Research, 2022; see also Kaufman et al. 2016) Failure to account for these supply shocks would bias my Column 1 estimate of the locality effect downwards. In addition, differential effects of the opioid epidemic may be a concern. Opioid overdose-related mortality rates are not significantly higher in rural areas on average, however mortality rates have grown much faster in rural areas. (Monnat and Riff, 2018) Perceptions of the opioid epidemic may affect physician willingness to attempt to alleviate pain through vertebroplasty. To the extent that rural physicians were becoming more willing to attempt vertebroplasty as an alternative to opioids, my Column 1 estimates of the effect of the locality rule would be biased upwards. In addition, rural physicians may differ from urban physicians on observable or unobservable dimensions. These differences could be due to selection issues, or they could develop over time as a physician adapts to local practice norms. (See for example Molitor 2016) If healthcare quality improves more quickly in urban areas and higher quality is associated with a greater responsivesness to evidence, failure to control for this trend would also lead to upward bias my Column 1 estimate (overstating the effect of the locality rule). Even if further controls are added to attempt to account for these differences, problems will arise if there isn't sufficient overlap in the support of the relevant variables between urban and rural physicians. In addition, Dingel et al. (2022) find that for less-common procedures, home-market effects tend to develop in larger regions, and more populous areas are net-exporters of healthcare services. [12] Regions that serve a larger patient population allow physicians to specialize and improve quality, especially for relatively rare procedures. If physicians who specialize in vertebroplasty have a higher opportunity cost of de-adoption, then urban areas would de-adopt more slowly, so the rural-urban difference in Column 1 could underestimate the true effect of the locality rule. Increasing returns to healthcare production could also result in dynamic effects that could lead to different trends in vertebroplsty rates between rural and urban areas, however this does not appear to be a problem from Figures 13 and 14.

Causal inference arising from the second difference in differences specification, where rural areas of locality states are compared to rural areas in national standard states, may also be problematic. Since standard of care definitions are not randomly assigned, locality standard states may differ from national standard states in ways that systematically affect vertebroplasty de-adoption trends. Medicare physician providers are assigned to a Medicare Administrative Contractor (A/B MAC), an entity which processes Part A and Part B Medicare claims, coordinates billing and handles preliminary coverage appeals. Medicare continued to reimburse physicians for vertebroplasy after the evidence reversal, however MACs could have made it more difficult for physicians to bill for vertebroplasty and examined patients' medical record more closely. In addition, between 2006 and 2010 CMS consolidated the number of A/B MACs, combining two or more states into a single MAC service area. With a few exceptions (for example for some large chain facilities) physicians are generally assigned to the A/B MAC that covers the state where they practice. If MACs varied in the time and effort costs of vertebroplasty reimbursement procedures and state MAC assignment was correlated with state standard of care definitions, this could bias the estimated effects of the locality rule. The author does not know of any evidence indicating that this likely occurred; nevertheless it remains a theoretical possibility. Since less than half of states have

[12] A home-market effect comes from theories of international trade, in which a country whose local demand for a good is high tends to also export larger quantities of the good to other countries.

locality-based standards of care, locality states may differ from national standard states on other dimensions by chance (for example the presence of liability reforms such as damage caps, joint liability reforms, mandatory pre-trial mediation or screening panels, statute of limitations and certificate of merit requirements). To the extent that these differences are constant over time and do not affect a state's response to the vertebroplasty evidence reversal, they are absorbed in the state fixed effects. However a difference or policy that changes over time and/or is correlated with physician decision-making may bias the estimates from this difference in differences specification. It is difficult to predict the direction of the effect that these policies would have on de-adoption. To the extent that these policies reduce medical malpractice pressure, physicians would incur fewer risks in performing vertebroplasties and these policies would likely slow de-adoption. To the extent that implementation was correlated with locality status and occurred during the study period, this would lead to upward bias in my Column 2 estimates.

A triple differences specification

Given the identification challenges inherent in each of the two individual difference in differences specifications alone, Column 3 of Table 1.6 proceeds with triple differences (equation 10). Using the first ("within locality states") difference in differences specification as a starting point, a triple differences specification can be constructed using rural vs urban differences within national standard states as a counterfactual difference. The rural vs urban difference in de-adoption within national standard states is reflected in the Column 3 coefficient -0.06. The triple difference serves as a robustness check for the validity of the estimated effect from the first difference in difference specification, and to the extent that rural and urban areas have statistically and substantively significant different trends, the triple differences specification improves upon the "within locality" difference in differences specification by addressing this concern. Trends for this placebo comparison of are displayed in Figure 17. If we instead begin with the second difference in differences specification comparing rural areas of locality states to rural areas of national standard states (Column 2 of Table 1.6), then the corresponding counterfactual difference would be to compare urban areas in locality states with urban areas in national standard states, which is reflected in the Column 3 coefficient of - 0.004. Similar to the reasoning above, this serves as a robustness check for the Column 2 difference in differences specification, and to the extent that locality states have different trends from national standard states for reasons unrelated to the locality rule, the triple differences specification improves upon the Column 2 difference in differences specification by "differencing out" this potential difference in trends. Trends for this counterfactual comparison are displayed in Figure 18.

Both of these counterfactual differences are small relative to the magnitude of the triple differences coefficient of 0.18, and neither is statistically significant. The interpretation of the triple differences coefficient in Column 3 of Table 1.6 is that after evidence reversal, local customs-based standards of care increase vertebroplasty rates by 0.18 per thousand relative to urban rates. During a period in which vertebroplasty rates were falling, this means that locality rules slow the speed of de-adoption by 0.18 procedures per thousand. The triple differences estimate of the effect of the locality rule (0.18) is larger than the differences in differences estimate in Column 1 (0.1) because in national standard states procedure rates fell more in rural areas than in urban ones. Similarly, since procedure rates in urban areas of locality states fell more than in urban areas of national standard states, the DDD coefficient is larger than the difference in differences estimate of 0.15 in Column 2. (Neither the comparison of rural vs urban procedure rates in national standard states, nor the comparison of procedure rates in urban areas of locality vs national standard states are statistically significant.) Figure 19 displays the event study version of this triple differences specification. None of the pre period coefficients are statistically significant and there is no clear anticipatory effect or pre trend. Figure 19 shows no evidence of dynamic effects - in mid 2010 the effect of the locality rule appears and remains fairly consistent through the end of 2014.

The average vertebroplasty rate before the evidence reversal was 0.376 per thousand Medicare beneficiaries, so the triple differences estimate of the effect of the locality rule in slowing de-adoption corresponds to 48% of the pre-evidence reversal mean. The effect of locality rules on de-adoption is one and a half times larger than the association between physician gender and de-adoption, and is over two times larger than the association between de-adoption and more than 30 years of physician experience.

1.6 Robustness

1.6.1 Measure proportional changes in procedure rates - Poisson specifications

Rather than modeling the effect of the locality rule on absolute procedure rates, modeling proportional changes in vertebroplasty rates alleviates concerns that the estimates in Table 1.6 reflect pre-reversal cross-state differences in vertebroplasty rates rather than the effect of the locality rule. Vertebroplasty rates declined substantially after the evidence reversal both in locality and non-locality regions, so the estimates in Table 1.6 may reflect pre-existing cross-sectional differences rather than a true effect of the locality rule. To address this concern we estimate the following model:

$$(5) VertRate_{ist} = e^{\alpha_0 + \beta TreatedPost_{ist} + \gamma_t + \delta_s + \mathbf{X}_{\mathbf{st}}^{'}\mathbf{\Gamma} + \epsilon_{ist}}$$

Where i indexes rural or urban regions within state s and t indexes a year and quarter. $VertRate_{ist}$ is the vertebroplasty rate for sub-state (rural or urban) region i in state s at time (year and quarter) t, measured in number of procedures per 1,000. $TreatedPost_{ist}$ is an indicator variable that equals 1 after the evidence reversal for rural areas in states that have a local standard of care for specialists. Covariates and fixed effects are as previously defined and standard errors are clustered at the state level. Compared to the more traditional approach of modeling the effect on $ln(VertRate_{ist})$, Poisson regression has the advantage of treating vertebroplasty rates of zero as part of the data generating process that influences the coefficient estimates. This is a realistic assumption in this setting: a vertebroplasty rate of zero likely reflects a situation where the local physicians could have performed a vertebroplasty but decided not to.

Table 1.7 displays the results of the Poisson regressions corresponding to the difference in differences and triple differences specifications in Table 1.6. The coefficients displayed in Table 1.7 are incidence-rate ratios, in other words e^{β}. The Column 1 within-locality-state comparison of rural and urban areas is not statistically significant and is small in magnitude. Under the identification assumptions of the difference in differences specification in Column 2 (where rural areas in locality states are compared to rural areas in national standard states), the locality rule causes vertebroplasty rates to decrease by 28% less. The triple differences coefficient in Column 3 is similar, implying an effect size of 29%.

1.6.2 Effects could be driven by differences in spatial distribution of elderly patients

Table 1.8 presents similar regressions to those in Table 1.5 where the level of observation is age group a in state s in quarter q of year y. Age fixed effects enable the estimation of locality effects within age groups. These estimates might differ substantially from the estimates in Table 1.5 if the Table 1.5 estimates were driven by differences in age composition between rural and urban areas of locality and national standard states. For example given that procedure rates fall more quickly for Medicare beneficiaries over 80 years old (Table 1.3 column 4), if rural areas in locality states have disproportionately fewer beneficiaries over 80 years old, the Table 1.5 estimates could partially reflect age composition differences rather than the effect of the locality-based standards. The estimates in Table 1.8 are qualitatively similar to those in Table 1.5.

1.6.3 Age-specific vertebroplasty rates

Columns 1-3 of Table 1.9 present difference in differences and triples estimates comparable to Table 1.6 while further disaggregating the data by age group and adding age group fixed effects. However once vertebroplasty rates are split by age group the variance increases dramatically and there are many quarters where some age groups in a given state's urban or rural regions have no vertebroplasties. In addition the denominator for vertebroplasty rates in the age-group specifications (total number of beneficiaries living in a rural/urban region r of state s in age group a in year y) becomes very small – about 7,043 on average. Considering that this denominator includes both men and women and elderly women far have greater risk for vertebroplasty, the true age-group exposure within a state's rural/urban region r is probably lower. Nevertheless, the confidence interval of the age-group triple differences coefficient contains the estimate of 0.18 from the state level triple differences specification in column 6. In addition, the age group specifications and the state level specifications display similar relationships between the difference in differences estimates, and between the coefficients in the triple differences specification.

1.6.4 Endogeneity in past standard of care reforms could persist over decades

Although most states have not changed their standard of care in over 2 decades, standard of care reforms that occurred in the 1980s or 1990s could have been affected by healthcare sector conditions that may persist in those states today, especially if the reforms arose through legislation rather than as a product of the judicial system . If legislatures passed national standard of care laws as part of an effort to modernize the healthcare sector, improve healthcare quality and increase accountability, and if this effort persisted through this study period, then including states with statutory standards of care would cause my estimates to overstate the true effect of the locality rule. To address this concern, Table 1.10 shows difference in difference and triples differences results when only states whose standard of care is solely based on case law-based are included. Both national standard and locality standard states were excluded if their standards of care were based on a statute. Across all Table 1.10 specifications the magnitude of the estimated effect of the locality rule is similar or even larger than the main results in Table 1.6. The coefficients of interest are not statistically significant however, due to the much smaller sample size: only Kansas, New York, and North Dakota have case law-based locality standards of care.

In order to gain more power Table 1.11 also includes locality states with statute-based standards of care. Because the vast majority of standard of care reforms since the 1970s involved a change from a locality based standard to a national standard, the potential endogeneity mainly affects states where pre-existing healthcare sector trends could have influenced national standard of care statutory reforms. Including states with longstanding local standard of care statutes should increase power and address this endogeneity concern. Table 1.11 displays the results of these specifications, where only states with statute-based national standards of care are excluded (Alabama, Alaska, Connecticut, Delaware, Georgia, Louisiana, Michigan, Missouri, Nevada, New

Hampshire, Oklahoma, Vermont, West Virginia and Wyoming). The results in column 1 are now statistically significant and are similar in magnitude to the estimates in Column 1 of Table 1.6. The triple differences estimate is slightly larger than the coefficient in Column 3 of Table 1.6 (though not statistically significant).

1.6.5 Robustness to excluding Rhode Island

In Figure 16 and Figure 17 rural areas of national standard states have some large jumps in procedure rates relative to rural areas of locality states or urban areas of national standard states, respectively. These jumps seem to be driven by idiosyncrasies in Rhode Island: once Rhode Island is excluded the trends become smoother (see Figure 20 and Figure 21). Excluding Rhode Island does not change the qualitative result that the locality rule slows de-adoption; however the coefficient of interest is smaller (about 0.10) and not statistically significant. (See Table 1.12 for further specifications.)

1.6.6 Independent validation of study classification of states' legal standard of care

A team of legal researchers conducted an independent analysis of the evolution of legal standards of care since 2000, and their classification of states was nearly identical to the classification proposed in this study.[13] The researchers agreed that the standard of care in Maryland and in Florida is ambiguous. (Section 1.4.2 discusses the cases of Maryland and Florida in more detail.) Hawaii was classified as a national standard state rather than as ambiguous, so this study adopted that approach. Lewis et al (2007) also classifies Hawaii as a national standard state as of 2007.

The only point of disagreement was how to categorize Utah. This study follows Lewis et al (2007) in classifying Utah as a national standard state, whereas the independent researchers classified Utah as a locality standard state. According to the researchers, Olsen v. Delcore, 2009 WL 3233712 (D. Ut. 2009) cites Swan v. Lamb, 584 P.2d 814 (Utah 1978) to state that Utah has a national standard, however this contradicts the substance of the Swan cases, which clearly articulated a similar locality standard of care. They conclude that the Swan citation in Delcore was a "mistaken interpretation," and that Utah has a locality standard of care. The Swan case does explicitly reject a same locality standard of care; the ambiguity lies in whether the court adopts a similar locality standard or goes farther in endorsing a national standard. In Swan, the court describes medical care in Utah as among the best in the United States, and cites the "outstanding reputation" of the University of Utah medical college in support of an argument that Utah physicians should

[13] I am extremely grateful to Professor Kyle Logue and the research librarians at the University of Michigan Law School for carrying out this independent analysis to confirm my classification.

be held to a high standard of care commensurate with their advanced training rather than held to a lower local standard of care. According to the court, "If this procedure is generally regarded to be unsatisfactory or dangerous, no doctor should escape responsibility merely because the local practice has not yet adopted it." A longer excerpt from Swan v Lamb is included in Appendix 2. Given the ambiguity in whether Swan abandons the same locality standard in favor of a similar locality standard or in favor of a national standard, this study accepted the interpretation of the U.S. District Court in Olsen v Delcore and categorized Utah as a national standard state. However the results in Table 1.6 are robust to classifying Utah as a locality state rather than a national standard state. (See Table 1.13.)

1.6.7 Kyphoplasty as a potential substitute

Optimal management of vertebral compression fractures was an active area of research during the study period. Unsurprisingly, the two vertebroplasty RCTs were not the only evidence that emerged around 2009 that could have affected vertebrplasty rates. Around the time of the vertebroplasty evidence reversal two studies were published that supported the use of kyphoplasty in patients with vertebral compression fractures: (1) Wardlaw et al was published the "FREE trail" in Lancet in 2009, and (2) Berenson et al published a study of kyphoplasty in cancer patients in Lancet in 2011. Kyphoplasty is an alternative surgical procedure to treat spinal compression fractures. As in vertebroplasty, spinal grade cement is injected into the fracture. Kyphoplasty differs from vertebroplasty because before the cement is injected into the fracture, the physician inserts and inflates a balloon to create a bone cavity and restore the normal height and shape of the collapsed vertebra. The cement is the injected into the cavity. At the time, kyphoplasty was a less established procedure so the publication of the Wardlaw et al and Berenson et al studies in 2009 could have had a significant effect on physicians' priors about kyphoplasty effectiveness. Given the timing of these studies it is possible that rather than responding to the vertebroplasty evidence reversal, physicians were reducing their use of vertebroplasty in response to the positive news about kyphoplasty.

If the declining vertebroplasty rates were due to a substitution effect because of the increasing appeal of the newer kyphoplasty procedure rather than a response to a vertebroplasty evidence reversal, one would expect to see increasing national kyphoplasty rates around and after 2009. However Figure 22 shows that national kyphoplasty rates did not increase in response to the vertebroplasty evidence reversal – if anything they may have decreased slightly. [14] If the differential

[14] The sharp changes in observed kyphoplasty rates before 2006 likely reflect the learning process among medical coders in response to coding changes that occurred in late 2004, when separate ICD-9 codes for vertebroplasty and kyphoplasty were introduced. Procedure rates increase again in early 2006, when the CPT code for kyphoplasty was introduced.

vertebroplasty de-adoption rates due to the locality rule were largely a byproduct of the improvements in kyphoplasty evidence, one might expect to see an increase in kyphoplasty rates in urban areas of locality states, and relatively flatter kyphoplasty rates in rural areas of locality states. Figure 23 suggests that the differential decline in vertebroplasty rates in Figure 5 was not due to substitution towards kyphoplasty: kyphoplasty rates did not increase in urban areas of locality states after the vertebroplasty evidence reversal. Figure 24 shows that after a sharp increase between 2004 and 2006, kyphoplasty rates in locality and national standard states stabilized and remained relatively flat during the study period. The possibility that the measured effects of locality rules on de-adoption reflect a combination of the direct effect of the vertebroplasty evidence reversal and a substitution effect caused by the kyphoplasty studies cannot be ruled out completely, but it seems unlikely that substitution effects driven by the kyphoplasty studies account for a significant portion of the measured effect of locality rules on de-adoption given the observed trends in kyphoplasty utilization .

1.7 Conclusion

This paper investigates whether and to what extent the medical liability system affects productivity in the healthcare sector. An optimally designed liability system should correct the incentive problems in the principal agent problem when a physician makes treatment decisions or recommendations for a less well-informed patient. However the evidence about whether the medical liability system is tailored to meet this objective in practice is mixed. To the extent that the medical liability system distorts physician decision-making away from the social optimum, this results in societal resources not being allocated efficiently and a loss of productivity (lower marginal health product per dollar of societal resources spent).

This paper focuses on medical malpractice standard of care laws within the broader medical liability system and on de-adoption as a specific aspect of productivity. Physicians are generally insulated from the more severe financial repercussions of malpractice lawsuits because most are insured and medical malpractice insurance is generally not experience rated. In light of these circumstances, if the threat of a lawsuit affects physician behavior it is likely to operate more strongly through other non-financial channels. Physicians report psychological distress when faced with the prospect of a lawsuit, and the process is time consuming. Physicians face a high opportunity cost of time as well. Many other aspects of the medical liability system that have been studied such as damage caps, collateral source rule reform and punitive damage bans directly affect the intensive margin of malpractice pressure by reducing the burden of a given case. They may impact the extensive margin by incentivizing or discouraging additional lawsuits, but this effect is less direct. In contrast, local customs-based standard of care laws directly impact malpractice pressure on both the intensive and extensive margin: they increase the difficulty of winning a case (intensive margin) and they serve as a "barrier to entry" for plaintiffs who must find an expert witness who is qualified to testify about local practice (extensive margin). If physicians are relatively more sensitive to non-financial aspects of malpractice pressure, then aspects of the liability system that impact the extensive margin may be important determinants of behavior since an individual lawsuits has a high fixed cost in terms of time, emotional energy, and stress.

I focus on de-adoption of medical practices after evidence reversal in order to avoid issues of cost-effectiveness or low value care. By choosing a setting where the new evidence indicated that the absolute benefits of the medical procedure were negligible, I can abstract away from philosophical differences and differences in values that may impact how a physician views cost-effectiveness measures as a tool to guide the practice of medicine. In addition given the historical context that gave rise to locality rules, legal scholars have theorized that locality rules not only decrease malpractice pressure, but that they do so in a very specific way: by enabling and abetting "laggards."

This study leverages state level variation in standard of care definitions to investigate whether and to what extent local customs-based standards of care result in slower de-adoption. I find that de-adoption does not occur more slowly overall in locality states. This does not mean that locality rules are completely irrelevant at the state level; it is possible that while locality rules do enable slower de-adoption at the margin, there are other factors that outweigh the locality rule in their influence at the state level. For rural areas, the locality rule slows de-adoption by 0.18 vertebroplasties per 1,000 people, or by 48% of the pre-reversal mean procedure rate. This finding speaks to the effects that tort law can have on the diffusion of innovation as well as how well legal institutions perform in aligning incentives among patients and physicians along this dimension.

1.8 Tables

Table 1.1: Effect of sample restrictions on sample size in 2008

Step	# beneficiaries remaining after step (%)
# Medicare Beneficiaries in 2008	9,690,866 (100%)
Drop if residence is not in one of the 50 US states	9,459,417 (97.6%)
Drop if age<65 or age>=95	7,794,944 (80.4%)
Drop if has ESRD or Disabled	7,111,223 (73.4%)
Must have both part A & B coverage, whole year	5,985,554 (61.8%)
No managed care coverage during the year	4,431,063 (45.7%)

This table shows the effect of various restrictions on the sample. There were 9,690,866 beneficiaries in 2008. Each line lists a sample restriction and the number of beneficiaries remaining after the restriction is imposed followed by the share of the original sample remaining (in parenthesis). For example, after beneficiaries who live outside the U.S. were dropped, 9,459,417 beneficiaries remained, representing 97.6% of the original sample.

Table 1.2: Vertebroplasty in the United States Medicare Fee for Service Population

Variable	2008 National	2014 National
# doctors who performed at least one vertebroplasty	1,474	907
# vertebroplasties performed in US (FFS Medicare)	4,809	2,182
National Annual Vertebroplasty rate (unweighted)	1.07 per thousand	.47 per thousand
Share of vertebroplasties done by ...		
Diagnostic Radiology	32.47%	27.56%
Vascular Interventional Radiology	20.51%	26.67%
Neuroradiology	13.42%	16.30 %
Neurological Surgery	5.68%	10.59%
Orthopedic surgery	7.29%	5.55%
Anesthesiology	6.66%	3.08%
Other specialty	13.97%	10.25%
Patient Characteristics...		
Average # vertebroplasties received among people who received at least one	1.13	1.10
Average age of vertebroplasty patient	81.28	80.78
# vertebroplasty patients age 65-69	250	166
# vertebroplasty patients age 70-74	538	280
# vertebroplasty patients age 75-79	891	391
# vertebroplasty patients age 80-84	1,089	472
# vertebroplasty patients age 85-89	963	413
# vertebroplasty patients age 90-94	440	226
# vertebroplasty patients age 95-99	88	40
% Black	1.17%	1.16%
% Female	75.21%	74.04%
% Rural	22.96%	23.29%

Table 1.3: Heterogeneity in de-adoption by patient characteristics (difference in differences)

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Vert rate	Vert rate	Vert rate	Vert rate	Vert rate
Age 70-74			0.0834***	0.141***	
			(0.0203)	(0.0502)	
Age 75-79			0.262***	0.394**	
			(0.0845)	(0.149)	
Age 80-84			0.415***	0.675***	
			(0.0921)	(0.228)	
Age 85-89			0.454***	0.609***	
			(0.0641)	(0.0942)	
Age 90-94			0.502***	0.666***	
			(0.0695)	(0.0925)	
Age 65-69 * Post				-0.0516***	
				(0.00842)	
Age 70-74 * Post				-0.145**	
				(0.0550)	
Age 75-79 * Post				-0.268**	
				(0.111)	
Age 80-84 * Post				-0.477**	
				(0.232)	
Age 85-89 * Post				-0.305***	
				(0.0732)	
Age 90-94 * Post				-0.320***	
				(0.0579)	
Post	-0.193***		-0.261***		
	(0.0514)		(0.0731)		
rural		-0.0768			
		(0.119)			
Rural*Post		-0.0102			
		(0.0726)			
Population % Male		-0.296			
		(0.192)			
Population % Black		-0.0558			-0.0224
		(0.0691)			(0.0438)
Female Patient*Post					-0.101***
					(0.0167)
Female Patient					0.277***
					(0.0437)
Observations	3,456	3,456	20,700	20,700	6,912
R-squared	0.094	0.108	0.028	0.028	0.064
Cluster by	State	State	State	State	State
Obs. Level	YrQ-State-Rural/Urban	YrQ-State-Rural/Urban	YrQ-State-Rural/Urban-Agegrp	YrQ-State-Rural/Urban-Agegrp	YrQ-State-Rural/Urban-Sex
Years	2006-2014	2006-2014	2006-2014	2006-2014	2006-2014
Mean	0.376	0.376	0.506	0.506	0.356

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Columns 1, 3 and 4 include quarter and state fixed effects. Columns 2 & 5 include year, quarter and state fixed effects.

Table 1.4: Heterogeneity in de-adoption by physician characteristics (difference in differences)

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Vert rate	Vert rate	Vert rate	Vert rate	Vert rate
Female Doc*Post	0.00540***				
	(0.00169)				
Female Doc	-0.00977***				
	(0.00193)				
Population % Male	-0.00693	-0.00686	-0.00684	-0.00690	-0.00477
	(0.00538)	(0.00537)	(0.00536)	(0.00538)	(0.00371)
Population % Black	-0.00248	-0.00250	-0.00250	-0.00249	-0.00227
	(0.00177)	(0.00178)	(0.00178)	(0.00177)	(0.00169)
% Population Ages 65-69	-0.00467	-0.00464	-0.00466	-0.00468	-0.00578
	(0.00376)	(0.00376)	(0.00377)	(0.00375)	(0.00400)
% Population Ages 70-74	-0.00450	-0.00446	-0.00448	-0.00449	-0.00467
	(0.00412)	(0.00411)	(0.00412)	(0.00411)	(0.00427)
% Population Ages 75-79	0.000899	0.000909	0.000874	0.000884	-0.00156
	(0.00363)	(0.00363)	(0.00362)	(0.00362)	(0.00311)
% Population Ages 80-84	-0.00983**	-0.00979**	-0.00981**	-0.00984**	-0.00889**
	(0.00445)	(0.00445)	(0.00445)	(0.00446)	(0.00395)
% Population Ages 85-89	0.00462	0.00461	0.00458	0.00463	0.00154
	(0.00599)	(0.00598)	(0.00597)	(0.00599)	(0.00537)
Yrs of Experience*Post		-5.87e-05			
		(6.25e-05)			
Yrs of Experience		-6.62e-05			
		(8.90e-05)			
Early Career*Post			0.00179*		
			(0.000959)		
Early Career			-0.000907		
			(0.00190)		
Experienced Doctor*Post				0.00283*	
				(0.00159)	
Experienced Doctor				-0.00503**	
				(0.00206)	
Orthopedic Surgery*Post					0.00722***
					(0.00156)
Neurosurgery*Post					0.00814***
					(0.00152)
Anesthesia*Post					0.00195
					(0.00156)
Orthopedic Surgery					-0.0129***
					(0.00263)
Neurosurgery					-0.0147***
					(0.00297)
Anesthesia					-0.00655***
					(0.00173)
Observations	27,882	27,882	27,882	27,882	22,698
R-squared	0.179	0.178	0.178	0.178	0.179
Cluster by	State	State	State	State	State
Obs. Level	Yr-Doctor	Yr-Doctor	Yr-Doctor	Yr-Doctor	Yr-Doctor
Years	2006-2014	2006-2014	2006-2014	2006-2014	2006-2014
Mean	0.0157	0.0157	0.0157	0.0157	0.0152

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include state and year fixed effects.

Table 1.5: Locality as a state level treatment

	(1)	(2)	(3)
VARIABLES	Vert rate	Vert rate	Vert rate
Locality*Post	-0.0404	0.0448	
	(0.0301)	(0.0713)	
Rural Locality Area*Post			0.109*
			(0.0595)
Rural Locality Area			-0.314**
			(0.155)
Rural		-0.0830	-0.0162
		(0.0787)	(0.0994)
Population % Male	-0.0732	-0.160	-0.160
	(0.0595)	(0.0960)	(0.0961)
Population % Black	-0.00844	-0.00201	-0.00180
	(0.0184)	(0.0515)	(0.0517)
% Population Ages 65-69	0.00758	0.103	0.103
	(0.0296)	(0.104)	(0.103)
% Population Ages 70-74	0.0102	0.0870	0.0879
	(0.0305)	(0.101)	(0.0986)
% Population Ages 75-79	0.0268	0.141	0.140
	(0.0243)	(0.0888)	(0.0907)
% Population Ages 80-84	-0.0195	0.267	0.267
	(0.0374)	(0.225)	(0.223)
% Population Ages 85-90	0.0736*	0.186*	0.187*
	(0.0375)	(0.103)	(0.0998)
Observations	1,728	3,456	3,456
R-squared	0.707	0.122	0.126
Cluster by	State	State	State
Obs. Level	Yr-Q-State	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban
Years	2006-2014	2006-2014	2006-2014
Mean	0.289	0.376	0.376

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

In Column 1 a state is the level of observation. Locality states are treated and national standard states serve as a control group. Column 2 disaggregates so that the level of observation is rural or urban areas within a state. Rural and urban areas of locality states are treated and all areas within national standard states serve as a control group. In Column 3, only rural areas of locality states are treated. Urban areas of locality states and all areas of national standard states serve as control groups. All specifications include year quarter and state fixed effects.

Table 1.6: Effect of locality rule on de-adoption, state-urban/rural level DD

	(1)	(2)	(3)
VARIABLES	Vert rate	Vert rate	Vert rate
Rural	-0.326**		0.0186
	(0.117)		(0.154)
Rural*Post	0.104**		-0.0569
	(0.0371)		(0.0973)
Locality*Post		0.153	-0.0412
		(0.129)	(0.0704)
Locality*Rural			-0.357*
			(0.194)
Rural*Locality*Post			0.180*
			(0.107)
Population % Male	0.0959*	-0.0794	-0.160
	(0.0453)	(0.147)	(0.0960)
Population % Black	0.0722	0.0258	-0.00192
	(0.0558)	(0.0860)	(0.0516)
% Population Ages 65-69	0.0569	0.205	0.103
	(0.0695)	(0.210)	(0.104)
% Population Ages 70-74	0.121**	0.187	0.0873
	(0.0545)	(0.204)	(0.101)
% Population Ages 75-79	0.0999	0.215	0.141
	(0.0607)	(0.173)	(0.0885)
% Population Ages 80-84	0.0410	0.646	0.267
	(0.0663)	(0.455)	(0.225)
% Population Ages 85-90	0.234***	0.239	0.186*
	(0.0736)	(0.207)	(0.103)
Observations	936	1,728	3,456
R-squared	0.547	0.192	0.127
Cluster by	State	State	State
Years	2006-2014	2006-2014	2006-2014
Mean	0.397	0.340	0.376

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The dependent variable is the vertebroplasty rate (per 1,000 beneficiaries) in rural or urban region r in state s in quarter q of year y. All specifications include year quarter and state fixed effects. In Column 1 rural areas of locality states are treated and urban areas of locality states serve as a control group. National standard states are excluded from Column 1. In Column 2 rural areas of locality states are treated and rural areas of national standard states serve as a control group. Urban areas of all states are excluded from Column 2. Column 3 is a triple differences specification.

Table 1.7: Effect of locality rule on de-adoption, Poisson models corresponding to Table 6

	(1)	(2)	(3)
VARIABLES	Vert rate	Vert rate	Vert rate
Rural	0.565***		0.544***
	(0.116)		(0.110)
Rural*Post	1.019		0.814**
	(0.0747)		(0.0673)
Locality*Post		1.284***	1.008
		(0.108)	(0.0483)
Locality*Rural			1.023
			(0.292)
Rural*Locality*Post			1.290**
			(0.132)
Population % Male	1.171	1.237*	0.997
	(0.126)	(0.145)	(0.0739)
Population % Black	1.016	0.909	1.026
	(0.0484)	(0.0750)	(0.0446)
% Population Ages 65-69	1.213	0.823	1.111
	(0.152)	(0.130)	(0.0926)
% Population Ages 70-74	1.183	0.778	1.076
	(0.136)	(0.121)	(0.104)
% Population Ages 75-79	1.021	0.731**	1.073
	(0.126)	(0.103)	(0.0782)
% Population Ages 80-84	1.196	0.905	1.081
	(0.154)	(0.166)	(0.114)
% Population Ages 85-90	1.565***	0.876	1.358**
	(0.167)	(0.178)	(0.169)
Observations	936	1,728	3,456
Cluster by	state	state	state
Years	2006-2014	2006-2014	2006-2014
Mean	0.397	0.340	0.376

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The dependent variable is the vertebroplasty rate (per 1,000 beneficiaries) in rural or urban region i in state s in quarter q of year y. All specifications include year quarter and state fixed effects. In Column 1 rural areas of locality states are treated and urban areas of locality states serve as a control group. National standard states are excluded from Column 1. In Column 2 rural areas of locality states are treated and rural areas of national standard states serve as a control group. Urban areas of all states are excluded from Column 2. Column 3 is a triple differences specification.

Table 1.8: Locality as a state level treatment, age group level data

VARIABLES	(1) Vert rate	(2) Vert rate	(3) Vert rate
Rural Locality Area*Post			0.133
			(0.0910)
Rural Locality Area			-0.434*
			(0.230)
Rural		-0.117	-0.0211
		(0.110)	(0.139)
Age 70-74	0.0561***	0.0834***	0.0834***
	(0.00599)	(0.0203)	(0.0203)
Age 75-79	0.159***	0.262***	0.262***
	(0.0163)	(0.0845)	(0.0845)
Age 80-84	0.292***	0.415***	0.415***
	(0.0297)	(0.0922)	(0.0922)
Age 85-89	0.408***	0.454***	0.454***
	(0.0454)	(0.0641)	(0.0641)
Age 90-94	0.444***	0.501***	0.502***
	(0.0479)	(0.0694)	(0.0697)
Population % Male	-0.0846	-0.382	-0.381
	(0.0819)	(0.255)	(0.254)
Population % Black	0.00216	-0.0725	-0.0700
	(0.0264)	(0.0822)	(0.0849)
Locality*Post	-0.0826*	0.0289	
	(0.0491)	(0.0992)	
Observations	10,368	20,700	20,700
R-squared	0.348	0.030	0.032
Cluster by	State	State	State
Obs. Level	Yr-Q-State-Age Grp	Yr-Q-State-Rural/Urban-Age Grp	Yr-Q-State-Rural/Urban-Age Grp
Years	2006-2014	2006-2014	2006-2014
Mean	0.395	0.506	0.506

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

In Column 1 an age group within a state is the level of observation. Locality states are treated and national standard states serve as a control group. Column 2 disaggregates so that the level of observation is an age group in a rural/urban area within a state. Rural and urban areas of locality states are treated and all areas within national standard states serve as a control group. In Column 3 only rural areas of locality states are treated. Urban areas of locality states and all areas of national standard states serve as a control group. All specifications include year quarter and state fixed effects.

Table 1.9: Effect of locality rule on de-adoption, state-urban/rural-age group level difference in differences

VARIABLES	(1) Vert rate	(2) Vert rate	(3) Vert rate
Rural	-0.472**		0.0335
	(0.172)		(0.225)
Rural*Post	0.160**		-0.0894
	(0.0652)		(0.155)
Locality*Post		0.174	-0.104
		(0.178)	(0.0902)
Locality*Rural			-0.524*
			(0.286)
Rural*Locality*Post			0.279
			(0.174)
Age 70-74	0.0779***	0.0845**	0.0834***
	(0.0192)	(0.0402)	(0.0203)
Age 75-79	0.222***	0.296*	0.262***
	(0.0434)	(0.170)	(0.0845)
Age 80-84	0.404***	0.420**	0.415***
	(0.0761)	(0.180)	(0.0922)
Age 85-89	0.595***	0.311***	0.454***
	(0.126)	(0.0843)	(0.0641)
Age 90-94	0.651***	0.385***	0.502***
	(0.148)	(0.118)	(0.0697)
Population % Male	0.0912	-0.447	-0.382
	(0.0559)	(0.507)	(0.255)
Population % Black	0.0635	-0.131	-0.0722
	(0.0838)	(0.157)	(0.0820)
Observations	5,616	10,332	20,700
R-squared	0.247	0.038	0.032
Cluster by	State	State	State
Obs. Level	Yr-Q-State-Rural/Urban-Age Grp	Yr-Q-State-Rural/Urban-Age Grp	Yr-Q-State-Rural/Urban-Age Grp
Years	2006-2014	2006-2014	2006-2014
Mean	0.548	0.451	0.506

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The dependent variable is the vertebroplasty rate (per 1,000 beneficiaries) in rural or urban region i in state s in quarter q of year y. All specifications include year quarter and state fixed effects. In Column 1 rural areas of locality states are treated and urban areas of locality states serve as a control group. National standard states are excluded from Column 1. In Column 2 rural areas of locality states are treated and rural areas of national standard states serve as a control group. Urban areas of all states are excluded from Column 2. Column 3 is a triple differences specification.

Table 1.10: Robustness to inclusion of only case law states

	(1)	(2)	(3)
VARIABLES	Vert rate	Vert rate	Vert rate
Rural	-0.482		0.120
	(0.450)		(0.256)
Rural*Post	0.137		-0.0891
	(0.0687)		(0.164)
Locality*Post		0.280	0.0449
		(0.283)	(0.153)
Locality*Rural			-0.614
			(0.451)
Rural*Locality*Post			0.246
			(0.177)
Population % Male	0.458**	0.0263	-0.203
	(0.0847)	(0.227)	(0.156)
Population % Black	0.395	0.0545	0.0284
	(0.385)	(0.177)	(0.0950)
% Population Ages 65-69	0.0188	0.387	0.185
	(0.253)	(0.393)	(0.193)
% Population Ages 70-74	0.332	0.437	0.223
	(0.217)	(0.487)	(0.242)
% Population Ages 75-79	0.347	0.370	0.238
	(0.239)	(0.325)	(0.162)
% Population Ages 80-84	0.223	1.035	0.420
	(0.200)	(0.754)	(0.380)
% Population Ages 85-90	0.314	0.489	0.330
	(0.181)	(0.437)	(0.216)
Observations	216	864	1,728
R-squared	0.458	0.198	0.116
Cluster by	State	State	State
Obs. Level	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban
Years	2006-2014	2006-2014	2006-2014
Locality Laws	case law only	case law only	case law only
National Laws	case law only	case law only	case law only
Mean	0.515	0.509	0.481

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The dependent variable is the vertebroplasty rate (per 1,000 beneficiaries) in rural or urban region i in state s in quarter q of year y. All specifications include year quarter and state fixed effects and include only states with case-law based standards of care. In Column 1 rural areas of locality states are treated and urban areas of locality states serve as a control group. National standard states are excluded from Column 1. In Column 2 rural areas of locality states are treated and rural areas of national standard states serve as a control group. Urban areas of all states are excluded from Column 2. Column 3 is a triple differences specification.

Table 1.11: Robustness to exclusion of national standard states whose reform was by statute

	(1)	(2)	(3)
VARIABLES	Vert rate	Vert rate	Vert rate
Rural	-0.326**		0.122
	(0.117)		(0.251)
Rural*Post	0.104**		-0.0926
	(0.0371)		(0.158)
Locality*Post		0.145	-0.0404
		(0.137)	(0.0839)
Locality*Rural			-0.460
			(0.277)
Rural*Locality*Post			0.216
			(0.164)
Population % Male	0.0959*	-0.0685	-0.178
	(0.0453)	(0.183)	(0.116)
Population % Black	0.0722	0.147	0.0635
	(0.0558)	(0.155)	(0.0906)
% Population Ages 65-69	0.0569	0.335	0.154
	(0.0695)	(0.337)	(0.164)
% Population Ages 70-74	0.121**	0.400	0.187
	(0.0545)	(0.416)	(0.206)
% Population Ages 75-79	0.0999	0.375	0.216
	(0.0607)	(0.288)	(0.140)
% Population Ages 80-84	0.0410	0.873	0.353
	(0.0663)	(0.667)	(0.330)
% Population Ages 85-90	0.234***	0.452	0.305
	(0.0736)	(0.386)	(0.186)
Observations	936	1,224	2,448
R-squared	0.547	0.194	0.124
Cluster by	State	State	State
Obs. Level	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban
Years	2006-2014	2006-2014	2006-2014
Locality Laws	both	both	both
National Laws	case law only	case law only	case law only
Mean	0.397	0.422	0.446

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The dependent variable is the vertebroplasty rate (per 1,000 beneficiaries) in rural or urban region i in state s in quarter q of year y. All specifications include year quarter and state fixed effects and exclude states that switched to a national standard of care by statute. In Column 1 rural areas of locality states are treated and urban areas of locality states serve as a control group. National standard states are excluded from Column 1. In Column 2 rural areas of locality states are treated and rural areas of national standard states serve as a control group. Urban areas of all states are excluded from Column 2. Column 3 is a triple differences specification.

Table 1.12: Robustness to excluding Rhode Island

Table 10

	(1)	(2)	(3)
VARIABLES	1	5	6
Rural	-0.326**		-0.0976
	(0.117)		(0.106)
Rural*Post	0.104**		0.0256
	(0.0371)		(0.0548)
Locality*Post		0.0248	-0.0695
		(0.0284)	(0.0550)
Locality*Rural			-0.241
			(0.158)
Rural*Locality*Post			0.0978
			(0.0702)
Population % Male	0.0959*	0.0902	-0.0794
	(0.0453)	(0.0642)	(0.0911)
Population % Black	0.0722	0.0451	0.00541
	(0.0558)	(0.0379)	(0.0280)
% Population Ages 65-69			0.0274
			(0.0341)
% Population Ages 70-74			0.0243
			(0.0394)
% Population Ages 75-79			0.0702**
			(0.0303)
% Population Ages 80-84			0.00249
			(0.0495)
% Population Ages 85-90			0.111***
			(0.0392)
Observations	936	1,692	3,384
R-squared	0.547	0.646	0.350
Cluster by	State	State	State
Obs. Level	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban
Years	2006-2014	2006-2014	2006-2014
Mean	0.302	0.174	0.240

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The dependent variable is the vertebroplasty rate (per 1,000 beneficiaries) in rural or urban region i in state s in quarter q of year y. All specifications include year quarter and state fixed effects. In Column 1 rural areas of locality states are treated and urban areas of locality states serve as a control group. National standard states are excluded from Column 1. In Column 2 rural areas of locality states are treated and rural areas of national standard states serve as a control group. Urban areas of all states are excluded from Column 2. Column 3 is a triple differences specification.

Table 1.13: Robustness to classifying Utah as a locality state rather than a national standard state

	(1)	(2)	(3)
VARIABLES	1	2	3
Rural	-0.328***		0.0292
	(0.108)		(0.159)
Rural*Post	0.108***		-0.0629
	(0.0346)		(0.0999)
Locality*Post		0.141	-0.0656
		(0.126)	(0.0689)
Locality*Rural			-0.371*
			(0.192)
Rural*Locality*Post			0.193*
			(0.108)
Population % Male	0.0994**	-0.0762	-0.160
	(0.0404)	(0.145)	(0.0955)
Population % Black	0.0587	0.0249	-0.00245
	(0.0550)	(0.0859)	(0.0515)
% Population Ages 65-69			0.103
			(0.104)
% Population Ages 70-74			0.0870
			(0.103)
% Population Ages 75-79			0.142
			(0.0883)
% Population Ages 80-84			0.266
			(0.227)
% Population Ages 85-90			0.185*
			(0.104)
Observations	1,008	1,728	3,456
R-squared	0.547	0.192	0.127
Cluster by	State	State	State
Obs. Level	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban
Years	2006-2014	2006-2014	2006-2014
Mean	0.300	0.220	0.262

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The dependent variable is the vertebroplasty rate (per 1,000 beneficiaries) in rural or urban region i in state s in quarter q of year y. All specifications include year quarter and state fixed effects. In Column 1 rural areas of locality states are treated and urban areas of locality states serve as a control group. National standard states are excluded from Column 1. In Column 2 rural areas of locality states are treated and rural areas of national standard states serve as a control group. Urban areas of all states are excluded from Column 2. Column 3 is a triple differences specification.

Table 1.14: Robustness to inclusion of all vertebroplasty claims (no procedure to drop duplicates)

	(1)	(2)	(3)	(4)
VARIABLES	2	3	4	5
Rural	-0.469***		-0.0257	-0.0486
	(0.151)		(0.201)	(0.133)
Rural*Post	0.165***		-0.0375	
	(0.0483)		(0.119)	
Locality*Post		0.116	-0.101	
		(0.112)	(0.0966)	
Locality*Rural			-0.459*	
			(0.250)	
Rural*Locality*Post			0.228*	
			(0.131)	
Population % Male	0.153*	-0.0896	-0.190	-0.190
	(0.0719)	(0.160)	(0.115)	(0.115)
Population % Black	0.113	0.0210	-0.00265	-0.00166
	(0.0878)	(0.0949)	(0.0601)	(0.0607)
% Population Ages 65-69	0.0359	-0.145	-0.0794	-0.0793
	(0.0781)	(0.204)	(0.114)	(0.116)
% Population Ages 70-74	0.112	-0.151	-0.103	-0.0985
	(0.0636)	(0.196)	(0.109)	(0.111)
% Population Ages 75-79	0.147	-0.00954	0.0433	0.0391
	(0.0826)	(0.163)	(0.0948)	(0.0939)
% Population Ages 80-84	-0.0368	0.322	0.0626	0.0662
	(0.0729)	(0.204)	(0.106)	(0.105)
% Population Ages 85-89	0.327***	-0.211	-0.0252	-0.0200
	(0.0867)	(0.278)	(0.161)	(0.164)
Rural Locality Area*Post				0.121**
				(0.0574)
Rural Locality Area				-0.394*
				(0.198)
Observations	936	1,728	3,456	3,456
R-squared	0.561	0.243	0.157	0.157
Cluster by	State	State	State	State
Obs. Level	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban	Yr-Q-State-Rural/Urban
Years	2006-2014	2006-2014	2006-2014	2006-2014
Mean	0.418	0.293	0.361	0.361

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The dependent variable is the vertebroplasty rate (per 1,000 beneficiaries) in rural or urban region i in state s in quarter q of year y. All specifications include year quarter and state fixed effects. In Column 1 rural areas of locality states are treated and urban areas of locality states serve as a control group. National standard states are excluded from Column 1. In Column 2 rural areas of locality states are treated and rural areas of national standard states serve as a control group. Urban areas of all states are excluded from Column 2. Column 3 is a triple differences specification. In Column 4, only rural areas of locality states are treated. Urban areas of locality states and all areas of national standard states serve as control groups.

1.9 Figures

Figure 1.1: Standard of care by state for specialists

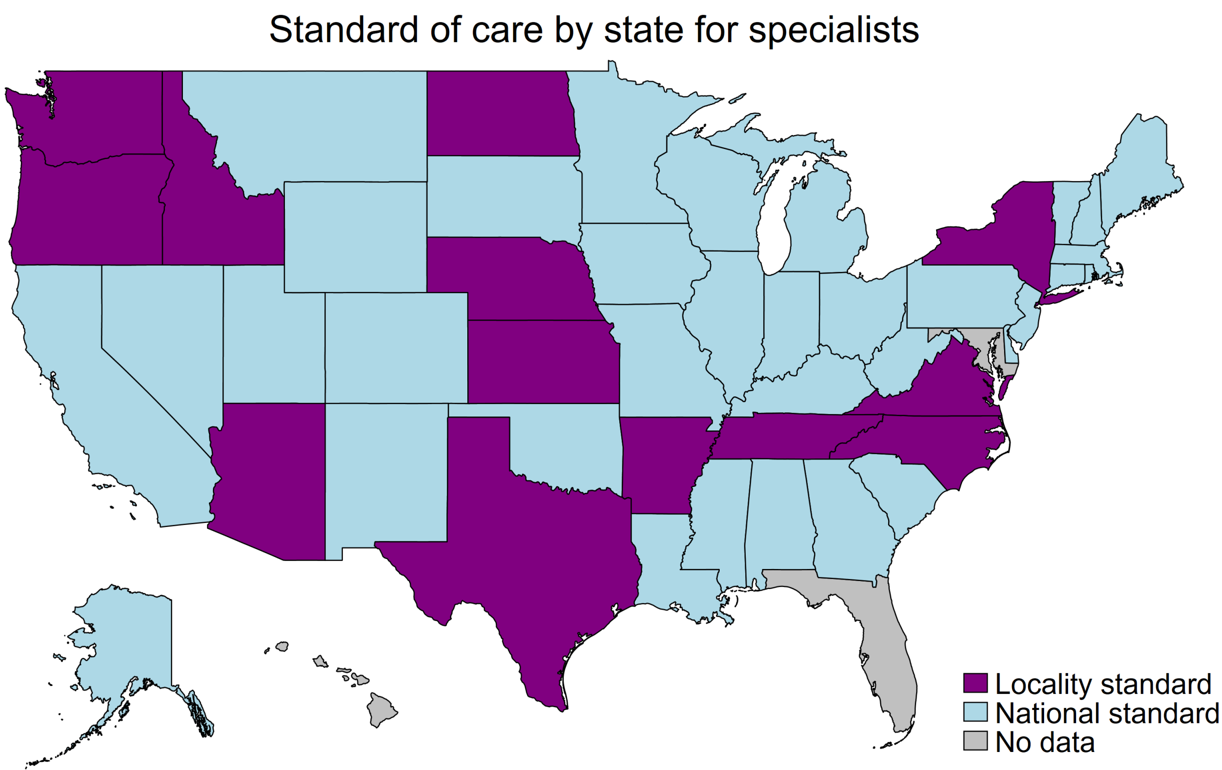


Figure 1.2: Standard of care by state for non-specialists

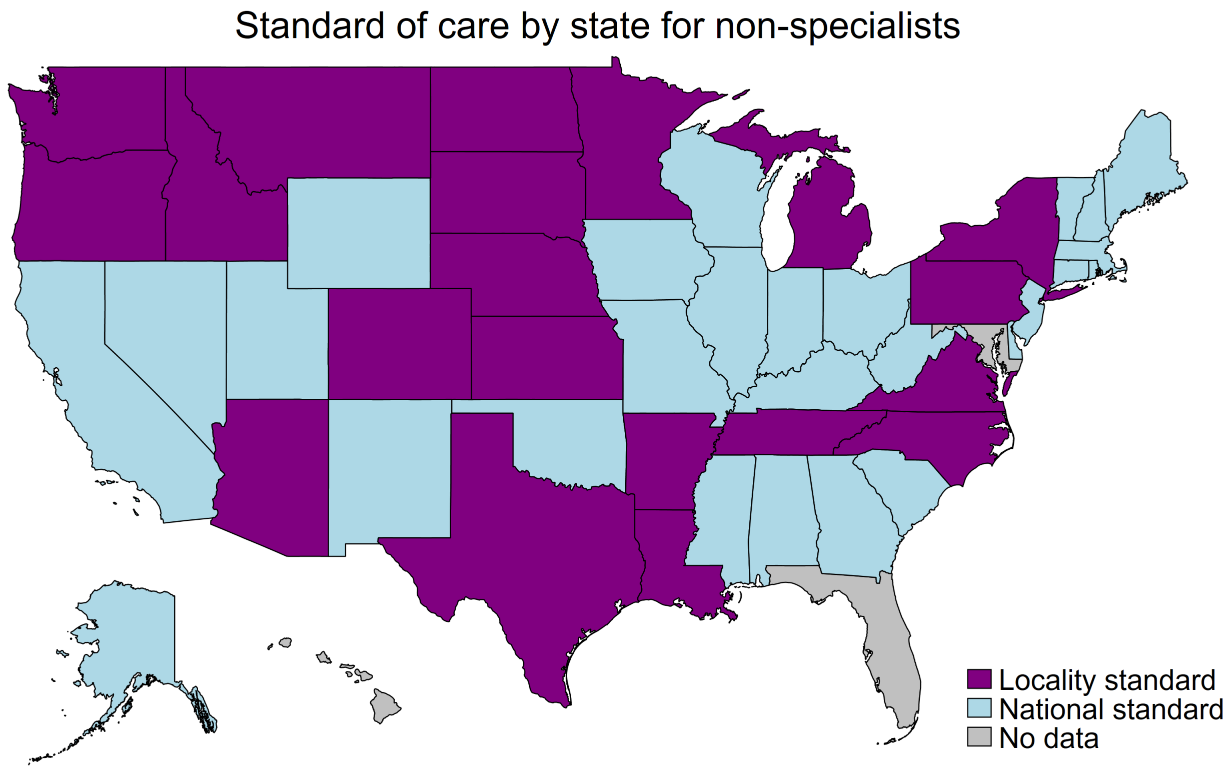


Figure 1.3: Physicians who performed at least one verteborplasty, 2008-2014

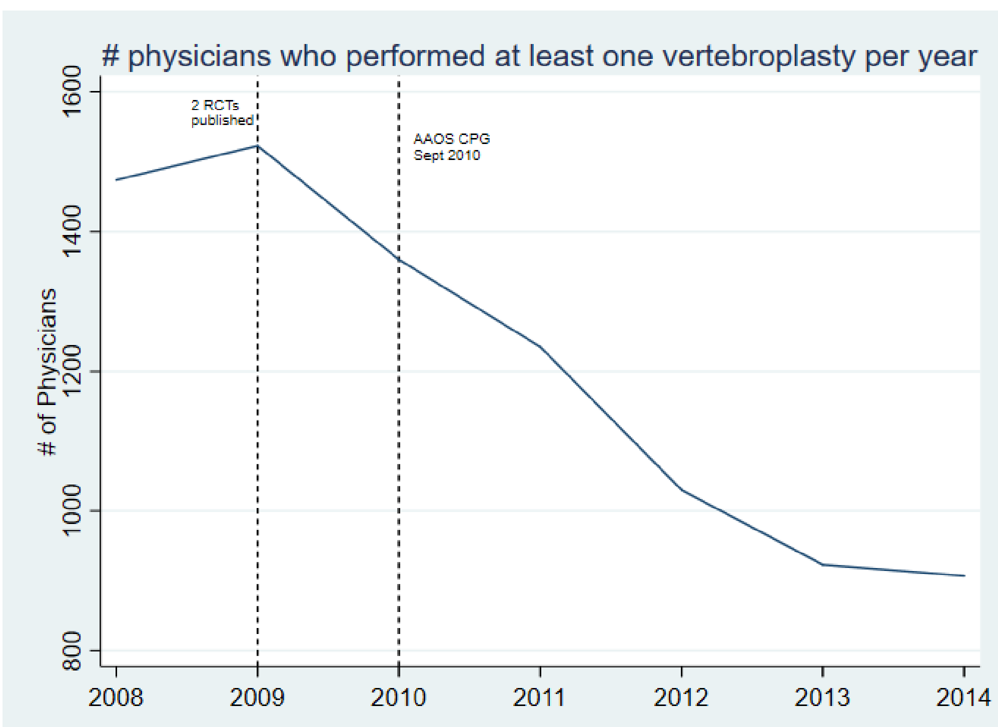


Figure 1.4: Specialty share of vertebroplasties performed, 2008-2014

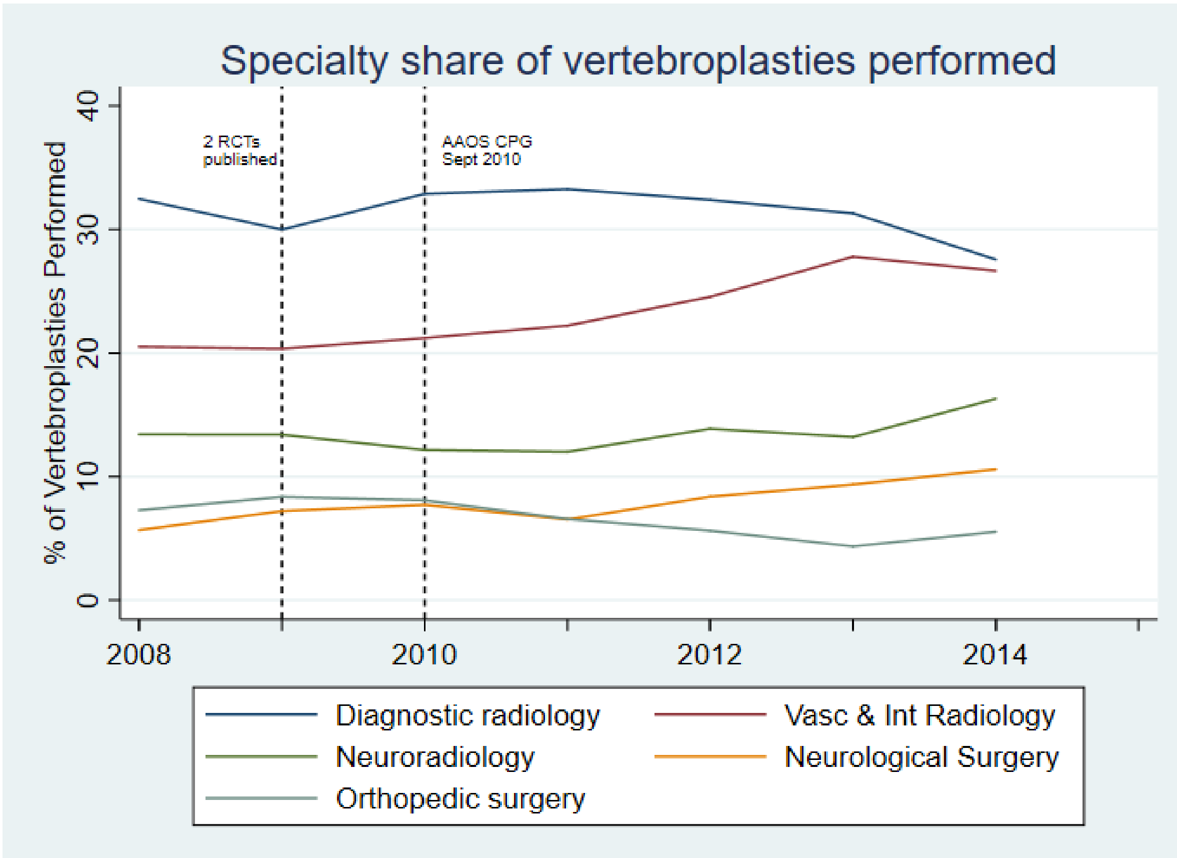


Figure 1.5: Vertebroplasty rates by standard of care, 2002-2014

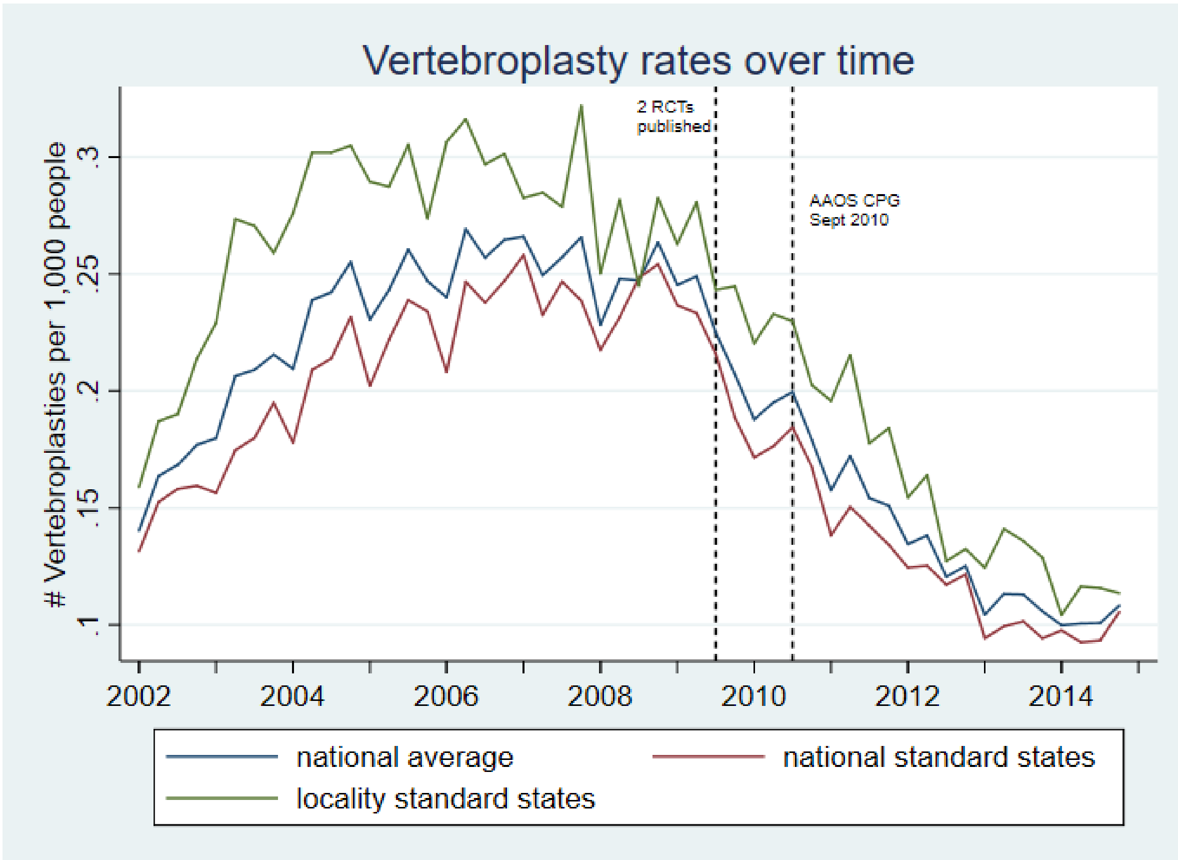


Figure 1.6: Vertebroplasty rates event study, female physicians

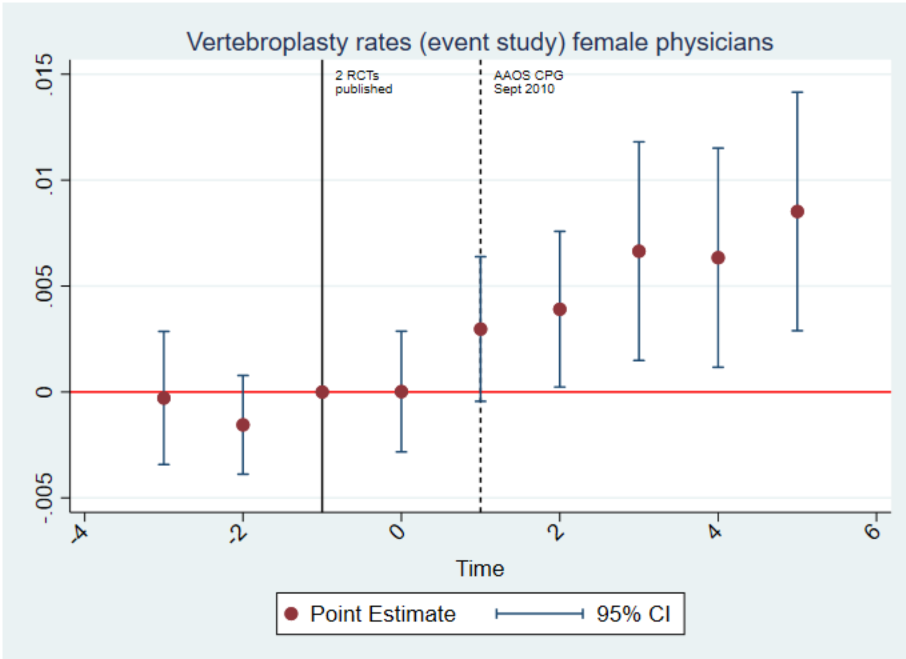


Figure 1.7: Vertebroplasty rates event study, early career physicians

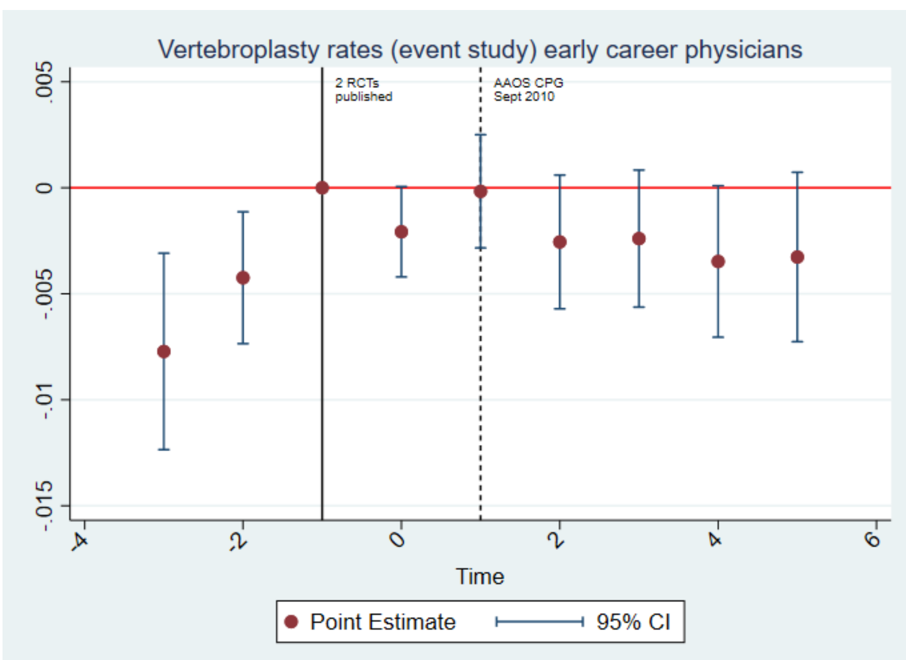


Figure 1.8: Vertebroplasty rates event study, experienced physicians

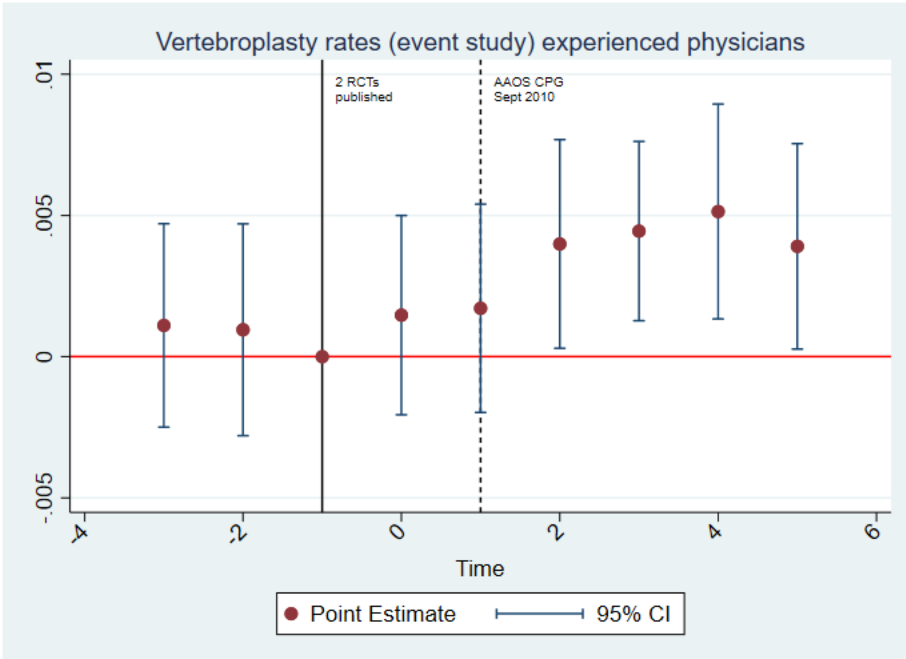


Figure 1.9: Vertebroplasty rates event study, anesthesiologists

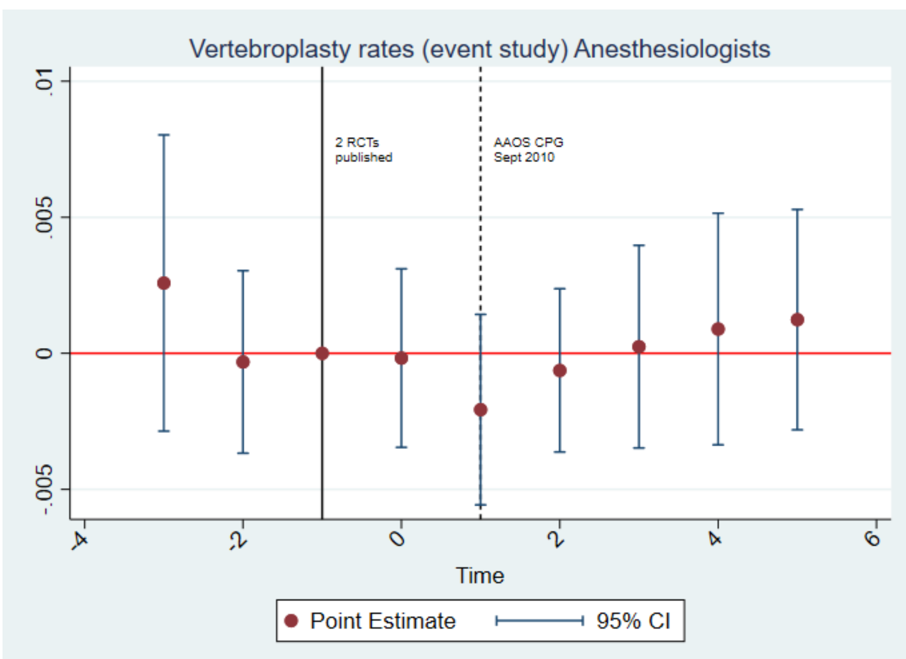


Figure 1.10: Vertebroplasty rates event study, orthopedic surgeons

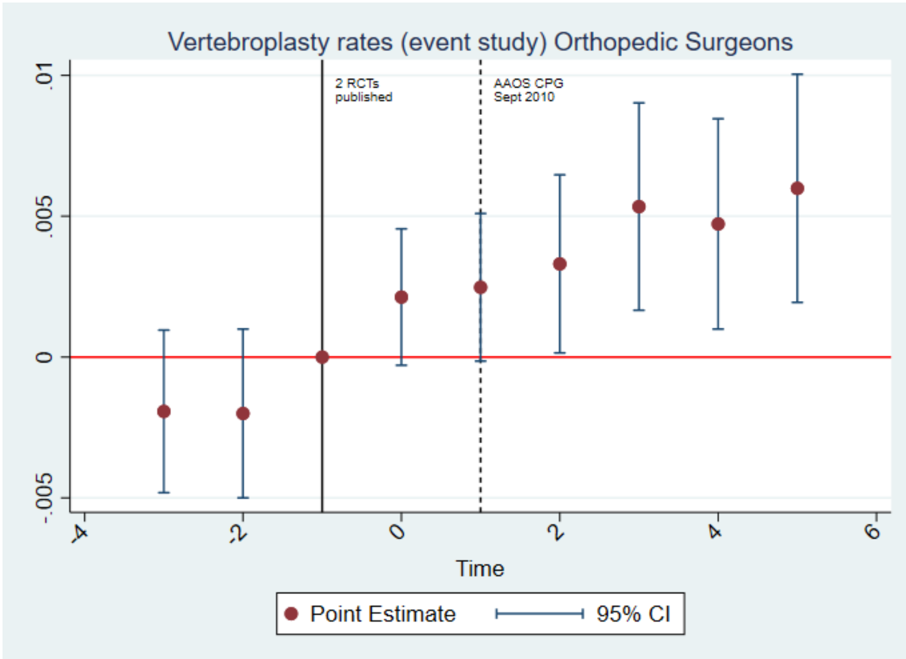


Figure 1.11: Vertebroplasty rates event study, neurosurgeons

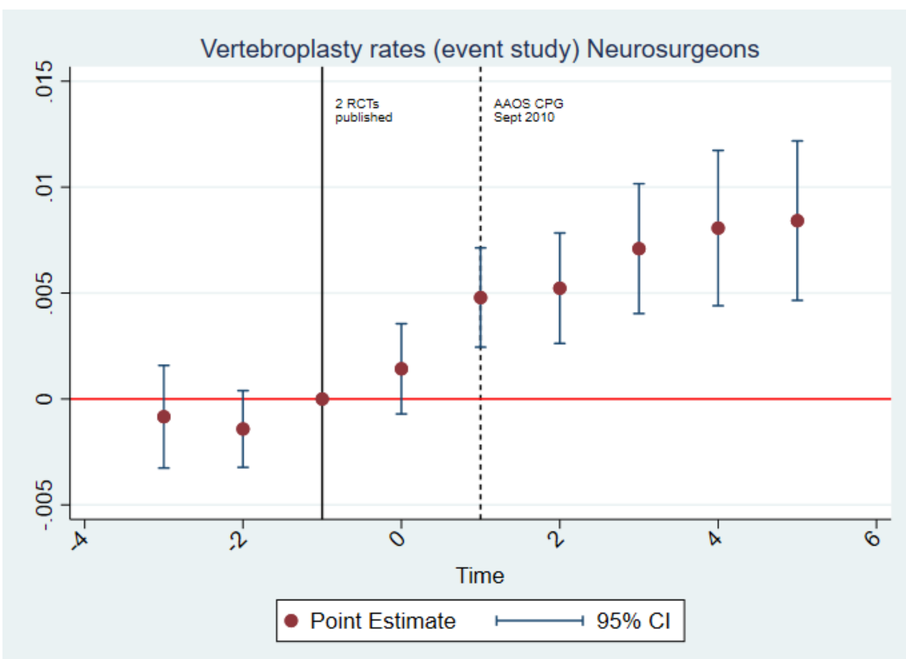


Figure 1.12: Vertebroplasty rates event study, locality vs national standard states

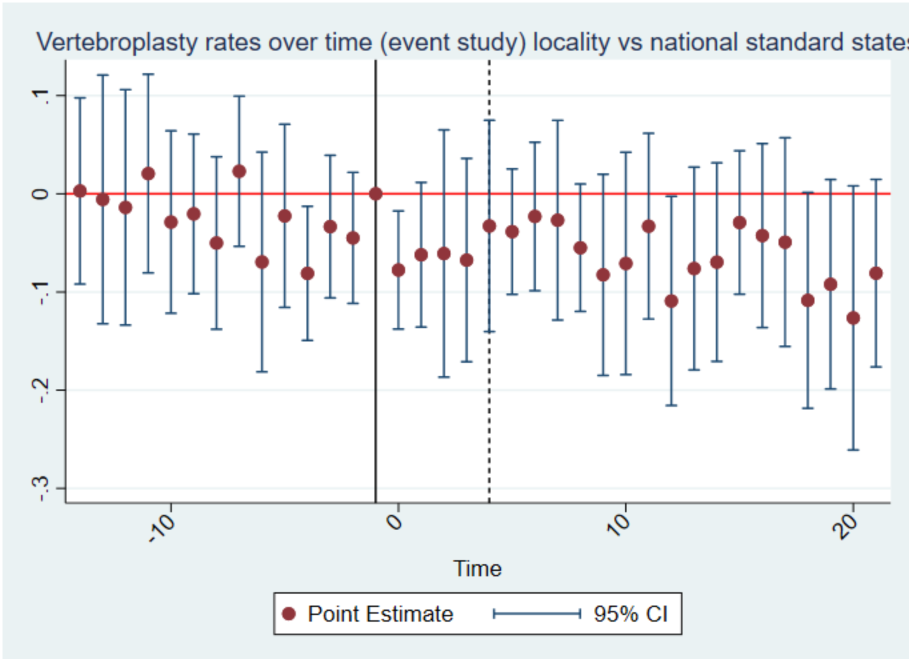


Figure 1.13: Rural and urban vertebroplasty trends, locality states

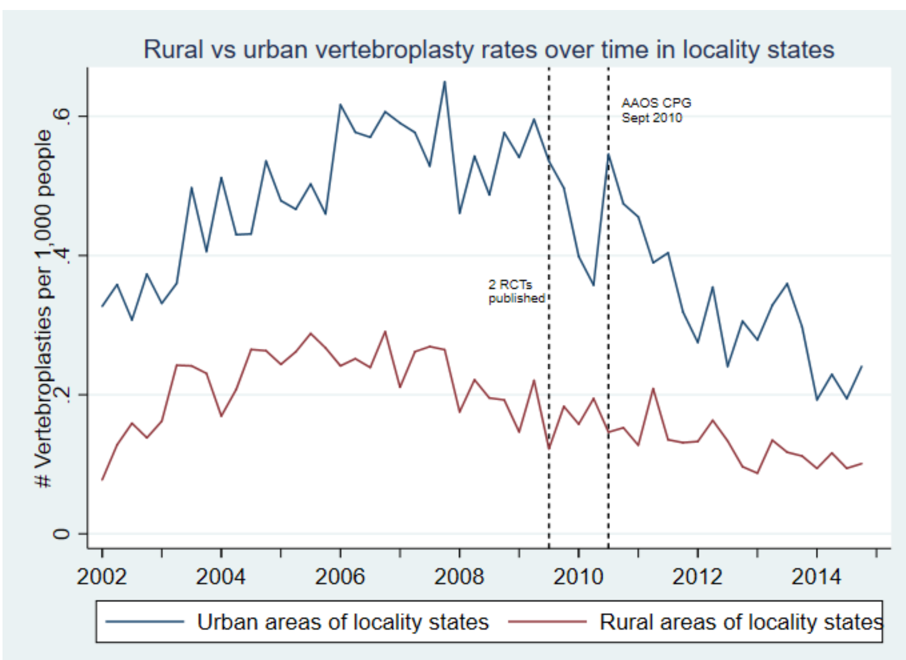


Figure 1.14: Vertebroplasty rates in locality states, event study of rural vs urban areas

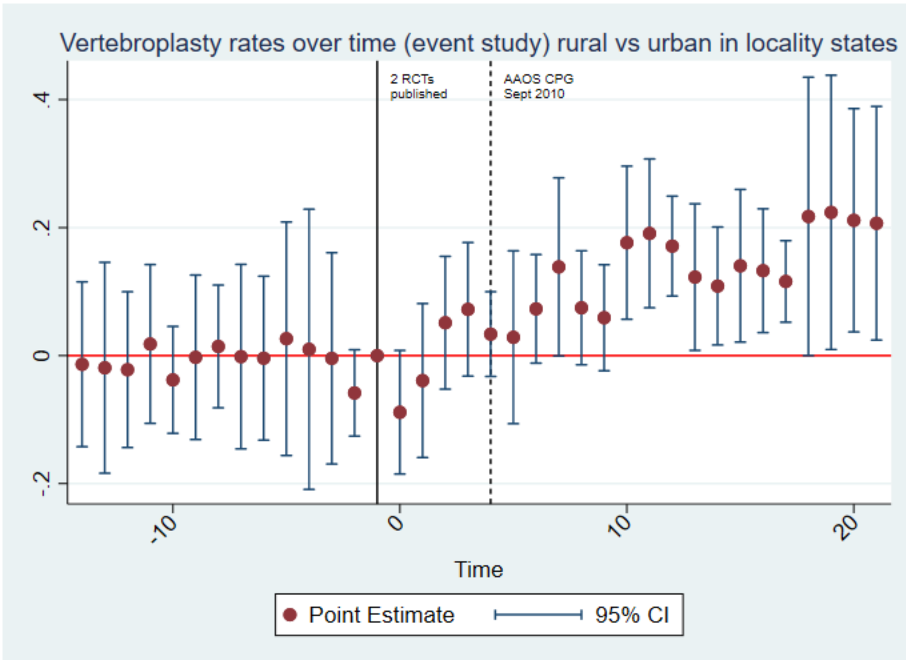


Figure 1.15: Vertebroplasty rates in rural areas, event study of locality vs national standard states

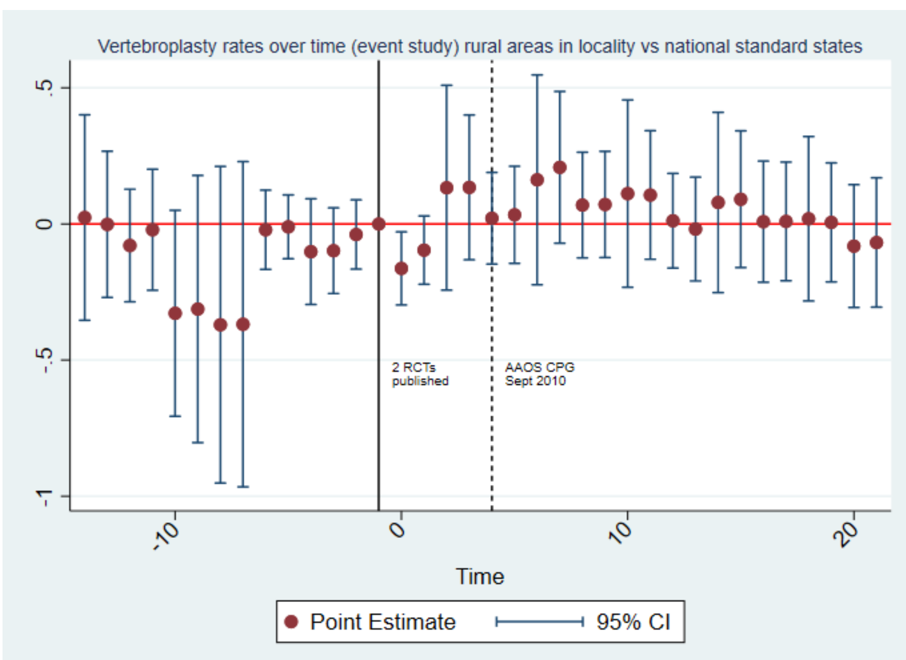


Figure 1.16: Rural vertebroplasty rate trends in locality vs national standard states

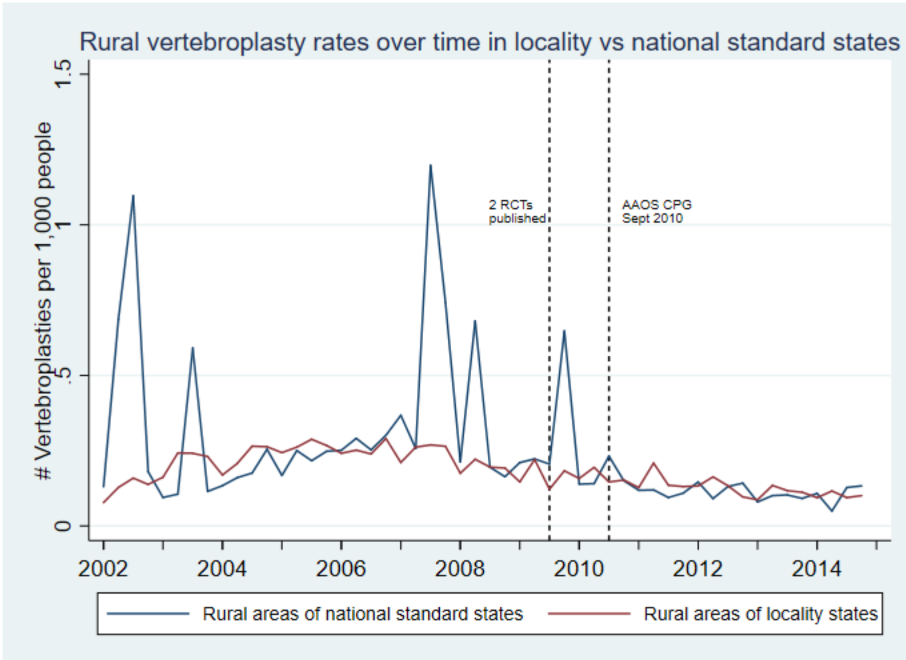


Figure 1.17: Rural vs urban vertebroplasty rate trends in national standard states

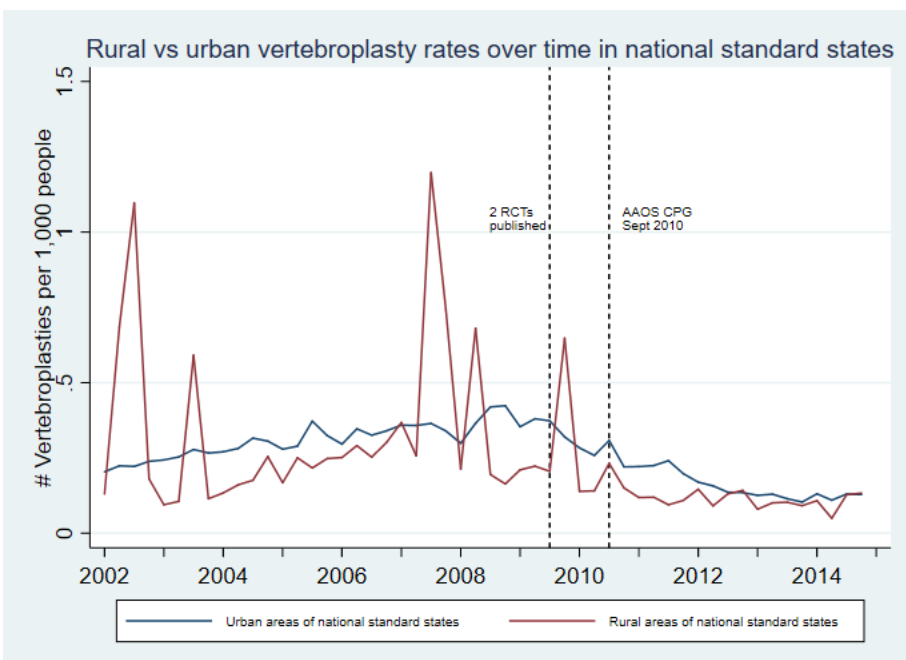


Figure 1.18: Urban vertebroplasty rate trends in locality vs national standard states

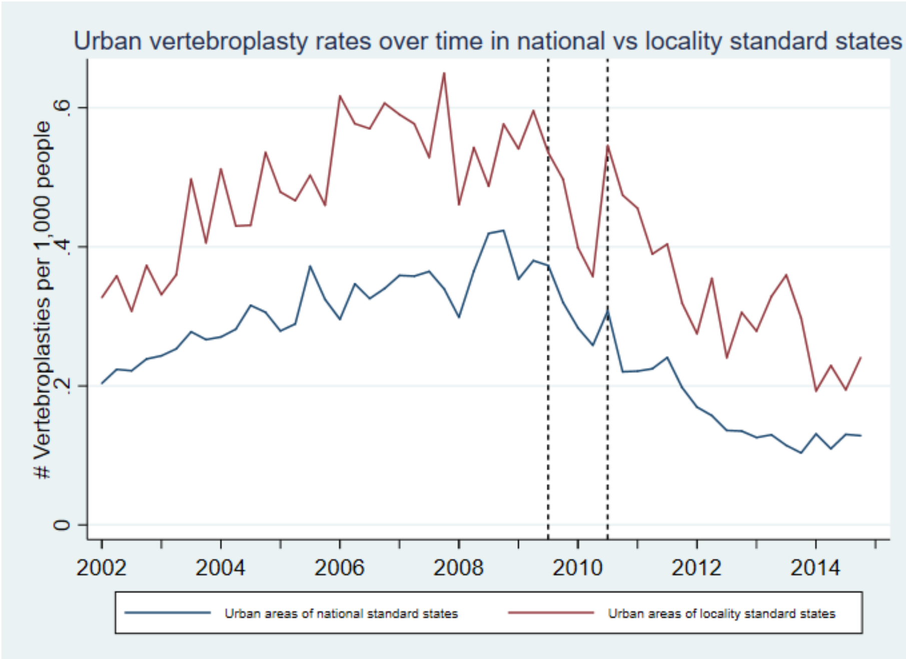


Figure 1.19: Triple differences event study, rural vs urban areas of locality vs national standard states pre and post reversal

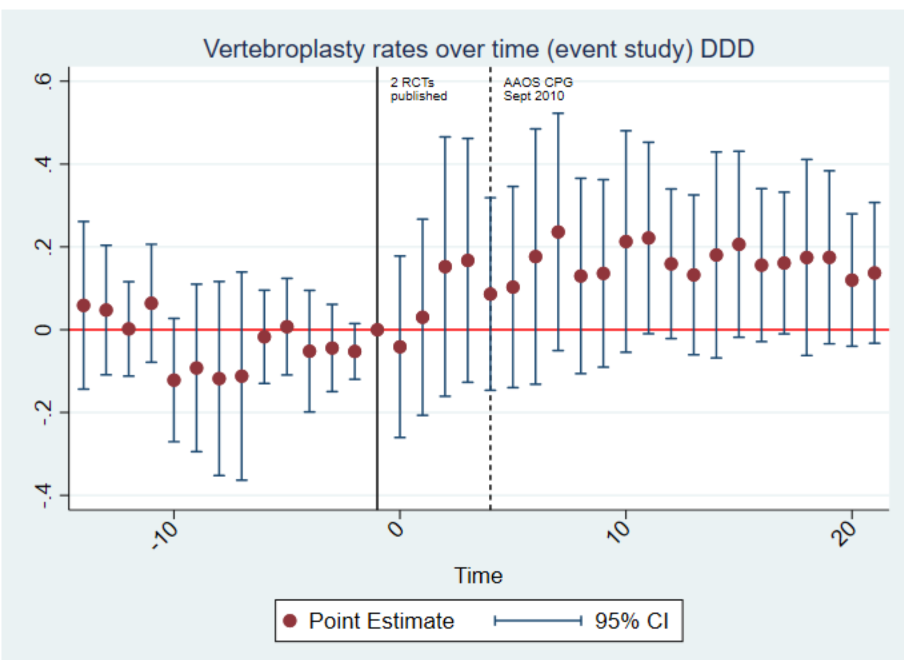


Figure 1.20: Rural vertebroplasty rate trends in locality vs national standard states, excluding Rhode Island

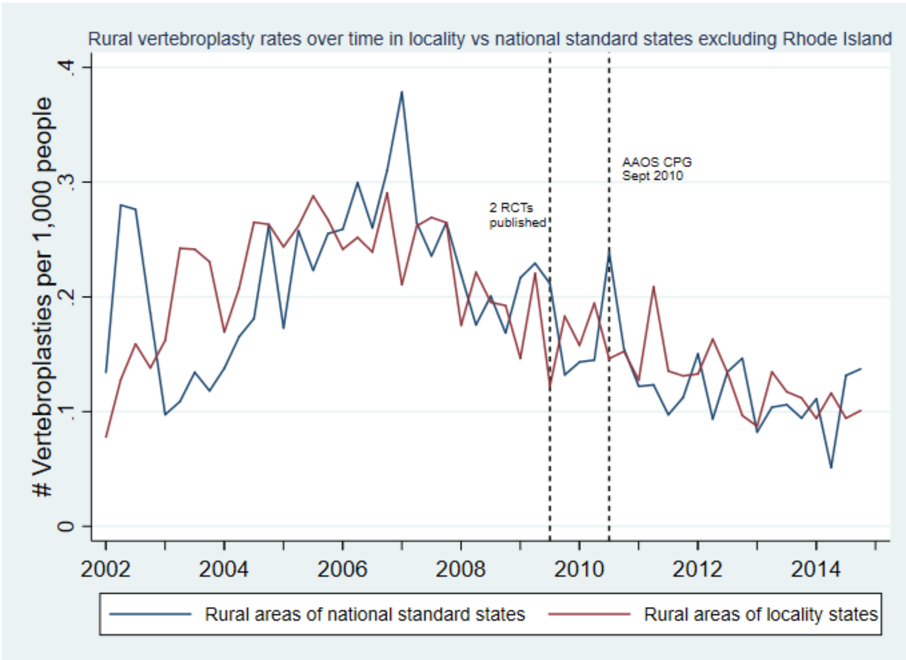


Figure 1.21: Rural vs urban vertebroplasty rate trends in national standard states, excluding Rhode Island

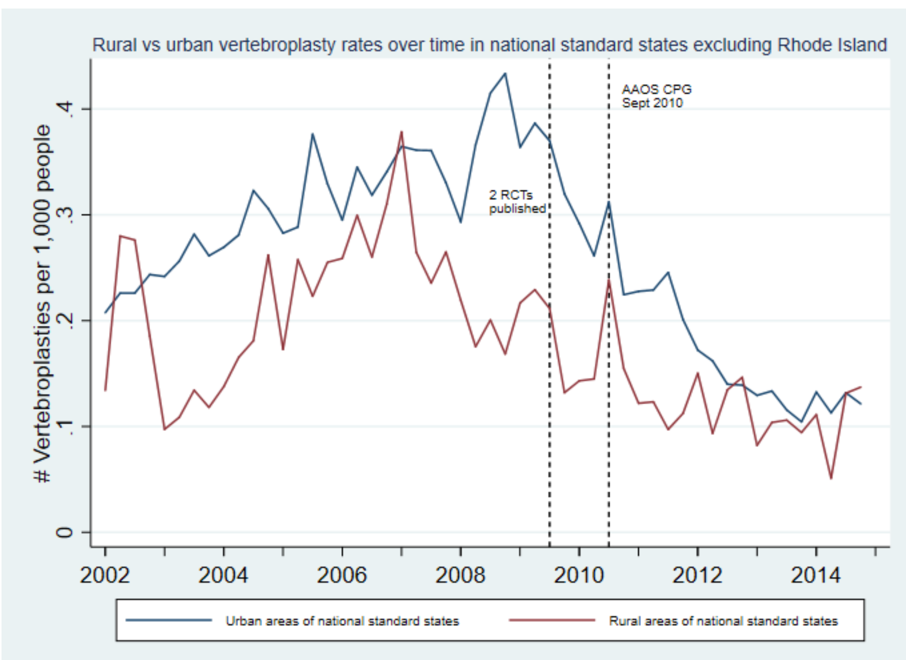


Figure 1.22: National kyphoplasty rates over time

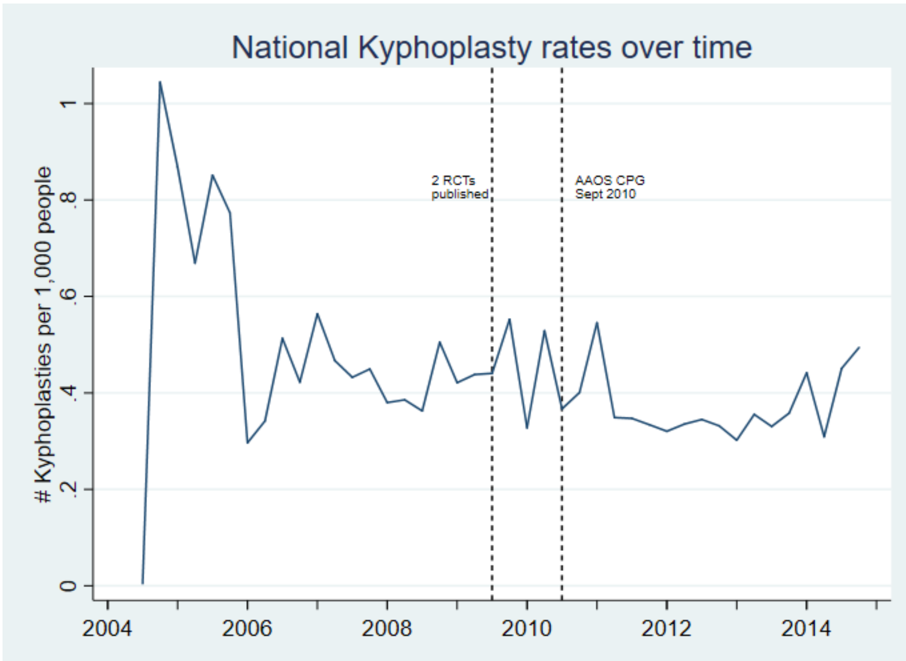


Figure 1.23: Kyphoplasty rate trends in rural vs urban areas of locality states

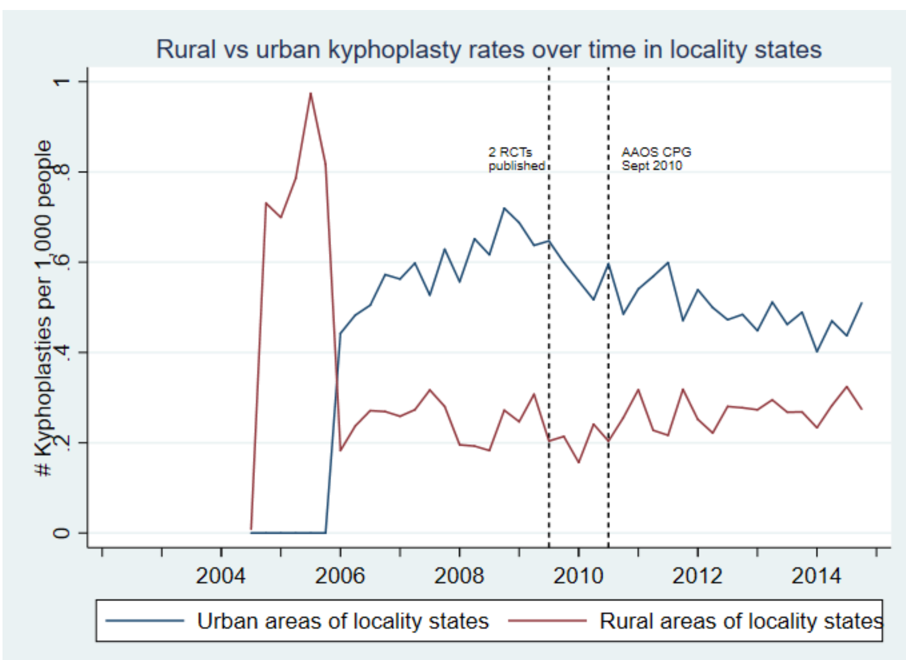
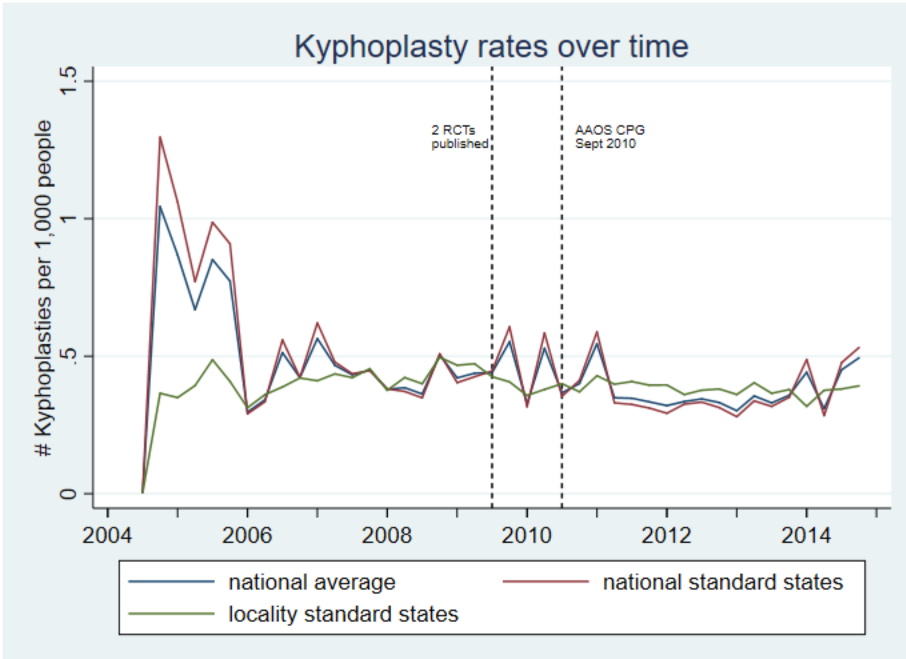


Figure 1.24: Kyphoplasty rate trends in locality vs national standard states



1.10 REFERENCES

Alexandru, Daniela and William So. 2012 "Evaluation and Management of Vertebral Compression Fractures." *The Permanente Journal* 16(4): 46-51. Albers, Sheri L. and Richard E. Latchaw. 2013.

"The Effects of Randomized Controlled Trials on Vertebroplasty and Kyphoplasty: A Square Peg in a Round Hole." *Pain Physician* 16:E331-348.
Barr, John D., Mary E. Jensen, Joshua A. Hirsch, J. Kevin McGraw, Robert M. Barr, Allan L. Brook, Philip M. Meyers et al. 2014. "Position statement on percutaneous vertebral augmentation: a consensus statement developed by the Society of Interventional Radiology (SIR), American Association of Neurological Surgeons (AANS) and the Congress of Neurological Surgeons (CNS), American College of Radiology (ACR)." *Journal of Vascular and Interventional Radiology* 25(2): 171-181.

Bekelis Kimon, Jonathan Skinner, Daniel Gottlieb and Philip Goodney. 2017. "Deadoption and exnovation in the use of carotid revascularization: retrospective cohort study." *British Medical Journal* 359(j4695) http://dx.doi.org/10.1136/bmj.j4695

Berenson J, Pflugmacher R, Jarzem P, et al. 2011. "Balloon kyphoplasty versus non-surgical fracture management for treatment of painful vertebral body compression fractures in patients with cancer: a multicentre, randomised controlled trial." *Lancet Oncology* 12:225-35.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *The Quarterly journal of economics* 119(1): 249-275.

Bottai, Vanna, Stefano Giannotti, Gloria Raffaetà, Maurizio Mazzantini, Francesco Casella, Gaia De Paola, Agnese Menconi, Francesca Falossi and Giulio Guido. 2016. "Underdiagnosis of osteoporotic vertebral fractures in patients with fragility fractures: retrospective analysis of over 300 patients." *Clinical Cases in Mineral and Bone Metabolism* 13(2):119-122.

Buchbinder R., R.H. Osborne, P.R. Ebeling, et al. 2009. "A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures." *New England Journal of Medicine*, 361: 557-568

Buchbinder R, Golmohammadi K, Johnston RV, Owen RJ, Homik J, Jones A, Dhillon SS, Kallmes DF, Lambert RGW. 2015. "Percutaneous vertebroplasty for osteoporotic vertebral compression fracture." *Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD006349.

Buchbinder R, Johnston RV, Rischin KJ, Homik J, Jones C, Golmohammadi K, Kallmes DF. 2018. "Percutaneous vertebroplasty for osteoporotic vertebral compression fracture." *Cochrane Database of Systematic Reviews*, Issue 11. Art. No.: CD006349.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-based improvements for inference with clustered errors." *The review of economics and statistics* 90(3): 414-427.

Centers for Disease Control and Prevention. “CDC Opioid Prescribing Guideline.” Accessed February 2022. https://www.cdc.gov/opioids/providers/prescribing/guideline.html

Chandra, Amitabh, David Malenka and Jonathan Skinner. 2014. “The Diffusion of New Non -Medical Technology: The Case of Drug -Eluting Stents.” In Discoveries in the Economics of Aging , edited by David A. Wise, 389- 403. University of Chicago Press.

Chandra, Amitabh David Cutler and Zirui Song. 2011. “Who ordered that? The Economics of treatment choices in medical care.” Handbook of Health Economics edited by Mark V. Pauly, Thomas G. Mcguire, Pedro P. Barros, 2:397-432.

Centers for Medicare and Medicaid Services. “National Health Expenditure Factsheet.” Accessed February 2022. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet

Centers for Medicare and Medicaid Services. 2019. “Rural-Urban Disparities in Health Care in Medicare.” Accessed July 2022. https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/Rural-Urban-Disparities-in-Health-Care-in-Medicare-Report.pdf

Committee on Quality of Healthcare in America of the Institute of Medicine. 2001. “Crossing the Quality Chasm: A New Health System for the 21st Century.” National Academy Press: Washington D.C.

Cummings Steven R and L Joseph Melton. 2002. “Epidemiology and outcomes of osteoporotic fractures.” *The Lancet* (359) 9319: 1761-1767.

Dingel, Jonathan I., and Joshua D. Gottleib. 2022 “Market Size and Trade in Medical Services.” Accessed July 2022. http://www.jdingel.com/research/DGLM_MSTMS.pdf

Doyle Joseph J., John A. Graves, and Jonathan Gruber. 2017 “Uncovering waste in US Healthcare: Evidence from ambulance referral patterns.” *Journal of Health Economics* (54): 25-39.

Ellis, R. P., and McGuire, T. G. 1986. “Provider behavior under prospective reimbursement: Cost sharing and supply.” *Journal of health economics*, 5(2), 129-151.

Fourney, Daryl. R., Donald F. Schomer, Remi Nader, Jennifer Chlan-Fourney, Dima Suki, Kamran Ahrar, Laurence D. Rhines, and Ziya L. Gokaslan. 2003. “Percutaneous vertebroplasty and kyphoplasty for painful vertebral body fractures in cancer patients.” *Journal of neurosurgery* 98(1 Suppl): 21–30. https://doi.org/10.3171/spi.2003.98.1.0021

Frakes, Michael. 2013. “The Impact of Medical Liability Standards on Regional Variations in Physician Behavior: Evidence from the Adoption of National-Standard Rules.” *American Economic Review* 103(1):257-276.

Frakes Michael and Anupam Jena. 2016. "Does medical malpractice law improve health care quality?" *Journal of Public Economics* 143:142-158.

Frakes, Michael D., Matthew B. Frank and Seth A. Seabury. 2017 "The Effect of Malpractice Law on Physician Supply: Evidence from Negligence-Standard Reforms." NBER Working Paper #23446

Freed, Meredith, Anthony Damico, and Tricia Neuman. 2021. "A Dozen Facts About Medicare Advantage in 2020." Accessed July 2022. https://www.kff.org/medicare/issue-brief/a-dozen-facts-about-medicare-advantage-in-2020/

Ginsberg, Marc D. 2013. "The Locality Rule Lives – Why? Using Modern Medicine to Eradicate an Unhealthy Law." 61 Drake L. Rev. 321.

Gray, Darryl T., William Hollingworth, Nneka Onwudiwe and Jeffery G. Jarvik. 2008. "Costs and State-Specific Rates of Thoracic 1 and Lumbar Vertebroplasty, 2001–2005." *Spine (Phila Pa 1976)* 33(17): 1905–1912. doi:10.1097/BRS.0b013e31817bb0a4

Greenberg, Michael D. 2009. "Medical malpractice and new devices: defining an elusive standard of care." *Health Matrix* 19(2): 423-445.

Hazzard, Matthew A., Kevin T. Huang, Ulysses N Toche, Beatrice Ugiliweneza, Chirag G Patil, Maxwell Boakye and Shivanand P Lad. 2014. "Comparison of Vertebroplasty, Kyphoplasty, and Nonsurgical Management of Vertebral Compression Fractures and Impact on US Healthcare Resource Utilization." *Asian Spine Journal* 8(5):605-614.

Howard David H. and E. Kathleen Adams. 2012. "Mammography rates after 2009 US Preventive Services Task Force breast cancer screening recommendation." *Preventive Medicine* 55:485-487.

Howard David H., Pamela R. Soulos, Anees B. Chagpar, Sarah Mougalian, Brigid Killelea and Cary P. Gross. 2016. "Contrary To Conventional Wisdom, Physicians Abandoned A Breast Cancer Treatment After A Trial Concluded It Was Ineffective." *Health Affairs* 35(7):1309-1315.

Howard David H., Guy David and Jason Hockenberry. 2016. "Selective Hearing: Physician-Ownership and Physicians' Response to New Evidence." *Journal of Economics and Management Strategy* 26(1):152-168.

Jena, Anupam B, Seth Seabury, Darius Lakdawalla and Amitabh Chandra. 2011. "Malpractice Risk Accordign to Physician Specialty." *New England Journal of Medicine* 365(7):629-636.

Kallmes D.F., B.A. Comstock, P.J. Heagerty, et al. 2009. "A randomized trial of vertebroplasty for osteoporotic spinal fractures." *New England Journal of Medicine*, 361: 569-579.

Kane C. 2010 "Policy research perspectives — medical liability claim frequency: a 2007-2008 snapshot of physicians." Chicago: American Medical Association: pages 1-7.

Kamano Hironori, Akio Hiwatashi, Nobuo Kobayashi, Sokun Fuwa, Osamu Takahashi, Yukihisa Saida, Hiroshi Honda, and Yuji Numaguchi. 2011. "New Vertebral Compression Fractures After Prophylactic Vertebroplasty in Osteoporotic Patients." *American Journal of Roentgenology* 197(2): 451-456

Kaptchuk, T. J., Goldman, P., Stone, D. A., and Stason, W. B. 2000. "Do medical devices have enhanced placebo effects?" *Journal of clinical epidemiology*, 53(8): 786-792.

Kaptchuk, Ted J. and Franklin G. Miller. 2015. "Placebo Effects in Medicine." *New England Journal of Medicine* 373(1):8-9. doi: 10.1056/NEJMp1504023.

Kaufman, Brystana G., Sharita R. Thomas, Randy K. Randolph, Julie R. Perry, Kristie W. Thompson, George M. Holmes and George H. Pink. 2016. "The Rising Rate of Rural Hospital Closures."*The Journal of Rural Health* 32: 35–43.

Keeton, Page. 1984 "Prosser and Keeton on the Law of Torts." West Publishing Company.

Klazen C.A., P.N. Lohle, J. de Vries, et al. 2010. "Vertebroplasty versus conservative treatment in acute osteoporotic vertebral compression fractures (Vertos II): an open-label randomised trial." *Lancet*, 376: 1085-1092.

Knavel EM, Thielen KR, Kallmes DF. 2009. "Vertebroplasty for the treatment of traumatic nonosteoporotic compression fractures." *American journal of neuroradiology* 30(2):323-7. doi: 10.3174/ajnr.A1356.

Kobayashi Nobuo, Yuji Numaguchi, Sokun Fuwa, Akihiro Uemura, Masaki Matsusako, Yuka Okajima, Mitsutomi Ishiyama and Osamu Takahashi. 2009. "Prophylactic Vertebroplasty: Cement Injection into Non-fractured Vertebral Bodies During Percutaneous Vertebroplasty." *Academic Radiology*, 16 (2):136-143

Kozhimannil Katy B., Pinar Karaca-Mandic, Cori J. Blauer-Peterson, Neel T. Shah, and Jonathan Snowden. 2017. "Uptake and utilization of practice guidelines in hospitals in the United States: the case of routine episiotomy." The Joint Commission Journal on Quality and Patient Safety 43:41-48.

Laakmann, Anna B. 2015. "When Should Physicians Be Liable for Innovation." *Cardozo Law Review* 36(3): 913-968.

Lad, Shivanand P., Chirag G. Patil, Eleonora M. Lad, Melanie G. Hayden and Maxwell Boakye. 2009. "National trends in vertebral augmentation procedures for the treatment of vertebral compression fractures." *Surgical Neurology* 71:580–585.

Laratta, J., Shillingford, J., Lombardi, J., Mueller, J., Reddy, H., Saifi, C., Fischer, C., Ludwig, S., Lenke, L., and Lehman, R. 2017. “Utilization of vertebroplasty and kyphoplasty procedures throughout the United States over a recent decade: an analysis of the Nationwide Inpatient Sample.” *Journal Of Spine Surgery*, 3(3): 364-370.

Leake, C. B., Brinjikji, W., Cloft, H. J., and Kallmes, D. F. 2011. “Trends of inpatient spine augmentation: 2001–2008.” *American journal of neuroradiology*, 32(8): 1464-1468.

Lenchik, Leon, Lee F. Rogers, Pierre D. Delmas and Harry K. Genant. 2004 “Diagnosis of Osteoporotic Vertebral Fractures: Importance of Recognition and Description by Radiologists.” *American Journal of Roentgenology* 183(4): 949–958. https://doi.org/10.2214/ajr.183.4.1830949

Lewis, Michelle Huckaby, John K. Gohagan and Daniel J. Merenstein. 2007. “The Locality Rule and the Physician’s Dilemma: Local Medical Practices vs the National Standard of Care.” *JAMA* 297(23):2633-2637.

Lopez CD, Boddapati V, Lombardi JM, Cerpa MK, Lee NJ, Mathew J, Sardar ZM, Lenke LG, and Lehman RA. 2020. “Medicare Utilization and Reimbursement for Vertebroplasty and Kyphoplasty: A National Analysis From 2012-2017.” *Spine (Phila Pa 1976)*: 45(24):1744-1750.

Mainor Alexander J., Nancy E. Morden, Jeremy Smith, Stephanie Tomlin and Jonathan Skinner. 2019. “ICD-10 Coding Will Challenge Researchers: Caution and Collaboration may Reduce Measurement Error and Improve Comparability Over Time.” *Medical Care* 57(7):42-46.

McConnell CT Jr, Wippold FJ 2nd, Ray CE Jr, Weissman BN, Angevine PD, Fries IB, Holly LT, Kapoor BS, Lorenz JM, Luchs JS, O’Toole JE, Patel ND, Roth CJ and Rubin DA. 2014. “ACR appropriateness criteria management of vertebral compression fractures.” *Journal of the American College of Radiology*: 11(8):757-63.

Mehio, A. K., Lerner, J. H., Engelhart, L. M., Kozma, C. M., Slaton, T. L., Edwards, N. C., and Lawler, G. J. 2011. “Comparative hospital economics and patient presentation: vertebroplasty and kyphoplasty for the treatment of vertebral compression fracture.” *American journal of neuroradiology*, 32(7), 1290–1294. https://doi.org/10.3174/ajnr.A2502

Meissner, K., Bingel, U., Colloca, L., Wager, T. D., Watson, A., and Flaten, M. A. 2011. “The placebo effect: advances from different methodological approaches.” *Journal of Neuroscience*, 31(45): 16117-16124

Melton, L. Joseph. 2000 “Who Has Osteoporosis? A Conflict Between Clinical and Public Health Perspectives.” *Journal of Bone and Mineral Research*. 15(12): 2309-2314.

Miller, Franklin G. 2003. “Sham Surgery: An Ethical Analysis.” *American Journal of Bioethics* 3(4): 41-48. DOI: 10.1162/152651603322614580

Molitor, David. 2016. “The Evolution of Physician Practice Styles: Evidence from Cardiologist

Migration." NBER Working Paper #22478.

Monico, Edward P., Chris L. Moore, and Arthur Calise. 2005. "The impact of evidence-based medicine and evolving technology on the standard of care in emergency medicine." *The Internet Journal of Law, Healthcare and Ethics* 3(2).

Monnat, Shannon M. and Khary K. Rigg. 2018 "The Opioid Crisis in Rural and Small Town America." Carsey Research National Issue Brief #135. https://lernercenter.syr.edu/wp-content/uploads/2018/06/The-Opioid-Crisis-in-Rural-and-Small-Town-America_18-1.pdf

Moseley, J. Bruce, Kimberly O'Malley, Nancy J. Petersen, Terri J. Menke, Baruch A. Brody, David H. Kuykendall, John C. Hollingsworth, Carol M. Ashton and Nelda P. Wray. 2002."A controlled trial of arthroscopic surgery for osteoarthritis of the knee." *New England Journal of Medicine* 347:81–88.

Nagaraja, Srinidhi, Hassan K. Awada, Maureen L. Dreher, Shikha Gupta, and Scott W. Miller. 2013 "Vertebroplasty increases compression of adjacent IVDs and vertebrae in osteoporotic spines." *The Spine Journal* 13:1872-1880.

Niven Daniel J., Gordon D. Rubenfeld, Andrew A. Kramer and Henry T. Stelfox. 2015. "Effect of Published Scientific Evidence on Glycemic Control in Adult Intensive Care Units." *JAMA Intern Med* 175(5):801-809

Office of the Surgeon General (US). "Bone Health and Osteoporosis: A Report of the Surgeon General." Rockville (MD): Office of the Surgeon General (US); 2004. PMID: 20945569 https://pubmed.ncbi.nlm.nih.gov/20945569/

Papaioannou Alexandra, Nelson B Watts, David L Kendler, Chui Kin Yuen, Jonathan D Adachi and Nicole Ferko. 2002. "Diagnosis and management of vertebral fractures in elderly adults." *The American Journal of Medicine*, 113(3): 220-228,

Parchomovsky, Gideon and Alex Stein. 2008. "Torts and Innovation." *Michigan Law Review* 17(2):285-316.

Peh Wilfred C.G., Peter L.Munk, Faisal Rashid and Louis A. Gilula. 2008 "Percutaneous Vertebral Augmentation: Vertebroplasty, Kyphoplasty and Skyphoplasty." *Radiologic Clinics of North America* 46: 611–635.

Peters Jr., Philip G. 2000. "The Quiet Demise of Deference to Custom: Malpractice Law at the Millenium." 57 Wash. & Lee L. Rev. 163

Prasad, V., and Adam Cifu. 2011. "Medical reversal: why we must raise the bar before adopting new technologies." *The Yale journal of biology and medicine*, 84(4): 471-478.

Prasad, V., Vandross, A., Toomey, C., Cheung, M., Rho, J., Quinn, S., ... and Adam Cifu. 2013.

"A Decade of Reversal: An Analysis of 146 Contradicted Medical Practices." *Mayo Clinic Proceedings*, 88(8): 790 – 798.

Schmidt, R., Cakir, B., Mattes, T., Wegener, M., Puhl, W., and Richter, M. 2005. "Cement leakage during vertebroplasty: an underestimated problem?" *European spine journal*, 14(5): 466-473.

Sheps Center for Health Services Research. 2022. "Rural Hospital Closures." Accessed March 1, 2022. https://www.shepscenter.unc.edu/programs-projects/rural-health/rural-hospital-closures/

Skinner, Jonathan and Douglas Staiger. 2015. "Technology Diffusion and Productivity Growth in Heath Care." *The Review of Economics and Statistics* 97, no. 5 (December): 951-964

Stock, Gregory. 2003. "If the Goal Is Relief, What's Wrong with a Placebo?" *American Journal of Bioethics* 3(4): 53-54. DOI: 10.1162/152651603322614616

Studdert DM, Mello MM, Sage WM, DesRoches CM, Peugh J, Zapert K and Brennan TA. 2005. "Defensive medicine among high-risk specialist physicians in a volatile malpractice environment." *JAMA*, 293(21):2609–2617.

Teng, G. G. 2008. "Mortality and osteoporotic fractures: is the link causal, and is it modifiable?" *Clinical and experimental rheumatology*, 26(5 0 51), S125.

Voormolen MH, Mali WP, Lohle PN, et al. 2007. "Percutaneous vertebroplasty compared with optimal pain medication treatment: short-term clinical outcome of patients with subacute or chronic painful osteoporotic vertebral compression fractures. The VERTOS study." *American Journal of Neuroradiology* 28:555–60.

Wang Michael T., Greg Gamble and Andrew Grey. 2015. "Responses of Specialist Societies to Evidence for Reversal of Practice." *JAMA Internal Medicine* 175 no. 5 (May):845-848.

Wang, Chao, Xiong Shu, Jianfeng Tao, Yanzhuo Zhang, Yue Yuan and ChengaiWu. 2021. "Seasonal Variation and Global Public Interest in the Internet Searches for Osteoporosis." *BioMed Research International* https://doi.org/10.1155/2021/6663559

Wardlaw, Douglas, Steven R Cummings, Jan Van Meirhaeghe, Leonard Bastian, John B Tillman, Jonas Ranstam, Richard Eastell, Peter Shabe, Karen Talmadge and Steven Boonen. 2009. "Efficacy and safety of balloon kyphoplasty compared with non-surgical care for vertebral compression fracture (FREE): a randomised controlled trial." *Lancet* 373 (9668): 1016-1024.

Weinstein, J. N. 2009. "Balancing science and informed choice in decisions about vertebroplasty." *New England Journal of Medicine*, 361(6): 619-621

West, Kirsten, Jason Devine, Bethany DeSalvo and Katherine Condon. 2010. "The Use of Medicare Enrollment Data in the 2010 Demographic Analysis Estimates." Working Paper No. 89.

Williams Jr. , John M. 2012. “A ’Familiar’ Standard of Care: What the Same or Similar Communities Standard Could Mean for Maryland.” *University of Baltimore Law Review* 41 (193)

Wright, N.C., Looker, A.C., Saag, K.G., Curtis, J.R., Delzell, E.S., Randall, S. and Dawson-Hughes, B. 2014. “The Recent Prevalence of Osteoporosis and Low Bone Mass in the United States Based on Bone Mineral Density at the Femoral Neck or Lumbar Spine.” *Journal of Bone Mineral Research* 29: 2520-2526.

Yang, Wencheng, Jianyi Yang and Ming Liang. 2019. “Percutaneous Vertebroplasty Does Not Increase the Incidence of New Fractures in Adjacent and Nonadjacent Vertebral Bodies.” *Clinical Spine Surgery* 32(2): E99-E106.

CHAPTER 2

Does Medicaid affect treatment intensity and mortality? Evidence from inpatient hospital stays

Hannah Bolder
Helen Levy

Abstract

We analyze the impact of health insurance coverage on the intensity of medical treatment and subsequent mortality for nonelderly adults hospitalized for serious health conditions. Data are from hospital discharge information for 316,000 inpatient stays in 14 states between 2011 and the third quarter of 2015. The Medicaid expansion provides variation in coverage, and restricting the analysis to admissions for serious health conditions allows us to isolate effects on the intensive margin of care and speak to the effects that coverage has on treatment intensity for seriously ill Medicaid patients. We find a significant increase in Medicaid coverage of hospital stays as a result of the Medicaid Expansion which is partially offset by declines in private coverage, leading to a small reduction in uninsurance. The net effect of these changes in insurance is that we find no statistically significant effects of Medicaid coverage on treatment intensity (number of procedures, length of stay and total list charges) or on mortality. Coefficients are imprecisely estimated because the analysis does not separately identify effects from two opposing channels: coverage gains from the uninsured vs the crowd-out of private insurance. Our results highlight the importance of isolating coverage gain channels from crowd-out channels when estimating intensive margin effects of insurance coverage on treatment intensity.

2.1 Introduction

A substantial body of work shows that at the population level, expanding publicly-subsidized health insurance improves access to medical care; see Antonisse et al. 2018 and Mazurenko et al. 2018 for reviews. Simply put, when more people have coverage, more care is consumed. But to what extent do these changes occur on the extensive versus the intensive margin? It is difficult to separate these two effects because the "natural experiments" that support causal inference about the impact of coverage typically induce changes on both margins. It is fairly clear from the evidence to date that increases in care occur on the extensive margin: more people are consuming medical care as a result of coverage expansion, which is reflected in improvements in reported access and increases in the total volume of care. It remains unclear, however, how insurance status affects the intensity of treatment for an individual patient, or what the impact of any change in intensity on outcomes might be.

In this paper, we use the Affordable Care Act's Medicaid expansion, which was adopted in some states but not others, as a source of exogenous variation in coverage. In order to isolate the effects of coverage on the intensity of treatment, our analysis focuses on hospitalizations for serious diagnoses for which the likelihood of hospitalization was plausibly unaffected by changes in coverage. To identify these conditions, we follow the method of Card et al. (2009) by identifying hospital stays that began as emergency room visits and for which there is no variation in the probability of admission across days of the week: "non-deferrable diagnoses." This approach allows us to isolate the intensive margin – these are patients who would have been hospitalized under any circumstances – as well as to say whether any changes in intensity affect patient health outcomes.

Using the State Inpatient Database discharge data from the Healthcare Cost and Utilization Project, we leverage variation across states and time in whether Medicaid was expanded in order to estimate the effects of the expansion on health insurance coverage. We then use the expansion as an instrument to estimate the effect of coverage on treatment intensity and outcomes, measured by length of stay, number of procedures, total list charges and mortality. We find that the expansions increased Medicaid coverage by 6.7 percentage points. The number of uninsured patients decreased, although the magnitude of the effect is imprecisely estimated. There was also a 2.5 percentage point decrease in the fraction of patients with private coverage. Although this effect is not statistically significant, it likely indicates some non-zero level of crowd-out. We found no statistically significant impacts of Medicaid coverage on treatment intensity or mortality. Our analysis does not separately identify effects from two opposing channels: coverage gains from the uninsured, which would likely have positive effects on treatment intensity, vs crowd-out of private insurance, which would likely have negative effects. Therefore to the extent that significant crowd-out effects exist, this would attenuate our estimates of the effects of Medicaid coverage on treatment intensity.

The paper proceeds as follows: Section 2.2 reviews background on our empirical setting, the effects of insurance coverage and the effects of the Medicaid expansion; Section 2.3 introduces the data, Section 2.4 presents our empirical strategy, Section 2.5 discusses our results and Section 2.6

concludes.

2.2 Background

Payer type has long been recognized as a source of variation in healthcare utilization and quality. Early studies found that payer type is correlated with rates of coronary revascularization among patients with ischemic heart disease (Langa and Sussman 1993), longer hospital length of stay and total charges (Arndt et al 1998), and type of medication provided at hospital discharge after heart attack (McCormick et al 1999), among other things. The Oregon Health Insurance Experiment was pursued in order to overcome the challenges associated with estimating causal effects using observational data: Medicaid coverage was randomly granted to a subset of people who were on a waiting list, and their utilization of healthcare services and health outcomes were followed over time using administrative records and surveys. Findings from this RCT indicate that Medicaid coverage increased the use of both outpatient services and inpatient admissions. Although Emergency Department use increased, the increase in inpatient admissions was driven by an increases in admissions that did not occur through the ED. (Taubman et al. 2014, Baicker et al. 2013, Finkelstein et al. 2012) Medicaid coverage improved self-reported health and resulted in lower rates of depression, but the authors found no statistically significant effects on physical health measures such as blood pressure and cholesterol levels. (Baicker et al., 2013) Buchmueller et al (2005), Freeman et al (2008) and Sommers et al (2017) evaluate this literature and conclude that there is strong evidence that insurance coverage increases utilization and improves health.

A huge volume of work has addressed the effects of the Medicaid Expansion on insurance coverage, access to care, utilization, healthcare quality, health outcomes and economic measures such as hospital finance, state budgets and economic growth. The evidence shows that the Medicaid expansion increased insurance coverage, access to care, utilization of many services, and improved healthcare quality. For a review and synthesis of the literature, see Antonisse et al. 2018, Mazurenko et al. 2018, and Guth et al. 2020). In addition, Miller et al. (2021) show that Medicaid coverage decreased annual mortality by 0.132 percentage points. More recent work has also emerged in the domains of behavioral health, reproductive health and racial disparities. (Guth and Ammula 2021) Many studies included in these reviews document significant increases in access to care as a result of Medicaid expansion under the Affordable Care Act. Most of these papers rely on a difference-in-differences design comparing states that did and did not expand Medicaid under the Affordable Care Act: a strong research design for identifying the impact of insurance expansions on these outcomes at the population level.

One feature of the papers discussed thus far is that they estimate the combined effect of changes on the extensive margin (do patients get care at all) and changes on the intensive margin (conditional on getting care, does coverage affect the intensity of treatment). For this reason, the difference-in-differences design at the population level is not ideal for understanding the effect of insurance coverage on treatment intensity, because changes in the composition of the patient pool may have direct effects on average treatment intensity. Moreover, it is hard to say which direction these effects might go. For example, people who have never accessed the system before may seek treatment as a result of gaining coverage; this might bring relatively sicker patients into the system, resulting in an increase in treatment intensity that is not directly related to any change in incentives the physician or hospital faces because the patient has insurance. On the other hand, improved access to care – in particular, better coverage for primary care - may result in patients seeking care at earlier stages, resulting in relatively healthier patients (for example, cancers might be detected at an earlier stage), which would result in less treatment intensity for a particular condition.

The literature now supports arguments in both directions. For example, Lin et al. report increased overall use of outpatient surgical care, and that "[m]ost of this increase represented patients who were newly treated rather than patients who converted from no insurance to Medicaid coverage." These marginal patients may well be sicker than existing patients as a result of having delayed care. In contrast, Loehrer et al. (2018) use data on approximately 300,000 patients aged 18 through 64 admitted to academic medical centers and affiliated hospitals with a variety of common, high-cost conditions (e.g. appendicitis, cholecystitis, and diverticulitis), and find evidence of "improved receipt of timely care." Specifically, they find that an increase in Medicaid coverage is associated with an increase in the probability of early uncomplicated presentation for these conditions. In other words, these patients may be less sick than they would have been in the absence of coverage. These two dynamics are, of course, not mutually exclusive; it is quite plausible that in the first months or years after expansion, the marginal patients who gained Medicaid are sicker than existing patients because of years of deferred medical care; while over time, the same patients are healthier than they would have been in the absence of expansion.

Prior work on the effects of insurance coverage on treatment intensity highlights other challenges related to the patient composition effects discussed above: overcoming selection effects and isolating the contribution of the "coverage gain" channel from the "crowd-out" channel. Several papers have developed methods to address these challenges, with different results. Card et al (2009) demonstrate the importance of focusing on a group of non-deferrable hospital admissions when estimating the effects of Medicare coverage on mortality, and our study borrows this important aspect of their empirical strategy. The authors leverage the fact that the majority of adults gain Medicare coverage at age 65 to construct a regression discontinuity design and estimate the effect of Medicare coverage on mortality and treatment intensity. Using a subset of non-deferrable diagnoses allows them to overcome identification problems due to patient selection: failing to exclude elective admissions would allow for the possibility that patients could wait until they gain

Medicare coverage and only then seek treatment, which would bias their treatment intensity estimates. Card et al (2009) find that Medicare coverage increases treatment intensity: total list charges increase by about 3%, the number of procedures increase by about 4% and 7-day mortality decreases by 1 percentage point. They also attempt to isolate the coverage gain channel from the crowd-out channel by separating patient zip codes into a "low insurance group" and a "high insurance group," however they do not find that treatment intensity estimates vary between the two groups. They derive an upper bound for the mortality effect that arises through the coverage gain channel, which is at most 40% of the magnitude of their estimated effect (averaging across both channels). Therefore they conclude that while some of their treatment intensity effects are due to gaining insurance, a substantial amount must be due to switching from either Medicaid or private insurance to Medicare.

Another closely related study that estimates the effects of Medicaid coverage on treatment intensity is Currie and Gruber (2001). Currie and Gruber (2001) examine the impact of expanding Medicaid eligibility during the late 1980s on treatment intensity during childbirth. Focusing on treatment for childbirth allows them to overcome selection problems because the vast majority of women in the United States give birth in hospitals and virtually all hospitals are required to treat any person who arrives at the hospital in labor, regardless of their insurance status. Currie and Gruber (2001) leverage variation in Medicaid coverage expansions over time and across states to identify the effect of Medicaid coverage on treatment intensity by constructing a simulated probability of being eligible for Medicaid for each unique age-race-education-marital status combination, and assigning this probability to each person who gives birth in their data. In order to separately identify coverage gain and crowd-out channel effects they use a combination of education and marital status as a proxy for pre-expansion insurance status. Results demonstrate that treatment intensity increased for women who likely gained coverage (from an uninsured state) and decreased for those who may have been crowded-out of private coverage. Finally, Doyle (2005) studies the effect of insurance on treatment intensity, focusing on severe automobile accidents in order to avoid selection problems. He finds that relative to privately insured patients, the uninsured receive about 20% less care and have mortality rates that are about 1.5 percentage points higher.

2.2.1 Provider Incentives

Since the focus of this study is intensive margin effects of insurance coverage, it is worth discussing physician and hospital financial incentives in the context of different kinds of patient insurance. Hospitals are reimbursed 1.5 times more generously by private insurance than by Medicaid for inpatient stays, and 9 times more generously for Medicaid patients than for the uninsured. (Nikpay et al. 2016) Estimates for what fraction of expenses are paid out-of-pocket by the uninsured vary from 20% (Coughlin et al. 2014) to 33% (Finkelstein, Hendren and Luttmer 2019). The remaining fraction, or uncompensated care, costs hospitals an average of about $800 per year per uninsured person (Garthwaite et al. 2018). Federal Disproportionate Share Hospital Payments and uncompensated care funds that exist in some states partially off-sett uncompensated care costs but large costs remain, and hospitals serve as "insurers of last resort" (Garthwaite et al., 2018). In fact, for every dollar of Medicaid spending, 60 cents offsets provider uncompensated care costs (Finkelstein et al. 2019).

Medicaid enrollees may participate in fee-for-service or managed care plans. In fee-for-service plans the state reimburses physicians and hospitals on a per-service basis, whereas for managed care the state pays an organization to manage the payment logistics and in return for a set fee per Medicaid participant. Managed care represents a growing share of Medicaid enrollees, however typically Medicaid managed care organizations also reimburse physicians and hospitals on a fee-for-service basis so the incentive structure is similar under both plans. In order to compare Medicaid reimbursement generosity it is useful to use Medicare reimbursement rates as a benchmark. According to the Medicaid and CHIP Payment and Access Commission (MACPAC), Medicaid physician fees are about 2/3 as generous as Medicare rates and Medicaid hospital reimbursement is similar or slightly higher than Medicare rates (MACPAC 2022). According to a Kaiser Family Foundation analysis, private insurance reimbursement rates for physicians are 143% of Medicare levels, and private insurance reimbursement rates for hospitals for inpatient care are 189% of the Medicare levels (Lopez et al 2020). This research confirms that hospitals and physicians receive higher compensation for providing healthcare to patients with private insurance than those with Medicaid, and the least compensation for uninsured patients.

2.3 Data

We use data on inpatient hospital admissions from the Healthcare Cost and Utilization Project (HCUP) database, produced and housed by the Agency for Healthcare Research and Quality (AHRQ). The inpatient hospital admissions data is in the State Inpatient Database, known as the SID.

Our data includes the following 10 states that expanded Medicaid: Arizona, Arkansas, Colorado, Iowa, Massachusetts, Michigan, Minnesota, New Jersey, New York, and Washington. Four states that did not expand Medicaid (during our period of analysis) are included in our data: Kansas, Nevada, North Carolina and Wisconsin. Our SID data includes the universe of hospital inpatient admissions for all the states listed above except Colorado from 2011 through the third quarter of 2015. For Colorado we have data for 2011 through the third quarter of 2015 except for the year 2012. No analytic criteria was applied to select which states to include in this study: any state data that was available to the researchers to re-use for no additional fee was included in the study,[1] with the exception of Florida and Maryland which were excluded because no data on month of admission or discharge was available for these states. (This information is necessary because the models below include month fixed effects.)

After applying our admissions selection criteria which we explain in more detail below, we are left with about 316,000 admissions across these 14 states. For each admission record, diagnosis and procedure codes are provided, as well as information about whether the patient died in the hospital, whether the admission occurred on a weekend or weekday, the length of stay and total list charges. The diagnosis codes and weekend-weekday admissions variables will allow us to select only those admissions that are deemed to be non-deferrable. We will measure utilization and treatment intensity for a given admission using the number of procedures, the length of stay, and total list charges, following Card et al (2009).

To date, the only variation in timing of the Medicaid expansion that we are able to exploit is due to Michigan's expansion in April of 2014. (The other states that expanded Medicaid did so in January of 2014.) In the future we hope to apply for more data, and this additional data should include some other states that expanded Medicaid after January 2014.

It should be noted that Medicaid coverage in Arizona, New York and Massachusetts before the expansion was already much better than in the average U.S state. In addition, New Jersey and Washington had some coverage for childless adults before the Medicaid expansion - unlike most states that restricted eligibility to adults with dependent children and pregnant women. The expansions did extend coverage to a larger group of childless adults in these states, but for some childless adults who were already covered by Medicaid before the Affordable Care Act the expansion represented a change in financing rather than a change on the extensive margin of coverage.

Our analytic sample is selective in two ways: it is a selective sample of states and also a selective sample of diagnoses. We will demonstrate that the selection based on states does not introduce bias. The selection of diagnoses is intentional. Table 1 displays mean characteristics of our sample, comparing expansion states and non-expansion states, before and after the expansion. Compared to patients in non-expansion states, patients in expansion states were less likely to be Black, more likely to be Hispanic, Asian or Pacific Islander, and slightly younger. Patients in expansion states lived in areas where the median income was slightly higher relative to the national distribution and relative to the distribution in their individual state. The characteristics of both expansion and non-expansion states remained very stable before and after the expansion, with only the fraction of patients who were female changing by more than one percentage point.

Figure 1 and Figure 2 use data from the American Community Survey to demonstrate that

[1] A state's data was available to re-use for no additional fee if other researchers at the University of Michigan had already purchased the data for a prior study.

our sample of 10 expansion and 4 non-expansion states is representative of expansion and non-expansion states in the United States overall. In both our sample and in the US as a whole, Medicaid coverage increases dramatically in expansion states. The fraction of people who are uninsured decreases in both expansion and non-expansion states, but decreases more in expansion states. Private non-group insurance coverage increases in expansion and non-expansion states but increases more in non-expansion states. Non-expansion states see modest increases in employer coverage. These patterns hold whether we compare our 10 expansion states with our 4 non-expansion states, or the complete group of expansion and non-expansion states. Figure 2 shows that relative to the US as a whole, our sample of 14 states has higher income, higher educational attainment and fewer Hispanic people. Although differences in levels exist between our sample and the US overall, they do not disproportionately affect only expansion states in our sample or only non-expansion states, but instead represent a level shift in both groups of states. The trends in these demographic variables are similar in our sample and in the US overall, so it is difficult to see how these differences in levels would introduce bias in our treatment intensity estimates, especially considering that trends in insurance coverage in our sample mirror trends in the US overall (Figure 1).

One limitation of our analysis is the potential for measurement error in our insurance coverage variable. The State Inpatient Database reports the "expected payer" for each hospital admission, rather than the entity who actually submitted payment. As such, the expected payer information is reported to HCUP by each state based on hospital level determinations rather than based on Medicaid program records or insurance company records. Expected payer classifications are often made by a hospital's business or finance department using information such as patient insurance cards, or registration or admission notes from in-take forms. According to a 2005 study, Medicaid managed care patients are sometimes misclassified under private managed care. (Chattopadhyay and Bindman, 2005) According to a 2018 HCUP methods report, the distinction between "expected payer" and the actual payer may be especially important in contexts such as this study, where a hospital must decide whether a patient who was uninsured at the time of admission will likely retroactively qualify for Medicaid. (HCUP, 2018) For successful Medicaid applicants state Medicaid programs generally cover medical care for the 3 months preceding the patient's application date. [2] In addition, the Affordable Care Act gave hospitals in all 50 states the power to choose to implement presumptive eligibility programs to temporarily enroll patients in Medicaid if they determined that they were likely to meet their state's eligibility requirements. Presumptive eligibility allows hospitals to receive timely payment for services for this population rather than waiting until the patient is well enough to take the time to apply for Medicaid and then the further delay while the application is reviewed. Some states had allowed hospitals to grant presumptive eligibility to a limited

[2] Iowa has a Section 115 waiver that exempts their Medicaid program from retroactive coverage obligations in some cases.

group of patients before 2014, which likely also decreased measurement error for the insurance coverage variable in those states. According to the HCUP 2018 methods report few studies have assessed the accuracy of the expected payer variable, however hospitals have a financial incentive to implement presumptive eligibility programs and to correctly classify patients as uninsured or likely eligible for Medicaid, and given the improvements in electronic medical and billing systems it seems likely that measurement error has fallen over time.

2.4 Empirical Strategy

2.4.1 Constructing a Non-Selective Sample of Admissions

The first step in our analysis is to construct a sample of hospital admissions based on conditions for which patient composition is unlikely to have changed as a result of Medicaid expansion.

We first restricted the sample to admissions for non-deferrable diagnoses. The approach of restricting admissions to only include non-deferrable diagnoses to overcome selection issues has been used in many studies, such as Dobkin (2003), Card, Dobkin and Maestas (2009); Mulcahy et al. (2013), Doyle, Graves, and Gruber (2017); and most recently Cooper et al. (2022). In Mulcahy et al. (2013), non-deferrable diagnoses are defined according to panel of 8 experts including physicians, social scientists and a health informatics expert. The remaining studies mainly rely on data-driven approaches. The data-driven methods rely on the insight by Dobkin (2003) that while elective or urgent but less critical admissions occur more often during weekdays, conditions that are truly non-deferrable (and not caused by accidents) will occur on all days of the week with equal probability. Card, Dobkin and Maestas (2009) use a similar approach, defining admissions as non-deferrable if the probability of weekend admission for the associated diagnosis is sufficiently close to 2/7. Doyle, Graves, and Gruber (2017) classify a diagnosis as non-deferrable if the weekend admission probability is as close or closer to 2/7 than the weekend admission probability for hip-fracture. Most recently, Cooper et al. (2022) use the non-deferrable diagnoses defined in Dobkin (2003), Card, Dobkin and Maestas (2009); and Doyle, Graves, and Gruber (2017), supplemented with the non-deferable diagnoses selected by the expert panel in Mulcahy et al. (2013). Note that although these studies share similar methodology for selecting non-deferrable diagnoses, they do not all focus on the effect of insurance. Doyle, Graves, and Gruber (2017) estimate the effect of hospital spending on healthcare quality and Cooper et al. (2022) investigate whether receiving care at a high-priced hospital reduced mortality.

Our study is most similar to Card et al. (2009) because like them, we are interested in the effect of insurance coverage on treatment intensity. Other studies of this question have addressed the problem of selection into treatment by focusing on a single condition for which the probability of treatment is essentially 100% (eg ischemic heart disease in Langa and Sussman (1993) or childbirth in Currie and Gruber (2001)). Following Card et al (2009), we test the hypothesis that the probability of being admitted on a weekend with a given diagnosis is 2/7, and we label a diagnosis "non-deferrable" if the t-statistic for this test is less than 0.429 in absolute value, where 0.429 is the cut-off for the bottom quartile in the distribution of the absolute value of the t-statistics. (We also drop diagnoses that have an estimated weekend admission probability of zero.) [3] Figure 3 shows a density plot of the number of ICD-9 diagnosis codes by probability of weekend admission, comparing our sample of non-deferrable diagnoses to 1) all diagnoses that occur in admissions through the ED and 2) all diagnoses that occur in the data for our population of interest. After excluding diagnoses for which the probability of being admitted on a weekend was very different from 2/7, we were left with about 1,500 diagnoses. The ten most common non-deferrable diagnoses and some accompanying descriptive statistics for these conditions are listed in Table 2; these diagnoses account for about one-third of the observations in our analytic sample. Figure 4 displays the evolution over time of total discharges for diagnoses that we classify as non-deferrable. The total number of non-deferrable discharges does not seem to increase as a result of the Medicaid expansion, whether in expansion or in non-expansion states, lending support to the idea that the diagnoses we examine were sufficiently serious that these patients would all have been hospitalized, regardless of insurance status.

The sample population includes adults between 26 and 64 years old, because the majority of adults are eligible for Medicare coverage at age 65. Under the Affordable Care Act, private insurance plans that offer coverage for dependents are required to include adult children under the age of 26 as eligible dependents so they are also omitted in this analysis. We also dropped admissions that appeared to be related to pregnancy or birth because Medicaid programs in most states covered pregnant women before the expansion. In the following paragraphs we first present a difference in differences strategy to estimate the effect of the expansion on Medicaid coverage. This becomes the first stage in an IV analysis of the effect of Medicaid coverage on treatment intensity, which we present next.

2.4.2 Estimating the effect of the expansion on Medicaid and other insurance coverage

We leverage variation in which states expanded Medicaid and timing in the expansions to estimate the following difference in differences specification:

[3] As a robustness check we classify all diagnoses that are in the bottom two quartiles as non-deferrable (a less conservative definition), and the results are similar (available upon request).

$$Medicaid_{ist} = \alpha_0 + \beta_1 Expansion_{st} + \gamma_t + \delta_s + \mathbf{X}_{\mathbf{it}}^{'}\mathbf{\Gamma} + \epsilon_{ist} \tag{2.1}$$

$Medicaid_{ist}$ is an indicator variable that equals one if the patient in admission i in state s at time t is covered by Medicaid. $Expansion_{st}$ is an indicator variable that equals 1 if admission i occurred in a state that had expanded Medicad at the time of admission. $\mathbf{X_{it}}$ is a vector of covariates that includes patient sex, age, age squared, indicators for race categories and indicators for the quartile in the national income distribution corresponding to median income in the patient's zip code. Race categories include White, Black, Hispanic, Asian or Pacific Islander, Native American and Other. We also include fixed effects for diagnosis groups as defined by the Clinical Classification System (CCS). The CCS groups reduces the over 14,000 diagnosis codes in the ICD-9 coding system into a smaller number of categories that remain clinically meaningful. Γ_s are state fixed effects, δ_t are fixed effects for year and month of admission, and standard errors are clustered at the state level. In addition to Medicaid insurance, we use this specification to investigate the expansions affected other types of coverage, such as private insurance, Medicare, and other payer. As noted earlier, currently the only state in our sample that did expand Medicaid but not in January 2014 was Michigan. We plan to add more states to the analysis, which will contribute to this aspect of the estimation.

In order for β_1 to be an unbiased estimate of the effect of the expansions on Medicaid coverage, the following assumption is required: had the expansions not occurred, the trends in Medicaid coverage would have evolved similarly in expansion and non-expansion states. In addition, the Medicaid expansion must be the only "treatment" that occurs at the event time that differentially affects the probability of Medicaid coverage in expansion and non-expansion states. If these assumptions are met, then the interpretation of β_1 is that the Medicaid expansion causes a $100 \times \beta_1$ percentage point increase in the probability of having Medicaid coverage.

Trends in the fraction of discharges by payer and state expansion status are shown in Figures 5 and 6. Event study graphs also allow a visual inspection for differential trends in the outcome variables, and can be used as a more flexible method to investigate the effects of the expansion on insurance outcomes. We use the following event study specification, where $\mathbf{X_{it}}$ is as defined as above and standard errors remain clustered at the state level:

$$InsuranceCoverage_{ist} = \alpha_0 + \sum_{k \neq -1} \tau_k \times D_{st}^k + \mathbf{X}_{\mathbf{it}}^{'}\mathbf{\Gamma} + \gamma_s + \delta_t + \epsilon_{ist} \tag{2.2}$$

Where $D_{st}^k = I(E_s - k = 1)$. E_s is the date that Medicaid was expanded in state s (January 2014 for all states that expanded Medicaid except for Michigan.) The indicator variable $I(E_s – k) = 1)$ is equal to 1 if the time to expansion is exactly k months. This is essentially the difference in differences specification in (2.1) with the $Expansion_{st}$ variable disaggregated into a series a leads and lags.

2.4.3 Estimating the effect of Medicaid coverage on treatment intensity and mortality

The next analysis uses equation (2.1) as the first stage in an instrumental variables strategy. Consider the following equation:

$$TreatmentIntensity_{ist} = \alpha_1 + \rho_1 \times Medicaid_{ist} + \mathbf{X}_{\mathbf{it}}^{'}\mathbf{\Omega} + \gamma_s + \delta_t + \epsilon_{ist} \tag{2.3}$$

where $TreatmentIntensity_{ist}$ is one of three outcome variables (the number of procedures per admission, total list charges or length of stay) and other variables are as previously defined. In general, insurance coverage is endogenous due to adverse selection. Low income people will be more willing to incur the costs associated with applying for Medicaid if they are in worse health. (These costs could include time, effort and potentially social stigma.) If having Medicaid (relative to being uninsured) is a signal of worse underlying health, then estimates of ρ would confound the effect of Medicaid coverage with the effects of worse underlying health. All else equal, worse underlying health should increase mortality risk and length of stay. Medicaid enrollees likely differ on unobservables from people with other types of insurance as well. Since estimates of ρ would likely be biased, we use the Medicaid expansion as an instrument for Medicaid coverage. The first stage is given in (2.1) above and the reduced form is:

$$TreatmentIntensity_{ist} = \alpha_3 + \beta_2 \times Expansion_{st} + \mathbf{X}_{\mathbf{it}}^{'}\tau + \gamma_s + \delta_t + \epsilon_{ist} \tag{2.4}$$

where γ, δ and $\mathbf{X_{it}}$ are as defined above, and standard errors are clustered at the state level. The effect of Medicaid coverage on the treatment intensity outcome variable is β_2/β_1, the reduced form estimate divided by the first stage. The central assumption that underlies this analysis is that the Medicaid expansion must only affect the outcomes of interest through its effect on Medicaid insurance coverage. For example, if the Medicaid expansion were bundled with other policies that affected hospital reimbursement rates, this assumption would likely be violated because hospital reimbursement rate policies would directly affect treatment intensity. An important point to address is the role of the health insurance exchanges, which were established around the same time as the Medicaid expansions. The implementation of the insurance exchanges resulted in gains in private coverage in both states that expanded Medicaid and in those that did not. To the extent that differential gains in private coverage (higher gains in non-expansion states than in expansion states) were a direct result of crowd-out from the Medicaid expansions, this does not bias the IV estimates even though private insurance coverage is correlated with higher treatment intensity. In order for the implementation of the health insurance exchanges to bias our IV estimates, it would need to have a direct effect on private insurance coverage that is correlated with the state expansion status instrument but that arises through an different channel. That is, the health insurance exchanges would need to induce a correlation between the Medicaid expansion status instrument and private insurance coverage other than through the direct Medicaid-coverage-induced crowd-out channel, which seems unlikely.

Changes in private insurance coverage that differ between expansion and non-expansion states should not bias our estimates as long as they arise through the Medicaid coverage channel, however they do affect the interpretation of our treatment intensity results. Changes in Medicaid coverage after the Medicaid expansions can arise though two channels - the "coverage gain" channel and the "crowd-out" channel. The Medicaid expansions induced many previously uninsured people to gain Medicaid coverage. However there may also have been crowd-out effects, as the expansion of a large public insurance program induced people to participate in Medicaid when they would have otherwise had private insurance. A substantial literature measures the size and determinants of these crowd-out effects. Buchmueller et al (2015) provides a comprehensive review, beginning with the seminal contribution made by Cutler and Gruber (1996) through more recent work. Our IV estimates represent the effects of Medicaid coverage that arise from both of these channels. Treatment intensity effects from the coverage gain channel are likely to be positive, while effects from the crowd-out channel are likely to be negative. Therefore to the extent that both of these channels play an important role, failing to separate these channels will result in attenuation bias.

2.5 Results

2.5.1 First Stage Results

We first present event study graphs for the probability that an admitted patient is covered by various types of insurance, corresponding to equation 2.2. As explained in Section 2.4, all event studies are the same as their corresponding difference in difference specifications except that $Expansion_{st}$ is replaced by a set of treatment indicator lags and leads D_{st}^k, where $D_{st}^k = 1$ when the expansion event is exactly k months away. The point estimates represent the τ coefficients on the D_{st}^k terms of the event study regressions and their corresponding standard errors are in the associated vertical bars.

The point estimates in Figure 7 show a sharp increase in Medicaid coverage after the expansion and this effect increases over time. Difference in differences estimates in Table 3 Column 1 (equation 2.1) confirm: the Medicaid expansion caused a statistically significant 6.7 percentage point increase in the probability of Medicaid coverage. Figure 7 shows no evidence of a pre-trend, supporting a causal interpretation of this effect. Figures 8-11 are event studies for the effect of the Medicaid expansion on other types of insurance coverage/payers: Medicare (Figure 8), private insurance (Figure 9), no insurance (Figure 10) and "other payer" (Figure 11). Although most of the pre-period coefficients in Figure 8 are not statistically different from zero, there does appear to be an upward trend in the magnitude of the coefficients, making inference about the effect of the Medicaid expansion on Medicare coverage using difference-in-differences methods problematic. Since our sample is restricted to admissions for patients between 26 and 64 years old, Medicare coverage in this context would apply mainly to people with disabilities who qualify through the Social Security Disability Insurance (SSDI) program. [4] Therefore we would expect the Medicaid expansion to affect Medicare coverage rates through an effect on SSDI applications and subsequent program participation. [5] As discussed by Schmidt and Watson (2020) the Medicaid expansions could cause an increase in disability applications through an information channel or an "employment lock" channel, as Medicaid provides an alternative to employer provided coverage, or could cause a decrease in applications through an "alternative source of health insurance" channel, since qualifying for Medicare via SSDI participation is a much more lengthy and onerous process than qualifying for Medicaid through income eligibility. Relying on a different identification strategy to overcome the challenges of difference-in-differences in this context, Schmidt and Watson (2020) find no economically meaningful impact of the Medicaid expansion on disability applications. Since disability applications serve as a first stage for the eventual effect of the Medicaid expansion on Medicare coverage rates, it seems highly unlikely that Medicare coverage was affected by the expansions.

Figure 9 shows the event study corresponding to the effect of the expansions on private insurance coverage. Most of the pre-period coefficients are not statistically different from zero. To the extent that the magnitude of these coefficients is drifting downwards, the trend is very slight. Difference-in-differences estimates in Column 2 of Table 3 indicate that the effect of the Medicaid expansion on private insurance coverage was not statistically significant and the magnitude of the coefficient is small, representing a decrease of 2.6 percentage points in the probability of private coverage. Most of the increase in Medicaid coverage came from gains from the uninsured rather

[4]People who have End Stage Renal Disease also qualify for Medicare before age 65, but this represents a very small share of Medicare beneficiaries (MEDPAC, 2021).

[5]Initial SSDI decisions are made 4 months after an application is submitted, and 30% of applications are approved at that time. Many more applications are approved upon appeal, which can take up to two years. (Schmidt and Watson, 2020) Therefore we would expect to observe a lag of at least 4 months after the expansions before Medicare coverage would be affected.

than crowd-out of private insurance. This is consistent with the trends displayed in Figure 5, where sharp increases in Medicaid coverage rates in expansion states are accompanied by a large drop in the fraction uninsured, and consistent with Figure 6, where private insurance coverage rates in expansion states fall by very little. In addition, the event study in Figure 10 shows that the Expansion was associated with a large decrease in rates of uninsurance. About 10 of the pre-period coefficients are statistically different from zero. Looking at the event study overall, there does not seem to be a trend in the pre-period coefficients; however, the data beginning about a year and a half before the Expansions does suggest a small downward anticipatory trend in Expansion states.

Figure 11 turns to the effect of the expansions on the probability of coverage by another payer, including Workers Compensation, charity care, Veterans Affairs, and Indian Health Service. None of the pre-period coefficients are statistically different from zero and there is no evidence of a pre-period trend. The event study shows that after the Medicaid expansion "other payer" coverage rates fell by between 1 and 2 percentage points and the effect remained constant through the third quarter of 2015. Column 5 of Table 3 confirms with an average effect of 1.7 percentage points. While this effect is not statistically significant, we would expect to see a decline in charity care as more uninsured people gain Medicaid coverage.

2.5.2 Treatment Intensity Results

According to our estimates, Medicaid coverage does not have a statistically significant impact on treatment intensity or mortality on the intensive margin. Table 4 contains estimates of the effects of the expansion on treatment intensity (the reduced form) and Table 5 contains IV estimates of the impact of Medicaid insurance on treatment intensity. The IV estimates in Table 5 come from dividing the reduced form estimates in Table 4 by the first stage in column 1 of Table 3. Unfortunately the standard errors on the treatment intensity estimates are very large so we cannot rule out economically significant effects. Figures 12-15 are event studies corresponding to the difference-in-differences specifications in Table 4 (event study versions of equation 2.4). The length of stay event study in Figure 12 shows that few of the pre-trend coefficients are statistically different from zero and there is no clear pre-trend. In the number of procedures event study (Figure 13) there does appear to be a slight negative pre-trend and a significant number of the pre-trend coefficients are statistically different from zero. The number of procedures appears to decrease as a result of the Medicaid expansion and this effect grows over time, however given the pre-trend this effect may not be causal. Figure 14 and Figure 15 show event studies for total list charges and for mortality, respectively. Neither event study shows evidence of a pre-trend nor evidence of a statistically significant impact of the Medicaid expansions on these treatment intensity outcomes. (Log specifications for total list charges available upon request.)

2.6 Conclusion

This study estimates the effects of Medicaid insurance on the intensive margin of treatment intensity. Confining the analysis to only serious inpatient non-deferrable conditions allows us to estimate these effects using difference in differences, event studies and instrumental variables methods without the interference of patient composition effects. The Medicaid expansion increased the fraction of patients among those with non-deferrable diagnoses who had Medicaid coverage. However in our data, this change may have been partially offset by a small amount of crowd-out of private coverage. This complicates the analysis because our treatment intensity estimates reflect a combination of (likely) opposing effects: the effect of gaining Medicaid coverage for the previously uninsured and effects from the crowd-out of private coverage. To the extent that we are primarily interested in intensive margin effects of Medicaid coverage through the coverage gain channel, crowd-out of private insurance will result in downward bias. While the effect of the Medicaid expansions on private coverage just missed the cutoff for statistical significance, we cannot rule out crowd-out effects that would be large enough to cause such bias.

Note that in our context there are two kinds of crowd-out: 'direct" crowd-out when people switch from private coverage to Medicaid, and "counterfactual crowd-out," which represents crowd-out that occurs when people who would have taken-up private insurance instead gain Medicaid coverage. Had the Medicaid expansion not occurred, private insurance coverage rates would have likely increased in both expansion and non-expansion states due to the Health Insurance Exchanges and employer mandates. A differential change in private insurance rates between expansion and non-expansion states before and after 2014 therefore reflects a combination of traditional crowd-out, when some eligible people abandon their pre-existing private coverage in favor of Medicaid, and "counterfactual crowd-out," when some people who would have *gained* private coverage had Medicaid not been available gain Medicaid coverage instead.

Some may wonder if an implication of our analysis is that on average for serious non-deferrable conditions, physician financial incentives to treat patients differently based on insurance status do not seem to have much effect on treatment intensity. Perhaps other variables are more important determinants of care in this setting, such as severity of the patient's underlying health condition, conditional on a given admitting diagnosis. Restricting the sample to non-deferrable diagnoses which often represent the most severe patient conditions has the advantage of allowing us to overcome selection issues, but at the cost of potentially limiting the role of financial incentives to the extent that physicians and hospitals will provide life-saving care regardless of insurance status due to professional ethics or laws such as the Emergency Medical Treatment and Labor Act (EMTLA). In the context of Medicaid coverage, it does seem likely that there is more room for financial incentives to matter in less urgent situations, where the benefits of interventions are less clear, or where effects of treatment accrue gradually over time. However many other studies find that financial incentives affect treatemnt intensity in accute inpatient settings. One such example that was discussed previously is Currie and Gruber (2001), who find that Medicaid coverage increases treatment intensity (relative to a lack of insurance), and the effect is stronger when financial incentives are steeper. The setting in Currie and Gruber (2001) was childbirth - a context where laws such as EMTLA are especially salient and medical malpractice risks for failing to provide adequate care are high. Considering that studies even in these circumstances have found that insurance coverage affects treatment intensity, it seems likely that what drives our statistically insignificant treatment intensity results is the inability to separately identify the coverage gain channel from the crowd-out channel. Separately identifying the contribution of these channels would be a fruitful avenue for future research.

2.7 Tables

Table 2.1: Summary statistics: Expansion and non-expansion states, pre and post 2014

	Non-expansion states pre-2014	Non-expansion states post-2014	Expansion states pre-2014	Expansion states post-2014
Black	.193044	.187608	.151886	.151171
Hispanic	.030549	.038738	.093767	.099522
Asian or Pacifical Islander	.00707	.00713	.0195	.022
Native American	.013832	.011427	.009719	.009114
Other race	.012936	.011113	.036195	.041612
Median household income, national quartile	2.0409	2.06047	2.53186	2.53776
Median household income, state quartile	2.27218	2.25678	2.30701	2.28865
Female	.448019	.43183	.427951	.416845
Age	48.2485	48.109	47.8843	47.9948

Notes: Mean characteristics of expansion states and non-expansion states, before and after expansion. Michigan expanded Medicaid in April 2014, so for Michigan post-2014 is defined as April 2014 or later and discharges from Jan-March 2014 are coded as pre-2014. Median household income quartile is the quartile of the median income in the patient's zipcode, averaged across discharges. In row 6 national income quartiles are used and in row 7 state income quartiles are used. National and state quartiles vary by year. In 2011, the national quartiles were 1 $1-$38,999 2 $39,000 - $47,999 3 $48,000 - $63,999 4 $64,000 and above.

Table 2.2: Summary statistics for the top 10 most common non-deferrable diagnoses

	# discharges	mean length of stay	mean # procedures	mean total list charges	mortality rate
ami inferior wall, init	23,153	5.17	1.69	$ 44,968	.05
facial weakness	19,550	3.27	7.44	$ 72,803	.02
ocl crtd art w infrct	16,508	4.48	.67	$ 28,180	.01
other alter consciousnes	10,844	3.5	.96	$ 26,794	.02
peritonsillar abscess	7,608	3.76	1.14	$ 30,483	.03
poison-antipsychotic nec	7,566	3.38	.84	$ 21,671	0
poisoning-opiates nec	5,214	3.27	1.05	$ 33,442	.01
poisoning-opium nos	5,077	2.11	1.07	$ 16,765	0
rhabdomyolysis	3,880	6.28	2.19	$ 60,298	.03
traumatic subdural hem	3,669	7.74	2.45	$ 78,119	.05
Totals	103069	430353	249385	$ 4,279,531,776	2,546

Notes: Statistics for the top 10 non-deferrable diagnoses. Bottom row represents sums across the diagnoses. Mortality rate is per 100 discharges except for the bottom row, where it is total # people who died.

Table 2.3: Difference in differences: Effect of Expansion on health insurance outcomes

Table 3: Differences in differences: Effect of Medicaid expansion on insurance outcomes

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Medicaid	Private Insurance	Uninsured	Medicare	Other payer
Expansion*Post	0.0665***	-0.0256*	-0.0374	0.0142	-0.0170*
	(0.0196)	(0.0122)	(0.0247)	(0.00907)	(0.00961)
Female	0.0257***	0.0202***	-0.0393***	0.0228***	-0.0290***
	(0.00562)	(0.00485)	(0.00536)	(0.00219)	(0.00291)
Age in years at admission	0.000559	0.00399***	-0.00429***	0.00202**	-0.00229***
	(0.00154)	(0.00111)	(0.000694)	(0.000768)	(0.000742)
Age*Age	-5.45e-05***	-2.27e-05	7.41e-06	5.30e-05***	1.70e-05*
	(1.50e-05)	(1.29e-05)	(8.03e-06)	(9.13e-06)	(8.15e-06)
Black	0.0913***	-0.120***	0.0157***	0.0126	0.00112
	(0.0128)	(0.0107)	(0.00361)	(0.0103)	(0.00339)
Hispanic	0.0915***	-0.134***	0.0513***	-0.0368***	0.0283**
	(0.0288)	(0.0131)	(0.0153)	(0.00414)	(0.0104)
Asian or Pacific Islander	0.0792*	-0.0189	0.0162*	-0.0808***	0.00491
	(0.0420)	(0.0352)	(0.00786)	(0.0125)	(0.00770)
Native American	0.0897***	-0.117***	-0.0205	-0.0271	0.0737*
	(0.0158)	(0.0188)	(0.0178)	(0.0158)	(0.0406)
Other race	0.0915***	-0.0883***	0.0314***	-0.0508***	0.0163
	(0.0230)	(0.00802)	(0.00770)	(0.00364)	(0.0127)
Median household income in second quartile	-0.0774**	0.0729***	-0.00320	-0.00174	0.00989*
	(0.0273)	(0.0111)	(0.00889)	(0.00667)	(0.00464)
Median household income in first third quartile	-0.121***	0.145***	-0.0116	-0.0233**	0.0106**
	(0.0255)	(0.0106)	(0.00863)	(0.00840)	(0.00471)
Median household income in fourth quartile	-0.176***	0.253***	-0.0220*	-0.0556***	0.00192
	(0.0287)	(0.00910)	(0.0124)	(0.0118)	(0.00612)
Observations	316,846	316,846	316,846	316,846	316,846
R-squared	0.105	0.119	0.076	0.088	0.042
mean	0.228	0.381	0.124	0.201	0.0640

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: All specifications include fixed effects for DXCCS diagnosis code group, state, year and month. Standard errors are clustered by state. Income quartiles are from the national income distribution.

Table 2.4: Difference in differences: Effect of Expansion on treatment intensity

Table 4: Differences in differences: Effect of Medicaid expansion on Treatment Intensity

VARIABLES	(1) # procedures	(2) length of stay	(3) total list charges	(4) mortality
Expansion*Post	-0.302	-0.0317	324.5	-0.000484
	(0.225)	(0.0598)	(558.1)	(0.00128)
Female	-0.180***	-0.122***	-1,424***	-0.00327***
	(0.00842)	(0.0228)	(248.3)	(0.000498)
Age in years at admission	0.00457*	-0.00698	8.002	-0.00123***
	(0.00226)	(0.0134)	(43.52)	(0.000229)
Age*Age	2.48e-05	0.000510***	2.366***	1.71e-05***
	(2.02e-05)	(0.000120)	(0.632)	(2.86e-06)
Black	0.0699**	0.477***	2,619***	-0.000497
	(0.0272)	(0.0460)	(513.2)	(0.000562)
Hispanic	-0.00207	0.0342	814.0	-0.00285***
	(0.0180)	(0.0442)	(851.4)	(0.000642)
Asian or Pacific Islander	0.175*	0.251	1,747	0.00370
	(0.0903)	(0.223)	(1,232)	(0.00255)
Native American	-0.0172	0.138	1,473	-0.00167
	(0.1000)	(0.109)	(1,386)	(0.00143)
Other race	0.256***	0.521***	5,007***	0.000390
	(0.0605)	(0.0505)	(694.6)	(0.00124)
Median household income in second quartile	-0.0207	-0.145*	414.5	-0.000249
	(0.0169)	(0.0680)	(426.4)	(0.000843)
Median household income in first third quartile	-0.00172	-0.138*	1,907**	-0.000409
	(0.0175)	(0.0674)	(659.1)	(0.000780)
Median household income in fourth quartile	-0.0122	-0.266***	3,472	-0.00228***
	(0.0289)	(0.0575)	(2,636)	(0.000738)
Observations	316,846	316,841	305,970	316,374
R-squared	0.342	0.066	0.192	0.039
mean	2.126	4.716	36291	0.0200

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: All specifications include fixed effects for DXCCS diagnosis code group, state, year and month. Standard errors are clustered by state. Income quartiles are from the national income distribution.

Table 2.5: IV estimates of the effects of health insurance on treatment intensity

Table 5: IV estimates of the effect of health insurance on treatment intensity

	(1)	(2)	(3)	(4)
VARIABLES	# procedures	length of stay	total list charges	mortality
medicaid	-4.543*	-0.477	4,873	-0.00727
	(2.648)	(0.909)	(8,277)	(0.0193)
Female	-0.0636	-0.109***	-1,551***	-0.00308***
	(0.0752)	(0.0323)	(264.8)	(0.000684)
Age in years at admission	0.00711	-0.00671	4.861	-0.00122***
	(0.00719)	(0.0129)	(39.60)	(0.000220)
Age*Age	-0.000223	0.000484***	2.632***	1.67e-05***
	(0.000161)	(0.000119)	(0.818)	(2.68e-06)
Black	0.485*	0.520***	2,177***	0.000168
	(0.275)	(0.102)	(783.2)	(0.00156)
Hispanic	0.414	0.0778	367.0	-0.00218
	(0.298)	(0.103)	(665.2)	(0.00181)
Asian or Pacific Islander	0.535	0.289	1,365	0.00428
	(0.399)	(0.272)	(870.7)	(0.00270)
Native American	0.391**	0.181	1,028	-0.00102
	(0.175)	(0.118)	(1,444)	(0.00257)
Other race	0.672**	0.564***	4,563***	0.00106
	(0.333)	(0.119)	(720.2)	(0.00237)
Median household income in second quartile	-0.372	-0.182	784.6	-0.000813
	(0.261)	(0.134)	(1,041)	(0.00197)
Median household income in first third quartile	-0.550	-0.195	2,499	-0.00129
	(0.369)	(0.150)	(1,617)	(0.00261)
Median household income in fourth quartile	-0.814	-0.350*	4,329	-0.00356
	(0.522)	(0.199)	(3,897)	(0.00389)
Observations	316,846	316,841	305,970	316,374
R-squared		0.063	0.192	0.038
mean	2.126	4.716	36291	0.0200

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: All specifications include fixed effects for DXCCS diagnosis code group, state, year and month. Standard errors are clustered by state. Income quartiles are from the national income distribution.

2.8 Figures

Figure 2.1: Comparing the study sample to the US overall: Insurance Outcomes

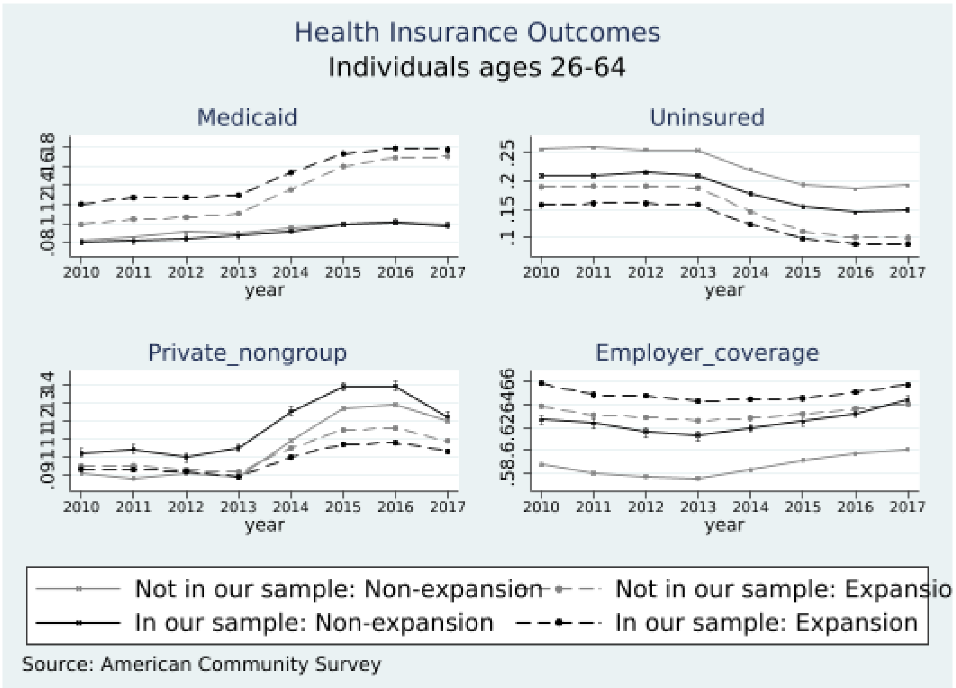


Figure 2.2: Comparing our sample to the US overall: Demographics

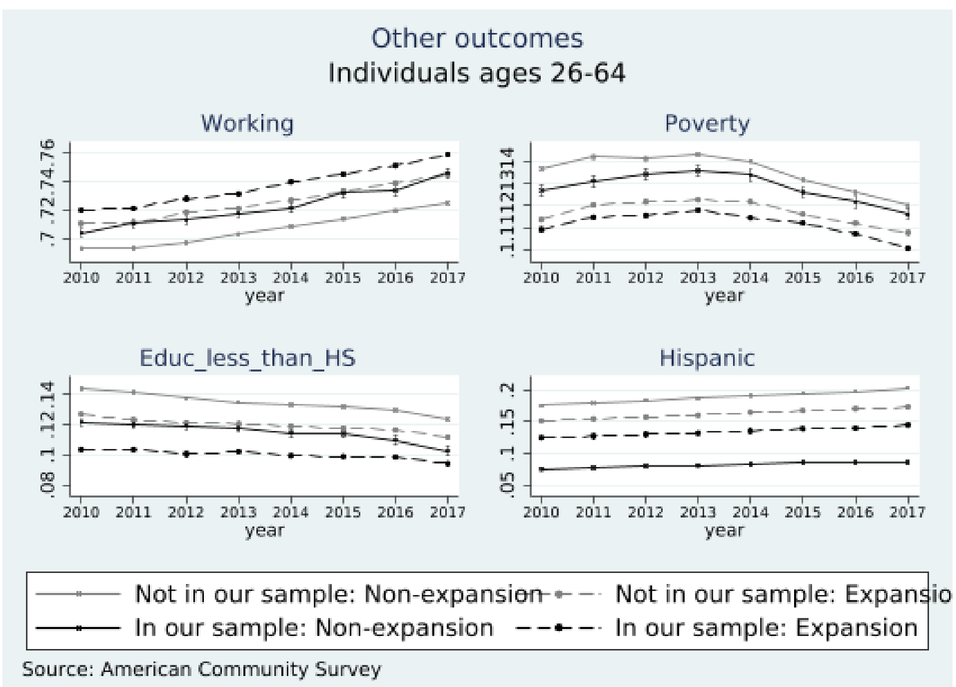


Figure 2.3: Density plot of the number of ICD-9 diagnosis codes by probability of weekend admission

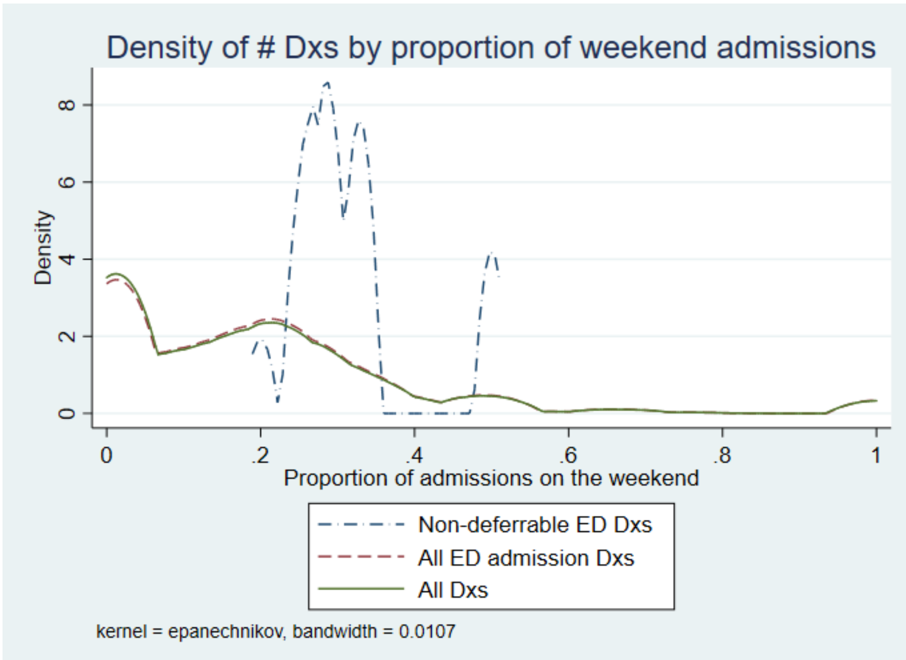


Figure 2.4: Trends in non-deferrable discharges in expansion and non-expansion states

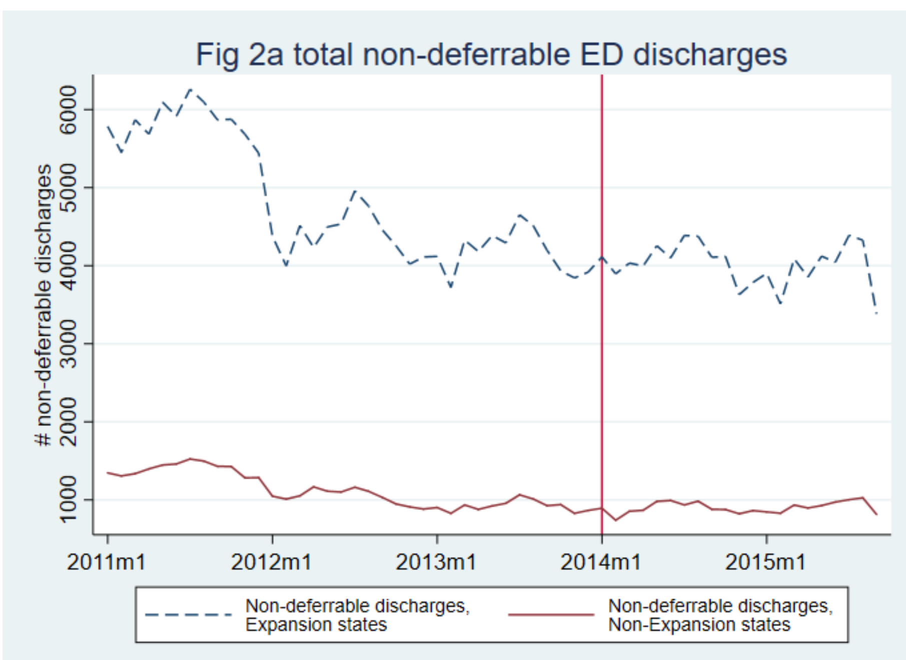


Figure 2.5: Trends in the fraction of discharges by state expansion status and payer: Medicaid vs uninsured

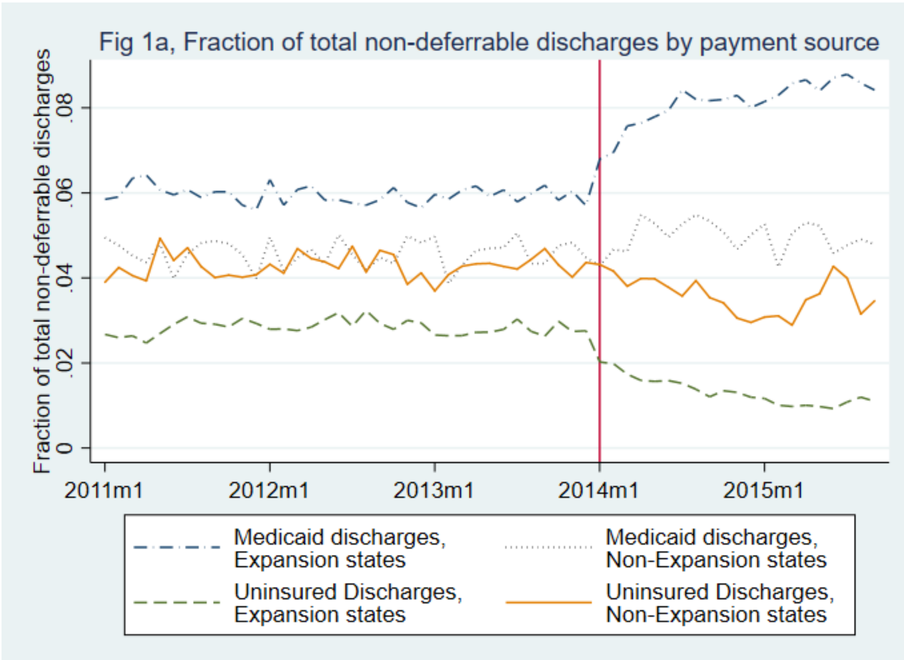


Figure 2.6: Trends in the fraction of discharges by state expansion status and payer: Medicare vs private insurance

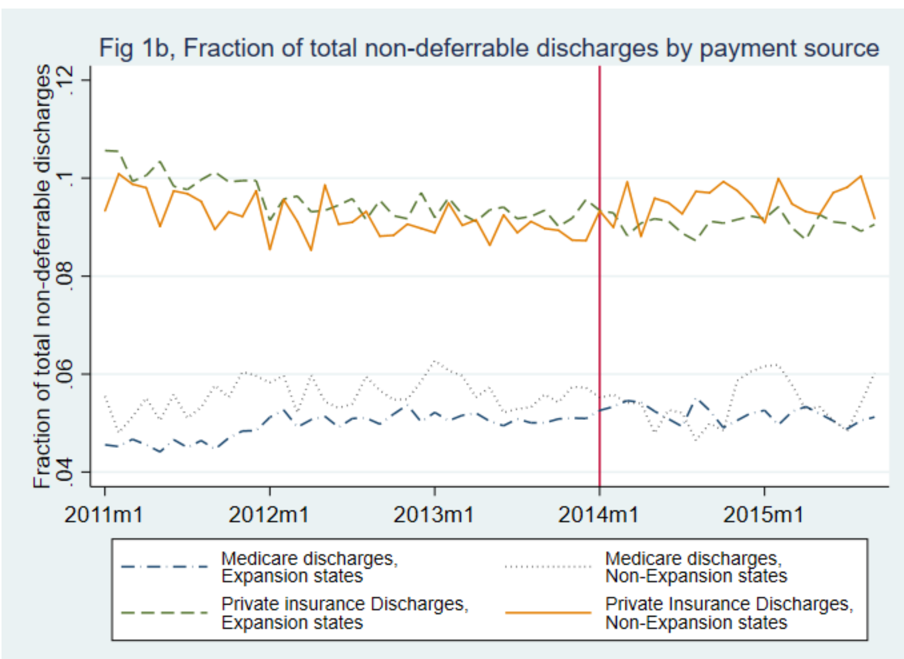


Figure 2.7: Effect of expansions on Medicaid coverage, event study

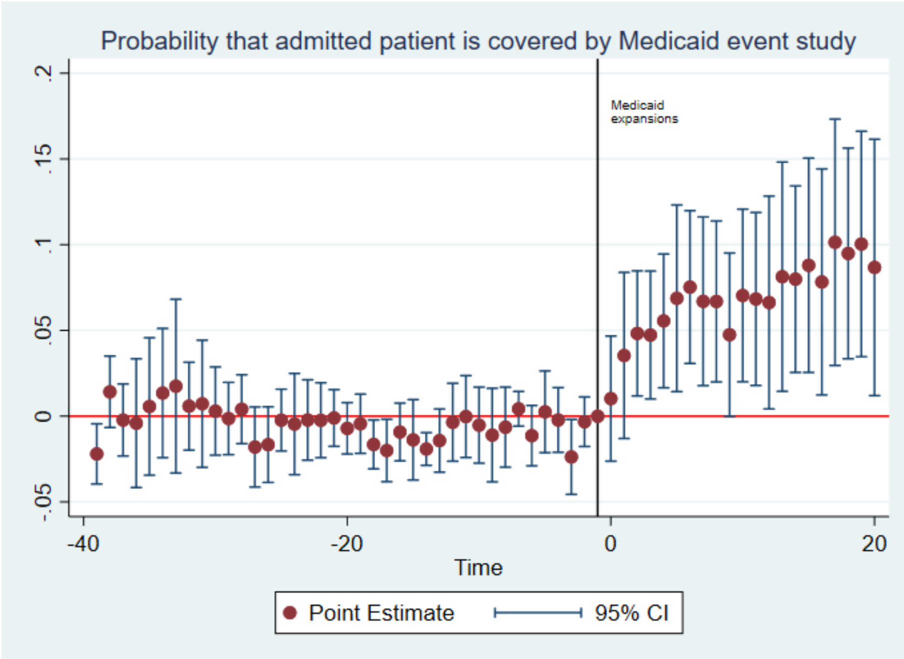


Figure 2.8: Effect of expansions on Medicare coverage, event study

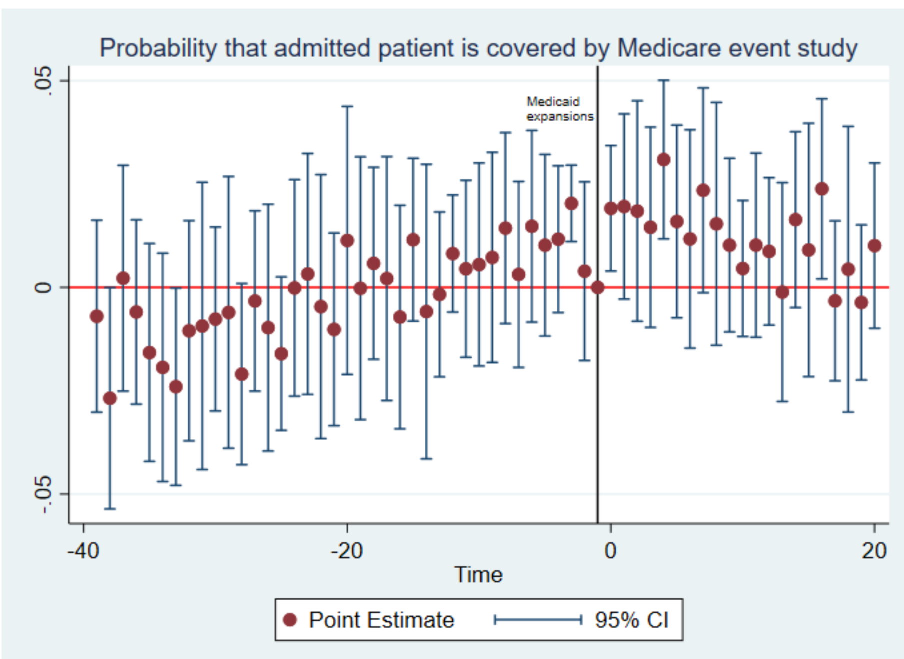


Figure 2.9: Effect of expansions on private insurance coverage, event study

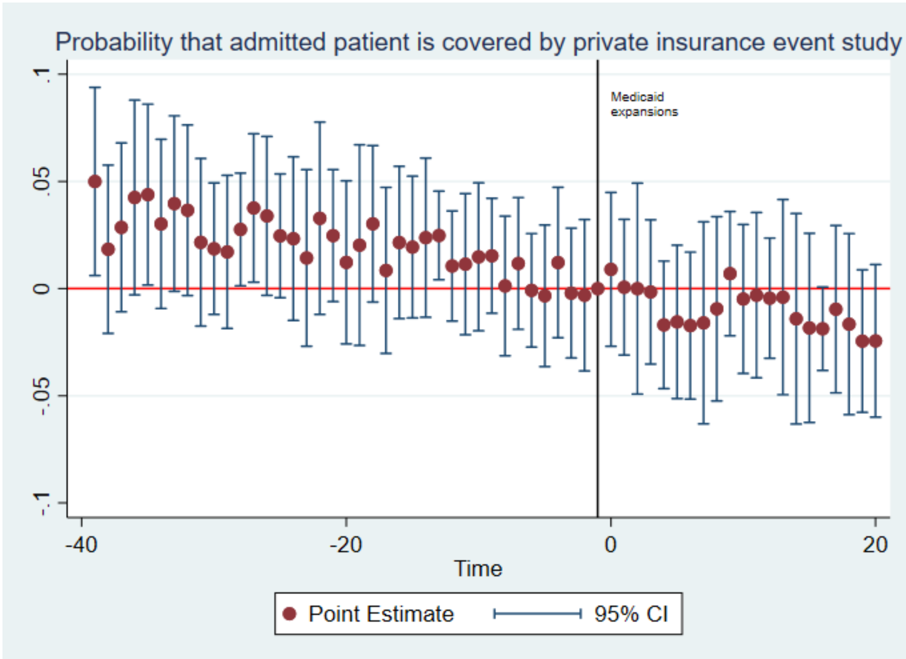


Figure 2.10: Effect of expansions on fraction uninsured, event study

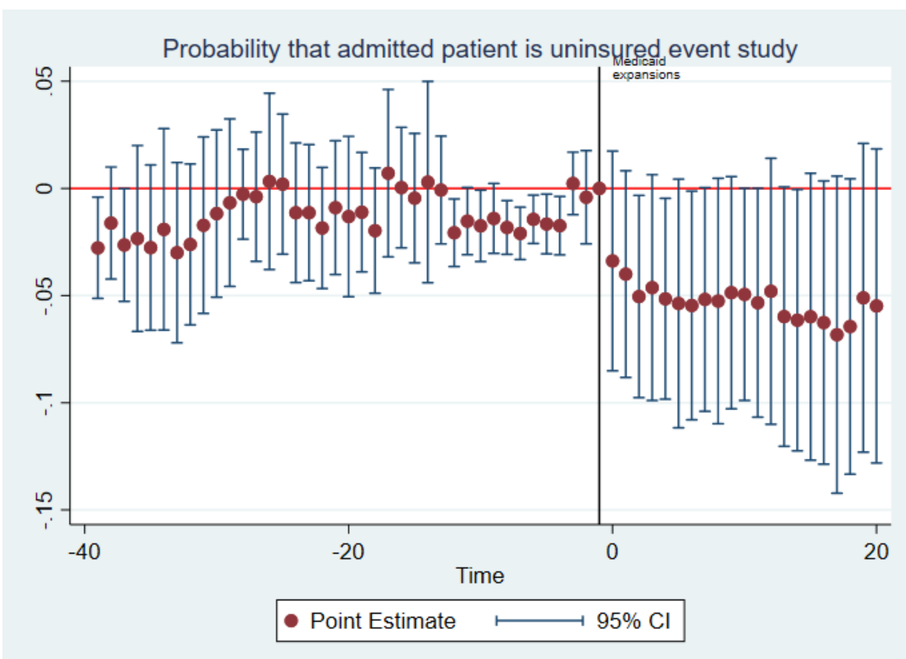


Figure 2.11: Effect of expansions on other coverage, event study

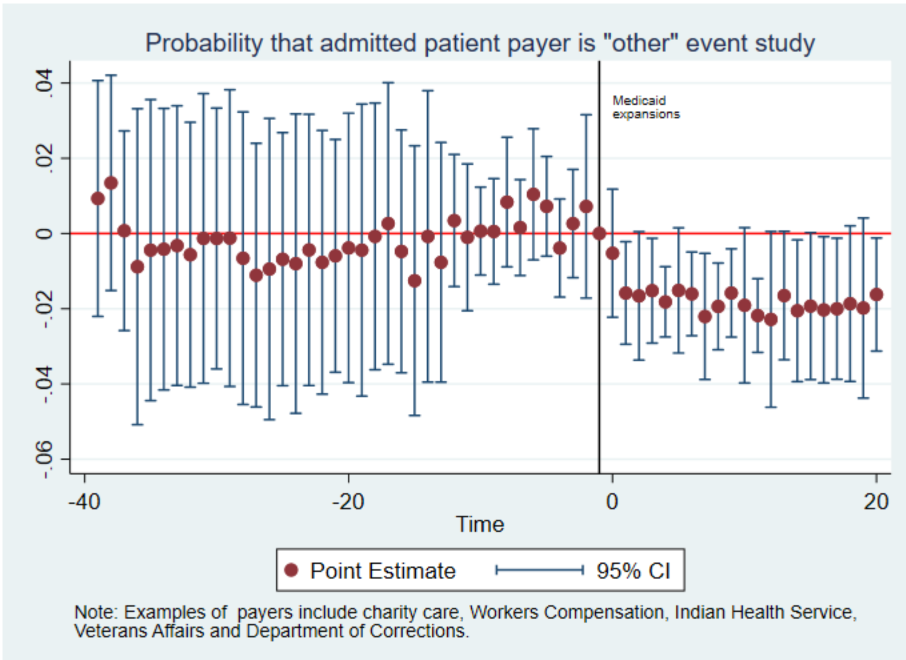


Figure 2.12: Effect of Medicaid expansion on length of stay, event study

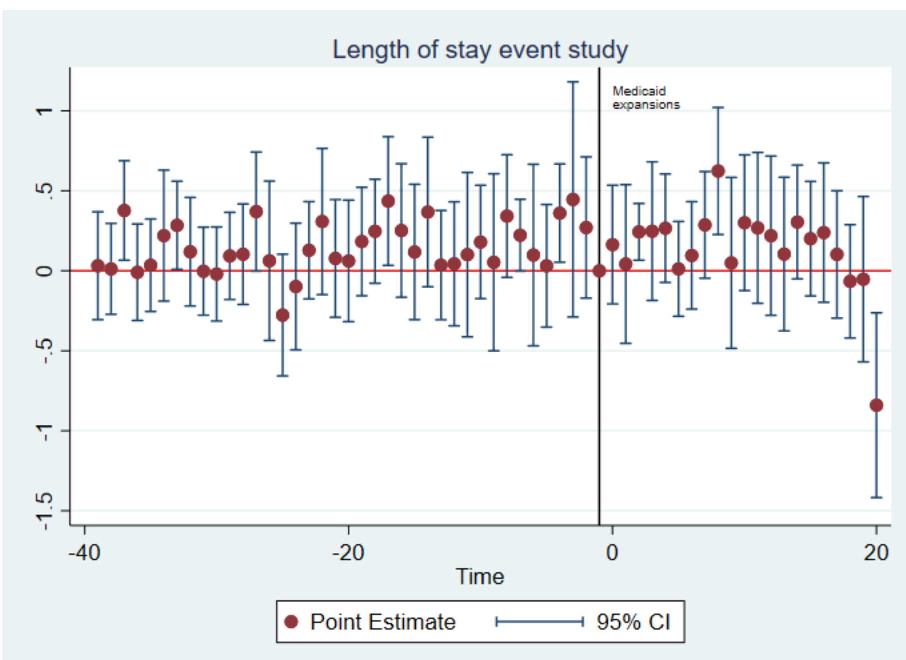


Figure 2.13: Effect of Medicaid expansion on number of procedures, event study

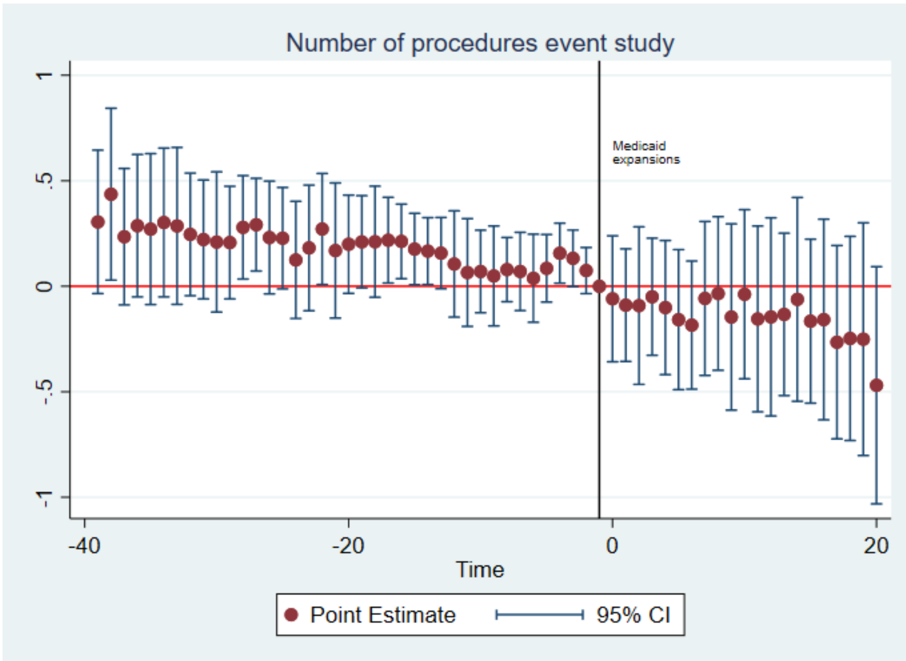


Figure 2.14: Effect of Medicaid expansion on total list charges, event study

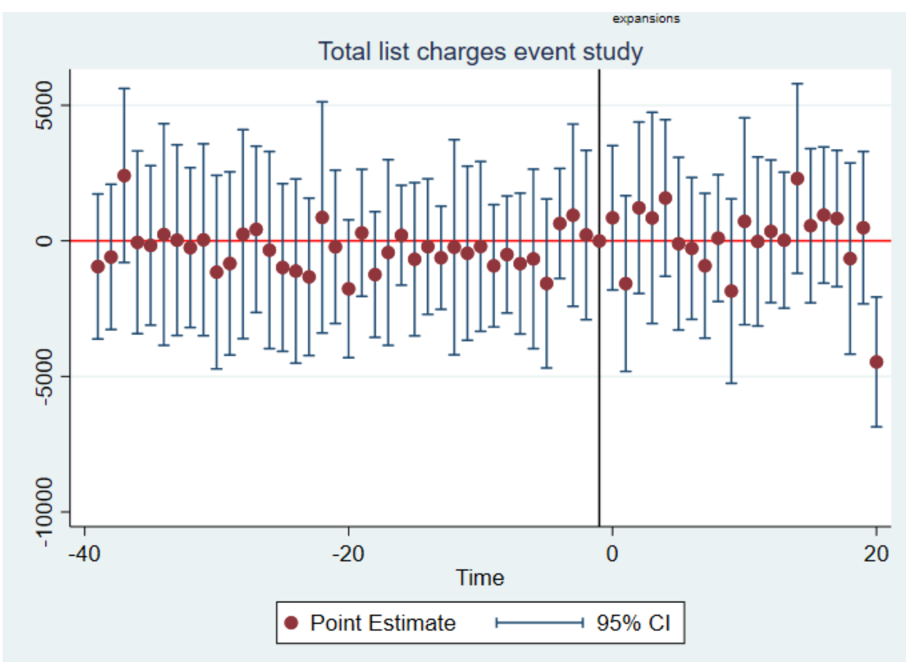
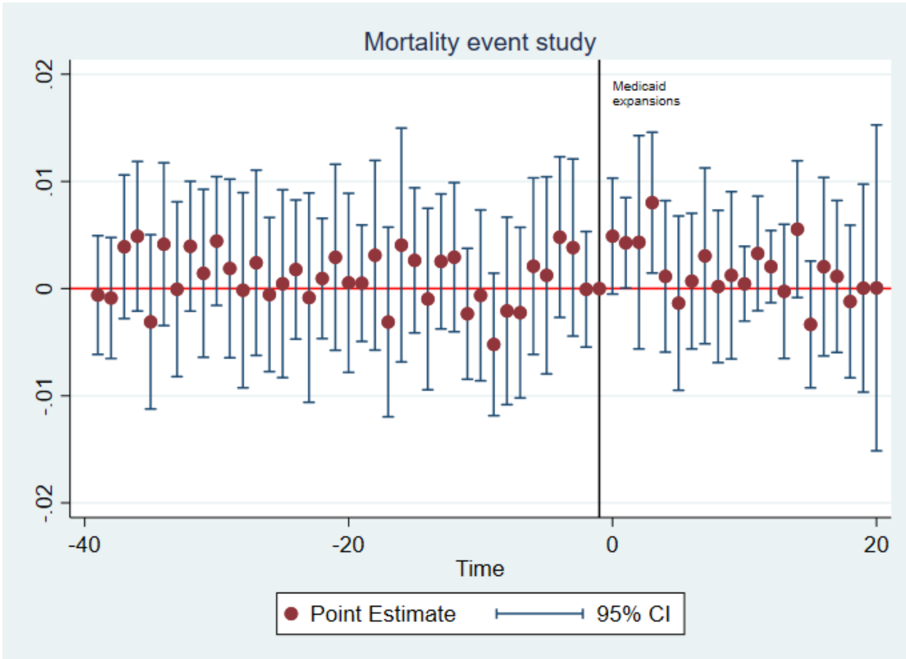


Figure 2.15: Effect of Medicaid expansion on mortality, event study



2.9 REFERENCES

Antonisse, Larisa, Rachel Garfield, Robin Rudowitz, and Samantha Artiga. 2018 (March). "The Effects of Medicaid Expansion under the ACA: Updated Findings from a Literature Review." Henry J Kaiser Family Foundation Issue Brief.

Arndt, Margarete, Robert C. Bradbury, Joseph H. Golec, and Paul M. Steen. 1998. "A comparison of hospital utilization by medicaid and privately insured patients." *Medical Care Research and Review* 55(1):32-53.

Baicker, Katherine, Sarah Taubman, Heidi Allen, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Eric Schneider, Bill Wright, Alan Zaslavsky, Amy Finkelstein, and the Oregon Health Study Group. 2013. "The Oregon Experiment - Effects of Medicaid on Clinical Outcomes." *New England Journal of Medicine* 368(18): 1713-1722.

Buchmueller, Thomas C., Kevin Grumbach, Richard Kronick, and James G. Kahn. 2005. "The effect of health insurance on medical care utilization and implications for insurance expansion: A review of the literature." *Medical Care Research and Review* 62(1): 3-30.

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. "Does Medicare save lives?" *The Quarterly Journal of Economics* 124(2): 597-636

Chattopadhyay, Arpita, and Andrew B. Bindman. 2005. "Accuracy of Medicaid Payer Coding in Hospital Patient Discharge Data: Implications for Medicaid Policy Evaluation." *Medical Care* 43(6) 586-591.

Cooper, Zack, Joseph J. Doyle Jr, John A. Graves, and Jonathan Gruber. 2022. "Do Higher-Priced Hospitals Deliver Higher-Quality Care?" No. w29809. National Bureau of Economic Research.

Coughlin, Teresa A., John Holahan, Kyle Caswell, Megan McGrath and The Urban Institute. 2014. "Uncompensated Care for Uninsured in 2013: A Detatiled Examination." Accessed July 2022. https://www.kff.org/uninsured/report/uncompensated-care-for-the-uninsured-in-2013-a-detailed-examination/

Currie, Janet and Jonathan Gruber. 2001. "Public health insurance and medical treatment: the equalizing impact of the Medicaid expansions." *Journal of Public Economics* 82: 63–89.

Cutler, David M and Jonathan Gruber. 1996. "Does Public Insurance Crowd out Private Insurance?" *The Quarterly Journal of Economics* 111(2), 391-430.

Dobkin, Carlos. 2003. "Hospital Staffing and Inpatient Mortality." Unpublished Working Paper. Accessed July 2022. https://people.ucsc.edu/c̃dobkin/Papers/Old%20Files/Staffing_and_Mortality.pdf

Doyle, Joseph J. Jr. 2005 “Health Insurance, Treatment and Outcomes: Using Auto Accidents as Health Shocks.” *The Review of Economics and Statistics* 87(2): 256–270.

Doyle, Joseph J. Jr. John A. Graves and Jonathan Gruber. 2017. “Uncovering waste in US healthcare Evidence from ambulance referral patterns.” *Journal of Health Economics* 54: 25-39.

Finkelstein, Amy, Nathaniel Hendren and Erzo F.P. Luttmer. 2019. “The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment.” *Journal of Political Economy* 127(6).

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group. 2012. “The Oregon Health Insurance Experiment: Evidence from the First Year.” *Quarterly Journal of Economics* 127(3): 1057-1106.

Freeman, Joseph D., Srikanth Kadiyala, Janice F. Bell and Diane P. Martin. 2008. “The Causal Effect of Health Insurance on Utilization and Outcomes in Adults: A Systematic Review of US Studies.” *Medical Care* 46(10):1023-1032.

Garthwaite, Craig, Tal Gross, and Matthew J. Notowidigdo. 2018. “Hospitals as Insurers of Last Resort.” *American Economic Journal: Applied Economics* 10(1): 1–39.

Guth, Madeline and Meghana Ammula. 2021. “Building on the Evidence Base: Studies on the Effects of Medicaid Expansion, February 2020 to March 2021.” Accessed July 2022. https://www.kff.org/medicaid/report/building-on-the-evidence-base-studies-on-the-effects-of-medicaid-expansion-february-2020-to-march-2021/

Guth, Madeline, Rachel Garfield, Robin Rudowitz and the Kaiser Family Foundation. 2020. “The Effects of Medicaid Expansion under the ACA: Updated Findings from a Literature Review.” Accessed July 2022. https://www.kff.org/medicaid/report/the-effects-of-medicaid-expansion-under-the-aca-updated-findings-from-a-literature-review/

Langa, Kenneth M., and Elliot J. Sussman. 1993. “The effect of cost-containment policies on rates of coronary revascularization in California.” *New England Journal of Medicine* 329(24): 1784-1789.

Lin, Saunders, Karen J. Brasel, Ougni Chakraborty, and Sherry A. Glied. 2020. “Association between Medicaid expansion and the use of outpatient general surgical care among US adults in multiple states.” *JAMA Surgery* 155(11):1058-1066.

Loehrer, Andrew P., David C. Chang, John W. Scott, Matthew M. Hutter, Virendra I. Patel, Jeffrey E. Lee, and Benjamin D. Sommers. 2018. “Association of the Affordable Care Act Medicaid expansion with access to and quality of care for surgical conditions.” *JAMA Surgery* 153(3):

e175568-e175568.

Lopez, Eric, Tricia Neuman, Gretchen Jacobson and Larry Levitt. 2020. "How Much More Than Medicare Do Private Insurers Pay? A Review of the Literature." Accessed July 2022. https://www.kff.org/medicare/issue-brief/how-much-more-than-medicare-do-private-insurers-pay-a-review-of-the-literature/

MACPAC. 2022. "Provider payment and delivery systems." Last accessed July 2022. https://www.macpac.gov/medicaid-101/provider-payment-and-delivery-systems/

Mazurenko, Olena, Casey P. Balio, Rajender Agarwal, Aaron E. Carroll, and Nir Menachemi. 2018. "The Effects Of Medicaid Expansion Under The ACA: A Systematic Review." *Health Affairs* 37(6): 944–950

McCormick, Danny, Jerry H. Gurwifz, Judith Savageau, Jorge Yarzebski, Joel M. Gore, and Robert J. Goldberg. 1999. "Differences in discharge medication after acute myocardial infarction in patients with HMO and fee-for-service medical insurance." *Journal of General Internal Medicine* 14(2): 73-81.

McDermott KW, Welch JA, Barrett ML, and Jiang HJ. 2018. "An Examination of CHIP and Medicaid Expected Payer Coding in HCUP Databases." Report # 2018-02. *HCUP Methods Series* Report # 2018-02 https://www.hcup-us.ahrq.gov/reports/methods/2018-02.pdf

MedPAC. 2021. "MEDPAC Databook: Beneficiary Demographics." Accessed July 2022. https://www.medpac.gov/wp-content/uploads/2021/10/July2021_MedPAC_DataBook_Sec2_SEC.pdf

Miller, Sarah, Norman Johnson, and Laura R. Wherry. 2021. "Medicaid and mortality: new evidence from linked survey and administrative data." *The Quarterly Journal of Economics* 136(3): 1783-1829.

Mulcahy, Andrew, Katherine Harris, Kenneth Finegold, Arthur Kellermann, Laurel Edelman, and Benjamin D. Sommers. 2013. "Insurance coverage of emergency care for young adults under health reform." *New England Journal of Medicine* 368(22): 2105-2112.

Nikpay, Sayeh, Thomas Buchmueller, Helen Levy and Simone R. Singh. 2016. "The Relationship between Uncompensated Care and Hospital Financial Position: Implications of the ACA Medicaid Expansion for Hospital Operating Margins." *Journal of Health Care Finance* 43(2).

Schmidt, Lucie, Lara D. Shore-Sheppard, and Tara Watson. 2020. "The impact of the ACA Medicaid expansion on disability program applications." *American Journal of Health Economics* 6(4): 444-476.

Sommers, Benjamin D., Atul A. Gawande, and Katherine Baicker. 2017. "Health insurance coverage and health - what the recent evidence tells us." *New England Journal of Medicine* 377(6): 586-593.

Taubman, Sarah, Heidi Allen, Bill Wright, Katherine Baicker, Amy Finkelstein, and the Oregon Health Study Group. 2014. "Medicaid Increases Emergency Department Use: Evidence from Oregon's Health Insurance Experiment." *Science* 343(6168): 263-268.

CHAPTER 3

Do Local Soda Taxes Affect Prices and Consumption? A Tale of Two Cities

Abstract

Many US cities have implemented or are considering soda taxes due in part to the growing literature about soda's negative health effects. As is the case with many public health interventions, measured effects of local soda taxes vary by city and also vary among studies that analyze the same city tax because estimates are sensitive to differences in methods and data sources. This study estimates the effects of local soda taxes in Berkeley and Philadelphia using the same methods and data in order to determine which measured differences can be attributed to local supplier and consumer responses rather than methodology. Comparing the cases of Berkeley and Philadelphia highlights which findings are constant across these very different cities and are likely to generalize to other local soda taxes. Berkeley and Philadelphia make an interesting comparative case study due to differences in motivation for the taxes (health vs revenue) and differences in demographics, city size and the size of the tax base. The results from both cities indicate incomplete pass-through rates that decline with beverage container size and moderate decreases in consumption of both diet and regular soda. Price and consumption effects were larger in Philadelphia than in Berkeley. Supply and demand elasticities can be separately identified using the tax as an instrument. [1]

3.1 Introduction

In 2015 Berkeley became the first US city to implement a soda excise tax. Since then many millions of dollars have been spent all over the country advocating for and against these taxes.

[1]Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

However despite the intense debate surrounding these measures and the many studies that analyze the impacts of local soda taxes, whether local soda taxes "work" and which effects are consistent across cities remains unclear in the public debate.

This discussion is complicated by differing policy objectives among soda tax proponents. Arguably the most frequently cited policy goal for soda taxes is to discourage soda consumption. Proponents see soda taxes as a modern successor to the cigarette taxes of the 1980's and 1990's, and hope that they will have similar effects. Similar assumptions about the effects of local soda taxes underlie the arguments of both health-based proponents and industry opponents. Both assume that prices will rise and that consumption will fall. Berkeley is an example of a soda tax that was primarily health-motivated. Other cities such as Philadelphia support a soda tax in order to help the local government raise revenue. Implicit in this motivation is the expectation that soda sales volume will remain "high enough." Still others believe that the purpose of soda taxes should be to correct negative externalities - a classic rationale for government intervention. Excess sugar consumption has been linked with increased risk of type 2 diabetes and other illnesses, which increase health insurance costs and decrease productivity. Liquid sugar consumption is thought to be particularly harmful. [2] A large change in consumption is not actually necessary in order for an externality-correcting tax to be deemed successful. Given the variety of objectives espoused by stakeholders it is hardly surprising that there is disagreement about whether soda taxes "work."

In order for soda taxes to cause a decrease in soda consumption, the cost of the tax must be passed on to soda consumers in the form of a price increase. [3] Even if regular soda consumption decreases however, demand for diet sodas could adjust to compensate if diet soda is not subject to the tax. This is an important margin of adjustment to investigate, because some researchers believe that the health consequences of artificial sweeteners are as bad or worse than natural sweeteners. See Yang (2010) for details. Consumers could also substitute towards beer, wine, or candy. Perhaps candy seems like a less natural substitute for soda, however there is a growing literature which suggests that sugar itself can be addictive. (See for example Avena et al (2008), who find that sugar is addictive in rats under some conditions, and support the theory that it may be for humans as well.) Sugars may be less detrimental to one's health in solid than in liquid form, but this is still an important factor to consider when evaluating the effectiveness of local soda taxes. Lastly, since these taxes only apply to products sold in a small geographic region, avoidance opportunities abound. If consumers choose to avoid the tax altogether by shopping in neighboring regions not

[2] Many propose that soda taxes should be set to not only correct for externalities, but also for internalities. See Allcott et al (2019) for a formal exposition of how the optimal commodity tax is a function of two terms: a term that represents the "corrective motive" (accounting for externalities and internalities) and a term that represents the "redistributive motive."

[3] Strictly speaking, soda consumption could decrease even if prices do not change, but that effect should arguably be attributed to the causal effect of the public awareness campaign surrounding the tax - not to the tax itself.

subject to the tax then local soda taxes will be less effective, whatever the objective of the local government.

Since 2015, many cities have continued to experiment with soda taxes: Philadelphia, San Francisco, Oakland (CA), Albany (CA), Boulder and Seattle followed in Berkeley's footsteps.[4] Industry lobbyists similarly continued their efforts and succeeded in having legislation passed in California and Michigan that prohibited new local excise taxes.

Understanding the effects of local excise taxes is important not only to inform the policy debate about soda taxes but also to improve our understanding of how these markets function. According to the classic theories of tax incidence with perfect competition, the pass-through rate of a soda tax depends on the relative elasticities of supply and demand for soda. If demand is relatively more inelastic, then a larger share of costs associated with the tax are passed on to consumers (and vice versa). In most cases pass-through will be incomplete; however predicting the pass-through rate of a tax ex-ante is difficult, and without accurate estimates of the supply and demand elasticities for soda in a given market, the pass-through rate could be anywhere between zero and one. In reality the sugar-sweetened beverages market is probably more accurately described as an environment with imperfect competition, in which case the theoretical prediction is even more ambiguous: the tax may be over-shifted to consumers (implying a pass-through rate greater than one). In addition to generating evidence to inform the policy debate, this paper shows how product-level variation in the soda tax rate can be leveraged to separately identify the theoretically important parameters of the elasticity of supply and the elasticity of demand - without needing to employ methods from structural estimation.

In order to investigate which market effects of soda taxes may be generalizable to other localities that are considering these proposals, this study compares the cases of Berkeley and Philadelphia. The tax bases differ between the two cities as well as the motivation for the tax. Berkeley's 1 cent per oz tax applies to regular but not diet soda. Philadelphia's tax was 1.5 cents per oz, and both regular and diet soda beverages are taxed. Berkeley and Philadelphia are also interesting to compare because the motivations for the taxes were different. In Berkeley proponents of the tax emphasized its potential health benefits, whereas in Philadelphia the primary objective of the tax was to raise revenue. While Philadelphia implemented a soda tax much later, it is a much larger city with different demographics and serves as an interesting comparative case study to determine which effects are consistent across both cities and which others can be attributed to the unique local public health context.

This study uses difference-in-differences to identify the effects of each of these taxes relative to a chosen control group city. The Nielsen scanner data used in this study has information on weekly prices and quantity sold for thousands of UPCs across many different product categories. Using the

[4]Cook County, IL passed a soda tax that was quickly repealed.

same data source and the same empirical strategies across both cities enables this study to compare the effects between the two cities without the confounding effects of different data sources and methods. In doing so, this study follows Rojas and Wang (2021) and Cawley, Frisvold and Jones (2020) who employ a similar approach in order to compare the effects of soda taxes across multiple jurisdictions. Rojas and Wang (2021) compare tax effects in Berkeley to those in Washington State using difference in differences and the same data source - Nielsen scanner data. They find full pass-through of the tax and a 4-6% decrease in soda consumption in Washington State, contrasting with less than 30% pass-through and no statistically significant change in consumption in Berkeley. Cawley, Frisvold and Jones (2020) focus on the effects of soda taxes on soda purchases in the largest cities that have implemented soda taxes: Philadelphia (PA), San Francisco (CA), Seattle (WA) and Oakland (CA). They find that across these cities, a 1 cent per oz increase in the tax rate reduces household consumption by 12%. Notably, this effect was mainly due to sharp decline in consumption in Philadelphia: no consumption effects were detected in the three other cities.

Recent literature has highlighted the importance of using consistent data sources and methods in order to compare effects across cities. In their analysis of the Philadelphia tax, Seiler, Tuchman and Yao (2021) engage in a detailed analysis of how different data and methods lead to differences between their estimates and those of Roberto et al. (2019), who also analyze the Philadelphia tax but reach different conclusions about its effectiveness. Seiler Tuchman and Yao (2021) estimate price effects that are 10-22 percentage points higher than Roberto et al (2019), and cross-border shopping effects that are twice as large. Effects of the Berkeley soda tax also differ across studies, likely reflecting a combination of differences in methods and data sources. Falbe et al (2015) surveys stores in Berkeley and estimates pass-through rates of 69% for soda and 47% for all sugar sweetened beverages. In contrast, Bollinger and Sexton (2017) estimate pass-through rates of less than 30%. Falbe (2016) estimates that consumption decreased by over 20%, whereas Bollinger and Sexton (2017) find minimal or no effects on consumption.

In addition to the gains from using uniform data and methods to compare two cities, this study contributes to the literature on soda taxes by examining differential pass-through by beverage size. An early study by Colchero et al (2015) finds higher pass-through for small beverage sizes and highlights this as an important dimension to consider since it increases relative incentives for consumers to purchase large beverage sizes and consume more, counter to the health motivation of many soda taxes. In Berkeley Cawley, Frisvold and Jones (2017) find that pass-through declines with beverage size. However Seiler, Tuchman and Yao (2021) do not find variation in pass-through by beverage size in Philadelphia. It remains unclear whether some characteristics of Philadelphia result in a different pass-through by beverage size relationship, or if this differences is a product of methodological differences.

Lastly, this is the first study to investigate potential substitution effects towards beer and wine after the Berkeley soda tax. Gibson et al (2021) finds no evidence of increased alcohol consumption after the Philadelphia tax, however Powell and Leider (2022) find increased consumption of beer after the Seattle soda tax. To my knowledge no paper has yet examined substitution to other alcoholic beverages after the soda tax in Berkeley.

3.2 Previous research on soda taxes

Several studies investigate the effects of national soda taxes. Grogger (2017) finds that after Mexico implemented a one peso per liter soda tax in 2014 (amounting to about 9% of the average price of soda), prices rose by about 12%. Berardi, et al (2016) use a differences and differences approach and find that within 6 months, the French soda tax was fully passed on to consumers. As for consumption of taxed soda, Colchero et al (2016) find a decrease of 6% after the Mexican tax was implemented. Another creative study on the impact of soda taxes on consumption is Fletcher, Frisvold and Tefft (2010), who identify these effects in the US by exploiting state-level variation in special excises taxes and whether soda qualifies for exemptions to the sales tax as food. They find that increased taxes result in modest reductions in soda consumption among youth. These studies provide evidence of the effects of macro-level soda taxes, however one might suspect that the effects of local soda taxes might differ due to differences in general equilibrium responses and avoidance responses. It is generally easier to avoid local taxes, and if re-optimization is costly for producers, they may choose to internalize some or all of the tax costs if they are sufficiently low. (See for example DellaVigna and Gentzkow (2019).)

A growing literature examines the effects of local soda taxes on prices and consumption. Most papers that examine price outcomes hand-collect price data from a set of retailers within and outside the tax jurisdiction, and those that examine consumption outcomes typically survey residents about their habits in person or by phone. (See for example Cawley et al (2017a), (2017b), (2019) and (2020a); Falbe et al (2015) and (2016), and some results from Silver et al (2017)) Hand-collected price data is by definition public, so researchers can release findings on heterogeneous responses by brand and conduct other interesting analyses such as examining whether pass-through increases as the distance from the store to the city limits increases (see Cawley et al (2017a) and (2020)). One advantage of survey-based consumption measures is that the researcher can differentiate between adult and child responses, which have different policy implications (see Cawley et al (2019)). If income data is collected, the researcher can also speak to the regressivity of the tax.

This study uses Nielsen scanner data to examine price and consumption responses. High frequency scanner data provides the opportunity to flexibly control for time trends and to use synthetic control methods in case the difference-in-differences parallel trends assumption is in doubt. Because hand-collecting price data is costly in time, money and effort, these studies typically measure prices twice (before and after the tax) and in a limited number of cities, which precludes this kind of analysis. Although scanner data does not provide an opportunity to analyze changes in individual-level consumption patterns, sales estimates from scanner data are less likely to be subject to response bias. Survey respondents might report less soda consumption after the tax because of increased awareness of its negative health effects and a reluctance to admit to engaging in a behavior that is newly-perceived as undesirable. Reliance on diverse data sources allows the literature to incorporate the strengths of each, thereby improving the quality of evidence about soda taxes.

A few recent studies use scanner data in their primary analysis. Bollinger and Sexton (2017) investigate the impact of the Berkeley soda tax on prices and consumption using Nielsen scanner data, focusing on heterogeneous effects by store type (drugstores vs supermarkets). They find no pass-through in drugstores and limited pass-through (about 20%) in a Berkeley supermarket. As for consumption effects, they find no effect on soda purchases made in drug stores and a decrease of about 7-12% for soda purchases made at a Berkeley grocery store. Seiler et al (2021) focus on the Philadelphia tax, estimating a pass-through rate of 97% and substantial consumption responses. In what must have a been a heroic effort, they also hand-collect nutrition information in order to estimate the effect of the tax on calorie and sugar intake. Rojas and Wang (2021) provide the first analysis of the price and consumption effects of the Washington state carbonated beverages tax, comparing it to the effect of the Berkeley SSB tax.

This study builds on the methodologies of these papers in also using scanner data. Like Rojas and Wang (2021), this paper adopts a comparative approach. Comparing the experiences of Berkeley and Philadelphia provides some preliminary evidence about which effects of local soda taxes may be idiosyncratic and which ones likely generalize across multiple localities. In addition, this paper shows how the tax may be used as an instrument to separately identify supply and demand elasticities by extending the method developed in Zoutman et al (2018). This will be discussed further in section 3.6.

3.3 Background and Data

Berkeley was the first U.S. city to pass and implement a soda excise tax. The tax (levied on soda distributors) was passed in November 2014 and was implemented in March, 2015. The wording of the ballot proposal was as follows:

"Shall an ordinance imposing a 1 cent per ounce general tax on the distribution of high-calorie, sugary drinks (e.g., sodas, energy drinks, presweetened teas) and sweeteners used to sweeten such drinks, but exempting: (1) sweeteners (e.g., sugar, honey, syrups) typically used by consumers and distributed to grocery stores; (2) drinks and sweeteners distributed to very small retailers; (3) diet drinks, milk products, 100% juice, baby formula, alcohol, or drinks taken for medical reasons, be adopted?"

During implementation of the tax, the "very small retailers" mentioned in the above proposal were defined as those with annual gross receipts less than \$100,000. Since the tax is levied on distributors and not retailers, this means that distributors do not have to pay tax on the SSBs that they sell to these very small retailers. Thus one should expect no pass-through in these small stores unless they decide to raise prices as a best response to the price increases of other larger retailers who are subject to the tax. Revenue calculations for stores in the Berkeley scanner data showed no stores that met the above definition of a very small retailer.

The Philadelphia soda tax was approved by the city council in June 2016 and implemented in January 2017. At a rate of 1.5 cents per oz, Philadelphia's tax is .5 cents per oz higher than Berkeley's and Philadelphia also taxes beverages with artificial sweeteners, such as diet soda. On other dimensions, Philadelphia's tax is similar to Berkeley's: milk, alcoholic beverages, and 100% fruit juice is exempt. As in Berkeley, the tax is imposed on distributors. At the time of this writing, the full text of the Philadelphia ordinance was available in the "Publications and Forms" section of the Philadelphia City Government website.

This analysis uses Nielsen retail scanner data from the Kilts Center for Marketing. The data set covers sales in over 35,000 stores and spans the years 2006 - 2017. Over 2.6 million UPC codes, or "product codes" are included, representing products ranging from paper towels to deli meat. The data are comprised of weekly sales numbers (prices and quantities) for these products. For example, one UPC code might correspond to a 12-pack of 12 oz Dr. Pepper cans. Importantly, one can observe both the quantity of the UPC product that was sold in a given week, as well as how the product is bundled - a 12 pack of 12 oz cans, a single 12 oz can, a 2 liter bottle, etc.

In order to investigate the impact of the Berkeley and Philadelphia soda taxes, it is necessary to determine which stores are in Berkeley and Philadelphia. However, the finest geographic label in the data is the first three digits of a store's zip code. Zip code boundaries do not necessarily correspond to city limits in general, and it is often the case that regions both within and outside city limits share the same first three digits of their zip codes. Luckily, this is not the case for Berkeley. All of the zip codes within Berkeley begin with 947 and all of the zip codes that begin with 947 are contained within Berkeley, with a few exceptions. Zip code 94706 is just outside the Berkeley city limits. Some other maps also show areas of 94707, 94708, 94720, 94704, and 94705 as being outside of Berkeley. In these maps, zip code 94707 contains a small area outside of Berkeley but this area appears to be mainly residential. For the zip codes 94708, 94720, 94704 and 94705, the areas in question are mainly comprised of the Tilden Nature Area, the Claremont Canyon Regional Preserve, and a small area within the University of California Berkeley. Therefore, although it seems that several zip codes cross the Berkeley city limits, most of the areas in these zip codes that are outside Berkeley are not likely to have many stores. The main zip code that could be problematic is 94706.

It is possible that one or more of the stores labeled here as "treated" are located just outside the Berkeley city limits and are therefore not subject to the tax, but given the small area of zip code 94706 and the geographic characteristics of the other areas, this is not likely to apply to many stores and is not likely to significantly impact the results. If some of the 947 stores are in fact located outside Berkeley, it would lead to an underestimate of the effects of the tax. Therefore this analysis includes all 9 stores in the 947– zip code. Fortunately, the 941– zip code area corresponds closely to San Francisco city limits as well, and San Francisco is used as a comparison group for some specifications.

Fortunately, in Philadelphia and Pittsburgh (used as a control) the zipcodes beginning with 191 and 152 overlap well with the zipcodes in Philadelphia and Pittsburgh (respectively).

In all specifications the analysis of post-tax conditions is restricted to a 3-month period: March-May 2015 for the Berkeley tax and January - March 2017 for the Philadelphia tax. In determining the appropriate interval there is a tradeoff between the desire to know long-run impacts of the tax and the plausibility of the parallel trends assumption. The longer the post-treatment window, the more likely it is that the estimation results would pick up other confounding factors.

For the results involving diet soda, regular soda and water, the analysis is also restricted to the most popular brands in order to improve the external validity of the results. In the scanner data there are some brands that are not well known and do not sell very frequently. Some of these may be regional brands only available in California or Pennsylvania and the process that determines prices for these brands may be very different depending on what niche of consumers constitutes their target market. In addition to limiting external validity, including these products would increase the variance of the estimates without adding useful information. The procedure to determine the most popular brands was as follows: first, sales revenue and volume sold (in oz) was calculated for each unique brand-flavor combination in the data (for example, Cherry Coke, Vanilla Pepsi, etc). The brand-flavor combinations were ranked by market share according to these two definitions, and the top ranked brands were identified. This process was repeated separately for Berkeley and Philadelphia. The top ranked brands were very similar, regardless of whether the volume-based or revenue based metric was used. For both Berkeley and Philadelphia, the top-selling brand-flavor combinations that are included in subsequent analyses account for about 90% of revenue and of volume sold (oz).

3.4 Estimation

Difference-in-differences (DiD) serves as the primary identification strategy to analyze the effects of the tax in both Berkeley and Philadelphia. It is convenient to compare the effects of these two taxes using the same identification strategy and given that most of the literature relies on DiD, this choice also facilitates comparison with other estimates in the literature.

Following Cawley and Frisvold (2017a), San Francisco is used as the difference in differences control group to investigate the effects of the Berkeley soda tax. Because Berkeley is so unique - both culturally and politically - one concern is that the unobservable trends and factors that led Berkeley to consider a soda tax in the first place might cause it to diverge from the control group in the aftermath of the tax. Using San Francisco as a control group should alleviate these concerns, because San Francisco and Berkeley are often thought to be similar: “two peas in a very liberal pod - both of them far-left, wacky cities filled with pot-smoking, quinoa-eating eccentrics whose mayoral candidates go by names like Chicken John (San Francisco, 2007) and Running Wolf (Berkeley, 2012),” according to an article in the SFGate (SFGate, Nov 7, 2014). (The article goes on to analyze what demographic differences or differences in strategic advocacy could explain the different electoral outcomes.)

In addition, San Francisco is an appealing control group because it voted on a similar soda tax proposal at the same time as Berkeley’s that was narrowly defeated. In San Francisco about 56% of votes were in favor and about 44% against; however the proposal (called “Local Measure E”) would have needed over 66 and 2/3 % in favor in order to pass because the tax revenue was earmarked instead of being allocated to the general budget. (See San Francisco Board of Elections (2014) for election results.)

It is less obvious which city would serve as a suitable control group for Philadelphia. This paper chooses Pittsburgh - the second largest city in Pennsylvania. Pittsburgh and Philadelphia have similar poverty rates (22 and 26%, respectively) and similar median incomes (about $44,000 and $41,000 respectively). A similar proportion of the population of each city graduated from high school, however Pittsburgh is a more educated city: about 42% of the population has a Bachelors degree or higher, where as in Philadelphia this number is 27%. The percentage of the population that is female, as well as the percentage over 65 years old are approximately equal; however the racial composition of the cities is different. Pittsburgh is about 67% white, 24% Black or African American, 6% Asian and about 3% Hispanic. Philadelphia is about 42% white, 43% Black or African American, 7% Asian and about 14% Hispanic. (See Census QuickFacts for Pittsburgh and Philadelphia.) Synthetic control methods are used as a robustness check in case the parallel trends assumption does not hold using Pittsburgh as a control.

For both Berkeley and Philadelphia, the specification for soda price and consumption difference-in-differences regressions is as follows:

(1) $pricepoz_{ijt} = \alpha_0 + \eta_j + \gamma_t + \psi_k + \theta BeverageSize_i + \beta T_{jt} + \epsilon_{ijt}$

The unit of analysis is product i purchased in store j in month t. A "product" is defined as a (regular or diet) x brand name x flavor x size category combination. $pricepoz_{ijt}$ is the sales-weighted price per ounce (in cents) of product i purchased in store j in month t, and in the main specifications $T_{jt} = 1$ for products sold in stores located in the taxed city after the SSB tax was implemented. The main specifications compare regular (diet) soda in the taxed city to regular (diet) soda in the untaxed city. Although diet soda was not taxed in Berkeley, it is not used as a comparison group because consumption and prices of diet soda may still be affected by the tax - for example, if regular and diet soda are substitutes. η_j are store fixed effects and γ_t are month and year fixed effects. $BeverageSize_i$ is the total volume in ounces of UPC i (12 oz for a 12 oz can and 144oz for a 12-pack of 12oz cans, for example). Quadratic and cubic beverage size terms are also included in some specifications. All price and soda consumption regressions include hand-coded brand fixed effects ψ_k, where a brand is defined as a brand name x flavor combination (for example Cherry Pepsi, Vanilla Pepsi, cola flavored Pepsi, etc). Given this specification, 100β is the average pass-through rate of the tax in Berkeley and $\frac{100\beta}{1.5}$ is the average pass-through rate of the tax in Philadelphia. In order to determine how and whether pass-through varies with beverage size, other specifications include treatment by beverage size interactions of various degrees.

To estimate the impact of the tax on consumption using difference in differences, a similar specification is used with a dependent variable of $lnconspercap_{kjt}$, or log consumption per capita of brand k from store j in month t. For the consumption regressions, the unit of analysis is brand k sold in store j at time t. (These specifications do not include beverage size covariates.) For the above regressions, standard errors are clustered at the store level.

3.5 Results

Table 1 shows difference-in-difference estimates of the effect of the Berkeley soda tax on prices-per-oz. With no interaction terms, the average pass-through rate of the tax (6%) is very low (column 1). (Recall that in this specification the pass-through rate is 100 times the treatment coefficient.) Although this pass-through rate is statistically indistinguishable from zero, it is not a dramatic departure from Cawley and Frisvold's (2017a) estimates of well below 50%. While pass-through rates are low on average, this seems to hide substantial heterogeneity by beverage size. In columns 2, 3, and 4 beverage size interaction terms are added incrementally and using the estimates from column 2, pass-through varies from 24% for 12oz sizes to 0% for 66oz sizes. Although the Berkeley tax only applied to regular sugar sweetened beverages, the tax was also passed on to consumers of diet beverages, with implied pass-through rates of 40% and 2% for 12oz and 66oz containers, respectively. Table 2 shows these effects in DiD specifications that compare diet soda prices in Berkeley to diet soda prices in San Francisco, before and after the Berkeley tax was implemented.

As seen in Table 3, the tax caused consumption per capita of regular soda to decrease by 7.6% for a given brand in a taxed store (preferred specification in column 2). Recall that some cities that passed soda taxes after Berkeley (including Philadelphia) decided to include diet soda in the tax base, however in Berkeley diet soda is exempt. In the wake of the tax, consumers may have seen diet soda as a cheaper and perhaps less socially stigmatized substitute for regular soda, so ex ante, one might expect that consumption of diet soda would increase. However, consistent with the diet soda price increases in Berkeley as well as the small impact of the tax on regular soda consumption, the results show no statistically significant or substantively significant impact of the tax on diet soda consumption (Table 3 column 5).

Philadelphia taxed both regular and diet sodas at a rate of 1.5 cents per ounce - .5 cents per ounce higher than Berkeleys tax rate. Philadelphias pass-through rates also decline with beverage size but are noticeably higher than Berkeleys: 77% for a 12oz container and 40% for a 66oz container for regular soda, and 81% and 53% for diet soda. (See Tables 4 and 5, preferred specifications in column 3.) Consumption of regular and diet soda for a given brand in Philadelphia fell by 28% and 33%, respectively. (Table 6) This larger consumption response is consistent with higher pass-through rates.

The remaining results explore other potential behavioral response margins: changes in consumption of related products and avoidance behaviors. This study fails to find evidence that consumers avoided the Berkeley tax by purchasing soda in neighboring regions. (Refer to DiD specifications in Table 7, where $T = 1$ in neighboring regions after the Berkeley tax was implemented and $T = 0$ in San Francisco, or before the tax was implemented.) These results should be interpreted with caution, however, because the neighboring zip code area (946–) is somewhat large and may include some stores that are too far from Berkeley to be a reasonable shopping alternative.

Although these estimates serve as a lower bound for the avoidance response, this null finding is not surprising given that the effect of the Berkeley tax on soda consumption was small. In addition, results from Tables 8 and 9 indicate that stores in neighboring regions increased their soda prices. Under the assumption that stores that are farther away from Berkeley are less likely to respond to Berkeley's tax, these estimates should be interpreted as a lower bound for the price responsiveness of neighboring stores to Berkeley's tax, which further supports the conclusion that cross-border shopping after the tax was minimal.

Stores in regions neighboring Philadelphia responded differently to the tax: they lowered prices. Regular soda prices in neighboring regions decreased by 27 cents per oz on average (Table 10 column 1) and diet soda prices decreased by 16 cents per oz on average (Table 11 column 1).

Consequently, regular and diet soda purchases per capita increased in these regions, by about 20% and 13%, respectively (Table 12).

If people are affected by soda taxes, they may substitute towards alternative consumption goods. Although diet soda or water might seem like more natural substitutes (when untaxed), consumers could also purchase increased quantities of beer or wine. These products are typically still more expensive than soda; however the soda tax would lower their opportunity cost. Given the small effect of the Berkeley tax on regular soda consumption, it is not surprising that this study finds no evidence of a consumption response for bottled water, wine or candy (Tables 13, 14, and 15, respectively). Table 16 does show that beer consumption decreased in Berkeley by 14%. After the Philadelphia tax, consumption of bottled water increased for both flavored and unflavored varieties (Table 17). Candy and beer consumption also increased (Tables 18 and 19) but both by less than 10%, and consumption of wine did not change (Table 20).

Although the parallel trends assumption necessary for DiD cannot be tested, event studies have been used to assess the plausibility of this assumption. If the trends are not parallel before the treatment intervention, it seems unlikely that they would have been parallel after the intervention, had the treated units not been treated. One advantage of using Nielsen Scanner Data is that there are enough pre-period observations to construct these event study graphs and additional robustness checks may be pursued if the trends appear to be questionable. Figures 1 and 2 show event study graphs for Berkeley - price per oz and log consumption per capita; and Figures 3 and 4 show event study graphs for these outcomes in Philadelphia. Neither of the Berkeley event studies suggests a blatant violation of the parallel trends assumption, although there seem to be some short-term anticipation effects in the consumption event study. The Philadelphia graphs are more questionable: although the price per oz and consumption event studies show clear and significant differences as a result of the tax, many pre-trend coefficients in both of the event studies are statistically different from zero.

This study addresses concerns that the parallel trends assumptions may not hold for the chosen control cities by using the Robbins, Saunders and Kilmer (2017) extension of Abadie et al's synthetic control method to construct a data-driven control group. In Abadie et al's synthetic control approach,[5] a weighted average of untreated units is chosen to minimize the distance to the treated unit along several chosen dimensions. This weighted group of untreated units serves as the "synthetic control group." Since the time path of the dependent variable in the synthetic control group is usually a very close match to the treatment group, the (untestable) parallel trends assumption is more credible. In addition, Abadie et al's method leaves less discretion to the researcher, allowing for a more objective and transparent understanding of how the control group was selected.

[5]See Abadie, Diamond and Hainmueller (2015); Abadie Diamond and Hainmueller (2010); and Abadie and Gardeazable (2003)

The Abadie et al synthetic control method has been used in many regional policy analyses. However, using their method in this context would have required aggregation of sales. The Robbins, Saunders and Kilmer (2017) method extends the Adabie et al synthetic control method so that multiple treated units may be used without sacrificing the richness of the micro-level data. Rather than aggregating the data to the city level, one can construct a control group using UPC level or store level data. By incorporating the additional information from micro-level data, the procedure constructed by Robbins, Saunders and Kilmer (referred to as RSK henceforth) results in synthetic control groups that more closely match the treated group - sometimes even exact matches.

Figures 5, 6, 7, and 8 present the RSK synthetic control results, as a robustness check for the DiD specifications. Synthetic control groups were constructed from candidate products sold in other California (Pennsylvania) cities, excluding regions that neighbor Berkeley (Philadelphia). The price per oz RSK graphs show a pass-through rate that is averaged across multiple size categories. Power was insufficient to construct synthetic control groups for each beverage size and analyze pass-through separately by beverage size. However the pass-through results appear to be consistent with what is suggested by the DiD specifications. The RSK consumption graphs also present treatment effects that are consistent with the results of the DiD specifications. In the right panel of these RSK graphs, the grey lines represent the results of 250 placebo tests. In each iteration, a placebo unit is chosen from within the untreated pool. Next, a synthetic control group is constructed for this placebo unit. Lastly, one computes the difference between the placebo unit and its synthetic control group. This process is then repeated 250 times. If the red "treatment" line appears outside or on the border of the mass of grey lines, then the treatment effect is statistically significant. Most treatment effects indicated by the RSK synthetic control procedure are similar to the DiD results, both qualitatively and in terms of magnitudes.

3.6 Estimating supply and demand elasticities

Two methods may be used to estimate the supply and demand elasticities for soda. In the first method DiD estimates of the pass-through rate, the percent change in prices, and the percent change in soda purchased are used in back-of-the-envelope calculations to yield supply and demand elasticities:

(2) $\eta^D = \frac{dQ^*/q}{dP^*/P^*}$ (Equation 1)

(3) $\eta^S = \frac{-\eta^D \rho}{1-\rho}$ (Equation 2)

where p^* is the equilibrium price, Q^* is the equilibrium quantity, and ρ is the pass-through rate of the tax, or dP^*/dt.[6] dP^*/P^* comes from a DiD regression with log prices as the dependent variable, controlling for beverage size and beverage size squared. ρ is the average pass-through rate, shown in Column 1 of Table 1 for Berkeley. These calculations yield demand and supply elasticities for regular soda in Berkeley of about -3 and .13. For Philadelphia, the demand elasticity for soda is -1.15 and the supply elasticity is .1.

Another method proposed by Zoutman et al (2018) uses an instrumental variables strategy to separately identify the supply and demand elasticities. Zoutman et al (2018) show how to estimate these elasticities for three cases: 1) an ad valorem tax levied on the demand side, 2) an ad valorem tax levied on the supply side, and 3) a specific tax levied on the demand side. This paper extends their method to the case of a specific tax levied on the supply side.

Suppose that there is an ad valorem tax τ_{it} that is levied on suppliers. Consider the following general equations for supply and demand:

(4) $y_{it} = \epsilon^S p_{it} + \eta z_{it} + \omega^S x_{it} + v_{it}^S$

(5) $y_{it} = \epsilon^D p_{it} + \gamma z_{it} + \omega^D x_{it} + v_{it}^D$

For now, let z_{it} be a general function of the tax rate. Lowercase y_{it} and p_{it} are logged prices and quantities, so $y_{it} = ln(Y_{it})$ and $p_{it} = ln(P_{it})$. Let the prices in these supply and demand equations represent the (log) observed list price, so $P_{it}^{-\tau} = (1-\tau_{it})P_{it}$ and $p_{it}^{-\tau} = ln(1-\tau_{it}) + p_{it}$. $P_{it}^{-\tau}$ or $p_{it}^{-\tau}$ are referred to as the net-of-tax price.

Since the tax is imposed on suppliers, assume that z_{it} does not directly enter the demand equation.

Standard exclusion restriction: If the tax τ_{it} is levied on the supply side, then $\gamma = 0$.

Given that prices are endogenous, $z_{it} = ln(1-\tau_{it})$ is used as an instrument for p_{it} in order to measure ϵ^D. Of course in order for the tax to be a valid instrument, the tax must only affect quantity demanded through its effect on the list price. This yields $\epsilon^D = \beta_{y,z}/\beta_{p,z}$, where $\beta_{y,z}$ is the coefficient on z_{it} when y_{it} is regressed on z_{it}, and $\beta_{p,z}$ is the coefficient on z_{it} when p_{it} is regressed on z_{it}.

To identify the supply elasticity, note that the tax enters the supply equation in a very specific

[6]Equation 2 comes from the standard tax incidence formula: note that $D(p^*) = S(p^* - t)$. Differentiating both sides with respect to t, and then solving for dP^*/dt yields the standard tax incidence formula, and Equation 2 results from solving instead for η^S.

way, given standard models of taxation. Because profit maximizing suppliers care about the net-of-tax price $p_{it}^{-\tau}$, assume that the tax only affects quantity supplied through the net-of-tax price $p_{it}^{-\tau}$. A \$1 decrease in the list price (all else equal) will have the same impact on suppliers as a \$1 decrease in $1-\tau$, all else equal.[7]

The supply equation is therefore re-written as:

(6) $y_{it} = \epsilon^S p_{it}^{-\tau} + \omega^S x_{it} + v_{it}^S$

(7) $= \epsilon^S p_{it} + \epsilon^S ln(1-\tau) + \omega^S x_{it} + v_{it}^S$

Therefore $z_{it} = ln(1-\tau_{it})$ can be used as an instrument for $p_{it}^{-\tau}$ in order to estimate ϵ^S.

Ramsey Exclusion Restriction: Supply depends on on the net-of-tax price. It follows that $z_{it} = ln(1-\tau_{it})$ and $\eta = \epsilon^S$.

Given this instrumental variables approach, $\epsilon^S = \beta_{y,z}/\beta_{p^{-\tau},z}$. Now consider a more general case, where $P_{it}^{-\tau} = (1-\theta_{it})P_{it}$. Since the Berkeley and Philadelphia soda taxes are specific taxes, $\theta_{it} = \tau_{it}/P_{it}$, where τ_{it} is the tax (in cents per oz) of product i. Because θ_{it} is a function of prices, it is endogenous and therefore $ln(1-\theta_{it})$ cannot serve as an instrument. We follow Zoutman et al (2018) in overcoming this difficulty by using lagged prices and period t tax rates to construct $ln(1-s_{i,t})$, which acts as a synthetic instrument for $ln(1-\theta_{it})$. $z_{it} = ln(1 \hat{-} \theta_{it})$ is used as an instrument to estimate the supply and demand elasticities.

This method provides alternate estimates of Berkeley's and Philadelphia's soda demand and supply elasticities that can be compared with the back-of-the-envelope ones presented above. The identification strategies underlying the back-of-the-envelope calculations above and the Zoutman et al method are different: the former relies on difference-in-differences assumptions and the later replies on instrumental variables. If there is significant tax avoidance in the form of cross-border shopping the IV assumptions will be violated and the elasticity estimates may be biased. As discussed above, there was no significant cross-border shopping in Berkeley, and the Zoutman et al IV methods yield demand and supply elasticity estimates of -1.4 (95% CI: [-1.91,-.91]) and .549 (95% CI: [.28,.82]), respectively. Given the amount of cross-border shopping in Philadelphia, it may not be surprising that the Philadelphia estimates appear suspect: -.93 (95% CI: [-1.04,.83]) for the elasticity of demand and -.3 (95% CI: [-.39,-.24]) for the elasticity of supply.[8] Examining the

[7] When the tax is levied on the demand side, this intuition is somewhat easier to understand. In that case the restriction is that the tax affects quantity demanded only through the list price. As Zoutman et al explain, the consumer's budget constraint is affected in the same way whether the list price increases due to an increase in the tax or due to an increase in the pre-tax price.

[8] These 95% confidence intervals are bootstrapped. The F statistics on all of the first stages are large, at least 80.

back-of-the-envelope elasticity calculations suggests that Berkeley and Philadelphia have similar supply elasticities. Differences in demand elasticities seem to be primarily driving the differences in pass-through rates and consumption between the two cities.

3.7 Conclusion

A growing number of cities in the US have implemented or are considering soda taxes. Much of the advocacy surrounding these proposals focuses on the health benefits of soda taxes - the potential for these taxes to lower soda consumption and reduce obesity and type 2 diabetes. Numerous studies examine the effects of these taxes on prices and consumption; however the diversity of data sources and methods among these studies makes it difficult to identify generalizable conclusions about the impacts of these measures and to understand why effects may differ across cities. This study compares the cases of Berkeley and Philadelphia using the same data source and difference in differences methods to estimate the effects of these taxes in two very different contexts. Berkeley and Philadelphia are interesting cases to compare because to the extent that tax effects differ between the two cities, Berkeley serves as an example for smaller progressive cities that may consider implementing health-focused excise taxes, whereas Philadelphias experience is more applicable to larger cities that view these taxes primarily as a source of revenue for city projects that are not explicitly health-related.

The results demonstrate that pass-through in both cities was incomplete. In both cities pass-through rates were higher for small product sizes and decreased with beverage size. This has important implications for the health consequences for such taxes: existing non-linear pricing already encourages consumers to purchase larger quantities of soda and soda taxes appear to steepen these incentives. Pass-through rates were higher in Philadelphia than in Berkeley, and in both cities pass-through was similar for both diet and regular soda, regardless of the taxed status of diet soda. Retailers might have believed that cross-border shopping was much more likely in Berkeley due to its small size and therefore might have been reluctant to raise prices and risk losing customers and sales of other goods that shoppers would have purchased in the same shopping trip. Approximately equal pass-through for diet and regular soda in both cities could reflect companies' long-term business or marketing strategy, similar to the non-linear pricing discussed above. In addition, Berkeley retailers may have decided to increase diet soda prices because it was too costly for them to identify every product's taxed status based on whether it contained added sugar; it may have been easier to spread tax costs across all products. Institutional details of the implementation process support this theory, and may have also contributed to the generally lower pass-through rates in Berkeley. Berkeley was the first city in the United States to implement such a tax, and initially existing city staff had to manage the implementation process until a program officer was hired through a grant to focus on the soda tax program full-time. (Falbe et al, 2020) With limited resources, the city focused their outreach efforts on distributors, who bear the statutory burden of the tax. (Falbe et al, 2020) Although distributors remit the tax, it is important for retailers to understand the mechanics of the tax in order for them to respond optimally. In a qualitative study of the implementation process, some retailers expressed confusion about whether diet soda and fruit drinks were subject to the tax. (Falbe et al, 2020). Rather than increasing transparency, invoices from distributors did not show how much tax was levied per beverage item, but instead displayed one aggregated SSB tax line item per order. Given the uncertainty in which beverages were subject to the tax, a simple heuristic that retailers could have followed would have been to spread the additional upstream beverage costs among a large base including both regular and diet soda in order to lessen the impact on each individual beverage.

In contrast, the Philadelphia government engaged in more active outreach with retailers through mailings, one-on-one interactions, and a website that explained which beverages are subject to the tax and included discussions of various sweeteners and subcases that might have led to confusion. (Holdbrook et al, 2019) According to Holdbrook et al (2019), many retailers also reported receiving formula sheets from their distributors that clearly indicated which products were taxed and by how much. Some retailers in the Holdbrook et al study believed that they were required to raise prices by the amount of the tax - perhaps a result of the detailed explanations and formula sheets provided to them. For stores in corporate groups, the tax was often already incorporated in automatic price updates so there was less room for managerial discretion and decision-making. Although some retailers in Philadelphia still reported that they received insufficient guidance, retailers in Philadelphia had more information and engagement than those in Berkeley.

It is interesting to note that using the estimates derived from back-of-the-envelope calculations, Berkeley and Philadelphia's supply elasticities are almost identical. This may not be surprising since the sample was restricted to the most popular brands in each city. Manufacturers of these brands most likely optimize with respect to a national market, and may have contracts with distributors and retailers that are relatively homogeneous across the US. However it is not a priori obvious that Berkeley and Philadelphia would have the same supply elasticities; these are very different cities and the soda supply chain has multiple levels, some of which may be affected by local conditions. Under the assumptions of perfect competition, the variation in the pass-through rate of the tax comes from differences in the demand elasticities, rather than through differences in both elasticities or differences in supply elasticities alone. In line with standard theories of taxation, demand is more elastic in Berkeley (both relative to supply and relative to demand in Philadelphia) so the pass-through rate is lower than in Philadelphia. In Philadelphia demand appears more elastic than supply which would imply a low pass-through rate, however these elasticity estimates fail to account for cross-border shopping, which was observed in Philadelphia. The finding that the supply elasticities are approximately equal suggests that in practice it may be sufficient for policymakers to rely on demand elasticity estimates to predict soda tax pass-through rates, at least for the most popular brands that represent over 90% of sales.

The results also suggest that if the tax is large enough, increases in consumption of other unhealthy foods such as candy and beer is a concern, however the effects are likely to be small in magnitude relative to the decline in soda consumption. If health advocates wish to incentivize consumers to switch to (untaxed) diet soda then it is important to provide retailers with detailed information about which products are subject to the tax in order to encourage them to adjust downstream prices accordingly - otherwise the increased tax costs are likely to be passed on to consumers of both regular and diet soda.

This study shows that store responses in neighboring regions are difficult to predict and not generalizable. Stores near Berkeley increased soda prices after the tax, while stores near Philadelphia decreased prices. It is possible that popular support for the Berkeley tax was viewed by stores in neighboring regions as a signal of higher willingness to pay, so they saw an opportunity to raise prices without losing sales. Given that the Philadelphia tax was less popular, neighboring stores may have anticipated that lowering prices would attract more soda sales and potentially spill over into increased revenue in other food categories. It is also possible that these effects come from differences at the distributor level. Since Berkeley is a much smaller geographic unit than Philadelphia, it may not be worthwhile for distributors to set different prices for Berkeley vs neighboring regions. Since Philadelphia is a much larger market, distributors may set Philadelphia-specific prices, which would have given retail stores in neighboring regions more flexibility to strategically adjust prices. Unfortunately this study was not able to find data on distributor prices, but this would be an interesting question for future research.

Differences in the effects of the Berkeley and Philadelphia soda taxes are likely due to a combination of factors, including the geographic and population size of the localities, administrative capacity of the local government and the motivations for enacting the taxes, related to the cultural and political environment of the cities. All public health is local, however comparing Berkeley and Philadelphias experiences with soda taxes and the factors that led to differences can help other cities considering these taxes prepare for what to expect.

3.8 Tables

Table 3.1: Effect of the Berkeley SSB tax on soda price per oz (cents)

VARIABLES	(1) Price per oz (cents)	(2) Price per oz (cents)	(3) Price per oz (cents)	(4) Price per oz (cents)
Treatment	0.0641	0.289**	0.518**	0.923*
	(0.0902)	(0.122)	(0.236)	(0.467)
Treatment*Beverage volume (oz)		-0.00440**	-0.0148*	-0.0422*
		(0.00193)	(0.00884)	(0.0243)
Treatment*Beverage volume (oz) squared			5.89e-05	0.000373*
			(4.13e-05)	(0.000223)
Treatment*Beverage volume (oz) cubed				-8.72e-07*
				(5.21e-07)
Beverage volume (oz)	-0.102***	-0.102***	-0.102***	-0.101***
	(0.0118)	(0.0118)	(0.0118)	(0.0119)
Beverage volume (oz) squared	0.000567***	0.000567***	0.000566***	0.000564***
	(0.000112)	(0.000112)	(0.000112)	(0.000113)
Beverage volume (oz) cubed	-9.65e-07***	-9.65e-07***	-9.65e-07***	-9.58e-07***
	(2.68e-07)	(2.68e-07)	(2.68e-07)	(2.71e-07)
Observations	10,246,074	10,246,074	10,246,074	10,246,074
R-squared	0.529	0.530	0.530	0.530

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and brand fixed effects.

Table 3.2: Effect of the Berkeley SSB tax on diet soda price per oz (cents)

VARIABLES	(1) Price per oz (cents)	(2) Price per oz (cents)	(3) Price per oz (cents)	(4) Price per oz (cents)
Treatment	-0.0362	0.484***	1.167***	2.343***
	(0.0652)	(0.114)	(0.191)	(0.415)
Treatment*Beverage volume (oz)		-0.00704***	-0.0310***	-0.0925***
		(0.000715)	(0.00385)	(0.0139)
Treatment*Beverage volume (oz) squared			0.000131***	0.000787***
			(2.18e-05)	(0.000133)
Treatment*Beverage volume (oz) cubed				-1.84e-06***
				(3.42e-07)
Beverage volume (oz)	-0.109***	-0.109***	-0.109***	-0.108***
	(0.0127)	(0.0127)	(0.0127)	(0.0128)
Beverage volume (oz) squared	0.000647***	0.000648***	0.000646***	0.000641***
	(0.000115)	(0.000115)	(0.000115)	(0.000116)
Beverage volume (oz) cubed	-1.19e-06***	-1.19e-06***	-1.19e-06***	-1.17e-06***
	(2.79e-07)	(2.78e-07)	(2.78e-07)	(2.81e-07)
Observations	6,331,079	6,331,079	6,331,079	6,331,079
R-squared	0.525	0.525	0.525	0.525

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
All specifications include store, month, year and brand fixed effects.

Table 3.3: Effect of the Berkeley SSB tax on consumption (oz) of regular and diet soda

VARIABLES	(1) Volume regular soda (oz) per capita	(2) log Volume regular soda (oz) per capita	(3) Volume regular soda (oz)	(4) Volume diet soda (oz) per capita	(5) log Volume diet soda (oz) per capita	(6) Volume diet soda (oz)
Treatment	-0.00535	-0.0790**	-362.7	-0.00394	-0.0534	302.3
	(0.00407)	(0.0335)	(599.8)	(0.00243)	(0.0538)	(419.8)
Observations	33,611	31,687	33,611	29,261	26,526	29,261
R-squared	0.453	0.928	0.428	0.367	0.908	0.443

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
All specifications include store, month, year and brand fixed effects.

Table 3.4: Effect of the Philadelphia SSB tax on regular soda price per oz (cents)

VARIABLES	(1) Price per oz (cents)	(2) Price per oz (cents)	(3) Price per oz (cents)	(4) Price per oz (cents)
Treatment	0.856***	0.792***	1.326***	1.493***
	(0.0843)	(0.0989)	(0.122)	(0.152)
Treatment*Beverage volume (oz)		0.00116	-0.0152***	-0.0228***
		(0.00101)	(0.00333)	(0.00579)
Treatment*Beverage volume (oz) squared			6.28e-05***	0.000130**
			(2.08e-05)	(5.15e-05)
Treatment*Beverage volume (oz) cubed				-1.30e-07
				(1.04e-07)
Beverage volume (oz)	-0.164***	-0.164***	-0.164***	-0.163***
	(0.0107)	(0.0107)	(0.0107)	(0.0108)
Beverage volume (oz) squared	0.000951***	0.000952***	0.000953***	0.000951***
	(0.000108)	(0.000108)	(0.000109)	(0.000110)
Beverage volume (oz) cubed	-1.52e-06***	-1.53e-06***	-1.53e-06***	-1.53e-06***
	(2.72e-07)	(2.72e-07)	(2.75e-07)	(2.77e-07)
Observations	19,869,841	19,869,841	19,869,841	19,869,841
R-squared	0.640	0.640	0.641	0.641

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and brand fixed effects.

Table 3.5: Effect of the Philadelphia SSB tax on diet soda price per oz (cents)

VARIABLES	(1) Price per oz (cents)	(2) Price per oz (cents)
Treatment	1.008***	0.751***
	(0.122)	(0.146)
Treatment*Beverage volume (oz)		0.00454***
		(0.00113)
Beverage volume (oz)	-0.161***	-0.161***
	(0.0116)	(0.0116)
Beverage volume (oz) squared	0.000967***	0.000969***
	(0.000113)	(0.000113)
Beverage volume (oz) cubed	-1.58e-06***	-1.59e-06***
	(2.90e-07)	(2.90e-07)
Observations	8,819,854	8,819,854
R-squared	0.682	0.682

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and brand fixed effects.

Table 3.6: Effect of the Philadelphia SSB tax on consumption (oz) of regular and diet soda

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Volume regular soda (oz) per capita	log Volume regular soda (oz) per capita	Volume regular soda (oz)	Volume diet soda (oz) per capita	log Volume diet soda (oz) per capita	Volume diet soda (oz)
Treatment	0.00124	-0.331***	-2,273*	0.00107**	-0.405***	-657.7
	(0.000964)	(0.0345)	(1,206)	(0.000485)	(0.0456)	(575.7)
Observations	101,466	96,518	101,466	79,785	64,259	79,785
R-squared	0.461	0.855	0.470	0.391	0.869	0.408

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and brand fixed effects.

Table 3.7: Effect of the Berkeley SSB tax on consumption (oz) of regular and diet soda in neighboring areas

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Volume regular soda (oz) per capita	log Volume regular soda (oz) per capita	Volume regular soda (oz)	Volume diet soda (oz) per capita	log Volume diet soda (oz) per capita	Volume diet soda (oz)
Treatment	0.0162	-0.0584*	6,842	0.00150	-0.0647	1,117
	(0.0129)	(0.0310)	(5,292)	(0.00324)	(0.0406)	(1,345)
Observations	38,918	36,442	38,918	34,307	30,536	34,307
R-squared	0.421	0.923	0.432	0.444	0.898	0.450

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and brand fixed effects.

Table 3.8: Effect of the Berkeley SSB tax on soda price per oz (cents) in neighboring areas

	(1)	(2)	(3)	(4)
VARIABLES	Price per oz (cents)	Price per oz (cents)	Price per oz (cents)	Price per oz (cents)
Treatment	-0.0654	0.0325	0.289**	0.731***
	(0.0798)	(0.0946)	(0.123)	(0.262)
Treatment*Beverage volume (oz)		-0.00143*	-0.00967***	-0.0361***
		(0.000860)	(0.00350)	(0.0125)
Treatment*Beverage volume (oz) squared			3.59e-05***	0.000322***
			(1.25e-05)	(0.000113)
Treatment*Beverage volume (oz) cubed				-7.32e-07***
				(2.61e-07)
Beverage volume (oz)	-0.110***	-0.110***	-0.110***	-0.109***
	(0.00828)	(0.00829)	(0.00830)	(0.00857)
Beverage volume (oz) squared	0.000645***	0.000645***	0.000644***	0.000632***
	(7.69e-05)	(7.69e-05)	(7.67e-05)	(7.98e-05)
Beverage volume (oz) cubed	-1.15e-06***	-1.14e-06***	-1.15e-06***	-1.12e-06***
	(1.82e-07)	(1.82e-07)	(1.81e-07)	(1.89e-07)
Observations	12,874,132	12,874,132	12,874,132	12,874,132
R-squared	0.571	0.571	0.571	0.572

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and brand fixed effects.

Table 3.9: Effect of the Berkeley SSB tax on diet soda price per oz (cents) in neighboring areas

VARIABLES	(1) Price per oz (cents)	(2) Price per oz (cents)	(3) Price per oz (cents)	(4) Price per oz (cents)
Treatment	0.0835	0.522***	1.170***	2.398***
	(0.0590)	(0.118)	(0.181)	(0.355)
Treatment*Beverage volume (oz)		-0.00548***	-0.0248***	-0.0850***
		(0.00106)	(0.00328)	(0.0120)
Treatment*Beverage volume (oz) squared			9.48e-05***	0.000711***
			(1.28e-05)	(0.000107)
Treatment*Beverage volume (oz) cubed				-1.63e-06***
				(2.68e-07)
Beverage volume (oz)	-0.115***	-0.115***	-0.115***	-0.114***
	(0.0102)	(0.0102)	(0.0103)	(0.0105)
Beverage volume (oz) squared	0.000710***	0.000710***	0.000709***	0.000694***
	(9.14e-05)	(9.12e-05)	(9.11e-05)	(9.38e-05)
Beverage volume (oz) cubed	-1.34e-06***	-1.34e-06***	-1.34e-06***	-1.30e-06***
	(2.20e-07)	(2.19e-07)	(2.18e-07)	(2.25e-07)
Observations	7,175,411	7,175,411	7,175,411	7,175,411
R-squared	0.543	0.543	0.544	0.544

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and brand fixed effects.

Table 3.10: Effect of the Philadelphia SSB tax on soda price per oz (cents) in neighboring areas

	(1)	(2)	(3)	(4)
VARIABLES	Price per oz (cents)	Price per oz (cents)	Price per oz (cents)	Price per oz (cents)
Treatment	-0.269***	-0.467***	-0.0755	-0.688***
	(0.0363)	(0.0817)	(0.136)	(0.128)
Treatment*Beverage volume (oz)		0.00234***	-0.00661**	0.0164***
		(0.000683)	(0.00319)	(0.00463)
Treatment*Beverage volume (oz) squared			3.33e-05**	-0.000150***
			(1.50e-05)	(4.48e-05)
Treatment*Beverage volume (oz) cubed				3.45e-07***
				(1.20e-07)
Beverage volume (oz)	-0.158***	-0.158***	-0.158***	-0.160***
	(0.0119)	(0.0118)	(0.0119)	(0.0122)
Beverage volume (oz) squared	0.000942***	0.000943***	0.000945***	0.000957***
	(0.000121)	(0.000121)	(0.000122)	(0.000126)
Beverage volume (oz) cubed	-1.57e-06***	-1.57e-06***	-1.58e-06***	-1.60e-06***
	(3.10e-07)	(3.10e-07)	(3.13e-07)	(3.22e-07)
Observations	21,079,606	21,079,606	21,079,606	21,079,606
R-squared	0.603	0.603	0.603	0.604

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and brand fixed effects.

Table 3.11: Effect of the Philadelphia SSB tax on diet soda price per oz (cents) in neighboring areas

VARIABLES	(1) Price per oz (cents)	(2) Price per oz (cents)	(3) Price per oz (cents)	(4) Price per oz (cents)
Treatment	-0.163***	-0.509***	-0.159	-0.628***
	(0.0299)	(0.0915)	(0.149)	(0.212)
Treatment*Beverage volume (oz)		0.00412***	-0.00375	0.0136**
		(0.000997)	(0.00367)	(0.00674)
Treatment*Beverage volume (oz) squared			3.03e-05*	-0.000109**
			(1.79e-05)	(5.45e-05)
Treatment*Beverage volume (oz) cubed				2.57e-07**
				(1.29e-07)
Beverage volume (oz)	-0.150***	-0.151***	-0.151***	-0.152***
	(0.00914)	(0.00913)	(0.00917)	(0.00959)
Beverage volume (oz) squared	0.000898***	0.000899***	0.000901***	0.000909***
	(8.80e-05)	(8.79e-05)	(8.87e-05)	(9.27e-05)
Beverage volume (oz) cubed	-1.43e-06***	-1.44e-06***	-1.45e-06***	-1.46e-06***
	(2.26e-07)	(2.26e-07)	(2.30e-07)	(2.39e-07)
Observations	11,934,940	11,934,940	11,934,940	11,934,940
R-squared	0.633	0.633	0.634	0.634

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
All specifications include store, month, year and brand fixed effects.

Table 3.12: Effect of the Philadelphia SSB tax on consumption (oz) of regular and diet soda in neighboring areas

VARIABLES	(1) Volume regular soda (oz) per capita	(2) log Volume regular soda (oz) per capita	(3) Volume regular soda (oz)	(4) Volume diet soda (oz) per capita	(5) log Volume diet soda (oz) per capita	(6) Volume diet soda (oz)
Treatment	0.00225	0.179***	370.9	-0.000455	0.126***	-1,686*
	(0.00137)	(0.0276)	(1,431)	(0.000842)	(0.0289)	(893.6)
Observations	94,716	84,734	94,716	78,192	61,918	78,192
R-squared	0.461	0.881	0.453	0.417	0.884	0.411

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
All specifications include store, month, year and brand fixed effects.

Table 3.13: Effect of the Berkeley SSB tax on consumption (oz) of water

VARIABLES	(1) Volume water (oz) per capita	(2) log Volume water (oz) per capita	(3) Volume water (oz)	(4) Volume water (oz) per capita	(5) log Volume water (oz) per capita	(6) Volume water (oz)
Treatment	0.00733***	0.0286	2,181*	0.0596*	0.0190	-6,787
	(0.00232)	(0.0449)	(1,245)	(0.0313)	(0.0601)	(4,607)
Treatment*Flavored				-0.163**	0.0298	27,934*
				(0.0727)	(0.114)	(16,347)
Observations	31,051	31,051	31,051	31,051	31,051	31,051
R-squared	0.324	0.719	0.316	0.326	0.719	0.316

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
All specifications include store, month, year and brand fixed effects.

Table 3.14: Effect of the Berkeley SSB tax on consumption (ml) of wine

VARIABLES	(1) Volume wine (ml) per capita	(2) log Volume wine (ml) per capita	(3) Volume wine (ml)
Treatment	-0.00128*	-0.0339	-127.6
	(0.000739)	(0.0303)	(144.7)
Observations	1,003,835	581,216	1,003,835
R-squared	0.094	0.201	0.026

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
All specifications include store, month, year and wine type fixed effects.

Table 3.15: Effect of the Berkeley SSB tax on consumption (oz) of candy

VARIABLES	(1) Volume candy (oz) per capita	(2) log Volume candy (oz) per capita	(3) Volume candy (oz)
Treatment	-1.89e-05***	-0.0426	-2.611***
	(5.76e-06)	(0.0262)	(0.913)
Observations	5,993,314	1,701,898	5,993,314
R-squared	0.036	0.294	0.042

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
All specifications include store, month, year and candy type fixed effects.

Table 3.16: Effect of the Berkeley SSB tax on consumption (oz) of beer

	(1)	(2)	(3)
VARIABLES	Volume beer (oz) per capita	log Volume beer (oz) per capita	Volume beer (oz)
Treatment	-0.00105	-0.159**	-120.7
	(0.000753)	(0.0681)	(108.1)
Observations	329,324	174,633	329,324
R-squared	0.063	0.314	0.059

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and beer type fixed effects.

Table 3.17: Effect of the Philadelphia SSB tax on consumption (oz) of water

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Volume water (oz) per capita	log Volume water (oz) per capita	Volume water (oz)	Volume water (oz) per capita	log Volume water (oz) per capita	Volume water (oz)
Treatment	0.000371	0.0461	2,178	-0.00397**	0.110***	618.4
	(0.00226)	(0.0337)	(3,306)	(0.00195)	(0.0341)	(2,430)
Treatment*Flavored				0.0153***	-0.224***	5,501
				(0.00461)	(0.0682)	(6,108)
Observations	55,899	55,899	55,899	55,899	55,899	55,899
R-squared	0.211	0.619	0.202	0.211	0.619	0.202

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and brand fixed effects.

Table 3.18: Effect of the Philadelphia SSB tax on consumption (oz) of candy

VARIABLES	(1) Volume candy (oz) per capita	(2) log Volume candy (oz) per capita	(3) Volume candy (oz)
Treatment	5.68e-06***	0.0446***	2.800***
	(9.21e-07)	(0.0106)	(0.861)
Observations	14,747,481	3,945,809	14,747,481
R-squared	0.040	0.269	0.063

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and candy type fixed effects.

Table 3.19: Effect of the Philadelphia SSB tax on consumption (oz) of beer

VARIABLES	(1) Volume beer (oz) per capita	(2) log Volume beer (oz) per capita	(3) Volume beer (oz)
Treatment	0.000111	0.0791**	126.4
	(6.99e-05)	(0.0311)	(92.15)
Observations	425,844	127,519	425,844
R-squared	0.052	0.189	0.052

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and beer type fixed effects.

Table 3.20: Effect of the Philadelphia SSB tax on consumption (ml) of wine

VARIABLES	(1) Volume wine (ml) per capita	(2) log Volume wine (ml) per capita	(3) Volume wine (ml)
Treatment	-0.00347**	0.181	-1,318
	(0.00175)	(0.125)	(1,564)
Observations	228,420	27,750	228,420
R-squared	0.097	0.372	0.096

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All specifications include store, month, year and wine type fixed effects.

3.9 Figures

Figure 3.1: Event study: Regular soda price per oz (cents), Berkeley

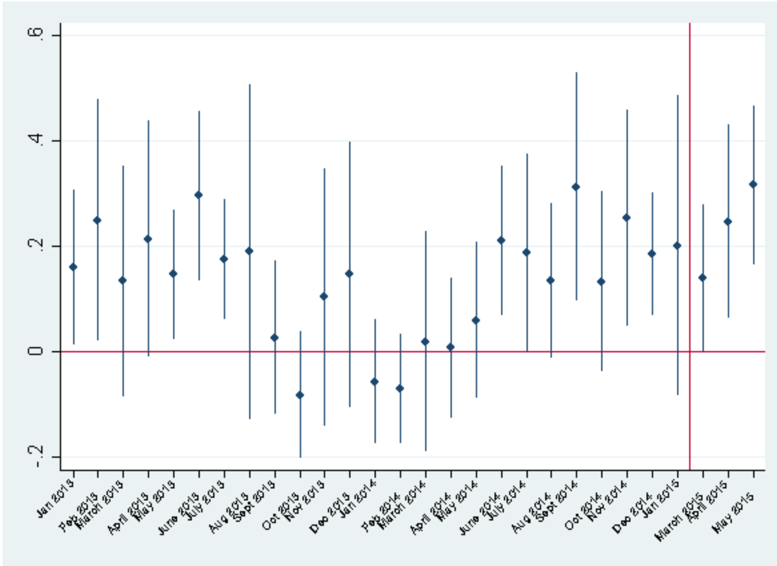


Figure 3.2: Event study: Log regular soda consumption per capita, Berkeley

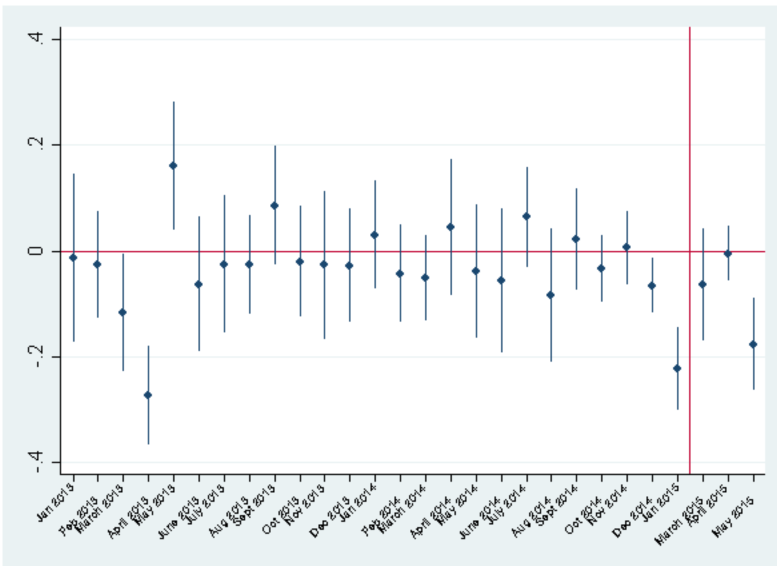


Figure 3.3: Event study: Regular soda price per oz (cents), Philadelphia

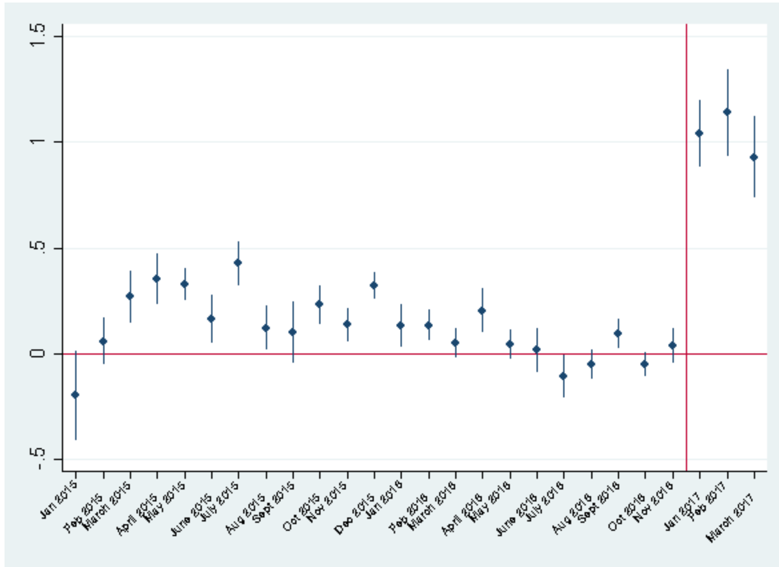


Figure 3.4: Event study: Log regular soda consumption per capita, Berkeley

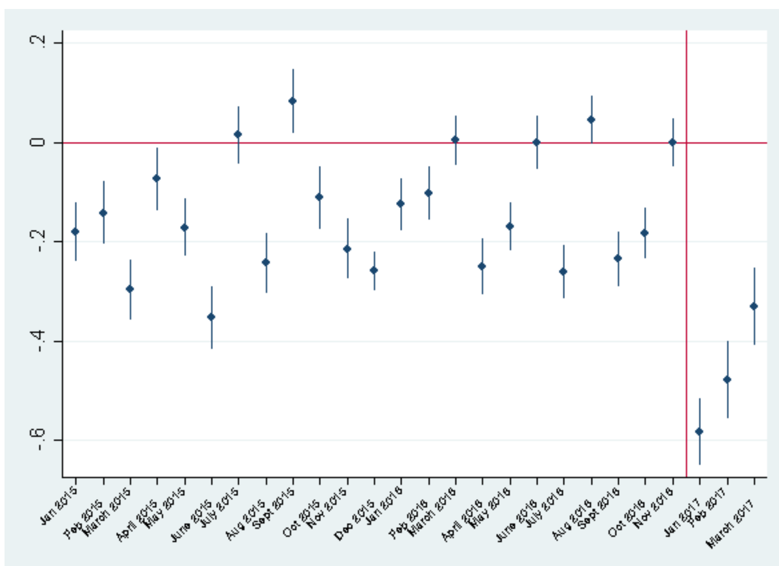


Figure 3.5: RSK Procedure: Effect of Berkeley tax on price per oz (cents)

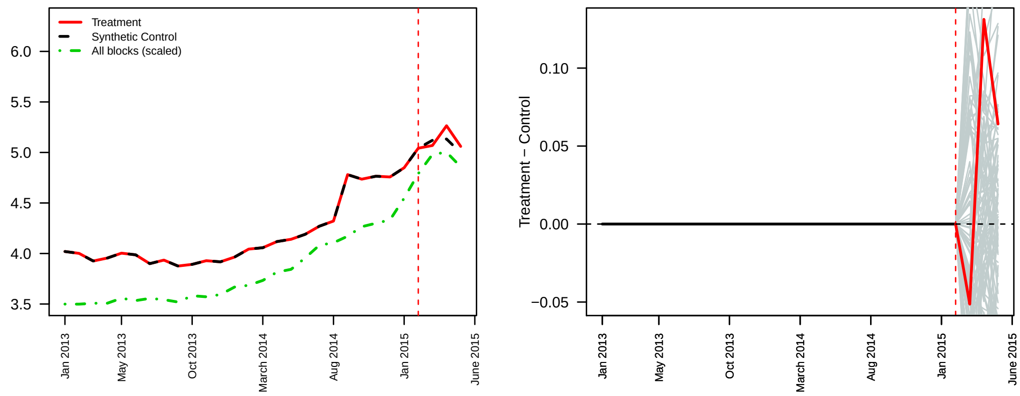


Figure 3.6: RSK Procedure: Effect of Berkeley tax on consumption per capita

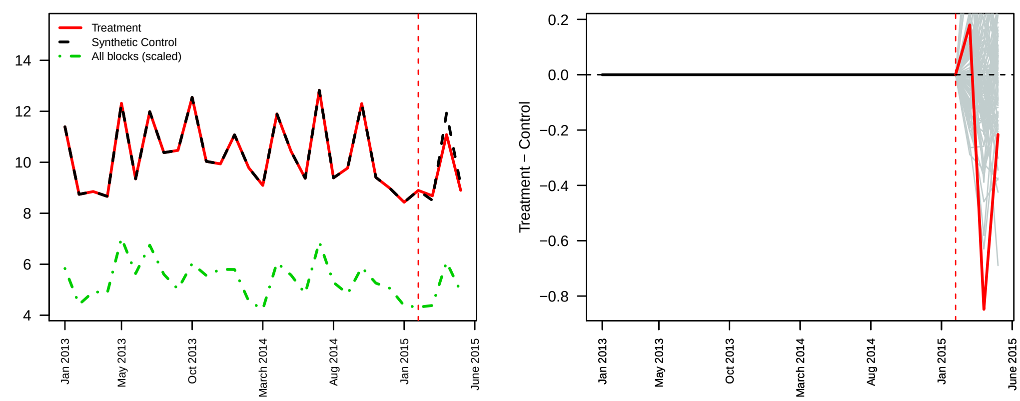


Figure 3.7: RSK Procedure: Effect of Philadelphia tax on price per oz (cents)

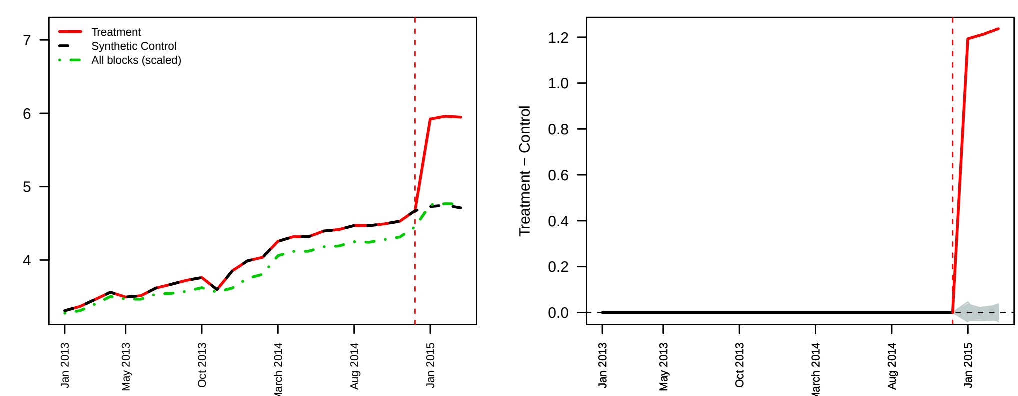
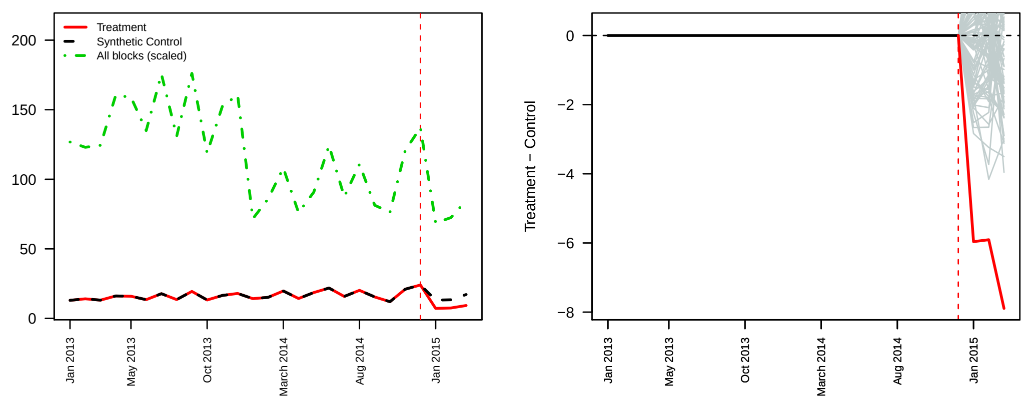


Figure 3.8: RSK Procedure: Effect of Philadelphia tax on consumption per capita



3.10 REFERENCES

Abadie Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105(490): 493-505.

Abadie Alberto, Alexis Diamond, and Jens Hainmueller. 2015. "Comparative politics and the synthetic control method." *American Journal of Political Science*, 59(2):495-510.

Abadie Alberto and Javier Gardeazabal. 2003. "The economic costs of conflict: a case study of the Basque Country." *American Economic Review*, 93(1):113-132.

Allcott Hunt, Benjamin B. Lockwood and Dmitry Taubinsky. 2019. "Regressive Sin Taxes with an Application to the Optimal Soda Tax." *The Quarterly Journal of Economics*, 134(3):1557-1626.

Avena, Nicole M., Pedro Rada and Bartley G. Hoebel. 2008. "Evidence for sugar addiction: Behavioral and neurochemical effects of intermittent, excessive sugar intake." *Neuroscience Biobehavioral Reviews*, 32: 20-39.

Berardi Nicoletta, Patrick Sevestre, Marine Tépaut and Alexandre Vigneron. 2016. "The impact of a 'soda tax' on prices: evidence from French micro data." *Applied Economics*, 48(41):3976-3994.

Bollinger Bryan and Steven E. Sexton. 2017. "Local Excise Taxes, Sticky Prices, and Spillovers: Evidence from Berkeley's Soda Tax." (January 12, 2018). Available at SSRN: https://ssrn.com/abstract=3087966 or http://dx.doi.org/10.2139/ssrn.3087966

Cawley, John and David Frisvold. 2017. "The Pass-Through of Taxes on Sugar-Sweetened Beverages to Retail Prices: The Case of Berkeley, California." *Journal of Policy Analysis and Management*, 36(2):303-326

Cawley, John and David Frisvold. 2017. "Pass-Through of a Tax on Sugar-Sweetened Beverages at the Philadelphia International Airport." *JAMA* 319(3): 305-306.

Cawley John, David Frisvold, Anna Hill and David Jones. 2019. "The Impact of the Philadelphia Beverage Tax on Purchases and Consumption by Adults and Children." *Journal of Health Economics*, 67:102225.

Cawley John, David Frisvold, Anna Hill and David Jones. 2020. "The Impact of the Philadelphia Beverage Tax on Prices and Product Availability." *Journal of Policy Analysis and Management*, 39(3):605-628.

Cawley John, David Frisvold and David Jones. 2020. "The Impact of Sugar-sweetened Beverage Taxes on Purchases: Evidence from Four City-level Taxes in the United States." *Health Economics*, 29(10): 1289-1306.

Census QuickFacts Pittsburgh city, Pennsylvania.
https://www.census.gov/quickfacts/fact/table/pittsburghcitypennsylvania/PST045218 Accessed July 29, 2019

Census QuickFacts Philadelphia city, Pennsylvania.
https://www.census.gov/quickfacts/fact/table/philadelphiacitypennsylvania/PST045218 Accessed July 29, 2019

Colchero M Arantxa, Juan Carlos Salgado, Mishel Unar-Munguía, Mariana Molina, Shuwen Ng and Juan Angel Rivera-Dommarco. 2015."Changes in Prices After an Excise Tax to Sweetened Sugar Beverages Was Implemented in Mexico: Evidence from Urban Areas." *PLoS ONE*, 10(12):e0144408.

Colchero M Arantxa, Barry M Popkin, Juan A Rivera and Shu Wen Ng. 2016. "Beverage purchases in stores in Mexico under the excise tax on sugary sweetened beverages: observational study." *British Medical Journal*, 352:h6704

DellaVigna Stefano and Matthew Gentzkow. 2019. "Uniform pricing in US retail chains." *The Quarterly Journal of Economics*, 134(4):2011-2084.

Falbe, Jennifer, Nadia Rojas, Anna Grummon, and Kristine Madsen. 2015. "Higher retail prices of sugar-sweetened beverages 3 months after implementation of an excise tax in Berkeley, California." *American Journal of Public Health*, 105: 2194-2201.

Falbe, Jennifer, Hannah R. Thompson, Christina Becker, Nadia Rojas, Charles E. McCulloch, and Kristine A. Madsen. 2016. "Impact of the Berkeley Excise Tax on Sugar-Sweetened Beverage Consumption," *American Journal of Public Health*: e1-e7.

Falbe Jennifer, Anna H. Grummon, Nadia Rojas, Suzanne Ryan-Ibarra, Lynn D. Silver and Kristine A. Madsen. 2020. "Implementation of the First US Sugar-Sweetened Beverage Tax in Berkeley, CA, 20152019." *American Journal of Public Health*, 110(9): 1429-1437.

Fletcher, Jason M., David E. Frisvold, and Nathan Tefft. 2010. "The Effects of Soft Drink Taxes on Child and Adolescent Consumption and Weight Outcomes," *Journal of Public Economics*, 94:967-974.

Gibson Laura A., Hannah G Lawman, Sara N. Bleich, Jiali Yan, Nandita Mitra, Michael T. LeVasseur, Caitlin M. Lowery, Christina A. Roberto. 2021. "No Evidence of Food or Alcohol Substitution in Response to a Sweetened Beverage Tax." *American Journal of Preventive Medicine*, 60(2):e49-e57.

Grogger, Jeffrey. 2017. "Soda taxes and the prices of sodas and other drinks: evidence from Mexico." *American Journal of Agricultural Economics*, 99(2):481-498.

Holdbrook Jeanette Dana Petersen, David Jones and David Frisvold. 2019. “How Retailers Responded to Taxes on Sweetened Beverages: A Tale of Two Cities.” *Mathematica Health Issue Brief* Available at:https://mathematica.org/publications/how-retailers-responded-to-taxes-on-sweetened-beverages-a-tale-of-two-cities

Knight, Heather. Nov 7, 2014. “Why Berkeley Passed A Soda Tax and S.F. Didnt.” SF-Gate. http://www.sfgate.com/bayarea/article/Why-Berkeley-passed-a-soda-tax-and-S-F-didn-t-5879757.php Accessed: April 17, 2017.

Powell Lisa M. and Julien Leider. 2022. “Impact of the Seattle Sweetened Beverage Tax on substitution to alcoholic beverages.” *PLoS One*, 17(1):e0262578.

Robbins Michael W., Jessica Saunders and Beau Kilmer. 2017. “A Framework for synthetic control methods with high-dimensional, micro-level data: evaluating a neighborhood-specific crime intervention.” *Journal of the American Statistical Association*, 112(517):109-126.

Roberto Christina A, Hannah G Lawman, Michael T. LeVasseur, Nandita Mitra, Ana Peterhans, Bradley Herring and Sara N. Bleich. “Association of a Beverage Tax on Sugar-Sweetened and Artificially Sweetened Beverages With Changes in Beverage Prices and Sales at Chain Retailers in a Large Urban Setting.” *Journal of the American Medical Association*, 321(18):1799-1810.

Rojas Christian and Emily Wang. 2021 “Do Taxes for Soda and Sugary Drinks Work? Scanner Data Evidence from Berkeley and Washington.”*Economic Inquiry*, 59(1):95-118.

San Francisco Board of Elections. “November 4, 2014 Official Election Results.” http://www.sfelections.org/results/20141104/ Accessed: April 17, 2017

Seiler Stephan, Anna Tuckman and Song Yao. 2021 “The Impact of Soda Taxes: Pass-Through, Tax Avoidance, and Nutritional Effects.” *Journal of Marketing Research*, 58(1):22-49.

Silver Lynn D, Shu Wen Ng, Suzanne Ryan-Ibarra, Lindsey Smith Taillie, Marta Induni, Donna R. Miles, Jennifer M. Poti, Barry M. Popkin. 2017. “Changes in prices, sales, consumer spending, and beverage consumption one year after a tax on sugar-sweetened beverages in Berkeley, California, US: A before and after study.” *Plos Medicine* 14(4):e1002283 https://doi.org/10.1371/journal.pmed.1002283

Yang, Qing. 2010. “Gain weight by ‘going diet?’ Artificial sweeteners and the neurobiology of sugar cravings.” *Yale Journal of Biology and Medicine*, 83: 101-108.

Zoutman Floris, Evelina Gavrilova and Arnt O. Hopland. 2018. “Estimating Both Supply and Demand Elasticities Using Variation in a Single Tax Rate.” *Econometrica* 86(2):763-771.

APPENDIX A

Evolution of the Standard of Care in Michigan - A Case Study

MCLS §600.2912a, the statute that describes the burden of proof in Michigan medical practice lawsuits, provides a great starting point for an analysis of the standard of care in Michigan. MCLS §600.2912a(1) reads

"Subject to subsection (2), in an action alleging malpractice, the plaintiff has the burden of proving that in light of the state of the art existing at the time of the alleged malpractice:

(a) The defendant, if a general practitioner, failed to provide the plaintiff the recognized standard of acceptable professional practice or care in the community in which the defendant practices or in a similar community, and that as a proximate result of the defendant failing to provide that standard, the plaintiff suffered an injury.

(b) The defendant, if a specialist, failed to provide the recognized standard of practice or care within that specialty as reasonably applied in light of the facilities available in the community or other facilities reasonably available under the circumstances, and as a proximate result of the defendant failing to provide that standard, the plaintiff suffered an injury."

Part a indicates that the standard of care for general practitioners in Michigan has a geographic dimension, specified as accepted practice or care in the same or similar community as the defendant. Court cases illustrate the importance of the same or similar community standard for general practitioners in the early 2000s, which continues to this day. In Robins v. Garg, 270 Mich. App. 519 (2007) a patient died of cardiac arrest in the office of their doctor. A lawsuit was filed on behalf of the patient, and the trial court granted a motion for summary judgment by the defense after the plaintiff's expert witness was prevented from testifying. The defense argued that the plaintiff's expert witness was not qualified to testify about the standard of care required of the defendant physician because he was not familiar with the standard of care in Oakland County, Michigan. The plaintiff's expert, Dr. Werlinsky, practiced in Palm Beak County, Florida. The Court of Appeals opined that as long as he is familiar with a community's standard of care, an expert may testify as to that community's the standard of care even if he has never practiced there. Dr. Werlinsky testified that the population size and number of hospitals and family practice physicians were similar in Palm Beach and Oakland County. He also testified that he "…interacted with general practitioners throughout the country and… he practiced medicine similarly to the way it was practiced in Michigan" so the Appeals Court overturned the trial court's ruling.

Two recent cases illustrate the continued relevance of locality to medical malpractice cases against general practitioners in Michigan. In Herrera v. Seiler, 2019 Mich. App. (2019) a podiatrist (classified as a general practitioner by Michigan courts) was sued when complications from a patient's broken foot resulted in an amputation below the knee. The trial court granted a motion for summary judgment for the defendants, because the plaintiff's witness could not demonstrate familiarity with the standard of care for podiatrists in Holland Michigan. The plaintiff appealed, arguing that a national standard of care applied and that his witness demonstrated familiarity with both the national standard and the local standard of care. The Court of Appeals rejected the plaintiff's argument that "a national standard of care should apply in this case because the medical treatment at issue in this case should not have varied by locality." Instead, the court stated that a local standard of care applied, and declined to overturn the trial court's ruling that the plaintiff's witness did not show sufficient familiarity with Holland Michigan to be allowed to testify. In a lengthy discussion of under what conditions a witness should be deemed qualified to testify about the standard of care, the court explained:

"An expert witness can become familiar with the local standard of care by talking with local doctors, reading about the community in question, and conducting other research of the community in question. Turbin, 214 Mich App at 218-219. An expert, however, must establish a basis for how he or she is familiar with the local standard of care about which the expert is called upon to testify… Dr. Marasco additionally stated that based on his unspecified research, Merrillville, Indiana and Holland, Michigan were similar in population and had similar availability of medical specialists, procedures, and technology. Specifically, Dr. Marasco stated that Merrillville, Indiana and Holland, Michigan had similar access to infectious disease specialists because of Holland, Michigan's proximity to Grand Rapids, Michigan. Dr. Marasco, however, failed to specify what his research entailed or how he came to these conclusions… The only basis for Dr. Marasco being qualified to offer such an opinion, however, was unspecified references to his 'research' comparing Holland, Michigan and Merrillville, Indiana. MRE 702 required the trial court to ensure, as a preliminary matter, that each aspect of Dr. Marasco's testimony was reliable before qualifying him as an expert witness. See Elher, 499 Mich at 22. As addressed earlier, MRE 702 also required the trial court to ensure a reliable basis for Dr. Marasco's knowledge of the standard of care in Holland, Michigan. Dr. Marasco failed to describe with any specificity why he believed that the communities of Holland, Michigan, and Merrillville, Indiana were similar."

The court of appeals further elaborates on the role of locality in determining the standard of care in Johnson v. Ziyadeh, 2019 Mich. App.. This case also involved a summary judgement for the defendant due to a plaintiff's expert who was deemed unqualified. The Court of Appeals did not overturn the summary judgement for the defendant, concluding that "Dr. Kelman [plaintiff's expert] denied having specific knowledge of Dr. Ziyadeh's practice, including the percentage of his Medicaid patients as opposed to his private-pay patients, or the amount of low- or high-income patients that Dr. Ziyadeh had served. Dr. Kelman also testified that he had never been to Michigan and that he did not recall the last time he had interacted with a Michigan dentist in any capacity. . . .Dr. Kelman testified at his deposition that he was unfamiliar with Dr. Ziyadeh's practice and spoke only vaguely about Wayne County and the surrounding area, making it clear that he was 'lumping in' all Michigan counties as similar to Wayne County. Dr. Kelman also merely speculated that he might have spoken with or received continuing education from Michigan dentists or graduates of Michigan dental schools, but had no specific recollections in that regard." While it is clear that the locality rule has played a significant role in establishing the standard of care for general practitioners in medical malpractice lawsuits, the place of the rule in lawsuits against specialists is less clear. The phrase in MCLS §600.2912b "applied in light of the facilities available in the community" is somewhat ambiguous: does this define a local standard of care, based on how specialists apply the standard of care for their specialty in the community in which they practice? Or does the statute establish a national standard of care, where locality only becomes relevant when the defendant is constrained by specific difficulties or resource constraints of his community?

Three Michigan Supreme Court cases from before 2000 affirmed a national standard of care for specialists. In 1970, the Supreme Court ruled in Naccarato v. Grob, 384 Mich. 248 that specialists must be held to a national standard of care. In this case, a pediatrician in Detroit was sued for failing to order a test to diagnose PKU when a child exhibited signs of mental decline. Expert witnesses for the defense testified that Detroit pediatricians do not customarily order the PKU test because PKU is very rare, raising the issue of locality. The court concluded:

"The reliance of the public upon the skills of a specialist and the wealth and sources of his knowledge are not limited to the geographic area in which he practices. Rather his knowledge is a speciality. He specializes so that he may keep abreast. Any other standard for a specialist would negate the fundamental expectations and purpose of a speciality. The standard of care for a specialist should be that of a reasonable specialist practicing medicine in the light of present day scientific knowledge. Therefore, geographical conditions or circumstances control neither the standard of a specialist's care nor the competence of an expert's testimony."

Nine years later, the Supreme Court upheld this ruling in Francisco v. Parchment Medical

Clinic, P. C., 407 Mich. 325 (1979), stating that "It was reversibly erroneous for the trial court to disallow Golomb's testimony because he was not familiar with the practice of surgeons in the community of Kalamazoo or similar communities. It does not matter whether the practice in Chicago and in Kalamazoo is similar; the standard for a specialist is a national standard, not a local one." In 1995, after MCLS §600.2912 was enacted, the Supreme Court continued to uphold a national standard of care for specialists in the oft-cited Bahr v. Harper-Grace Hosps., 448 Mich. 135 (1995). However in 2002, in Cox v. Bd. of Hosp. Managers, 467 Mich. 1 the Supreme Court interpreted MCLS §600.2912a as setting a community standard of care for both general practitioners and specialists: "The term 'national,' however, is not an accurate description of the statutory standard of care for specialists… Under the plain language of the statute, then, the standard of care for both general practitioners and specialists refers to the community." Later courts interpreted the Cox decision as establishing a two-step procedure. In Smith v. Joy, 2005 Mich. App. the Court of Appeals overturned a trial court's determination that an expert witness was not qualified to testify about the standard of care for specialists in Charlevoix because he practiced in an urban area and had never visited the Charlevoix Area Hospital (located in a rural area). The Court of Appeals stated that under MCLS §600.2912a the fact that specialists must comply with the standard of care within their specialty points to a national standard of care; however local conditions must be considered when they are relevant in individual cases. The Court of Appeals did not deem that in Smith v Joy locality should prevent the urban expert witness from testifying about administering antibiotics in Charlevoix.

Although parties in cases involving specialist defendants are sometimes required to demonstrate that their witness is familiar with the standard of care for specialists in the same or similar community (see for example Lowery v. Beer, 2009 Mich. App. (2009)), in most cases this does not seem to be required (see for example Bryson v. Genesys Reg'l Med. Ctr., 2018 Mich. App. (2018)). Locality is much less contested in cases involving specialists than in cases involving general practitioners. (See for example Herrera v Seiler and Johnson v. Ziyadeh discussed above for comparison.)

From an analysis of statutes and case law from Michigan, it appears that from 2000 through 2015, the standard of care for general practitioners had a "same or similar community" locality component. For specialists, the standard of care can be described as a national standard default, with accommodations possible if there are some particular local circumstances that need to be accounted for.

APPENDIX B

Excerpt from Swan v. Lamb, 584 P.2d 814 Supreme Court 1978

The following is an excerpt from Swan v. Lamb, 584 P.2d 814 Supreme Court 1978, referenced in Section 5 Robustness above.

Our quality of medical care in Utah rates with the best in the nation. Our hospitals are among the finest with the most recent technology, and the medical college at the University of Utah enjoys an outstanding reputation. In addition, doctors practicing their profession here come from various medical colleges throughout the nation. Medical journals are available nationally as are seminars and workshops. There is no need for doctors here to have a lower standard of care than that of other doctors who are practicing in similar localities. Indeed, it is doubtful that any physician in the State of Utah would be willing to admit that his skill and knowledge is not equal to any other physician trained in his field, or that his ability is less than that of doctors trained and practicing in other cities.

True it may be that doctors practicing in small rural communities cannot be expected to have the facilities or the equipment to perform equally as well as can physicians in Salt Lake City; however, they have the same quality of training and should know enough to refuse to undertake operations or to treat patients if they are not in a position to successfully administer the needed treatment - save perhaps in emergency cases.

If surgeons throughout the nation consider it improper to allow foreign substances that have been injected into the spinal canal to remain there after completing a myelogram, it beggars the imagination to think a doctor in Salt Lake City could escape responsibility for harm done to his patient by failing to remove the substance merely because the local custom is to leave the substance in the canal so that it will be absorbed by the body. If this procedure is generally regarded to be unsatisfactory or dangerous, no doctor should escape responsibility merely because the local practice has not yet adopted it.