Three Essays on the Economics of Consumerism in Health Policy

by

Elizabeth Querna Cliff

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Health Services Organization and Policy)
in the University of Michigan
2019

Doctoral Committee:

        Professor Richard A. Hirth, Chair
        Professor Thomas C. Buchmueller
        Professor A. Mark Fendrick
        Professor Helen G. Levy

Elizabeth Querna Cliff

bqcliff@umich.edu

ORCID iD: 0000-0002-0120-7439

Dedication

To Kellie Bach, Angie Bishop, Thabani Muller, Thomas Waldon, Laurie Gianateimpo, Patricia Carley, Tara Haight, Tommy Silva, Jacqueline Chino, William Gibbons, Carolyn Burch, Suzy Combs, Greg Pierce, Kathleen Dolan, Nancy Pierce, Galina Alva, Erica Harris, Steve and Carol Tupper, Jamie Israel, Lisa Collins, Jack Triperinas, Ken Flory, Ed and Jo Lane, Roger Olson, and other sources who shared with me, during my time as a reporter, their stories of navigating and struggling in the U.S. health system.

Acknowledgements

University of Michigan, H2D2 at the University of Michigan, Middlebury College, and The Dartmouth Institute. I benefited from seminar participants and discussants in attendance.

I am thankful for the people who allowed me the time and energy to work by caring for my children: the wonderful staff at Community Day Care and Preschool—Laura Stidham, Sandra Swan, Beth Meza, Megan Thompson and Katie Davison, among many others —Ms. Fern Jackson at Ann Arbor Open's after school program, Brandon Jordan, and Lissandre Dufresne.

I am indebted to musicians who I do not know personally but who, despite being in a challenging industry, continue to write music that punctuate life's highs and make the lows less lonely. Completing my doctoral degree would have been more much difficult without the company of their songs.

Finally, I am grateful to family and friends for encouragement and perspective. My parents, Kit and Christie Querna, and parents-in-law, Tom and Joan Cliff, have been endlessly loving and giving. Many thanks to Katie Querna, Emily Querna Strizich, Garrett Strizich, Brian Cliff, Marni Rothman, Maura Cliff, Margaret Crandall, Alex Ralph, Lizzie Hutton, Kate Andrias, Nick Allen, Mike Caferella, Torri Still, Sarah Snip Stroup, Sunny Bowles, Leslie Plaisted, Lucy Ricca, Amanda Manger, and Kim Berglund.

I would not be where or who I am without my beloved husband, Patrick, and my sons, Charlie and Gus. My biggest thanks are to them for enriching every one of my days.

Table of Contents

List of Tables

List of Figures

Abstract

Consumerism, the notion of giving people the ability and responsibility to make choices about cost and quality, is an increasingly popular strategy to contain health spending. Yet we lack robust knowledge about how recent consumer-based initiatives affect key components of spending: service utilization and price. This dissertation examines spending outcomes of three popular initiatives that leverage principles of consumerism to promote efficiency in health care: value-based insurance design, high-deductible health plans, and releasing information on medical prices. In each case, I find shortcomings in the way the initiative is currently deployed. First, I find that an intervention using value-based insurance design as part of a strategy to encourage high-value preventive services unintentionally increased the use of services *without* evidence of clinical benefit. In my second essay, I find that state-based efforts to make prices less opaque through the release of pricing information in all-payer claims databases does not change mean price and may increase prices at the low end of the pricing distribution. Finally, I show that increasing people's financial responsibility through high-deductible health plans does not increase their likelihood of shopping for a lower-priced hospital for childbirth. Taken together, these essays show that consumer-based incentives, at least as they are currently deployed, are not sufficient to reduce prices paid or improve the efficiency of health service utilization. Policymakers should consider additional supports to assist people in discerning value in the healthcare system or use provider incentives in conjunction with these consumer-based efforts to mitigate growth of medical spending.

Chapter 1 : Introduction

Levels and growth of health care spending in the United States are among the highest in the developed world, raising concerns about both national expenditures and affordability for individual families. On a national scale, there is a widespread recognition that higher spending has not translated to better health outcomes relative to other nations. For individuals and families, the cost of both health insurance and medical services is consistently among the chief complaints about the health system, with a significant percentage of the population reporting high financial burdens from health care.

One of the most popular strategies to control health spending—and the most popular among employers—has been to increase the responsibility of individuals for paying for their own care and seeking out high-value providers. Over the past decade, levels of enrollee cost sharing in health insurance plans have grown quickly and steadily. At the same time, recent years have seen a proliferation of tools to assist individuals in discerning value and seeking health information: price shopping tools, health information tracking apps, and quality comparison websites are all becoming widespread.

Despite the popularity, we still lack knowledge about whether these initiatives and whether, more generally, treating people as consumers in the health system can work to control spending without impacting health. Economic theory on this point is enticingly clear; more choices, greater financial responsibility, and access to information should help people wring value from

the system. It works in other markets. But, the ways people make choices in health care and the information they value may be so different from other markets that lessons from those market may not apply. Previous work has established that people respond to lower out-of-pocket costs by increasing the quantity of medical care they use but on other facets of consumerism— shopping for price, discerning high- from low-value health services, using comparison websites—they have remained remarkably inert. There is a growing body of literature on how to structure cost sharing, assist consumers in making decisions, and align the incentives of patients with their providers. But, given the proliferation of efforts and amount of money involved—both on a national scale and, increasingly, for individuals and families—the importance of a better understanding of the role of consumerism in health care is crucial.

This dissertation aims to address gaps in our current knowledge of the effects of consumerism in health care, specifically looking at price transparency and two types of insurance benefit design. These consumer-based strategies can be used independently though are often used in tandem and, ideally, in ways that complement each other. In related essays, this dissertation examines the impacts on medical service price and utilization of three strategies currently used throughout the health care system: value-based insurance design, price transparency, and high-deductible health plans.

In the first essay, I examine whether encouraging valuable, recommended preventive services through health insurance benefit design affects services shown not to improve health, so-called low-value services. While decreasing out-of-pocket costs for enrollees has been shown to increase use of recommended preventive services, it could also inadvertently increase use of low-value services if these services are delivered in conjunction with recommended services. I measure prevalence of both recommended and low-value medical services following the

implementation of an insurance benefit design change for a group of state employees that reduced out-of-pocket costs for, and required use of, recommended preventive services. I find increases in both recommended services, which was intended by the intervention, and in low-value preventive care services, which were not targeted. The increase in low-value services is an unintended spillover from the intervention, and undermines potential efficiency gains from greater use of recommended preventive care. The results imply that consumers, providers, or both do not discern between recommended and low-value preventive services in a case in which recommended services were specifically encouraged.

In the second essay, I focus on state efforts aimed at making health care prices in states more transparent. Price transparency is popular among both consumers and policymakers as a way to encourage pricing pressure on providers. However, given limited consumer use to date and the lack of competition in health care markets, price transparency may have perverse anti-competitive effects. I describe a mechanism for these effects through bargaining ability of providers and test it empirically by looking at an intervention that made price information available to competing hospitals in several states. I find there is no change in mean price of inpatient hospital services, at least in the short-run, but that prices at the lower tails of the price distribution increase. The results imply some hospitals are able to use information to change their bargaining position, particularly those that were previously inexpensive. Given data limitations, this finding is suggestive rather than conclusive. However, it represents an early attempt to tackle what thus far has been an understudied aspect of policies to make health care prices more transparent.

In the final essay, I examine whether enrollees who have high-deductible health plans (HDHPs), and thus significant upfront cost sharing for most health care services, pay lower prices for one

shoppable service, childbirth. Current evidence is coalescing around a lack of shopping in

HDHPs, yet there remain open questions about the generalizability of previous findings as well

as whether there are important mediators of propensity to price shop, such as living in a market

with more choices. To mitigate issues of selection into HDHPs, I use a large national sample of

claims, and focus on situations in which it is clear an individual switched from a traditional

insurance plan to an HDHP due to the decision of an employer. I then use a difference-in-

differences design to measure whether enrollees whose employers switched all employees into

an HDHP were more likely than a control group to pay lower prices, or to use providers with

below-median prices in the market. I find that, after HDHP enrollment, women in markets with

more choices pay lower prices (about 3%) after HDHP enrollment than they did prior to having

an HDHP, but that this difference can largely be explained by provider drops in prices rather

than by active choices by enrollees. Consistent with previous evidence, my results show limited

shopping behavior by enrollees in HDHPs, even in markets in which there are multiple provider

choices.

In sum, these three essays suggest that using popular consumer-based strategies, at least in their

current form, fall short of promises to create more efficiency in the health system. In the first

case, lowering cost sharing for high-value services does increase their use, but not without

significant spillover in the form of increased use of services that do not improve health. In the

second case, making price information available in a way that allows competitors to see it may

have perverse effects by giving lower-priced providers fodder to raise prices. And, finally, the

third essay shows that, even in situations where enrollees have incentives and ability to price

shop, they do not do so. Taken together, these essays rebut the unfettered optimism with which

some policymakers and payers have embraced consumer-based strategies. The policy remedies

should include better supports to help consumers discern value among services and providers, along with complementary provider incentives to deliver it.

Chapter 2 : Hitting a Bulls-Eye?:
Evidence of Spillovers in Interventions Targeting Utilization of Evidence-Based Preventive Service

INTRODUCTION

Better aligning medical service use with evidence of medical benefit it is a major focus of efficiency improvement in the U.S. health system. There is substantial documentation of underuse of high-value services, those with clear benefits for patients, including age-appropriate cancer screenings and recommended preventive care (Smith et al. 2018). At the same time, there is a growing awareness of overuse of health care with few or no clinical benefits, often called low-value care (Schwartz et al. 2014; Colla et al. 2015; Charlesworth et al. 2016). Low-value care is worrisome not only because it represents wasteful spending but, by leading to overtreatment, can cause patient harm (Armstrong 2018). To date, encouragement of high-value, recommended medical services and deterrence of low-value services have largely been studied in isolation. Yet, these services are often delivered in similar settings to similar populations, and policies intended to affect one type of service may have spillover effects on the other. Understanding how policies that aim to change the use of either high-value or low-value care impact *both* types of services is crucial to understanding the overall efficiency impacts.

In this paper, I assess changes in high- and low-value service use that result from a 2011 insurance benefit design change intended to promote high-value preventive services for

employees of one state. Engaging consumers in alignment of service value through insurance benefit design is increasingly popular, as employers and insurers shift more financial risk onto consumers (Bundorf 2012; Claxton et al. 2017). Using a panel of continuously enrolled employees from one year before and two years after implementation of the intervention, I compare service use by employees in the intervention state with a control panel of employees from other states that had stable benefit plans over the study period. Compared with the control group, I find the likelihood of receiving a high-value preventive service in the first year of the intervention increased 11 percentage points, a relative 18% from baseline, and, in the second year increased 5.2 percentage points, a relative 9% from baseline. The likelihood of receiving a low-value service increased 7.9 percentage points, a relative 24% from baseline, in the first year, and 2.7 percentage points, a relative 8% from baseline, in the second year. To explore heterogeneity in the types of low-value services used, I split low-value services into preventive screening services delivered to asymptomatic adults (e.g. vitamin D screening, cardiac imaging without indication) and imaging used to diagnose a symptomatic concern (e.g. imaging for uncomplicated back pain, imaging for uncomplicated headache). I find the overall increase in low-value service use is entirely due to increases in likelihood of low-value preventive screenings for asymptomatic adults; likelihood of low-value imaging for symptomatic concerns declined slightly. As low-value preventive services are most likely delivered in conjunction with high-value preventive services, this result suggests economic complementarity between high- and low-value preventive care. Additionally, I explicitly test whether preventive office visits are associated with use of other preventive services by decomposing the increase in high- and low-value services into the portion attributable directly to the increase in preventive office visits. Here, I find that the majority of the increase is due to an increase in preventive care visits,

suggesting that the promotion of preventive care office visits in this intervention increased both high- and low-value care use.

I also estimate own- and cross-price elasticities for high- and low-value services, respectively, to measure the relative sensitivity of demand response to the intervention. I estimate the own-price elasticity of demand for high-value services to be -0.24, in the first year of the intervention and -0.11 in the second year. The cross-price elasticity of demand for low-value services is estimated to be -0.29 in the first year and -0.11 in the second year. These estimates align with previous work showing similar own-price sensitivity for both high- and low-value health care (Baicker, Schwartzstein, and Mullainathan 2015), and imply a lack of discernment about the clinical value of different preventive services.

This study is the first to provide evidence that an enhancement of generosity of health insurance specifically intended to improve uptake of high-value preventive services is associated with spillover effects on low-value services. The results show that changes in service use in response to cost-sharing changes can be blunt, cutting across the range of clinical benefits, even in cases in which out-of-pocket costs are explicitly tailored to encourage only beneficial services. While I do not propose a behavioral mechanism for why enrollees do not discern between high- and low-value services in this context, others have emphasized informational deficiencies (Pauly and Blavin 2008) or other biases that may cause people to make mistakes in service value (Baicker and Levy 2015; Baicker, Schwartzstein, and Mullainathan 2015). My results suggest one structural mechanism for the increase in low-value preventive services is the substantial increase in preventive office visits after the intervention. These visits are associated with use of other primary preventive services, both high- and low-value. Efforts that explicitly encourage

preventive office visits should be cognizant of the potential for increasing waste, and could take steps to explicitly discourage low-value preventive care.

These results should not be taken as a rebuke of aspirations to increase utilization of high-value medical care. Evidence-based preventive care remains an essential component of high-quality health systems. Yet, these results do have potential implications for efficiency arguments embedded in efforts to increase primary or preventive care use. There are myriad efforts to engage consumers in correcting the alignment of service use to better match evidence of clinical benefit including through high-deductible health plans, value-based insurance design and direct primary care initiatives. This paper suggests that these efforts may fall short if they do not take into account the complementarities between recommended high-value and low-value services and, thus, may have unintended consequences. Efforts to promote preventive care should adopt comprehensive evaluation measures that incorporate both high- and low-value services, and should consider incentives, for both consumers and providers, that specifically target low-value services.

BACKGROUND AND CONCEPTUAL MODEL

Cost sharing in insurance benefit design has long been used to support efficient service use. Work on optimal cost sharing is robust and dates back decades, recent additions include empirical investigations of specific types of cost-sharing structures. There is also a nascent but quickly growing body of work investigating the prevalence of low-value health service utilization. This study draws from both these bodies of literature in an assessment of how benefit structure interacts with the type of health care service used.

Optimal Copayment Design

The seminal Rand Health Insurance Experiment empirically established that increases in cost sharing reduced demand for medical services (Manning et al. 1987). Previous theoretical work had shown that, with a non-zero elasticity of demand, optimal health insurance includes some cost sharing to mitigate moral hazard (Pauly 1968; Zeckhauser 1970). An implication of both the experimental and theoretical work is that efficient use is attained by making cost sharing higher in cases in which the demand curve more elastic, and thus differs based on the medical good (Cutler and Zeckhauser 2000).

Later work considered the effect of imperfect information, noting that coinsurance amounts should vary not only based on responsiveness of the demand curve, but also on the difference between patient beliefs about the value of a given medical service and expert knowledge of that service (Pauly and Blavin 2008). The key element of this formulation is that it takes into account the "mistakes" people make in valuing medical care, overusing services without evidence of benefit and underusing beneficial services. Pauly and Blavin view this as an information problem. When patient demand curves exceed the demand based on expert opinion, coinsurance should be higher than it would under perfect information; when patient demand curves fall short of optimal demand based on expert opinion, optimal coinsurance is lower than in the case of full information. This logic is extended to take into account a range of behavioral mistakes in a formal model of welfare optimizing copay design by Baicker, Mullainathan and Schwartzstein (Baicker, Schwartzstein, and Mullainathan 2015). Their theory of behavioral hazard incorporates a range of behavioral mistakes people make in judging the value of medical care into the welfare calculation of optimal service use. It provides a theoretical framework for

differing incentives for specific medical products, such as is used in value-based insurance design (V-BID), which aligns cost sharing to the clinical value of a medical service (Chernew, Rosen, and Fendrick 2007; Fendrick and Chernew 2006).

In the context of V-BID plans, evidence has found an increase in demand response after lowering cost sharing with some evidence of offsets (substitutes) in downstream health care services. (Agarwal, Gupta, and Fendrick 2018; Maciejewski et al. 2014; Chernew et al. 2008). In a previous study using the same population used in this essay, eliminating cost sharing for recommended preventive services and requiring their use to avoid financial penalties substantially increased utilization of the recommended services (Hirth et al. 2016). That study found substantial increases in per enrollee cost after the intervention, though it was unclear if these effects were due to price or utilization trends.

Taking into account only own-price elasticities in setting optimal cost sharing ignores the fact that nearly all medical services are delivered in conjunction with other medical services, either contemporaneously or intertemporally. As medical goods may have co-varying demand curves, using only own-price effects may lead to incorrect estimates of the overall impact of a change in cost sharing. Thus, a complete model should incorporate the cross-price elasticity of medical services (Ellis and Manning 2007; Goldman and Philipson 2007). For medical goods that have substitutes, lowering cost sharing leads to a reduction in use of those other services. Conversely, for complementary goods, a decrease in cost sharing for one service leads to an increase in the demand for other services. When this economic theory is applied with specific medical services in mind, a result is the over-subsidization (relative to classic theory) of medical services with the potential to offset later medical costs (Newhouse 2006). Newhouse notes that in some cases, notably the Rand Health Insurance Experiment, subsidization of outpatient visits resulted in the

use of more care, including inpatient care, suggesting a complementary relationship and an increase in overall costs. More recently, a study of a high-deductible plan that included a health account available to pay outpatient, but not inpatient, care found increased outpatient spending led to a corresponding increase in inpatient spending (Kaestner and Sasso 2015). In instances like these, Newhouse argues, subsidization of services was not specific enough to minimize future medical costs. This paper tests an intervention in which only specific services were subsidized, perhaps extending Newhouse's question to ask how specific subsidization needs to be.

Low-Value Services

Recently, efforts to improve efficiency of service utilization have focused on the reduction of low-value care, which are medical tests and treatments used in patients unlikely to meaningfully benefit. The Choosing Wisely campaign, which has brought together a number of specialty societies to identify low-value services was formed in 2010 (Morden et al. 2014). Most work measuring the prevalence of low-value care is relatively recent. Studies to date are largely cross-sectional and have found that low-value care is common across a range of populations and care settings (Colla et al. 2015; Hong et al. 2017; Schwartz et al. 2014). Very little work has examined specific interventions that may reduce low-value care. As exceptions, Schwartz and co-authors found providers that enter into an accountable care organization have seen modest reductions (2015), and Gruber and co-authors found that a V-BID program that increased cost sharing for low-value services reduced utilization of those services (2016). The Gruber paper is somewhat similar to mine; a key difference is that they test own-price elasticities where my primary goal is to test the cross-price elasticities between high- and low-value services.

A goal of insurance benefit design is to maximize the use of recommended high-value services and minimize the use of low-value services. However, optimally adjusting copayments to take into account not only behavioral mistakes leading to misuse of each type of service but also cross-price elasticities between services is difficult. These two ideas may be at odds with one another, such as in the case where a high-value medical service is a complement to a low-value service, or a low-value service substitutes for another low-value service. Reconciling how to best take into account these trade offs requires knowledge of the nature of these relationships. To date, there is little information on the magnitudes of cross-price elasticities in an environment in which behavioral mistakes are accounted for through more generous subsidization of high-value services.

Conceptual Model

Using the literature cited above, I begin from a standard model of demand for medical services described by Arrow (1963)and Pauly (1968):

$$D\,(Q_1) \;=\; f(P_1, X, B(h), \varepsilon)$$

where demand for a medical service D (Q_1) depends on the price of that service (P_1), patient characteristics (X), the benefit of the service (B(h)), which is dependent on the underlying health condition, and an idiosyncratic preference for the medical service (ε). The presence of health insurance distorts this relationship by modifying the price of the good from the consumer perspective to $p^* = \delta P_1$, where δ is in the set [0, 1] such that it reduces the cost to the consumer to $p^* < P_1$. Consumers demand care when $p^* <= B(h)$ and moral hazard can arise in situations in which the demand curve has elasticity and consumption occurs past the point at which it would if consumers were subject to the full price, $p^* = P_1$.

13

This chapter reconsiders the standard model of demand, adding an element to the standard equation and relaxing an assumption made in the traditional model. First, consider the demand equation above with multiple medical goods:

$$D\,(Q_1\,) \;=\; f(P_1, \boldsymbol{P_2}, X, B(h), \varepsilon)$$

In this model, where P_2 is the price of other medical goods, demand for a medical service depends not just on the price of the service itself but also on the price of other medical services. That sets up the possibility that medical services could be economic complements or substitutes of each other; one iteration of this idea has preventive care substituting for more intensive treatment (Ellis and Manning 2007; Goldman and Philipson 2007). When we make $\boldsymbol{P_2}$ the price of recommended medical services, we can then ask how changing that price affects the cross-price elasticities for low-value services ($\frac{\partial Q_1}{\partial P_2}$). The sign and magnitude of this ratio compared with the own-price elasticity of demand for recommended services determine how changes in cost sharing affect the overall demand for health services.

Second, the idea that consumers would use low-value care at all comes from relaxing the assumption implicit in the neoclassical model that consumers use services when $p^* <= B(h)$. That is, the model does not necessarily take evidence of demand for a good as evidence that the good is beneficial from a social welfare (i.e. population health) point of view. In this context, that means people can make mistakes about the value of medical services and demand care for which they are unlikely to benefit.

SETTING, DATA AND ANALYSIS

Intervention: The state employee program

I examine a northeastern U.S. state that launched a new health benefit program in October 2011 for all employees and adult dependents. The program was part of an effort by the state to address a nearly $4 billion budget gap in that year. It was put together by both employees' unions and the Office of the State Comptroller, with the goals of both improving enrollee health and generating savings for the state. It continues to present day.

To address underuse of evidence-based preventive services and chronic disease care, the state's program incorporates principles of V-BID in which enrollee out-of-pocket (OOP) costs are aligned with the clinical value of medical services. The program reduces or eliminates consumer cost sharing for certain office visits and medications used for management of chronic disease. A novel feature of the program is a requirement that members receive specific services, including preventive care office visits and age-appropriate primary preventive services such as cancer screenings. Enrollee compliance with these requirements is evaluated annually. Enrollees can be removed from the plan for non-compliance, and be switched into an alternative plan without the program's requirements but with higher cost sharing and premiums. To encourage use of evidence-based services, first dollar coverage (i.e., no consumer cost sharing) is provided for specific primary preventive services. In general, these targeted services followed guidelines from the U.S. Preventive Service Task Force (USPSTF) at the time of program implementation (US Preventive Services Task Force 2010). Figure 2-1 is a screen shot from the program's webpage showing the requirements and how they are presented to enrollees.

Enrollment in the program is voluntary, though members received a $100 discount per month on premiums and elimination of the deductible, which could save $350 for an individual and $1,400 for a family. Ninety-eight percent of employees enrolled in the first year and 98% of those remained in compliance in the program's first two years.

15

Data

I use administrative claims data from intervention state employees and their dependents from one plan year before the intervention (July 2010-June 2011) and two years post intervention (July 2011-June 2013). The data include service dates and claims-level procedure and diagnostic codes for all medical services, though we focus here on outpatient only. From an initial cohort of approximately 190,000 enrollees, I exclude those who were not continuously enrolled, those with negative claim amounts, and those not eligible for the intervention (children and retirees), and those who may have been eligible for Medicare during our study period (older than 61 in 2010). The final cohort includes 64,165 individuals.

I compare outcomes in this group to a control group of state employees from six Eastern states over the same time period with stable benefit designs from the Truven MarketScan database, with the same exclusion criteria (n=215,314). A full explanation of the cohort creation for the intervention population and control sample is in Figure 2-2.

After exclusions, I use all state health plan enrollees regardless of participation in the program. I did this for several reasons. First, I am interested in the population-wide effect of the program on service use. Second, as a practical matter, it is impossible to identify those members of the control group who would not have participated in the program if it had been offered in their plan. While these inclusions mean that I interpret results as intent-to-treat estimates, in practice, as more than 95% of the population remained in the program after two years, these statistics are likely similar in magnitude to an average treatment on the treated effect.

Outcome Measures

Using procedure and diagnostic claims codes, I define a set of 13 low-value services using 2011 grade D recommendations from the USPSTF, guidelines from the Choosing Wisely initiative, and previous literature (Table 2-1) (Colla et al. 2015; Charlesworth et al. 2016; US Preventive Services Task Force 2010). I chose services commonly delivered in a primary care setting or as a direct referral from it. I include four common preventive screenings for asymptomatic individuals (vitamin D screening, prostate-specific antigen (PSA) screening, cervical cancer screening in women under 21, and cardiac screening in low-risk, asymptomatic individuals). I additionally included nine low-value services done in response to a symptomatic concern, primarily imaging studies. For each of these low-value services, I use published coding methods to exclude cases in which the service are deemed clinically appropriate. (Charlesworth et al. 2016; Schwartz et al. 2014; Colla et al. 2015) (For technical coding details see Table 2-2).

After data were collected, federal guidelines for one low-value service, prostate-specific antigen (PSA) screening for prostate cancer, were updated from a grade D to a grade C, which removed them from low-value, instead reflecting that men should make a personal decision in consultation with their physicians. I left it in the low-value category for this study, despite the recent update, as the recommendations against screening were in effect for our entire study period.

The primary dependent variable is the use of any low-value service in the first and second years after program implementation. Additionally, I split low-value services into low-value screenings, done on asymptomatic individuals, and low-value tests for specific symptomatic concerns. I create this split for a couple of reasons. First, much of the existing evidence on low-value care in the commercial population has focused on low-value treatments for symptomatic concerns (e.g. imaging for uncomplicated back pain) and I wanted to generate estimates in this

population that could be compared to others. Second, I am looking for changes in services concurrent with increases in high-value preventive screenings done on asymptomatic individuals. Thus, low-value screenings for asymptomatic individuals are more closely related, and perhaps may be differentially impacted, compared with those done for symptomatic concerns.

For comparison, I also generate a measure of the likelihood of recommended, high-value service use. These are defined by the program and have been previously measured (Hirth et al. 2016). The set of high-value services includes the four preventive disease screenings required at regular intervals to remain compliant in the program (colon cancer screening for individuals over 50, cervical cancer screening for women older than 21, mammography beginning at 35, lipid screening at age-specific intervals). I note that these screenings approximate, but deviate slightly from, USPSTF recommendations. Nevertheless, I define leave them in the high-value category in this case because they are quite close to USPSTF recommendations, and because the program defined them as such. I also measure preventive office visits, required once every two years for an average member of our population, and separate these visits from other high-value services to study their relationship to other services. Preventive office visits are an important component of the program, both because they are a required service and because they are a referral point for other required and non-required services.

Analysis

I use a difference-in-differences research design to measure the change in both high-value and low-value service use after the implementation of the intervention state's program relative to changes in the control group with stable benefit designs. Following the theoretical model

discussed above, I assess change in Q_1 as likelihood of getting a specific type of service in a year, based on the opportunity to enroll in program, which changes the price of high-value services (P_2.) Other factors that may change likelihood of service use are controlled using a claims from the same time period from a group of employees from other states as well as demographic and health status controls.

I estimate the likelihood of service use with the following regression equation:

$$Pr(Y_{it} = 1) = f(\alpha + \beta_1(Tx_i) + \beta_2(Post_{t=1}) + \beta_3(Tx_i \bullet Post_{t=1}) + \beta_4(Post_{t=2}) +$$

$$\beta_5(Tx_i \bullet Post_{t=2}) + \beta_6 X_{it} + \epsilon_{it})$$

In this equation, Y_{it} is the outcome, likelihood of a specific type of service, for each enrollee in each year. It includes 3 time periods, the pre period, post year 1 and post year 2. In keeping with common difference-in-differences regression models, Tx_i is an indicator for whether an individual is eligible to enroll in the state's new benefit program; $Post_{t=1}$ is an indicator for whether the observation is in the first post-period year and $Post_{t=2}$ is an indicator for whether the observation is in the second post period year. The main effects are measured by the interactions of these dummies with the parameters β_3 and β_5. Notably, these parameters are independent of each other, such that they both measure the effect relative to the pre-period. The main specifications use linear probability models ($f = 1$) though I use the probit function in sensitivity analyses, ($f = \Phi$). All models are estimated with robust standard errors.

Relating the equation to the conceptual model, I include P_2 through the treatment indicator and changes to P_2 are measured by the interaction of the treatment indicator with the indicator variables for each year after the intervention. P_1 is not included in the equation because the intervention did not affect these prices. The equation also includes, X_{it} a vector of demographic

characteristics (age, gender, dependent status). Last, B(h), demand for medical care due to health status is included in two ways. First, I calculate a Charlson score for each enrollee, a measure of comorbidities and include it in the equation. Second, in constructing the measures for low-value services, I excluded diagnoses in which the service is no longer considered low-value. For example, cardiac imaging without indication is a low-value service but there are a number of indications (e.g. previous diagnosis of heart disease, chest trauma) that warrant the imaging. These diagnoses are excluded in all cases of the service, not just at baseline. To the extent that the intervention increased medical care and diagnosed medical conditions, excluding those new diagnoses for which low-value services should be used mitigates the extent to which the value of a service may have increased in the population after the intervention.

I perform sensitivity analyses using the same dependent and explanatory variables using maximum likelihood estimation with a probit function. As in the main equation, the key variables of interest in these models are the interaction terms between post year and an indicator for the intervention group, which estimates the change in the treatment group relative to trends over the same time period in the control group. All models control for baseline differences between groups, year trends in utilization, age (in approximately 10-year categories), gender and dependent status.

Finally, to explore a potential mechanism behind observed changes in service use, I decompose the main results into the portion associated with changes in receipt of preventive office visits and the portion associated with an increase in high- or low-value service use conditional on a preventive office visit. This decomposition is performed by computing counterfactual changes in high- and low-value service use that would have occurred in the post-intervention years if the percent of the population using a high- or low-value service conditional on a preventive visit had

20

remained unchanged from baseline levels. I compare the counterfactual change to the actual change to compute the share attributable to changes in the prevalence of preventive visits.

Internal Validity: Pre-Period Trends and Matching

An advantage of difference-in-differences research designs is the ability to control not only for baseline differences, but for differences over time. However, it assumes that, absent treatment, those differences are constant over time. This parallel trends assumption is key to the identification strategy; differing trends in outcomes between the treatment and control prior to treatment will bias estimates. Ideally, I would examine at least several years of pre-period data; in this case I have only one year. I thus examine trends in high- and low-value service use in the months prior to the intervention. Figure 2-3 shows the monthly population prevalence and de-meaned prevalence of high- and low-value service use in the pre-period. (For the de-meaned prevalence, the overall mean for each group subtracted from the monthly prevalence.) As shown in these graphs, the HEP population has consistently higher baseline utilization of both high- and low-value services than the control sample, with similar, slightly increasing trends. For high-value services, the de-meaned lines nearly overlap, suggesting very similar trends. For low-value services, the de-meaned trend line is slightly steeper for the control sample; if that trend continued through the post-period it would bias these estimates toward the null. Formal tests of the parallel trends assumption follow previous literature to assess differences in outcome trends in the pre-intervention period (Table 2-3) (Ryan 2009; Dimick and Ryan 2014). In these tests, the magnitude of the interaction term between the time trend and the treatment group is less than 1 percentage point in all cases and most are not statistically significant.

The fact remains that the pre-period data are inadequate for a robust test of pre-period trends. Further, there are some clear differences in levels of pre-period differences, namely a higher percent of the intervention population has a co-morbid condition, is male, and has a pre-period office visit (Table 2-4). These baseline differences in levels raise potential concerns about violation of the parallel trends assumption, namely that certain characteristics may influence selection into the treated group and, thus, bias estimates. Matching on observable characteristics so that pre-intervention levels of important variables are similar between the treated and control group can ameliorate this concern (Ryan, Burgess, and Dimick 2015). I matched the eligible intervention population (n=64,165) to the eligible sample group (n=215,314) using 1-to-1 nearest neighbor matching without replacement, a caliper of 0.1 and enforcing common support to exclude those treatment observations with a propensity score outside the total distribution of the control group. (This last restriction turns out to be non-binding.) I matched treated to control observations on gender, Charlson comorbidity score, age and whether the person had a preventive diagnostic visit in the pre-period. The final matched sample includes the entire treatment group, and 64,110 individuals from the control group. Matching was performed using psmatch2, a user-written command in Stata (Leuven and Sianesi 2003). Descriptive statistics for the matched sample are in Table 2-5.

Sample Description

The median age at baseline is about 42 for both the intervention and control groups, and the control group has a slightly higher percentage of females (Table 2-4). The intervention population has a higher percentage of enrollees with a positive Charlson score, a measure of comorbidities, as well as slightly more enrollees with a chronic condition. A high percent of

both groups have a medical claim in the baseline year, and the intervention population is more likely to have a claim.

At baseline, enrollees in the intervention group are more likely to receive a preventive office visit as well as any high-value or low-value service (Table 2-6). About one third of intervention state enrollees received at least one of the 13 measured low-value services in the baseline period, compared with about one fourth of those in the control group. For both groups, the likelihood of receiving a high-value service was higher; about 60% in the intervention group and 55% in the control group. In the baseline year, about half of enrollees in intervention who had a preventive office visit also received a low-value service compared with one-third of enrollees in the control group.

RESULTS

Changes in High-Value and Low-Value Care

In the first and second years after the intervention, the likelihood of receiving a low-value service increased. The percent of intervention state enrollees versus the control group receiving a low-value service increased by 7.9 percentage points in year 1, a 24% relative change, and 2.7 percentage points in year 2, an 8% relative change, compared with to baseline (Table 2-6). The likelihood of receiving a recommended, high-value service increased by 11.0 percentage points in the first year and 5.2 percentage points in year 2. Figures 2-4 and 2-5 show these results graphically and underscore that changes in low-value services occur coincident to changes in high-value services.

23

There are distinct patterns in changes in types of low-value services used. The increase in low-value services is entirely driven by increases in the likelihood of a low-value primary preventive service (vitamin D screening, PSA testing, cardiac imaging without indication and cervical cancer screening for women under 21). The percent of the population that receives a low-value primary preventive service increases by 8.8 percentage points, a relative 31%, in the first year and by 3.1 percentage points, relative 11%, in the second year compared with baseline (Table 2-6). By contrast, low-value testing for symptomatic concerns actually declines slightly (0.5 percentage points) in each year. The magnitudes of these declines are much lower, as is the relative decline, about 6%. Full coefficients and statistics for each regression are in Table 2-7. These results are also shown graphically in Figures 2-6 and 2-7.

The change in low-value service use varied by type of service, with PSA tests and cardiac imaging for low-risk asymptomatic individuals increasing most in year 1, and vitamin D tests and cardiac imaging increasing most in year 2. A full set of estimates for every service measured is shown in Table 2-8.

Changes in Care Associated with Preventive Visits and Other High-Value Care

To further explore the mechanism behind increases in low-value care, I measure whether medical encounters are more likely to include high- and low-value services after the intervention. Figure 2-8 shows that, after the intervention, the percent of people receiving either just a required high-value service (including a preventive visit) increased to nearly 35% of the intervention state population in the third quarter after the intervention. The percent of the population receiving both a high- and low-value service also increased, with rates mirroring those for high-value only

services. The percent of the population receiving only a low-value service, however, remained

flat suggesting the increase in these services is driving by instances in which they are delivered

in conjunction with high-value services, including preventive office visits.

In the decomposition of the change in service use attributable to the increase in preventive office

visits, I find that more than half of the increase in high- and low-value service use is associated

with an increase in preventive visits (Table 2-9). Using counterfactual rates of the increase in

high- and low-value service use that assume the likelihood of preventive visits increased and the

rate of high- and low-value service use conditional on a preventive visit remains constant, I find

an 8.1 ppt increase in likelihood of high-value service use in year 1 and a 2.9 ppt increase in year

2, representing 74% and 56% of the total increase respectively. The rate of low-value service

use increases 4.5 ppts in year 1 and 1.6 ppts in year 2, representing 57% and 56% of the total

increase.

Sensitivity Analyses

Results using a probit function were similar to the main OLS model (Table 2-10). Compared

with the control group, the probit model estimated a 12.0 percentage point increase in use of

high-value services in the first year relative to baseline, and a 5.5 percentage point increase in the

second year. The model estimated a 6.7 percentage point increase in likelihood of low-value

services in the first year and a 2.2 percentage point increase in likelihood of low-value service

use. Other outcomes are similar to main results.

For the matched sample, results for high- and low-value services align with the main results

(Table 2-11). Use of a high-value service increases by 11.2 percentage points in the first year

and by 4.8 percentage points in the second year. Use of a low-value service increases by 7.9

percentage points in the first year and 1.9 percentage points in the second year. As with the main

results, the effect appears driven by an increase in the use of low-value preventive services, with

slight decreases in use of low-value tests for specific symptoms.

Elasticity Estimates

Elasticity of demand was measured by taking average total price of the bundle of recommended

services for each gender/age cohort. To calculate this number, I sum across all enrollees in the

treatment group *(i)* (n=64,165) to find the average price for each high-value preventive services

(j) (n=6) for each year *(t)*:

$$AvPrice_{jt} = \frac{\sum_{i=1}^{n} OOPpayments_{ijt}}{\sum_{i=1}^{n} ServicesUsed_{ijt}}$$

I then divided the population up into 8 gender/age categories based on services and service

frequency required in the program. For example, a female between the ages of 40 and 49 is

required to have a preventive visit every 2 years, a cholesterol screen every 2 years, and a

cervical cancer screen every 3 years. Average annual high-value service price for each

demographic group *(k)* in year *(t)* is the sum of the expected annual costs:

$$OOP\_Price_{kt} = \sum_{j=1}^{6} (\frac{1}{YearsBetween\ Service}) AvPrice_{jt}, \text{for all } i \text{ in category } k$$

Then, average price for the intervention population is calculated by taking the weighted average

of each group's cost in each year using the intervention population prevalence as the weight.

$$OOP\_Price_{t} = \sum_{k=1}^{8} (PopulationPrevalence_{k}) (OOP_{Price\ kt})$$

As a final step, the own- and cross-price elasticities are calculated by taking the change in the quantity, measured by the predicted prevalence of the population receiving at least one high- or low-value service annually, compared with the change in out-of-pocket price for high-value services.

$$Elasticity_t = \frac{\frac{ServiceUse_t - ServiceUse_{t=1}}{ServiceUse_{t=1}}}{\frac{OOP\_Price_t - OOP\_Price_{t=1}}{OOP\_Price_{t=1}}}$$

Table 2-12 shows the average amount individuals paid for all high-value services is relatively small at baseline $2.10, reflecting that most of these services were likely covered at 100% prior to program implementation. The cost decreased to an average of $0.40 in the first year and to $0.57 in the second year. Elasticity estimates were relatively similar for both high- and low-value services. In the first year, low-value service elasticity is -0.29; high-value service elasticity is -0.24. In the second year of the program, both high-and low-value services had a measured elasticity of -0.11 reflecting that the relative change in use of both services was equivalent.

Before moving on, it's important to offer a word of caution about over-interpreting these estimates. I estimated high-and low-value service elasticity to show a comparison of relative changes in both services and the most important takeaway is the estimates comparison with each other. These estimates are not indicative of long-run changes in service use for several reasons. First, estimates in the first year may represent pent up demand for preventive services as people utilize those services for which they were not in compliance with the program's requirements. Second, consistent with our main analysis, these elasticities are from the entire population—not just those enrolled in the state's new benefit program—and thus may be different from those for the treated population. Finally, given the small absolute changes in out-of-pocket service costs,

it's unclear if these price changes are driving behavior. It is possible, perhaps even likely, that the program requirements drive behavioral change more than the reduction in service prices. Thus, the high- and low-value elasticity estimates should be compared with each other in this setting, and may not generalize into other settings where cost sharing is reduced for specific services.

CONCLUSION

This paper finds that an intervention intended to increase use of recommended, high-value preventive services unintentionally increased use of low-value medical services. Coincident with increases in high-value services that include preventive office visits and recommended cancer screenings, I find increases in low-value services including cardiac screening for low-risk, asymptomatic individuals, vitamin D tests and prostate-specific antigen screening. I estimate similar own- and cross-price elasticities between both types of services, suggesting that high- and low-value preventive services are complements. This result calls into question the ability of consumers to make nuanced discernments about service value, at least in the context of enhanced benefits and greater use of preventive care.

Additionally, I find that the majority of the increase in high- and low-value service use is due to an increase in preventive care visits. Preventive office visits are an important referral point, and can lead to additional service use. It is possible that, in this intervention, promoting use of preventive office visits exacerbated spillover into low-value services. Further studies should examine the mechanisms underlying the concurrent increases in high- and low-value care, including the role of preventive visits in encouraging both.

While these results showed an increase in low-value services in both program years, utilization increases were stronger in the first compared to the second year. Most of the measured high-value services and preventive office visits are not recommended every year (e.g. Pap tests, colonoscopies), so these results could reflect a natural wane as people completed their requirements and then were exempt in the second year.

These findings are largely consistent with – but directionally opposite of– analogous research showing that increases in consumer cost sharing for all but evidence-based primary preventive services cause reductions in use of both high- and low-value services (Brot-Goldberg et al. 2017a; Zheng et al. 2016). The results also align with cross-sectional research reporting that patients who have more contact with the medical system receive a higher number of low-value services per patient (McWilliams and Schwartz 2017). Finally, these results add to literature showing little discernment among providers or consumers about the value of medical services (Baicker, Schwartzstein, and Mullainathan 2015; Baicker and Levy 2015).

This study has several limitations. First, as noted above in the sensitivity analyses, the research design allows us to control for baseline differences between groups, and differences that remain constant over time. However, as with all difference-in-differences designs, it relies on the assumption that utilization trends unrelated to the treatment are parallel between groups. Given that there is only have one-year of pre-period data and that many preventive services are done annually or less, the ability to look for differences in pre-period trends is limited. A second limitation is that our estimates relate to a single group of employees in one state. Thus, the estimates may have limited generalizability to other populations or other types of interventions. In extrapolating to other populations, I note that this population is similar to our control in baseline demographic characteristics. Extrapolating to other types of interventions may be more

difficult. As preventive care was required to remain on the health plan with lower cost sharing, this intervention had a stronger nudge toward such care than interventions that rely solely on reduced cost sharing. However, the care delivery model is not different in this state than in other states and there is no reason to think that other interventions that successfully induce demand for preventive care would not see similar directional effects, even if the magnitudes differ. Third, the Affordable Care Act was signed into law in March 2010 just before the study period begins (July 2010). One provision of that law eliminated cost sharing for high-value preventive services for all new health plans, including those targeted by the intervention and potentially our control group plans depending on their status. That may mean that the increase in high-value preventive services, and potentially complementary effects to low-value services, reflect the role of participation requirements rather than of a reduction in cost sharing, though I cannot directly separate these effects. Last, the use of administrative medical claims to classify high- and low-value services is imprecise. While this method has advantages over other possible methods, including the ability to easily look at a large population, claims are not comprehensive enough to capture every nuance of a medical interaction, and there may be variation in coding among providers. However, I do not think there are systematic differences that would bias estimates of changes in prevalence of high- and low-value services between our treatment and control samples over time and classical measurement error of this type on the dependent variable does not change estimates.

While this was one specific intervention, the results may have implications for other contexts. Many interventions induce demand for preventive care and specifically preventive visits, including through health insurance cost-sharing structures, direct primary care initiatives, and alternative payment contracts tied to quality metrics. This study suggests that high-value

services may be bundled with low-value services, and that these efforts could increase wasteful spending. Health system leaders and policymakers should consider designing programs that would reduce unwanted consequences, either through more precise targeting of specific high-value services or, preferably, by including deterrents to low-value care in benefit design and provider incentives. Additionally, assessment efforts should evaluate the impacts of consumer-focused interventions to change the service utilization on both high- and low-value services, regardless of how incentives are targeted.

Table 2-1: Low-value service measures and rationale

SERVICE	Source and Recommendation	Notes
Vitamin D Screen	Choosing Wisely "Don't routinely measure 1,25-dihydroxyvitamin D unless the patient has hypercalcemia or decreased kidney function."	
Prostate-specific antigen screen	USPSTF; "The USPSTF recommends against prostate-specific antigen-based screening for prostate cancer	Guidelines updating PSA screening were released in Fall 2011, changing the grade of recommendation from I to D. It was updated, and the recommendation changed from a D to C, in 2017
Cervical Cancer Screen in women under 21	USPSTF "The USPSTF recommends against screening for cervical cancer in women younger than age 21 years"	
Cardiac screening in asymptomatic individuals	USPSTF, Choosing Wisely, American College of Cardiology; "Don't order annual electrocardiograms (EKGs) or any other cardiac screening for low-risk patients without symptoms."	
Imaging for uncomplicated headache	Choosing Wisely, American College of Radiology "Don't do imaging for uncomplicated headache."	
Imaging for uncomplicated low back pain	Choosing Wisely, American College of Physicians; "Don't obtain imaging studies in patients with nonspecific low back pain"	
Electroencephalogram for Headaches	Choosing Wisely, American Academy of Neurology "Don't perform electroencephalography (EEG) for headaches."	
CT scan for acute uncomplicated rhinosinusitis	Choosing Wisely, American Academy of Allergy, Asthma and Immunology; "Don't order sinus computed tomography (CT) or indiscriminately prescribe antibiotics for uncomplicated acute rhinosinusitis".	
Imaging for plantar fasciitis	Choosing Wisely, American College of Occupational and Environmental Medicine, "X-ray is not recommended for routine evaluations for plantar fasciitis."	
Head imaging for syncope	Choosing Wisely, American College of Physicians. "In the evaluation of simple syncope and a normal neurological examination, don't obtain brain imaging studies (CT or MRI)."	
Abdominal CT combined studies	Centers for Medicare and Medicaid Services, National Quality Forum; "The evidence base indicates that a CT Thorax scan be performed either without or with contrast but not both.'	
Simultaneous brain and sinus CT	Agency for Healthcare Research and Quality, National Quality Measures Clearinghouse; "simultaneous CT sinus and brain imaging for headache without suspected complications is generally considered inappropriate"	
Thorax CT combined studies	Centers for Medicare and Medicaid Services, National Quality Forum "The indiscriminate use of combined Thorax CT studies defined as those that are performed both without and with contrast agents for the evaluation of solid organs and body cavities represents a serious inefficiency of practice and a patient safety issue."	
Sources: U.S. Preventive Services Task Force (uspreventiveservicestaskforce.org); Choosing Wisely (choosingwisely.org); Centers for Medicare and Medicaid Services (cms.gov); Agency for Healthcare Research and Quality (ahrq.gov); Charlesworth CJ, Meath TH, Schwartz AL, McConnell KJ. Comparison of low-value care in Medicaid vs commercially insured populations. JAMA internal			

medicine. 2016 Jul 1;176(7):998-1004; Colla CH, Sequist TD, Rosenthal MB, Schpero WL, Gottlieb DJ, Morden NE. Use of non-indicated cardiac testing in low-risk patients: Choosing Wisely. BMJ Qual Saf. 2014 Aug 5.

Table 2-2: Coding specifications for high- and low-value services

Condition	Source of Recommendation/ Technical Specification	Inclusion Criteria	Exclusion Criteria	Notes
High Value Services				
Colonoscopy/F ecal occult blood testing	Program criteria/ American Gastroenterologi cal Association; AAPC	Colonoscopy: (CPT/HCPCS: G0105, G0121, G0120, 45378, 45380, 45383, 45384, 45385) Fecal occult blood test: (CPT/HCPCS) G0328, 82270, 82274	Males and females, < 50 years	Criteria for inclusion defined through the state's program, and are approximately (though not exactly) in line with 2011 U.S. Preventive Services Task Force A and B recommendatio ns.
Mammography	Program criteria/HEDIS	CPT/HCPCS: G0202, 77052, 77057	Males, Females < 35 years	
Cervical Cancer Screen	Program criteria/HEDIS	CPT/HCPCS: G0123, G0143, G0144, G0145, G0147, G0148, P3000, P3001, 88141, 88142, 88143, 88147, 88148, 88152, 88154, 88166, 88167, 88174, 88175	Males, Females >= 21 years	
Lipid testing	Program criteria/ CMS	CPT/HCPCS: 80061, 82465, 83700, 83701, 83704, 83716, 83718, 83719, 83721, 84478, G8725, G8767	None	
Low Value Services				
Vitamin D Screen	Blue Cross Blue Shield Excellus codes for appropriate use	CPT/HCPCS: 82306, 82652	Diagnosis of (ICD-9): parathyroid gland disorder (252.00 – 252.1); rickets (268.0);	

	of vitamin D screening		osteomalacia (268.2); unspecified vitamin D deficiency (268.9) disorder of phosphorus metabolism (275.3); hypocalcemia (275.41); hypercalcemia (275.42); chronic kidney disease (585.3 – 585.6); secondary hyperparathyroidism (588.81); osteoporosis, unspecified, idiopathic, disuse or other (733.00, 733.02, 733.03, 733.09) disorder of bone and cartilage, unspecified (733.90)	
Prostate-specific antigen screen	Schwartz; USPSTF for exclusion criteria applicable to commercial population	CPT/HCPCS: G0103; 84152-84154 with diagnosis V700 (adult exam without abnormal findings) or diagnosis V7644 (screening for malignancies in prostate)	ICD-9 codes: 185 (diagnosis of prostate cancer); V1046 (reported history of prostate cancer)	
Cervical Cancer Screen	USPSTF /HEDIS	CPT/HCPCS: G0123, G0143, G0144, G0145, G0147, G0148, P3000, P3001, 88141, 88142, 88143, 88147, 88148, 88152, 88154, 88166, 88167, 88174, 88175	Designation of low-value service applies to females < 21 years	
Cardiac screening in asymptomatic individuals	Colla (2014)	Advanced Cardiac imaging (CPT/HCPCS: 0144T, 0145T, 0146T, 0147T, 0148T, 0149T, 0150T, 75552-75559, 75561,	Excluded any procedures in individuals at high risk and, in low risk individuals, any procedures associated with cardiac or other situations that might	

| | | 75562, 75565, 75571-75574, 78451-78454, 78459, 78460, 78461, 78464, 78465, 78478, 78480, 78481, 78483, 78491, 78492, 78494, 78496, 78499; Stress test: 75560, 75563, 75564, 93015 – 93018, 93024, 93350 – 933052; Echocardiogram: 93303, 93304, 93306, 93307, 93308, 93320, 93321, 93325; Electrocardiogram: 3120F, 93000, 93005, 93010 | justify the test. Exclusions in low-risk individuals for individual tests examined diagnosis codes in first 5 diagnosis fields and follow Colla (2014) including cardiac disease, malignant hypertension and injuries. Following Colla (2014), we define as high risk any individual with at least 2 dates of service within a plan year with primary, secondary or tertiary diagnosis codes related to following Hierarchical Condition Categories: 1(HIV/AIDS); 7-10 (Cancer); 15-19 (Diabetes, over age 40 only); 52 (Drug/Alcohol Dependence); 79 (Cardio-Respiratory Failure and Shock); 80 (Congestive Heart Failure); 81(Acute Myocardial Infarction); 82 (Unstable Angina or Other Acute Ischemic Heart Disease); 83(Angina Pectoris/Old Myocardial Infarction); 92(Specified Heart Arrhythmias); 96 (Ischemic or Unspecified Stroke); 100 (Hemiplegia/Hemiparesis); 104 (Vascular Disease with Complications); 105 (Vascular Disease); 108(Chronic Obstructive Pulmonary | |

			Disease) 113 (Renal Failure)	
Imaging for uncomplicated headache	Schwartz	CPT: 70551-70553, 70450, 70460, 70470; with diagnosis of chronic headache (ICD-9: 784.0, 307.81, 346.xx, 339.xx)	Diagnosis of (ICD-9): cancer (230.xx – 239.xx, 140.xx – 209.xx), cerebrovascular disease (430.xx – 439.xx), head or face trauma (850.xx – 854.xx, 800.xx-804.xx, 870.xx - 873.xx, 95901, 910.xx, 920.xx, 921.xx), post-traumatic or thunderclap headache (: 33920 – 33921, 33943), migraine with hemiplegia or infarction (346.3x, 346.6x), giant cell arteritis (ICD-9 4465). epilepsy or convulsions (345.xx, 780.3x), altered mental status, nervous and musculoskeletal system symptoms (ICD-9 78097, 781.xx, 7820, 7845, 79953), personal history of stroke, TIA or cancer (ICD-9 V1254, V1000-V1091)	
Imaging for uncomplicated low back pain	Schwartz	CPT: 72158, 72148, 72010, 72020, 72052, 72100, 72110, 72114, 72120, 72131, 72132, 72133, 72141, 72142, 72146, 72147, 72149, 72156, 72200; with diagnosis of low back pain from ICD-9: (7213, 72190, 72210, 72252, 7226, 72293, 72402, 7242,	Diagnosis of (ICD-9): cancer (140.xx – 209.xx, 230.xx-239.xx), trauma (800.xx – 839.xx, 850.xx-854.xx, 860.xx-869.xx, 905.xx – 909.xx, 929.xx, 952.xx, 958.xx, 959.xx, 92611, 92612), IV drug abuse (304.0x, 304.1x, 304.2x, 304.4x, 305.4x, 305.5x, 305.6x, 305.7x) neurologic impairment (34460, 7292), endocarditis	

		7243, 7244, 7245, 7246, 72470, 72471, 72479, 7385, 7393, 7394, 846.x, 8472)	(4210, 4211, 4219), septicemia (038.xx), tuberculosis (010.xx), osteomyelitis (730.xx), fever, weight loss, malaise, anemia not due to blood loss (780.6x, 783.0, 783.21, 783.22, 285.9) myelopathy, neuritis and radiculopathy (72142, 72191, 72270, 72273, 7244)	
Electroencephalogram for Headaches	Charlesworth	CPT: 95812 95813 95816 95819 95822 95827 95830 95957; ICD-9 codes: 30781, 339.xx, 346.xx, 7840	ICD-9 codes: 345.xx, 78030-78039; 7810 (epilepsy or convulsions)	
CT scan for acute uncomplicated rhinosinusitis	Charlesworth, Schwartz	CPT: 70486 – 70488; ICD-9: 461.xx, 473.xx (acute sinusitis)	ICD-9: 27700-27709, 042.xx, 07953, 279.xx (immune disorders), 373.xx, 37600 (eyelid/orbit inflammation), 800.xx-804.xx, 850.xx-854.xx, 870.xx-873.xx, 95900-95909, 910.xx, 920.xx-921.xx (head or face trauma)	
Imaging for plantar fasciitis	Charlesworth	CPT:73620 73630 73650 73718 73719 73720 76880 76881 76882	First foot imaging claim occurring within 14 days of first diagnosis of plantar fasciitis (ICD-9 72871, 7294)	
Head imaging for syncope	Charlesworth	CPT: 70551-70553, 70450, 70460, 70470 with ICD-9: 7802, 9921	345.xx, 78030-78039 (epilepsy or convulsions);, 430.xx-438.xx (cerebrovascular diseases); 800.xx-804.xx , 850.xx-854.xx, 870.xx-873.xx, 95900-95909, 910.xx, 920.xx-921.xx (head or face trauma), 78097, 7810, 781.xx, 7820,	

			78450-78459 (altered mental status, nervous or musculoskeletal system symptoms, V1254 (personal history or stroke or TIA)	
Abdominal CT combined studies	Charlesworth	CPT: 74170, 74178	ICD-9: ICD-9: 5939 (kidney unspecified); 5990, 5999, 595.xx, 597.xx (urinary disease); 1200, 59970, 59971, 59972 (hematuria); 2508x, 2510, 2511, 2512, 2703, 5770, 5771 (pancreatic); 2559, 194.xx, 237.xx, 227.xx (adrenal mass); 1550, 1551, 1552, 1977, 20972, 2115, 2308, 2353 (liver lesion); 188x, 2337 (bladder neoplasm); 1570-1579, 1890, 2116, 2117, 2230 (pancreas neoplasm); 863.xx-- 869.xx, 902.xx, 929.x (blunt abdominal trauma); 59010-59019, 59080-59089, 5909 (kidney infection); 7824 (jaundice)	
Simultaneous brain and sinus CT	Charlesworth	CPT: Brain CT: 70450, 70460, 70470 on same service data as sinus CT: 70486-70488	ICD-9: 140-239 (neoplasms), 800.xx-839.xx,850.xx-854.xx,860.xx-869.xx,905.xx-909.xx, 926.11, 926.12, 929.xx, 952.xx, 958.xx-959.xx (trauma), 376.01 (orbital cellulitis), 324.0 (intracranial abcess)	
Thorax CT combined studies	Charlesworth	CPT: Thorax with and without contrast: 71270	ICD-9: 860.xx-869.xx (chest, abdomen, pelvis injury), 901.xx-902.xx (injury to blood vessels), 926.xx, 929.xx (crush injury)	

References for Coding Specifications

Charlesworth CJ, Meath TH, Schwartz AL, McConnell KJ. Comparison of low-value care in Medicaid vs commercially insured populations. JAMA internal medicine. 2016 Jul 1;176(7):998-1004.

Colla CH, Sequist TD, Rosenthal MB, Schpero WL, Gottlieb DJ, Morden NE. Use of non-indicated cardiac testing in low-risk patients: Choosing Wisely. BMJ Qual Saf. 2014 Aug 5:bmjqs-2014.

Hirth RA, Cliff EQ, Gibson TB, McKellar MR, Fendrick AM. Connecticut's value-based insurance plan increased the use of targeted services and medication adherence. Health Affairs. 2016 Apr 1; 35(4): 637-46.

Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM. Measuring low-value care in Medicare. JAMA internal medicine. 2014 Jul 1;174(7):1067-76.

Vitamin D screening guidelines from Excellus Medical Policy:
https://www.excellusbcbs.com/wps/wcm/connect/e4b9a340-9564-4284-ae52-80f5d664b15d/acupunc+mpc3+18.pdf?MOD=AJPERES&CACHEID=e4b9a340-9564-4284-ae52-80f5d664b15d

Table 2-3: Estimates from pre-intervention trends test

	Linear time trend estimates (ppt change at month level)	Interaction between HEP and linear time trend (ppt change at month level)
High-value service	0.11	-0.015
Any low-value service	0.0038	-0.012
Preventive office visit	0.22	-0.14**
Notes: These estimates result from a model that tests whether pre-intervention trends differ between the treatment and control groups. To do so, we regress each outcome in the left-most column on a linear time trend (observed in months), a group main effect (HEP/control) and an interaction between the two. All models use data from baseline year (pre-intervention) only. *p<0.05 **p<0.01		

Table 2-4: Descriptive characteristics of the sample

	Intervention Population	Control Sample	P-value of t-test of differences between groups
Female (%)	53.8	57.8	<0.001
Age (Mean)	41.6	42.2	<0.001
Dependent (%)	41.0	30.9	<0.001
Charlson Score = 1 (%)	14.8	11.4	<0.001
Charlson Score = 2 (%)	6.0	4.5	<0.001
Chronic Condition (%)	42.3	39.6	<0.001
Enrollees with any claim (%)	93.3	88.7	<0.001
Preventive office visit	53.0	36.9	<0.001

Notes: Baseline is the plan year before the implementation of the state's new health benefit program , July 2010–June 2011. Chronic conditions measured include diabetes, heart disease, COPD, asthma, hypertension, and hyperlipidemia. The p-value refers to the probability that the value for the treatment group is equal to that of the comparison group using a t-test**.**

Table 2-5: Matching analysis results—descriptive statistics

	Intervention Population	Control	p value from ttest sample overlap
Female	54%	54%	0.669
Age	41.6	41.6	0.598
Dependent	41%	34%	<0.001
Charlson Score = 1	15%	15%	0.361
Charlson Score = 2	6%	6%	0.891
Preventive Diagnostic Visit	53%	53%	0.889
OOP Preventive Diagnostic Visit	$ 2.65	$ 6.53	<0.001
OOP Office Visit	$ 51.68	$ 108.85	<0.001
OOP High Value Services	$ 0.04	$ 8.68	<0.001
OOP Low Value Services	$ 0.20	$ 7.36	<0.001
Number of Observations	64,165	64,110	

Notes: Baseline is the plan year before the implementation of the state's new health benefit program, July 2010–June 2011. Control group is matched to treatment group using one-to-one nearest neighbor matching on age, gender, Charlson score and pre-period preventive office visit use. The p-value refers to the probability that the value for the treatment group is equal to that of the comparison group using a t-test.

Table 2-6: Rates and regression-based estimates of changes in high- and low-value services in baseline and post-period years

Percent of Intervention Population and Control Sample Using Following Services	Baseline Intervention Population (%)	Baseline Control Sample (%)	Percentage Point Change: Year 1	Percentage Point Change: Year 2
Preventive Office Visits	53.0	36.9	13.5**	4.8**
High-Value Service	59.8	55.0	11.0**	5.2**
Low-Value Service	33.6	23.4	7.9**	2.7**
Low-Value: Primary Preventive Services	28.8	18.5	8.8**	3.1**
Low-Value: Referral for Symptomatic Concerns	7.8	6.8	-0.5**	-0.5**
Notes: Percentage point change are from a linear probability model with service use as the dependent variable and the interaction between post period and treatment group estimating the effect of interest. All models controlled for age (in approximately 10-year bands), gender, dependent status, Charlson comorbidity score, plan year and baseline differences between treatment and control group in outcome measures. Statistical significance refers to the p value at a level that allows us to reject the null hypothesis of no difference in the intervention population compared with the comparison group. **p<=0.01 *p<=0.05				

Table 2-7: Difference-in-differences estimates and coefficients for main outcomes of interest

	Likelihood of preventive diagnostic visit	Likelihood of targeted high-value service	Likelihood of low-value service	Likelihood of low-value screening	Likelihood of low-value symptomatic referral
Intervention (baseline)	17.9% (0.0021)**	6.2% (0.0020)**	10.2% (0.0020)**	10.3% (0.0019)**	0.9% (0.0012)**
Post Year 1	1.7% (0.0014)**	0.3% (0.0014)*	0.1% (0.0013)	0.1% (0.0012)	0.1% (0.0008)**
Post Year 2	1.9% (0.0014)**	-0.2% (0.0014)	-0.5% (0.0013)**	-0.3% (0.0012)*	-0.2% (0.0008)**
Tx * Post Year 1	13.5% (0.0029)**	11.0% (0.0027)**	7.9% (0.0029)**	8.8% (0.0028)**	-0.5% (0.0017)**
Tx * Post Year 2	4.8% (0.003)**	5.2% (0.0028)**	2.7% (0.0029)**	3.1% (0.0027)**	-0.5% (0.0017)**
Female	31.0% (0.0010)**	26.4% (0.0010)**	-0.2% (0.0009)*	-1.1% (0.0009)**	1.3% (0.0006)**
Age: 30 to 39	5.3% (0.0017)**	16.2% (0.0017)**	7.8% (0.0014)**	6.4% (0.0012)**	2.3% (0.0009)**
Age: 40 to 49	7.2% (0.0016)**	25.4% (0.0016)**	16.2% (0.0014)**	14.9% (0.0012)**	3.1% (0.0008)**
Age: 50 and older	9.1% (0.0016)**	34.7% (0.0016)**	25.3% (0.0014)**	24.4% (0.0013)**	3.4% (0.0008)**
Dependent	-3.5% (0.0011)**	-5.8% (0.0011)**	-0.3% (0.0011)*	-0.4% (0.0010)**	0.2% (0.0006)**
Charlson Score = 1	-1.6% (0.0016)**	13.4% (0.0015)**	5.6% (0.0016)**	3.6% (0.0015)**	3.2% (0.0010)**
Charlson Score = 2	-2.3% (0.0024)**	15.1% (0.0021)**	6.0% (0.0025)**	2.5% (0.0024)**	5.3% (0.0017)**
Constant	14.1% (0.0017)	17.2% (0.0017)	8.1% (0.0015)	5.2% (0.0013)	2.9% (0.0009)
Number of observations	838437	838437	838437	838437.	838437
R squared	0.143	0.172	0.066	0.072	0.007
Fstatistic	14471.63	19662.20	5401.66	5772.83	462.90

Notes: Percentage point change are from a linear probability model with service use as the dependent variable and the interaction between post period and treatment group estimating the effect of interest. All models controlled for age (in 10-year bands), gender, dependent status, Charlson comorbidity score, plan year and baseline differences in outcome measures. Robust standard errors in parentheses. *p<0.05 **p<0.01

Table 2-8: Difference-in-differences estimates for individual high- and low-value services

	Low Value Screenings				Low value tests for specific symptoms					
	Vitamin D Screening	Prostate-Specific Antigen Screening in men	Cardiac Imaging without Indication	Pap test in women < 21	CT scan for uncomplicated headache	Low back imaging for uncomplicated low back pain	Imaging for plantar fasciitis	Head imaging for syncope	Abdominal CT combined studies	Simultaneous brain and sinus CT
Intervention (baseline)	6.1% (0.0014)**	3.7% (0.002)**	7.0% (0.0016)**	-1.0% (0.0067)	0.23% (0.0006)**	0.4% (0.0009)**	0.59% (0.0005)*	-0.051% (0.0002)**	-0.36% (0.0003)**	-0.033% (0.0001)*
Post Year 1	0.3% (0.0008)**	0.3% (0.0013)*	-0.2% (0.0009)	-3.5% (0.0053)**	0.054% (0.0004)	0.095% (0.0006)	0.00010% (0.0002)	0.015% (0.0001)	0.03% (0.0003)	-0.012% (0.0001)
Post Year 2	0.7% (0.0008)**	0.6% (0.0013)*	-1.1% (0.0009)**	-8.6% (0.0049)**	-0.0011% (0.0004)	-0.033% (0.0006)	-0.071% (0.0002)*	0.003% (0.0001)	-0.10% (0.0003)**	0.017% (0.0001)
Tx * Post Year 1	4.1% (0.0022)**	4.4% (0.0031)*	6.5% (0.0023)**	-0.4% (0.0093)	-0.073% (0.0008)	-0.28% (0.0013)*	0.10% (0.0007)	-0.012% (0.0003)	-0.17% (0.0004)**	0.007% (0.0002)
Tx * Post Year 2	1.8% (0.0021)**	0.17% (0.0030)	3.2% (0.0022)**	-0.3% (0.0084)	-0.20% (0.0008)*	-0.082% (0.0013)	0.068% (0.00065)	-0.019% (0.0003)	-0.16% (0.0004)**	-0.023% (0.0002)
Female	6.8% (0.0006)**	--	-0.5% (0.0007)**	--	0.84% (0.0003)**	0.26% (0.0004)	0.27% (0.0002)*	0.016% (0.0001)	0.08% (0.0002)**	-0.022% (0.0001)**
Age: 30 to 39	3.6% (0.0009)**	1.4% (0.0007)**	5.0% (0.0009)**	--	0.30% (0.0005)**	1.6% (0.0007)**	0.39% (0.0003)*	-0.003% (0.0002)	0.21% (0.0002)**	-0.078% (0.0001)**
Age: 40 to 49	5.9% (0.0009)**	10.7% (0.0011)**	10.1% (0.0009)**	--	0.13% (0.0004)**	2.2% (0.0006)**	0.64% (0.0003)*	0.027% (0.0001)	0.35% (0.0002)**	-0.085% (0.0001)**
Age: 50 and older	8.7% (0.0009)**	22.8% (0.0013)*	15.2% (0.0010)**	--	-0.27% (0.0004)**	2.6% (0.0006)**	0.61% (0.0002)*	0.058% (0.0002)	0.51% (0.0002)**	-0.085% (0.0001)**

Dependent	-0.4% (0.0007)**	-0.04% (0.0011)	-0.9% (0.0008)**	5.1% (0.0242)*	0.045% (0.0003)	0.2% (0.0005)**	-0.075% (0.0002)**	0.026% (0.0001)*	0.04% (0.0002)	0.026% (0.0001)**
Charlson Score = 1	4.9% (0.0012)**	-0.04% (0.0018)*	-2.5% (0.0011)**	1.4% (0.0073)	0.97% (0.0005)**	1.7% (0.0008)**	0.26% (0.0003)*	0.15% (0.0002)**	0.33% (0.0003)**	0.060% (0.0001)**
Charlson Score = 2	8.8% (0.0020)**	-2.9% (0.0028)*	-8.2% (0.0015)**	1.9% (0.0225)	1.4% (0.0008)**	2.4% (0.0013)**	0.25% (0.0005)*	0.24% (0.0003)**	1.4% (0.0007)**	0.1% (0.0002)**
Constant	-2.3%	-1.5%	1.8%	7.9%	0.7%	1.7%	-0.031%	0.129%	0.27%	0.2%
Number of observations	838437	361617	838437	26995	838437	838437	838437	838437	838437	838437
R squared	0.046	0.091	0.047	0.014	0.003	0.004	0.002	0.00003	0.003	0.0002
Fstatistic	2958.44	4279.98	3421.51	64.95	155.32	289.50	164.39	15.35	167.46	9.74

Notes: Percentage point change are from a linear probability model with service use as the dependent variable and the interaction between post period and treatment group estimating the effect of interest. All models controlled for age (in 10-year bands), gender, dependent status, Charlson comorbidity score, plan year and baseline differences in outcome measures. Services used by less than 0.1% of the population (electroencephalogram for headaches, CT scan for acute uncomplicated rhinosinusitis, and thorax CT combined studies) eliminated for easier presentation. $*p<0.05$ $**p<0.01$

Table 2-9: Results from decomposition of high- and low-value service increases

	Post-Tx Year 1: High Value Services	Post-Tx Year 1: Low-Value Services		Post-Tx Year 2: High-Value Services	Post Tx Year 2: Low-Value Services
Rate of Increase	0.110	0.079		0.052	0.027
Rate associated with increase in preventive office visits	0.081	0.045		0.029	0.016
Share associated with preventive office visits	0.74	0.57		0.56	0.59

Notes: This decomposition is derived by using the share of the population that has both a preventive visit and high- or low-value service in the baseline year, and the share of the population that has a high- or low-value service without a preventive visit in the baseline year to create counterfactual rates of increase in post-period years. To do so, we apply those baseline shares of high- and low-value service receipt conditional on whether or not a preventive office visit was received to the share of the population with and without preventive visits in post period years.

For example, consider the calculation of the counterfactual increase in high-value services in year 1 post-intervention that holds the rate of service use conditional on a preventive visit constant. The share of the population at baseline with a preventive visit and a high-value service is 0.878; the share with a high-value service and no preventive visit is 0.283. Using the baseline preventive visit prevalence (0.53) and percentage point increase in preventive visits from the regression (13.5 ppts; Table 2-6) we create the counterfactual increase in preventive visits with the following equation: $((0.665 *0.878) + (0.283*(1-0.665))) = 0.679$. We compare this counterfactual increase in preventive visits to the actual increase predicted by the model (Table 2-6: $(0.598 + 0.11) – 0.679 = 0.081$), and divide the result by the measured total increase to find the share associated with the increase in preventive visits $(0.081/0.11 = 0.74.)$ This procedure is repeated for low-value services in Year 1, and high- and low-value services in Year 2.

Reference for decomposition method: Thorpe KE, Florence CS, Howard DH, Joski P. The Impact Of Obesity On Rising Medical Spending. Health Affairs. 2004 Nov 1;23(6):283.

Table 2-10: Difference-in-differences estimates from regressions estimated with a probit function

	Likelihood of preventive diagnostic visit	Likelihood of targeted high-value service	Likelihood of low-value service	Likelihood of low-value screening	Likelihood of low-value symptomatic referral	Likelihood of high-value service conditional on preventive visit	Likelihood of low-value service conditional on preventive visit
Intervention (baseline)	18.0% (0.0059)**	6.1% (0.0060)**	10.3% (0.0060)**	10.6% (0.0063)**	0.9% (0.0084)***	0.1% (0.0114)	13.3% (0.0085)**
Post Year 1	1.7% (0.0040)**	0.3% (0.0041)*	0.2% (0.0043)	0.1% (0.0045)	0.1% (0.0058)	-0.8% (0.0090)**	-1.0% (0.0066)**
Post Year 2	1.8% (0.0040)**	-0.3% (0.0041)	-0.5% (0.0043)**	-0.2% (0.0045)	-0.2% (0.0059)**	-0.8% (0.0091)**	-1.5% (0.0066)**
Tx * Post Year 1	13.4% (0.0085)**	12.0% (0.0087)	6.7% (0.0084)**	6.9% (0.0087)**	-0.5% (0.0119)**	3.5% (0.0157)**	3.6% (0.0115)**
Tx * Post Year 2	4.6% (0.0084)**	5.5% (0.0086)**	2.2% (0.0085)**	2.4% (0.0087)**	-0.4% (0.0120)**	1.9% (0.0162)**	0.7% (0.0117)
Female	29.9% (0.0030)**	25.1% (0.0030)**	0.02% (0.0030)	-0.8% (0.0032)**	1.3% (0.0043)**	9.4% (0.0066)**	-18.1% (0.0049)**
Age: 30 to 39	5.6% (0.0052)**	15.5% (0.0051)**	10.3% (0.0061)**	9.4% (0.0067)**	2.9% (0.0083)**	9.6% (0.0089)**	10.2% (0.0091)**
Age: 40 to 49	7.5% (0.0048)**	24.0% (0.0048)**	19.1% (0.0056)**	18.5% (0.0061)**	3.7% (0.0077)**	15.1% (0.0087)**	22.3% (0.0084)**
Age: 50 and older	9.4% (0.0048)**	33.2% (0.0048)**	27.1% (0.0056)**	26.7% (0.0061)**	3.9% (0.0077)**	18.4% (0.0090)**	32.3% (0.0084)**
Dependent	-3.4% (0.0033)**	-5.8% (0.0033)**	-0.2% (0.0034)	-0.2% (0.0036)*	0.2% (0.0048)**	-4.8% (0.0063)**	1.1% (0.0051)**
Charlson Score = 1	-1.6% (0.0045)**	13.7% (0.0048)**	5.3% (0.0045)**	3.5% (0.0047)**	2.8% (0.0059)**	1.5% (0.0101)**	2.7% (0.0067)**
Charlson Score = 2	-2.3% (0.0069)**	16.3% (0.0078)**	5.3% (0.0068)**	2.3% (0.0071)**	4.2% (0.0085)**	1.9% (0.0170)**	0.3% (0.0102)
Constant (predicted)	43.3% (0.0054)	58.1% (0.0053)	27.1% (0.0061)	22.4% (0.0066)	7.0% (0.0085)	89.5% (0.0106)	39.2% (0.0097)
Number of observations	838437	838437	838437	838437	838437	363693	363693
R squared	0.111	0.136	0.067	0.078	0.013	0.145	0.105
Fstatistic	116494.4	137774.5	64560.5	68642.7	5621.7	31806.9	45886.9

Notes: Percentage point change are marginal effects from a probit model with service use as the dependent variable and the interaction between post period and treatment group estimating the effect of interest. All models controlled for age (in 10-year bands), gender, dependent status, Charlson comorbidity score, plan year, and baseline differences in outcome measures. *p<0.05 **p<0.01

Table 2-11: Difference-in-differences estimates from matched sample

Percent of Intervention Population and Control Sample Using Following Services -- Matched Sample	Baseline Intervention Population	Baseline Control Sample	Percentage Point Change Year 1	Percentage Point Change Year 2
Preventive Office Visits	53.0	53.0	26.0**	16.2**
Any Targeted Preventive High-Value Service	59.8	61.5	11.2**	4.8**
Low-Value Service	33.6	27.3	7.9**	1.9**
Low-Value Preventive Services	28.8	22.7	11.0**	4.2**
Low-Value Referral Services	7.8	6.9	-0.5*	-0.6**
Number of Observations	64,165	64,110		
Notes: Percentage point change are from a difference-in-differences linear probability model with service use as the dependent variable and the interaction between placebo post period and treatment group estimating the effect of interest. Both models controlled for age (in 10-year bands), gender, dependent status, Charlson comorbidity score, plan year and baseline differences in outcome measures. All models use data from baseline year (pre-intervention) only. *p<0.05 **p<0.01				

Table 2-12: Elasticity estimates

				(%ΔQ/%ΔP)	
Total Costs for High Value + Preventive Services	$ 2.10	$ 0.40	$ 0.57	Year 1	Year 2
Predicted Percentage of High-Value Service Use (Including Preventive Visits)	66%	79%	72%	-0.24	-0.11
Predicted Percentage of Low-Value Service Use	34%	42%	37%	-0.29	-0.11
Notes: Based on changes in the average price for the weighted average of the program-eligible population in each year. Changes in likelihood of service use estimated using the predicted prevalence of service use from a difference-in-differences OLS model that controls for age, gender, dependent status, Charlson comorbidity score, and time trends in service use.					

Figure 2-1: Basic program requirements as shown to enrollees

	Birth to Age 5	Ages 6 to 17	Ages 18 to 24	Ages 25 to 29	Ages 30 to 39	Ages 40 to 49	Ages 50 +
Preventive Visit	**One per year	Once every other year	Once every 3 years	Once every 3 years	Once every 3 years	Once every 2 years	One per year
Vision Exam	N/A	Once every 2 years	Once every 2 years	Once every 2 years	Once every 2 years	Once every 2 years	Once every 2 years
Dental Cleanings (if enrolled on a Dental plan with UHC or Cigna)	N/A	Two per year	Two per year	Two per year	Two per year	Two per year	Two per year
Cholesterol Screening	N/A	N/A	Once every 5 years starting at age 20	Once every 5 years	Once every 3 years	Once every 2 years	One per year
Clinical Breast Exam	N/A	N/A	Once every 3 years	Once every 3 years	Once every 3 years	Once every 3 years	Once every 3 years
Mammogram (Females)	N/A	N/A	N/A	N/A	One screening between ages of 35 and 39. Otherwise, as Recommended by Physician	As Recommended by Physician	As Recommended by Physician
Cervical Cancer Screening (Females)	N/A	N/A	Once every 3 years starting at age 21	Once every 3 years	Once every 3 years	Once every 3 years	Once every 3 years
Colorectal Screening	N/A	N/A	N/A	N/A	N/A	N/A	Annual fecal occult blood test OR Colonoscopy (w/CT) every 10 years

¹ Birth to Age 1 requires visits at Months 1,2,4,6,9 and 12 per American Academy of Pediatrics Guidelines

Source: State program website

Figure 2-2: Exclusion criteria for enrollees and final sample calculation
The chart below describes the exclusion criteria and process of creation of the entire sample. The steps are sequential; thus, someone

could not be dropped for more than one reason and the total number dropped will add up to (initial sample – final sample). I excluded

anyone not enrolled for the full period of the study. I also excluded children (who were not subject to benefit changes in our treatment

group) as well as anyone older than 61 at baseline, who would be likely to transition to Medicare during the following 3 years.

Initial Intervention Population: 190, 367	Initial Control Sample: 600,003
Dropped because of negative pay amounts: 2,369	Dropped because negative pay: 24,887
Dropped because not enrolled for full period: 38,681	Dropped because not enrolled for full period: 251,233
Dropped because > 61 or < 18: 62,802	Dropped because > 61 or < 18 : 81,236
Dropped because retired: 22,350	Dropped because retired: 27,333
Final Intervention Population: 64,165	Final Control Sample: 215,314

Figure 2-3: Pre-trend graphs :high- and low-value services

54
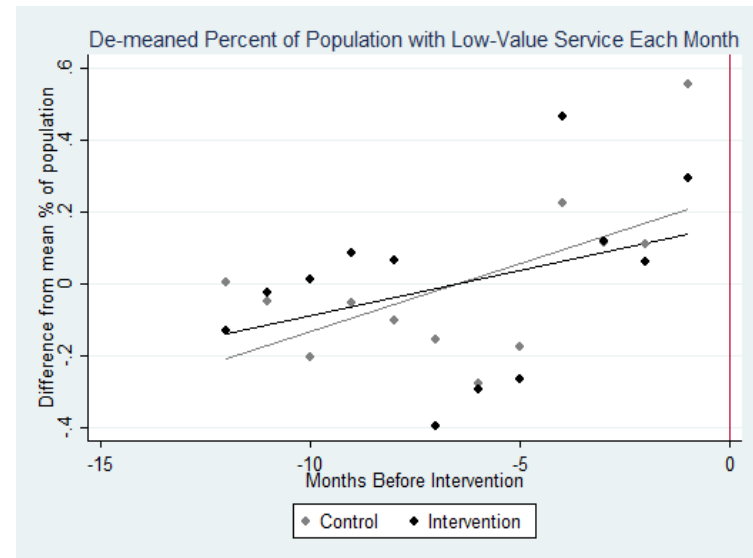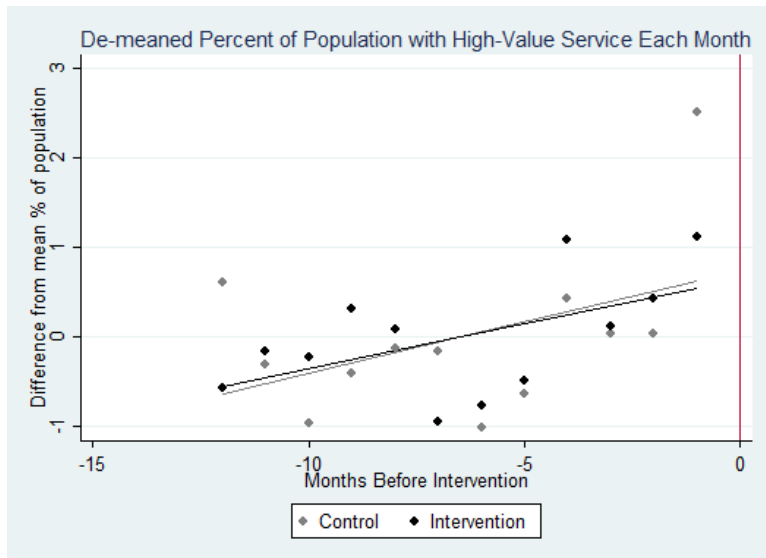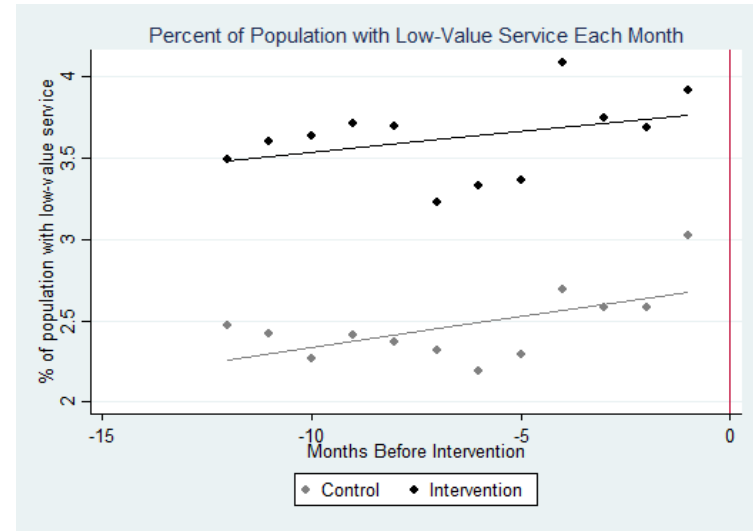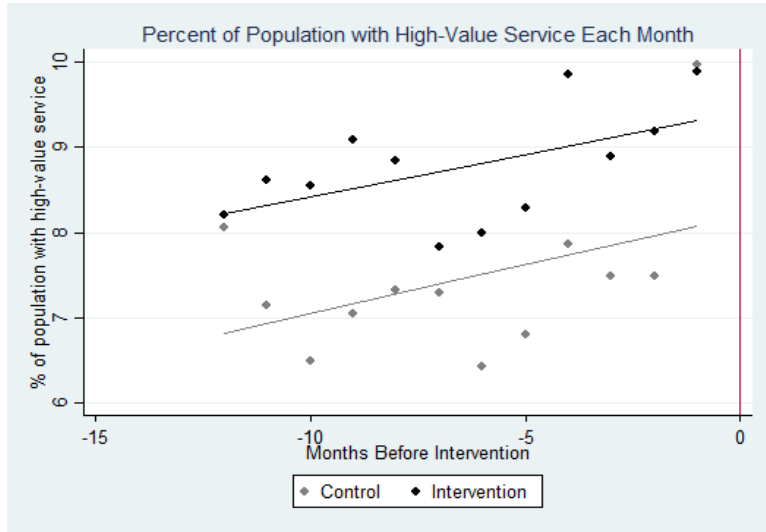
Figure 2-4: Average percent of intervention population and control sample receiving any high-value service
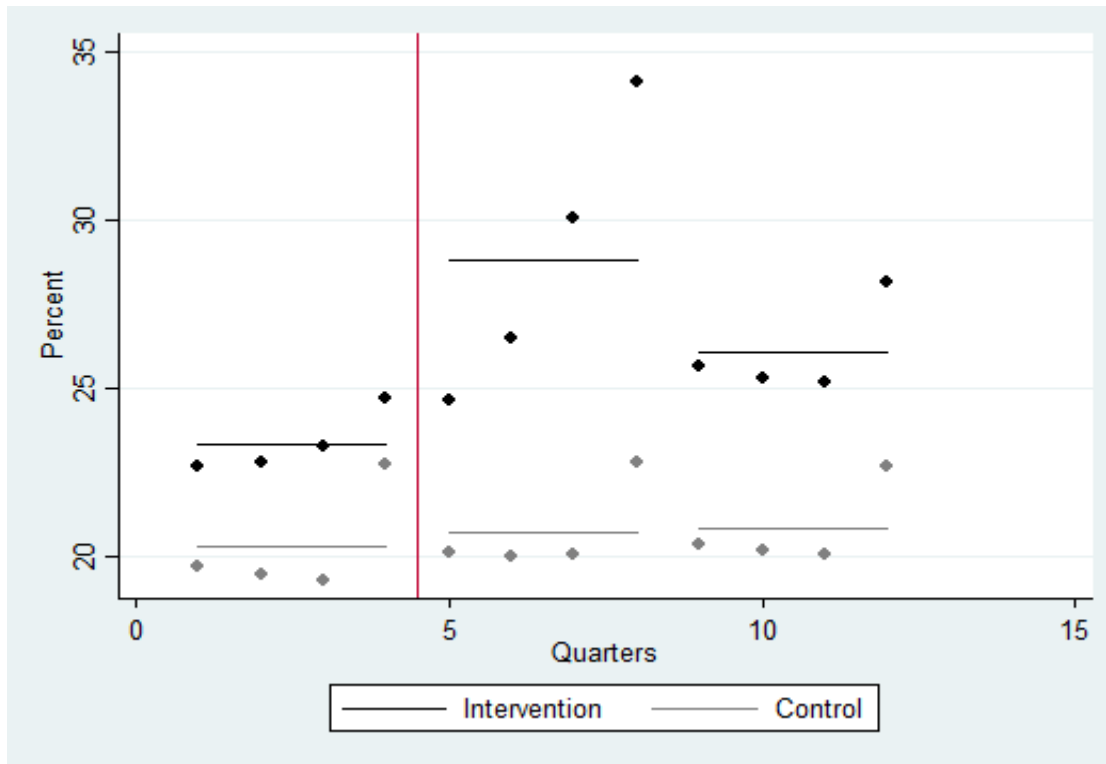
Notes: Horizontal axis shows quarter of study period; intervention occurs after quarter 4.
Each dot represents the per-quarter average percent of the population who receive a service
and lines represent the average across quarters.

Figure 2-5: Average percent of intervention population and control sample receiving any low-value service
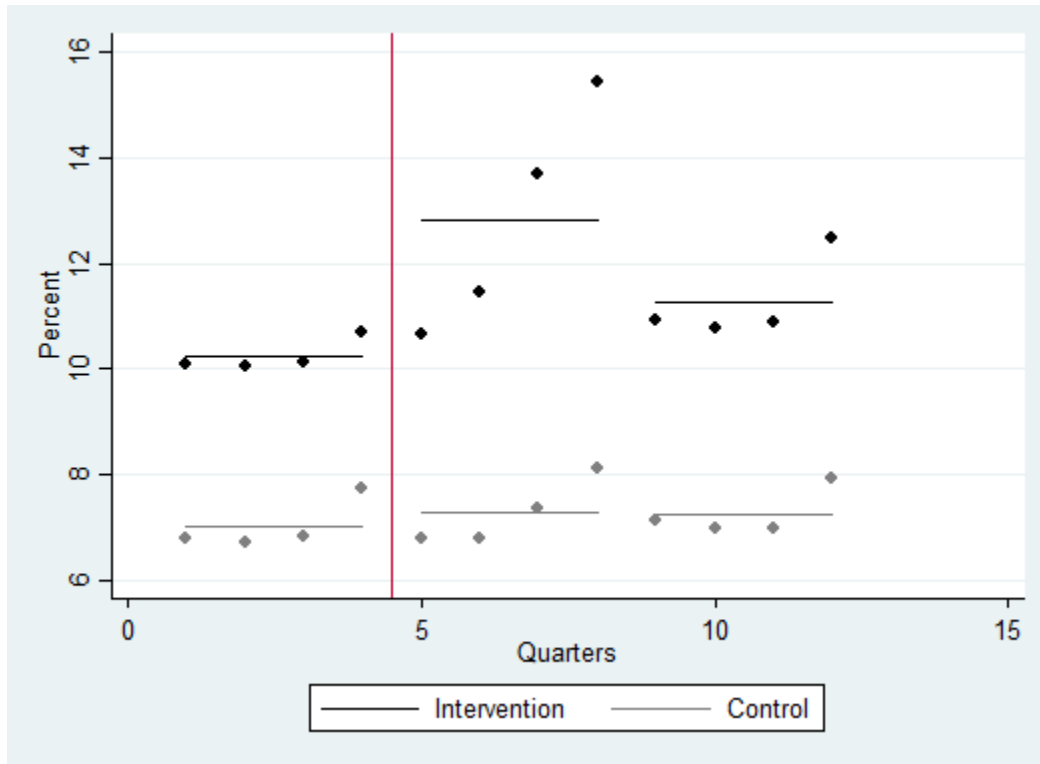
Notes: Horizontal axis shows quarter of study period; intervention occurs after quarter 4.
Each dot represents the per-quarter average percent of the population who receive a service
and lines represent the average across quarters.

Figure 2-6: Average percent of intervention population and control sample receiving low-value screening service
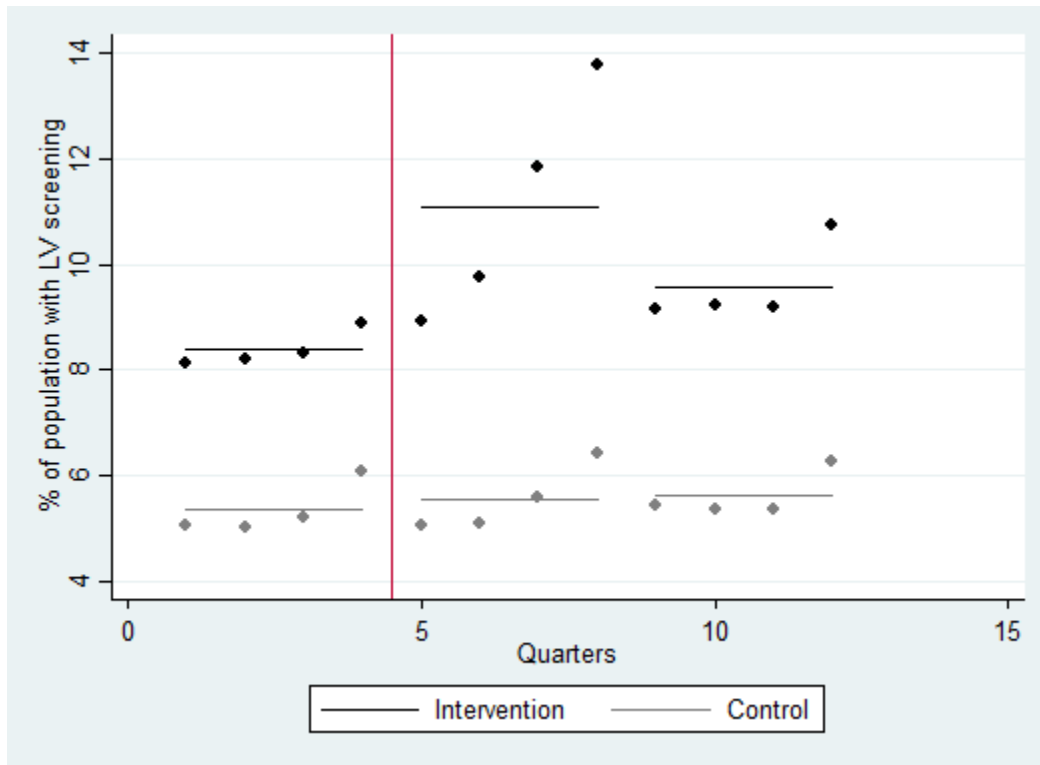
Notes: Horizontal axis shows quarter of study period; intervention occurs after quarter 4.
Each dot represents the per-quarter average percent of the population who receive a service
and lines represent the average across quarters.

Figure 2-7: Average percent of intervention population and control sample receiving low-value referral service
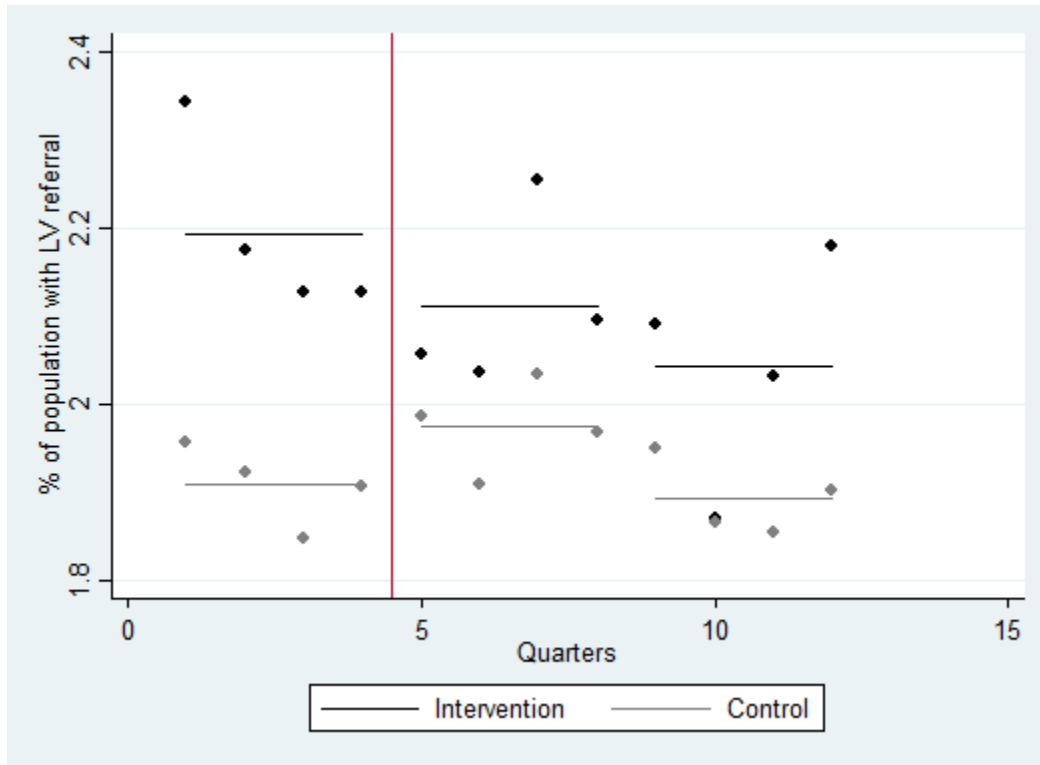
Notes: Horizontal axis shows quarter of study period; intervention occurs after quarter 4.
Each dot represents the per-quarter average percent of the population who receive a service
and lines represent the average across quarters.

Figure 2-8: Percent of the intervention population receiving high- and low-value services as a percent of the population with any medical encounter
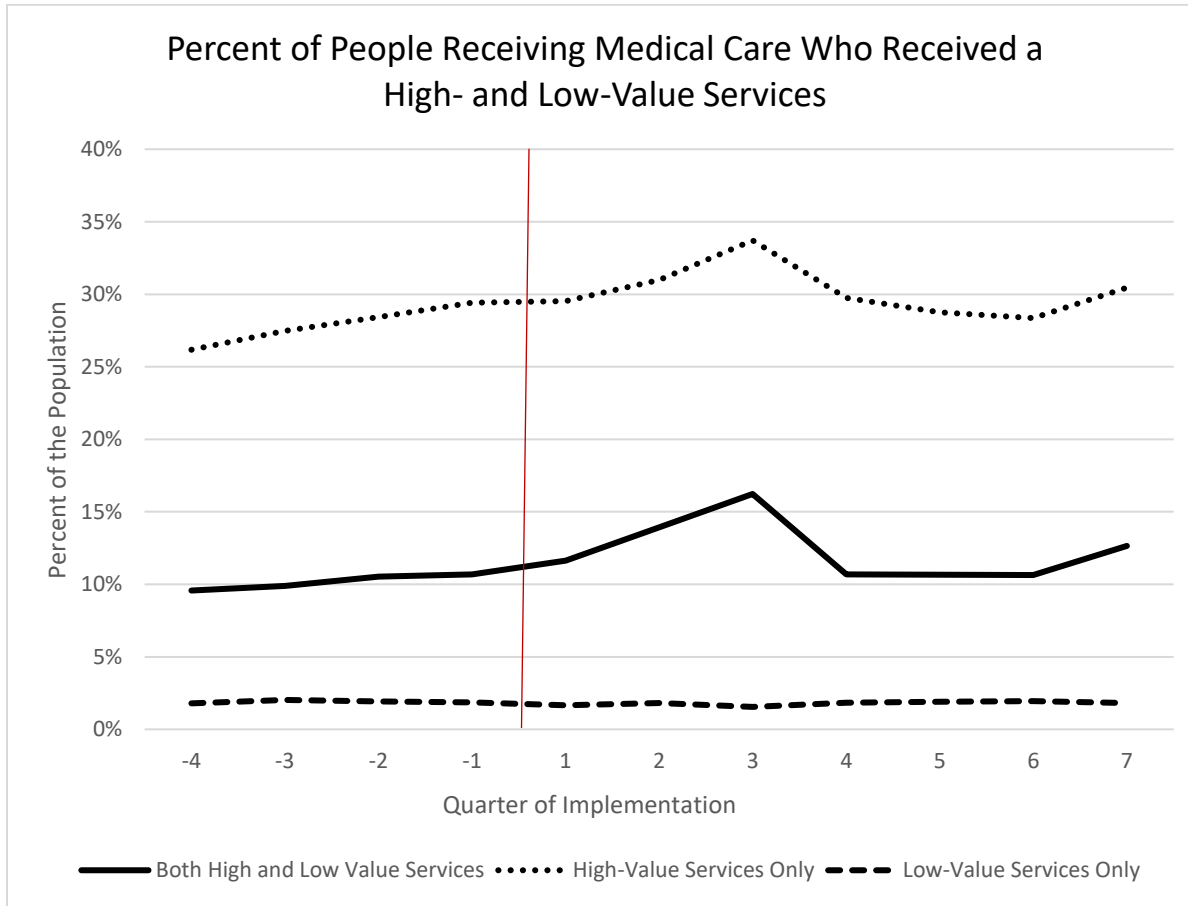
Notes: Horizontal axis shows quarter of study period; intervention occurs after quarter 0.
Lines show percent of population who received any high-value service (including a preventive office visit), low-value service or both types of services as a percent of all people who had a medical encounter in the quarter.

Chapter 3 : Public Reporting with Negotiated Prices: The Effect of Information on Competitors' Prices for Inpatient Hospital Procedures

INTRODUCTION

In the commercial insurance market, historically opaque pricing for medical services has enabled high prices and wide variation both within and across hospital service markets. Prices have been shown unrelated to quality, and thus are often considered a primary cause of inefficient healthcare spending. (Capps, Dranove, and Satterthwaite 2003; White, Reschovsky, and Bond 2014; Coakley 2011). Price transparency is a potential correction to this issue. Harnessing consumer demand through price transparency, ideally in tandem with quality metrics where applicable, is a potential approach for lowering total spending and reducing individual financial burdens without sacrificing quality of or access to care. While price transparency has typically been heralded by policymakers and journalists as a boon for consumers, economists have been less encouraged about welfare effects based on how it might be used by payers and providers. In cases where prices are negotiated and each side has some market power, as is common in hospital-insurer negotiations, information on prices can be used strategically by either side. Comparing it to the use of most-favored nation clauses, in which an insurer signs a formal agreement with a hospital that the insurer will receive the lowest available price, some have

argued that average hospitals prices will increase if hospitals can no longer offer these secret deals (Cutler and Dafny 2011). Other work has emphasized the ability of price information about competitors to facilitate collusion or simply to give lower-priced providers information about payers' willingness to pay higher prices, thus resulting in higher negotiated prices (Ginsburg 2007; Sinaiko and Rosenthal 2011). On the other hand, some experts have worried about increased competition driven by transparency efforts as instigating such harsh pricing competition that may drive down prices, but also harm quality and service lines at hospitals (Altman, Shactman, and Eilat 2006).

To date, there are few analyses of the effect of price transparency initiatives on equilibrium prices, particularly in the absence of consumer response. Changes in equilibrium prices are more impactful than whether individual consumers can or do shop for lower prices because they affect all consumers and a potentially larger percent of services. The transaction prices negotiated by commercial payers, who insure approximately two-thirds of the non-elderly population, determine the unit price for this population, either directly through coinsurance or as a primary input into plan premiums.

Using a common model of hospital-insurer negotiations, this paper proposes bargaining ability as a mechanism through which through which price transparency may influence the transaction price of hospital services, and empirically tests predictions from the model by estimating the effect of introducing comprehensive price information in a state. Briefly, it posits that the availability of information about competing hospitals' prices may affect the hospital's ability to bargain, especially for hospitals with potentially questionable value for an insurer's network. Information can affect bargaining ability through a number of channels and may lead to price change under certain conditions.

Specifically this paper uses the introduction of state all-payer claims databases (APCDs) that aim to collect transaction prices for all insured health care encounters in a state and, in some cases, publicly release these prices. The paper compares states in which an APCD was implemented and pricing information was publicly released to states that did not implement a publically released APCD between 2010 and 2014. While I use the availability of data as a quasi-experimental intervention, I do not actually use data from APCDs in this study. Instead, I use a large administrative claims dataset available from the Health Care Cost Institute, covering the years 2010-14. The advantage of using this dataset is that I can observe prices in all U.S. states, including those that do not publicly release prices. However, unlike an APCD, I observed only a subset of prices. Though I describe the dataset in detail more later, briefly, the dataset includes claims line information from three major commercial health insurers, including the transaction prices. Within the years covered in our dataset, Colorado, Massachusetts and Oregon implemented and publicly released information from APCDs.

On a narrow policy level, whether the release of prices to stakeholders through an APCD alters price dispersion or price levels in a market is relevant for the future development of APCDs. A number of states currently have APCDs or are in the process of developing them. As of April 2019, California, Tennessee and Florida are all implementing a statewide APCD (www.apcdcouncil.org). Whether or not to release data, and which data to release, remains a key

question for policy makers.[1] More generally, the release of prices through an APCD informs the larger question about how firms with market power, particularly hospitals, react to the comprehensive release of such information. As the number of places to get information about health care prices grows and prices become less opaque in a market, questions of how those involved in price negotiations use that information take on more relevance. If the release of information has a pro-competitive effect, it may signal positive impacts of price transparency initiatives, even in the absence of a consumer response. On the other hand, if releasing information enables collusion or increases in price levels, policymakers may want to be more reticent about how such information is released in the future.

Overall and stratified by market concentration, I find a small positive increase in mean prices, though estimates are imprecise and I am unable to distinguish the effect from 0. Within different quantiles of the price distribution, I find prices in the lower half of the price distribution increase 1 to 2 percent after implementation of an APCD, which is statistically significant or nearly so, at some quantiles. That increase appears to be driven by changes primarily in the least concentrated markets, though there are some changes in more highly concentrated markets as well. Looking at the dispersion of prices across markets, I estimate a drop in price dispersion by 0.034 points, a relative 12%, though standard error is not significant by conventional standards

---

[1] A March 1, 2016 Supreme Court decision, Gobielle v Liberty Mutual Insurance Company held that ERISA pre-empts state laws—there, Vermont's—that would compel self-insured employers to report claims information to APCDs. The ruling mostly made note of the potential for multiple regulatory environments to create administrative burdens and costs. The federal government took the position that states could compel the release of such information given their important role in cost-containment, implying APCDs encourage such containment, and the majority opinion left open the option that the Department of Labor could amend ERISA to compel that information to be disclosed in cases in which states had set up APCDs. While the DOL proposed a rule that would have improved reporting by self-insured plans shortly after the ruling, it was never finalized. Some companies are continuing to voluntarily provide their information, and some APCDs—notably Colorado—have explicit pathways for them to do so. It is not currently known how many self-insured employers are providing information to APCDs.. Further, it's unclear to the extent that self-insured plans differ in prices from those for which the payer bears risk. If self-insured and risk-bearing contracts do not differ in negotiated transaction prices within payers, the loss of self-insured clients may not be relevant for price benchmarking.

(p=0.087). Consistent with changes in price levels, dispersion drops more in less concentrated markets.

While my findings are not conclusive, they are indicative of a pattern of behavior. My results suggest that price transparency initiatives change the dispersion of prices by increasing prices near the bottom of the price distribution. This effect seems most prominent in markets with less concentration, though there is some indication of price changes at the bottom of the price distribution in more concentrated markets as well.   However, there appears to be no effect on mean price and few effects on prices at the higher end of the distribution. The finding is consistent with the predictions from theory that improvements in bargaining ability are most likely to affect negotiations in markets with a number of substitutes. It is also consistent with earlier predictions, notably by Cutler and Dafny (2011), that price transparency may raise prices at the lower end of the price distribution. While data limitations may attenuate the number of conclusions that can be firmly drawn from these estimates, they do point toward an area that should be investigated further and with a longer panel.   While the results suggest that price transparency may not impact most markets, the effect at the lower end of the distribution and in the least concentrated markets may be particularly important as these markets that offer consumers the greatest opportunities for saving money.

BACKGROUND: HOSPITAL NEGOTIATIONS AND THEORETICAL FRAMEWORK

Negotiated Hospital Prices

Hospitals are price takers from large government programs such as Medicare or Medicaid, but, with commercial insurers, negotiate payment rates—the transaction prices that people with commercial insurance pay for medical services. According to industry sources, these contracts are negotiated, on average, every one to three years. (Weiss 2012). Payers and providers typically do not negotiate a price for each individual procedure or service. Instead, payment rates are based on entire service lines, and consultants encourage hospitals to think about where they might receive higher rates. (*Becker's Hospital CFO Report* 2011). Rates are usually based on a percentage increase from the amount the provider is paid for Medicare patients, or a percentage decrease from the amount on a hospital's chargemaster, a publicly available list of prices for hospital services (Batty and Ippolito 2017; Clemens and Gottlieb 2013). Using charges and Medicare payment rates as a basis for negotiating sets up a floor and ceiling for negotiations, though there is substantial room between these two price points to allow a wide variation in prices across both insurers and providers (Cooper et al. 2018).

Prior to the advent of selective contracting in the 1980s and 1990s, hospitals had much more leverage under which to set their own prices (Devers et al. 2003). However, as Devers et al found, once payers began excluding providers and steering patients to providers with whom they could negotiate favorable terms, the balance of power shifted. In response, providers consolidated and, today, the relative power of payers and providers in each negotiation is determined in large part by the relative value of each side to the other (Ho and Lee 2013). The balance of power can be influenced by a number of things including the size of each entity, the selectivity of a payers' network, the proximity of comparable providers and the features of the provider including reputation, capacity limitations and facility attributes (Ho 2009; Clemens and Gottlieb 2013; Gowrisankaran, Nevo, and Town 2015). Consultants recommend providers

understand the market environment under which they are operating, including the rates of other hospitals and the rates across payers. Both industry sources and prior academic work note the use of information as a bargaining tool to be used in bettering a contract for a provider (*Becker's Hospital CFO Report* 2011; Tu and Gourevitch 2014; Town and Vistnes 2001). They also note that, in contrast to national payers or large health systems, community and regional hospitals often have the hardest time with contracting, because of a lack of dedicated personnel for the task (*Becker's Hospital CFO Report* 2011). The suggestion of the strategic use of market environment suggests a role for information about competitors in establishing a bargaining position and in influencing the balance of power in negotiations.

Empirical Tests of Price Transparency

Empirical evidence for how price transparency could affect equilibrium health care prices is limited but growing quickly. Early work on drug and optometry services found that price advertising lowered prices in those markets (Cady 1976; Kwoka 1984). Some recent studies have also supported the idea that publishing prices can effectively lower prices. Two recent papers find price transparency initiatives put downward pressure on equilibrium prices through reduced searching costs or increased demand elasticity from newly informed consumers (Brown 2016; Whaley 2015). Both papers find reductions in transaction prices after the introduction of pricing information; Brown's intervention is New Hampshire's APCD and Whaley's is a transparency tool deployed by insurers and employers for health plan enrollees.

Other studies, using hospital prices, have found no effect on market-wide prices when price information has been made available. An analysis of California's price transparency initiative,

which required hospitals to post charges online, found that modal price increased and no change in dispersion of prices for childbirth (Austin and Gravelle 2007). However, the initiative made charges—not the actual prices paid—public, which may have limited the initiative's relevance (Reinhardt 2006). Another analysis, using as an intervention price transparency legislation that mandated charge disclosure, found evidence of decreases in hospital charges though no decreases in mean transaction price, suggesting no market-level movement on average price (Christensen, Floyd, and Maffett 2017). Notably, these authors attribute their results to the reputational effect of being perceived as an over-charging hospital rather than a consumer shopping on hospital charges. In New Hampshire, which has a well-developed website for pricing information, a study found no change in price variation a year after the state implemented its website (Tu and Lauer 2008). However, a more recent qualitative analysis over a longer time span, suggests that consumers have been slow to use the state's price transparency information but that it has been used by hospitals and payers in negotiating prices causing a change in prices, at least anecdotally (Tu and Gourevitch 2014).

Work in an adjacent market, hospital purchasing of medical supplies, found that access to price information had little effect on price at the mean though did affect the right tail of the distribution, that is it decreased prices paid by the hospitals (buyers) who were previously paying the highest prices (Grennan and Swanson 2016). The authors note that two mechanisms could be working to lower prices through access to greater information: an information asymmetry model in which price transparency reduces uncertainty and equilibrium dispersion of prices or, second, an agency model in which hospital managers can better monitor the actions of those employees specifically tasked with negotiating prices. Notably, and different from the market

studied in our paper, this initiative made prices available only to one party (the buyer) in the negotiation.

In other industries, publishing information on negotiated transaction prices has increased average prices in a manner akin to the concerns raised by Cutler and Dafny (2011). Prices for Danish ready-mix concrete increased 15-20% following the publishing of those prices; the authors concluded that suppliers were less likely to give purchasers discounts following the transparency initiative (Albæk, Møllgaard, and Overgaard 1997).

Theoretical Predictions of Price Transparency and Hospital Negotiation Models

In the commercial market, oligopolistic hospitals and insurers engage in bilateral negotiations to form networks of hospitals. These models are inherently more complex than markets that have posted prices as each player has some degree of market power and thus some ability to influence price. Prices are unique, or nearly so, to each provider-payer pair, and perhaps even to subunits within provider-payer pairs, such as different lines of business within an insurer. As such, papers that model such negotiations make a number of simplifying assumptions to gain traction.

To date, much of the work in this area, particularly in the healthcare setting, has examined the effect of changing market structure on provider prices. Using models first developed by Horn and Wolinsky (1988) these papers estimate the effects of mergers (Gowrisankaran, Nevo, and Town 2015; Lewis and Pflum 2015), insurer competition (Ho and Lee 2013) or system membership (Dafny, Ho, and Lee 2016). The basic structure of these papers, and others that build on the Horn and Wolinsky framework (e.g. Gaynor, Ho, and Town 2015) is to model the objective functions for both the hospital and insurer. I'll refer to this model as the GNT model

for simplicity, though these researchers are certainly not the only economists to model

negotiations in this fashion (Gowrisankaran, Nevo, and Town 2015).

Following GNT, a hospital's objective function (π_h), in which the goal is profit maximization,[2]

depends on prices negotiated with a specific payer (s), quantities associated with those price

negotiations, and the hospital's marginal cost[3]:

$$\pi_h = f(\vec{p}, \vec{q}\ mc) = \sum_{s \in S} q_{s,h}(q_{-s}, p_{-s})[p_{hs} - mc_h] \tag{1}$$

where profit is a function of quantity (dependent on the number of potential patients that go to

other systems and the prices they pay), negotiated price and the hospitals marginal cost. Here, I

assume that non-agreement with the insurer substantially decreases the number of patients who

use the hospital from that insurer. This outside option is normalized to 0.

The insurer's objective function (π_s) aims to maximize the difference between consumer welfare

(W_s) and total cost (TC), and relates the value of having specific hospitals in its network to the

total cost of having that network. GNT specify as:

$$\pi_s = V_{H\{1\ldots h\}} = f(H\{1\ldots h\}, \vec{p}) = \tau W_s(H, \vec{p}) - TC(H, \vec{p}) \tag{2}$$

where network value, V_H, is a function of hospitals in the network, and the price paid to each

one. In this equation, τ is the weight a payer puts on consumer welfare, $\in [0,1]$. It follows from

---

[2] Including for non-profit hospitals

[3] GNT allow for variable marginal costs across MCOs and weighted by patient severity. I specify constant mc for simplicity and because it does not change my primary conclusions.

this equation that the value including any one hospital is the difference between the value of the

network with the hospital in it (V_h) and the value of it without the hospital V_{-h}.

The optimal price is the price that solves the Nash bargaining solution for each payer-hospital

combination, modeled as a non-cooperative game with alternating offers:

$$p_{h,s} * = \arg max_{P_{h,s}}(NB^{s,h} (p_{hs} |p_{h-s})) =$$

$$\arg max_{P_{h,s}} [\pi_h(p_{hs}, q_s, q_{-s}, mc)]^b \times [\pi_s(p_{hs}, V_h, V_{-h})]^{1-b} \tag{3}$$

where the equilibrium price is determined through a joint maximization of the objective function

of the hospital and payer. The parameter $b \in [0,1]$ represents the relative bargaining power or

bargaining ability of the hospital in the negotiations. It is this parameter that determines the

surplus division from the negotiation; in this case it is typically the amount down from charge

price or amount up from Medicare payment for each service. In this paper, much of the focus of

the role of information will be on this parameter.

To determine the relative importance of the bargaining parameter and make predictions about

how it will influence behavior, we first derive the first order conditions with respect to each

objective function.

$$\frac{\partial p*}{\partial \pi_{h*}} = \pi_s^{1-b_h} b\pi_h^{b_h-1} = 0 \tag{4}$$

$$\frac{\partial p*}{\partial \pi_s} = \pi_h^{b_h}(1 - b_{h)}\pi_s^{-b_h} = 0 \tag{5}$$

Solving for π_h, we find that hospitals' ability to optimize its objective function depends directly on its bargaining ability parameter and the insurer objective function, the difference between its value relative to competing hospitals:

$$\pi_h(p_{h,s}) = \frac{b_h \pi_s(p_{hs}, V_h, V_{-h})}{1 - b_h} \qquad (6)$$

Information can enter these models at a number of points. Note, however, that information does not have to vary in order to cause variations in price. Even if insurers' objective functions for a given hospital are all the same, the model predicts changes in hospital prices through differences across b_h in hospital-insurer pairs.

Both Brown (2018) and Whaley (2015) use this same framework to model the effect of an increase in consumer attention to prices and a resultant drop in prices. Whaley modeled price transparency as reducing search costs for consumers, leading them to place higher value on providers with lower price, holding quality constant. Thus, price transparency enhanced the value (V_h) of low-cost providers relative to competition. Brown uses the bargaining model to show theoretical ambiguity around effects of price transparency, potentially making consumer demand more elastic but, at the same time, reducing incentives for insurers to exclude high-priced providers from their network. Using information from a state-based effort to introduce price transparency in New Hampshire, he estimates a drop in prices if all consumers became informed and shopped for price for medical imaging.

In contrast to Whaley and Brown and in line with recent evidence, I assume consumer inattention to prices revealed through transparency initiatives (Kullgren et al. 2018; M. Chernew et al. 2018; Desai et al. 2016). Thus, in contrast with previous literature on hospital prices, the insurers' objective function is unchanged with price transparency, i.e. consumer valuation of

hospitals is unchanged. One potential objection to this assumption is that there is good evidence that consumers respond to drops in co-insurance or premiums, both of which are a function of price. Note that this assumption does not state that consumers do not care about price, which is an important component of their objective function, only that transparency initiatives do not cause increased attention to price.

Instead, I look at changes in the bargaining ability (b_h) of hospitals (and insurers) due to changes in information. Several recent papers have taken this same approach to examine changes to information or the market environment in hospital services or closely related markets (Grennan and Swanson 2016; Lewis and Pflum 2015). Bargaining ability is the skill a hospital (or insurer) has at using their position to secure gains from the negotiation. Lewis and Pflume explicitly distinguish it from bargaining position, the relative size of a hospital or insurer in a market, and model it as a linear additive function of hospital and market attributes. Grennan and Swanson use changes in perceived strength as a determinant of price, without modeling the function directly.

I model bargaining ability as a function of a number of fixed and variable hospital attributes:

$$b_h = f(\gamma_h, \partial_i, \alpha_{-h}, \delta_h, \varepsilon) \tag{7}$$

where γ_h are fixed hospital attributes (e.g. academic medical center), ∂_i is the agency or skill of individual employees involved in the negotiation, α_{-h} is the availability of close substitutes, δ_h is information about the opposing party's willingness to pay (discount), and ε is an error term capturing idiosyncrasies of the negotiation.

Previous literature models price offer and acceptance as a repeated, non-cooperative game with alternating offers (Gowrisankaran, Nevo, and Town 2015; Binmore, Rubinstein, and Wolinsky 1986). Agreement with an insurer always increases a hospital's objective function as long as price is above marginal cost of care (Equation 1). Thus, we assume that hospitals will continue offering prices until they get to p_{mc}. The game structure may look something like this:
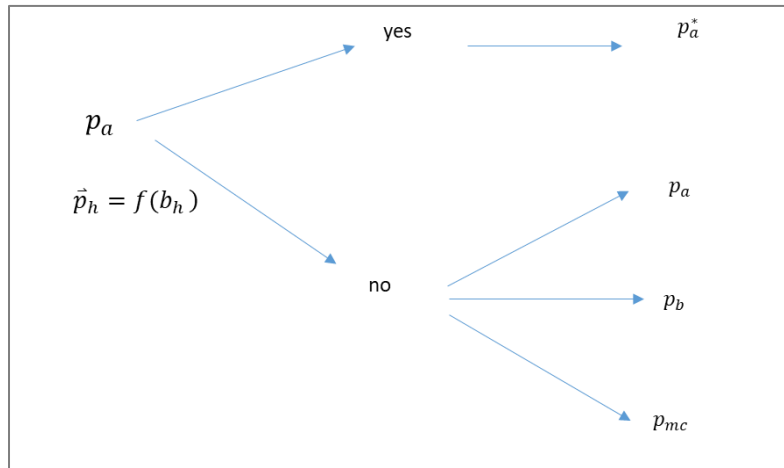
In this figure $p_a > p_b > p_{mc}$ and hospitals start by offering their highest price, p_a. If that is accepted, it causes agreement and solves the Nash bargaining problem. If it is not, then hospital must make a choice about the next price to offer, which includes the original price. Its ability and willingness to offer either p_a (again) or p_b depends on its bargaining ability.

Information on other hospitals can enter bargaining ability by giving hospitals information on how close a substitute a competitor might be, providing evidence to support an ask of higher price, or by giving hospital administrators information about the success of their employees, who negotiate on behalf of the hospital.[4] There are several predictions from this model of the bargaining game:

---

[4] Grennan and Swanson (2018) make this same point about the ability of information to solve a principal-agent problem.

1) Increases in bargaining ability for hospitals increases price levels; increases in bargaining ability for insurers decreases price levels.

2) Dispersion of prices will drop if information improves either hospital or insurer bargaining power as parties are less likely to accept unfavorable offers.

3) Bargaining ability matters more when there are alternatives (i.e. least concentrated markets), more likely leading to more rounds of the game.

Price transparency initiatives will only improve bargaining ability if they offer new information to hospitals. It is probable that these initiatives do offer some new information to some hospitals, as there is evidence hospitals use it. However, prior to a price transparency initiative, information about prices is likely variable among hospitals. Price transparency initiatives, then, can be thought of as raising the baseline level of information about prices in the market. The bargaining ability of hospitals with at least baseline level information about prices will be unaffected by introduction of this pricing information.

SETTING: ALL-PAYER CLAIMS DATABASES

APCDs are state-based entities and include negotiated transaction prices from all types of payers, drawn from administrative medical claims. As such, they also include provider identifiers and detailed information about the medical procedures, co-morbidities, and procedures for each patient. Some APCDs also include a consumer-facing website. New Hampshire has an extensive state-based consumer price transparency tool, for instance, that is based on its APCD. Colorado has also made provider-specific price information readily available for some

procedures. Through 2014, when I can observe prices in this study, four procedures (childbirth, Cesarean section childbirth, hip replacements, knee replacements) were included on the consumer website, though more are now available (www.civhc.org). Datasets available for release in these states include information on all (or nearly all) insured procedures. This feature means that payers and providers, who presumably have the sophistication to analyze such data, have better information than consumers in this setting.

Maine was the first state to implement an APCD in 2003. To date, at least 22 states have created APCDs, or significant multi-payer claims databases, through legislation or strong voluntary efforts (Figure 3-1). Several other states have APCDs in implementation or have introduced legislation to create such databases.

A key distinction in implementation across states is the degree to which each state publicly releases the data. Because these databases collect the actual amounts paid by payers to health care providers, APCDs contain proprietary pricing information. Both payers and providers have noted the release of this information could affect hospital-payer negotiations, sometimes expressing concern. In some states—Maine, Oregon, Vermont, New Hampshire, Massachusetts, and Colorado—legislation or regulation allows data to be released to third party requesters, including payers or providers. In other states—Kansas, Texas, Washington—data from the APCD is strictly controlled and state employees may be the only ones allowed to access the APCD. Three states implemented an APCD and released data publicly during the study period. Colorado implemented its APCD in 2013, with the first data release requests granted that year. Massachusetts first released data from its APCD in the summer of 2012, the same year Oregon first released data from its APCD.

75

Many hospitals and payers use information from the APCDs when it is available. In Maine, for example, a 2014 request from St. Mary's Health System and St. Joseph Hospital (both members of Covenant Health) notes that the hospitals are requesting the APCD information "to determine if [hospital] pricing is in line with the market" (mhdo.maine.gov). Several other requests in Maine, including a 2016 request from CIGNA HealthCare, note the use of the data for either strategic or reimbursement purposes.

There is evidence of the APCD being used in hospital pricing strategies in all 3 states studied here. In Colorado, both payers and providers use the APCD to benchmark themselves (personal communication, Tracey Campbell, Center for Improving Value in Health Care). In New Hampshire, payers noted explicit use of the APCD and publicity surrounding its publication to lower prices at an outlier hospital (Tu and Gourevitch 2014). Oregon releases data for "healthcare operations" for purposes that include provider performance and cost management (Oregon Statute 442.466).   The public use file in that state contains the amount paid to a provider by payer and patient, which, in many cases, represents the negotiated transaction price. (Slightham 2016) Though the public use dataset in Oregon contains limited information about which entity provided a service, the limited use data file is also available to healthcare entities and contains more information. A state report confirms providers have used the APCD to benchmark prices. ("Oregon All Payer All Claims Database (APAC) Use Case Document" 2017). Massachusetts' release policy operates under a different paradigm. There, prices are released to payers and providers so long as the entity can show doing so is in the "public interest". ("Overview of the Massachusetts All-Payer Claims Database" 2014) To date, at least two payers have shown that using the information to benchmark and keep insurance products competitive is in the public interest (chiamass.gov).

DATA AND ESTIMATION

Overall, the goal of this paper is estimate changes in hospital procedure prices associated with changes in information about competitors' prices among hospitals. As the theoretical model shows, hospital prices vary directly with hospital bargaining ability and the value of a hospital to an insurer's network relative to other potential hospitals in the network. I test the 3 predictions from the theoretical model above, looking at whether changes in bargaining ability change market level price dispersion, price levels, and heterogeneity of effects across markets.

To identify relevant effects, I exploit variation in the timing of state-based initiatives to release hospital prices through APCDs. I assume that these initiatives are independent of pre-existing trends in hospital prices in each state. The goal is to isolate bargaining ability from other parameters that might influence price. I'm particularly concerned about adjustments for patient severity, changes in hospital ownership, or insurer changes in market share. However, because of the complexity medical procedure price, this cannot be accomplished using regression alone. So, I undertake a series of steps to construct prices that reflect, as much as possible, differences between hospitals in price associated with bargaining ability. Specifically, I do the following: 1) pull procedures using a limited set of procedure and diagnostic codes to mitigate heterogeneity between hospitals in procedure complexity; 2) adjust prices for patient characteristics and service setting; 3) average prices to the hospital-procedure-year level to equally weight hospitals; 4) match treatment hospitals to control group hospitals with similar facility and market characteristics to better balance pre-treatment characteristics; and 5) estimate a multivariate regression with additional hospital and market controls. Each of these steps is described in greater detail below.

Data

This paper uses data from the Health Care Cost Institute (HCCI) database of commercial claims for the years 2010-14. Data come from three health insurers: Aetna, Humana, and UnitedHealthcare. The data contain claims line information for more than 50 million individuals, representing about 25 percent of all individuals younger than 65 with employer-sponsored insurance. The percent of individuals captured in any one state, however, varies. For example, all Blue Cross and Blue Shield plans are excluded; states with large shares of enrollees in these plans have less coverage in the HCCI data. For the states that are the focus of my primary analysis here, Colorado includes approximately 37 percent of all individuals younger than 65 with employer sponsored insurance in 2014, Massachusetts includes 14 percent, Oregon includes 17 percent. Effectively this variation in coverage means that, in estimation, Colorado makes up the majority of the sample. For each procedure, I require minimum volume thresholds; because of the larger sample in Colorado, more hospitals each year meet these volume thresholds in the state (even given Massachusetts slightly higher population).

Importantly, the HCCI database includes the transaction prices for each procedure. The data allow us to separate prices by provider using a National Provider Identifier (NPI), however, I cannot separate prices by insurer as there are no variables relating to the specific identity of the payer. While other studies have attempted to infer the identify of specific contracts (Cooper et al. 2018), in this paper I treat all contributors as a single entity and do not try to observe changes in prices among payers as a result of price transparency.

I merged HCCI data with the American Hospital Association Survey Dataset (AHA survey) for each of the relevant years on facility-level NPI. The AHA survey data contain a number self-

reported hospital characteristics and, relevant for this study, number of beds, whether the facility

is an academic medical center, location of the facility, and system membership.

Procedure Measures

The data allow us to pull out prices for discrete procedures. I focus on 5 hospital procedures that

are common in this population, that are nearly always done in a hospital, and that differ in the

extent of potential consumer information about price. I identified medical services for inclusion

based on several criteria. First, I used DRG and CPT codes to define homogeneous hospital

procedures and exclude similar procedures with complications. Second, I search all commercial

(and under 65) HCCI claims in relevant states to ensure sufficient population volumes for

reliable price reporting. Last, I examined service setting and chose only those procedures for

which >90% were done in a hospital to mitigate the differential effect across markets of

ambulatory surgery center competition for these procedures. Using these criteria, I look at prices

for uncomplicated normal childbirth, uncomplicated Cesarean section, hip replacement,

laparoscopic appendectomy and cholecystectomy (gall bladder removal). (See Table 3-1 for

exact codes and percent in each care setting.)

Any claim line containing the relevant procedural and diagnostic codes was included; claim lines

are summed to the claims level. Claims from people in unknown or short term health insurance

plans, who are older than 64 (35 for birth claims), male for birth claims, and younger than 18 are

excluded. Claims are also excluded when the enrollee state of residence did not match the

provider state (more likely out-of-network) to capture only those claims reflecting a negotiated

price.

Construction of Price Adjusted for Patient Characteristics

Prices are adjusted for patient characteristics to mitigate changes in price due to changes in patient or illness complexity. I used standard risk adjustment procedures to compare expected prices from a multivariate generalized linear regression incorporating patient age (in bands), gender, length of stay, Charlson comorbidity score, and service setting (inpatient, outpatient, emergency department). Following this methodology, each price is constructed as follows for each procedure separately:

$$P_i = \frac{P_i \ (observed)}{P_i \ (expected)} * population \ mean \ price_p \tag{8}$$

Thus, what I deem price in this case is actually the variance from average procedure price not explained by the observable patient-specific factors above. Similar methods of using the residual variation as a measure of price variation has been previously used with these data. (Cooper et al. 2018) and are often used to risk-adjust prices across service settings (e.g. Osborne et al. 2015). The adjusted compared to raw average transaction prices for each procedure are in Table 3-2. I collapse data to the hospital-year-procedure level so that each observation reflects an average annual price for a provider in a state. Facilities are removed from the sample if they did not have data for at least 4 of the 5 years included in our dataset and did not meet a minimum volume threshold for each procedure or episode (10/yr for births; 5/yr for other procedures). I cut any individual claims that were at the bottom 1% of the total price distribution, as these reflect likely coding errors (e.g. negative or implausibly low prices). For inpatient procedures,I also excluded those episodes for which length of stay was unreasonably long for an uncomplicated episode (> 5 days for childbirth; >7 days for surgeries).

80

Hospital and Insurer Size and Bargaining Power

A key covariate is the relative size of the hospital-insurer pairs, which is a major determinant of bargaining outcome. Because I have an almost fully balanced panel of hospitals, to the extent that relationships are fixed across the short panel (5 years), I am able to control for differences between them. Ideally, I would be able to observe the relative size of each hospital-insurer bargaining pair as well. The extent to which hospital size changes due, for example, to merger and acquisition activity, or to which insurer size changes, perhaps due to higher enrollment relative to other carriers, may lead to changes in price and confound our estimates. A difficulty with these covariates is the danger of reverse causality; changes in prices could make a merger more likely or influence the benefit package offered by an insurer to attract (or dissuade) enrollees. Given the choice between not including measures of size in the analysis and including them, I opt for the latter to adjust the estimates for these associations. However, these particular covariates should not be considered causal.

One of the primary ways in which a hospital would change its size and bargaining power would be by joining or forming a health system. Using system identifiers in the AHA data, I include such controls. Relative changes in insurer size and power are more difficult to measure. Each of the three data contributors to HCCI may have different relative amounts of negotiating power depending on the market. However, it is not feasible to separate the three contributors with the data structure as I do not have payer or contract identifiers. I can measure the combined share of all HCCI contributors (UnitedHealthcare, Aetna, Humana) at each hospital in each year. Changes across shares in each year, then, reflect the combined gain or loss of market share among the insurers but mask offsets among the three insurers. While this measure is far from

ideal, nearly all data suffer from similar issues. Even if I could separate the data into the three companies, differences in contracting across business lines within insurers may mask fluctuations in relative power within insurers. To create the measure, I count all inpatient admissions and outpatients visits per hospital per year in the HCCI data, and divide this number by the total number of admissions as reported in the AHA survey for the year. The average fraction of admissions in a state in 2010 (baseline year) is reported in Table 3-3.

Matching and Estimation

Hospitals and hospital markets vary across the nation, and are not similarly distributed across states. Table 3-4 shows the comparison of hospital and market characteristics of hospitals in treated states (Colorado, Massachusetts, Oregon) to hospitals not in those states. From the table, it's clear that treatment state hospitals tend to be bigger and more likely in concentrated markets, which may affect both their baseline levels and, more crucially, price trends over time.
To better control for trends in hospital prices over time, I matched each hospital or hospital market in a treatment state to a control hospital or hospital market for individual procedures using a nearest-neighbor matching approach with two matches for each hospital. For price regressions, I matched on facility and system HHI, the baseline fraction of HCCI claims in a hospital, the total number of beds and whether hospitals were academic medical centers or belonged to systems. As Table 3-4 shows, the matching of hospitals is particularly helpful in creating a control group with adjusted prices much more closely matched to those in the treatment group.
For measures of dispersion across markets, I matched on similar fields at a market level: facility and system HHI at baseline, the percent of hospitals in a system and the number of academic

82

medical centers in a market. Main estimation results use matched samples, though I constructed some outcomes on the full sample for comparison purposes.

I estimate effects by exploiting the panel nature of the dataset to compare prices at hospitals before and after the release of an APCD. These prices are compared against states that either had an APCD throughout the study period, or implemented an APCD but did not release information from that database. As noted above, states comprising the treatment group include Colorado, Oregon and Massachusetts. All other states form the control group with two exceptions. I exclude both Maryland and West Virginia, as both have longstanding legislation that constrains hospital price negotiations in the commercial market.

Prices for these procedures tend to be right-skewed (Figure 3-2) arguing for a transformation of the dependent variable (Deb, Norton, and Manning 2017). For my main price estimation, I log the price and used a fixed effect linear regression model to take advantage of the panel structure and control for time invariant price shifters. The interpretation of results is in semi-elasticities. To determine the exact specification, I test the primary regression specifications, for log price and coefficient of variation, for inclusion of covariates. Specifically, I used the AIC and BIC information criteria to determine whether to use HHI at the facility level, system level, or both, and whether to interact the percent of admissions from HCCI contributors with measures of market structure. Once I narrowed down the specifications to two nested candidates, I used a Wald test to determine whether including interaction terms enhanced explanatory power (Norton 2016).

For the main estimation, I used the following fixed-effects linear regression, where the primary specification was the natural log of the adjusted price:

$$\ln(adj\_price_{jkt}) = \alpha + \beta_1 APCD_{release} + \beta_2 X_{jkt} + B_3 Z_{kt} + \gamma_{kj} + \delta_t + \epsilon_{jkt} \tag{9}$$

In this equation:

- adj_price_{jkt} is the adjusted price of procedure j in hospital k in a given year t; in the primary specification price is logged to mitigate outlier effects
- $APCD_{release}$ is an indicator for any year *after* a state gains an APCD with info release
- X_{jkt} are hospital-procedure-year level covariates: HHI Index at the facility and HRR-system level
- Z_{kt} are hospital-year covariates: percent of total admissions from HCCI data contributors, whether hospital is part of a system
- γ_{kj} are hospital-procedure fixed effects
- δ_t are year fixed effects

The indicators for the gain of an APCD lags each effect by one year as I assume that prices would be negotiated and could begin to change in the year after information is available. Thus, Oregon and Massachusetts, which implemented in 2012, have two years of post-period data (2013, 2014) and Colorado, which implemented in 2013, has one year of post-period data (2014). Each regression is run on the full matched sample, and on subgroups based on level of market concentration, split into the bottom quartile, middle two quartiles, and top quartile of concentration of markets in the treatment group. Standard errors are clustered at the facility level.

To measure dispersion of prices, I look at the coefficient of variation at the market level, which I define as a Hospital Referral Region (HRR), regional healthcare markets that perform tertiary care defined by The Dartmouth Atlas. The United States is divided into 306 HRRs; there are 13 either wholly or mostly located in treatment states. HRRs that cross state lines were included when more than half of the hospitals in the HCCI sample were located within a treatment state. I constructed the procedure-specific coefficient of variation (COV) for each HRR in each year.

HRRs with only one provider for a given service are dropped from this analysis (6.6% of the

sample). As with the regressions of price, I used a generalized difference-in-differences

structure to compare the COV for markets. I used AIC and BIC information tests to determine

the structure of covariates, specifically whether or not to interact the count of hospitals in an

HRR with the percent of total admissions from data contributors. The exact specification of the

regression is:

$$cov_{mjt} = \beta_1 APCD_{release} + \beta_2 X_{jkt} + B_3 Z_{kt} + \gamma_{kj} + \delta_t + \epsilon_{jkt} \qquad (10)$$

- cov_{mjt} is the market-level coefficient of variation for procedure j in market m in year t
- $APCD_{release}$ is an indicator for any year *after* a state gains an APCD with info release
- X_{jkt} are hospital-procedure-year level covariates: HHI Index at HRR-system level
- Z_{kt} are hospital-year covariates: number of hospitals that are part of systems, a count of hospitals, percent of total admissions from HCCI data contributors, and an interaction between hospital count and percent of total admissions from HCCI data contributors
- γ_{kj} are market-procedure fixed effects
- δ_t are year fixed effects

In perfectly competitive markets with homogeneous goods, price dispersion is 0. However, in

markets measured in these data, which are both differentiated and imperfectly competitive,

markets that are relatively less concentrated have higher levels of dispersion at baseline. To get

an idea of whether the variance in prices is changing in different types of markets and to create a

basis to compare the dispersion results to price results, I report estimates for the full sample and

stratified by market along the same definitions as the price results. Standard errors are clustered

at the market level.

Finally, to measure whether potential changes in dispersion occurred at an end of the price distribution and, if so, which end, I used a similar difference-in-difference strategy with a quantile regression. For this specification, I use the same specification as the main price results, with a logged price, but instead of regressing at the mean, I regress for an effect at different quantiles of the distribution (.10, .15, .25, .5, .75, .85, .90). Additionally, I use an interaction with baseline market concentration in these equations to examine results across various market concentrations. The final, interacted equation is:

$$\ln\left(adj_{price_{jkt}}\right)_q = \alpha + \beta_1 APCD_{release} * HHI_{jkt} + \beta_2 X_{jkt} + B_3 Z_{kt} + \gamma_{kj} + \delta_t * HHI_{jkt} + \epsilon_{jkt} \quad (11)$$

This equation uses the same covariates as the price equation above, and clusters standard errors at the hospital-procedure level.

RESULTS

Prices: Cross-Sectional and Descriptive Results

To get an idea of whether prices in states that publicly release prices are higher or lower than in other states, I first look cross-sectionally using 2014 data at all states with public datasets based on an APCD compared to those that have no such dataset. That includes treatment states, and additionally New Hampshire and Maine, which released information to hospitals from their APCDs prior to 2014. That regression shows that prices are higher in these states by an estimated $1,300 (Table 3-5), and that publicly releasing prices is correlated with higher prices for hospital procedures. However, given that there may be selection bias in the types of states

that chose to release prices from APCDs (perhaps the price differential was higher than $1,300

prior to APCD implementation), this correlation does not suggest a causal relationship.

To examine changes in price, and move closer to estimating causality, we first begin by looking

at prices descriptively, without using a formal estimation model.   Prices increase gradually in all

states from 2010 to 2014 (Figure 3-3) and are always higher in treatment states then in the full

control group. Prices in the matched control group, as the graph shows, are much closer in

levels, and trends diverge beginning after 2012, the year Oregon and Massachusetts released

prices from their APCDs. Split by relative level of market concentration (Figure 3-4), the

increase in prices is most apparent in the markets in the bottom quartile of concentration (HHI <

2700). Prices increase much more slowly markets in the middle two quartiles (2700-5400) and

in the highest quartile of market concentration (HHI > 5400). Figure 3-5 splits price levels by

procedure, which shows average levels of prices in treatment and matched control hospitals for

each year.

Prices: Estimation Results

Using Equation 9, I estimate whether there is a change in price in treatment states using a model

with fixed effects for each hospital-procedure pair and year, and controlling for the covariates as

described above. In levels, I find a positive estimate ($103.74 increase in post period) but a large

standard error that does not allow us to rule out no effect or even a drop in price (Table 3-6).

For the main log specification, I again find a positive estimate (0.019; about a 2 percent increase

in prices in the post period). However, a large standard error, as with the previous estimate,

means I cannot distinguish this result from 0 or even a negative number.

Sensitivity checks on these main estimates are shown in Tables 3-7 and 3-8. Table 3-7 uses the full control sample (all hospitals but treatment state and WV or MD) and finds the same thing, with smaller estimates and larger p-values, suggesting no detectable results. Table 3-8 uses a standard difference-in-difference framework with one pre and one post period instead of fixed effects. All show positive but imprecise estimates on the main effect of post-APCD prices.

Stratification by Market Concentration and Shoppability

Stratified by market concentration, I find the positive main results are driven by increases in prices in the lowest and middle concentrations of markets (Table 3-9). For the least concentrated markets, I find a 2.2% increase in prices; for moderate concentration I find a 3.6% increase. Both results, however, lack precision enough to confidently distinguish the estimate from 0. In the highest concentration markets (HHI > 5400) the change in prices is estimated to be negative but, again, is imprecise. These results do suggest, however, that there may be different effects of price transparency depending on the structure of the market.

Using Colorado's data only, I also stratified procedures according to whether their prices were published on a consumer-friendly website (childbirth, Cesarean sections, hip replacement) or available through dataset request only (appendectomy, gall bladder removal). As Table 3-10 shows, I estimate prices for shoppable procedures dropped in the post period by 5.4 percent compared to those procedures for which pricing information was not available to consumers. Prices for the non-shoppable procedures rise by 4.3 percent, though as with the other estimates, it is not statistically significant. A comparison of prices for shoppable and non-shoppable services directly strongly suggests a difference in effects (p=0.067), in which shoppable procedures show slightly negative results and non-shoppable procedures increase in price. While these results are

only suggestive, as the confidence intervals run through 0 and sample sizes are small, they do suggest a difference in how hospitals respond to transparency based on how prices are released.

Quantile Regression: Changes at Non-Mean Points

I used quantile regression with a similar structure (Equation 11) to look for changes in the log price distribution at points off the mean. These models show prices increase at the lower ends of the price distribution post-APCD much more than they do at higher ends of the distribution. Specifically, the model estimates increases in prices at the lower end of the price distribution post-APCD for all hospitals, with estimates of an increase in prices between 1.4 and 1.7 percent in the 10th through 50th quantiles of the distribution (Table 3-11). Statistical significance varies for each of these estimates. Price increases at the 15th quantile are statistically significant and at the 25th quantile are marginally statistically significant (Figure 3-6, panel a).

Stratified by market concentration, the increase log price at lower parts of the distribution is most evident in the least concentrated markets, particularly at the 15th and 25th quantile of price distribution where prices are estimated to increase 3.4 and 3.3 percent respectively (Table 3-11). Estimates for moderately and highly concentrated markets are generally lower and less precise at this part of the distribution, although there is some increase in the most concentrated markets at the lowest end of the distribution (Figure 3-6, panels b-d). For moderately concentrated markets, prices appear to increase post-APCD at the highest part of the distribution. However, sensitivity checks with placebo cutoff points in earlier years show prices increase at the high end of the distribution in earlier years for these markets (Table 3-12), and it is likely this result is part of a differential trend in prices in treatment hospitals in moderately concentrated markets rather than a true effect of the APCD price release. By contrast, in the low concentration markets, estimates

89

in the pre-period for hospitals at the lower end of the distribution are small and not significantly different from 0, suggesting the release of prices through the APCD may have changed the price distribution in these markets, shifting prices up at the lower end.

Coefficient of Variation

A second primary outcome is the coefficient of variation (COV), a measure of dispersion of prices for each procedure across each market. COVs are based on the risk-adjusted price and the sample unit for these results is the procedure-market-year; these samples are smaller than the samples used in the price models. Additionally, as described above, I matched the control sample at the market level so that it may include different hospitals than the sample above. Raw COVs for each year and each procedure are shown in Figure 3-7; a lower COV means that prices are clustered more tightly around a mean.

As with prices, I first examine the cross-sectional differences in COV for markets with publicly released prices (CO, MA, OR, NH, ME). Results from that regression (Table 3-5) show that public release states have a COV 0.028 points lower than states that do not have an APCD. However, as above, this regression only shows a correlation between lower price spread and public release of prices; it does not establish whether the APCD has an effect on that dispersion. To look at whether the APCD effects price dispersion, we first examine the change in the COV over time (Figure 3-8). That figure shows that, without adjustment, dispersion in treatment states drops more quickly than dispersion in either the full set of control markets or the matched control. It appears, however, to begin to drop before the implementation of the first APCDs in 2012. Split by market concentration (Figure 3-9), both the least and most concentrated markets show some drops in price dispersion. In those markets with moderate concentration, the

treatment markets do not appear to vary from control throughout the data period. Note these figure also show variance in initial COV by market; less concentrated markets have higher COVs, on average, than more highly concentrated markets.

Estimates from a fixed effects model, with fixed effects at the procedure-market and year levels finds a decrease in the COV in the post period of 0.034 points, though the large standard error does not allow us to rule no, or even a small positive effect (Table 3-13). Sensitivity checks on this main effect (Table 3-14), also show small negative and imprecise estimates. Additionally, running placebo cutoff points (beginning all post in 2011 or 2012), we find larger and statistically significant effects in these years, suggesting the estimates may represent differential trends in COV, and that it may be hard to disentangle the main effect (Table 3-15). Stratified by market concentration, we find a drop in COV of 0.071 points in the least concentrated markets (Table 3-16), a relative 19% decrease in dispersion, though it is not statistically significant at conventional levels (p=0.068). For moderately and highly concentrated markets, estimates are negative and smaller. Standard errors indicate these estimates are imprecise and do not allow rejection of the null hypothesis of no effect.

DISCUSSION

These results indicate that releasing price information does not affect mean equilibrium prices, at least in the short term. The effects for the main outcome—mean price—are small and imprecise, indicating either that there are few effects or that they could not be detected in this study. However, looking across quantiles of price, there is suggestive evidence that prices in the lower half of the price distribution increase slightly following release of an APCD, and that these effects may be primarily driven by hospitals in relatively less concentrated markets. Dispersion

of prices in markets shrinks following release of the APCD, although sensitivity checks indicate this trend may predate, and thus not be caused by, APCD implementation.

Relating these results back to the predictions from the bargaining model, the estimates do suggest that some hospitals, particularly those at the lower end of the distribution, are able to use information to gain bargaining advantage. I don't find evidence of lower prices for those hospitals that had initially high prices, suggesting insurers either do not gain new information or are unable to use it. It may be that these hospitals have other attributes that override the effects of increased transparency, perhaps the attributes that gave rise to higher prices in the first place. Additionally, I find stronger effects on prices and dispersion in the markets that are least concentrated (i.e. more alternative providers), where information was predicted to have the largest effect. Finally, these results suggest hospitals are selectively raising these prices where consumers are least likely to notice them, in surgeries that are not typically considered shoppable procedures and for which consumer information is not readily available. This behavior is consistent with strategic use of the information in a bargaining setting.

While the results imply a pattern of behavior by hospitals, they should be considered suggestive because of a number of data limitations. The most important is that the panel may be too short to show meaningful change. Within the time period covered by the data, only 3 states switch to an APCD and, in the state with the most claims, we have only one year of post-period data. Hospital-insurer contracts are typically negotiated every 1-3 years, so 1 year of post-period data may not be enough to capture the effect of potential prices changes. Longer panels of data would be helpful as would data with more coverage across all states. Second, it would be ideal to have additional states, as some states have confounding effects. In Massachusetts, for example, legislation passed in 2012 implements price controls and encourages the formation of alternative

payment methodologies, among other things. However, the former provision lacks enforcement power until 2016, and state analyses show the health care cost growth rate was approximately the same as the full U.S. through 2014 (Bump 2017). Finally, we do not know what level of information about competitors' prices hospitals had in the pre-period, nor exactly which hospitals accessed the new information in the post-period. In that way, the intervention can be thought of as giving access to information, rather than a test of whether use of information causes price change. Whether access to or use of information is more important from a policy perspective is arguable; theoretical predictions from economics tend to focus on the use of information to change behavior rather than access to it. Nevertheless, some measure of whether hospitals used the information provided, and attributes of hospitals more likely to use it, would be helpful in future analysis.

Despite these limitations, this paper provides some of the first evidence that price transparency efforts may be used strategically by hospitals to increase their bargaining ability and raise prices for hospital services. These effects are most apparent in procedures that consumers are unlikely to shop for and in among lower-priced providers in less concentrated markets. However, these may be the markets we may want to protect as they represent the best opportunity for consumers to save money within the health care system. More research should be done to confirm these findings and, if confirmed, efforts may want to focus on how to release information from APCDs or other accessible databases in a way that does not allow for the potential for increased prices for consumers.

Table 3-1: Procedures examined, medical codes and percent in each setting

	DRG Codes Used	CPT/ICD9CM, HCPCS Codes Used	ICD-9 Codes	Percent Inpatient 2010	Percent Inpatient 2014	Percent Outpatient 2010	Percent Outpatient 2014	Percent ASC 2010	Percent ASC 2014
Normal Childbirth*	775 (vaginal delivery without complications)	7359 (manually assisted delivery), 7569 (repair obstetrical laceration)		99.7	99.5	0.024	0.013	0.39	0.44
Cesarean Section Childbirth	766 (Cesarean delivery without complications)	741 (Cesarean section)		99.9	99.9	0.035	0.033	<0.001	0.0001
Appendectomy*		4701 (laproscopic appendectomy). 44970 (laproscopic procedure of the appendix)	5409 (acute appendicitis), 540 (acute appendicitis),541 (appendicitis NOS)	55.7	33.8	43.5	65.7	0.75	0.47
Gall Bladder Removal		5123 (cholecystectomy, laproscopic surgical), 47563 (cholecystectomy with cholangiography), 47564 (cholecystectomy with exploration of common duct)		28.7	27.5	62.6	65.1	8.8	7.4
Hip Replacement	470 (Major joint replacement or reattchment of lower extremity)	8151 (total hip replacement)		99.7	99.9	0.29	0.07	<0.001	0.01

*For normal childbirth, the DRG was combined with one of the two procedure codes to mitigate price differences by intensity of service use. **For Cesarean section childbirth, DRG plus the CPT was required ***For appendectomy, one of two procedure codes was combined with at least one of the ICD-9 codes ****For gall bladder removal, we only required one of the listed procedure codes. The diagnostic reason for the removal did not change price in tests or price listings

Table 3-2: Coefficient of variation, raw prices and adjusted prices in baseline year

	Treatment (CO, MA, OR) Hospitals	Control Hospitals	Matched Control Hospitals	p-value (Difference between full control and treatment)	p-value (Difference between matched control and treatment)	Number hospitals treatment	Number of hospitals control	Number of hospitals matched control
Childbirth-delivery								
HRR Coefficient of Variation	0.26	0.23	0.23	0.10	0.28	70	1180	89
Adjusted Baseline Price	5409.66	5092.63	5391.51	0.13	0.95	75	1249	95
Raw price (not risk adjusted)	5442.04	5060.33	5333.73	0.07	0.68	75	1249	95
Cesarean Section Delivery								
HRR Coefficient of Variation	0.27	0.22	0.21	0.01	0.03	37	632	51
Adjusted Baseline Price	9430.86	7356.25	8164.84	<0.001	0.02	46	708	59
Raw price (not risk adjusted)	9616.29	7299.27	8064.71	<0.001	0.01	46	708	59
Appendix Removal								
HRR Coefficient of Variation	0.60	0.47	0.44	<0.001	0.00	36	714	53
Adjusted Baseline Price	6473.69	6067.21	6578.58	0.51	0.88	40	783	61
Raw price (not risk adjusted)	5720.99	6638.02	7558.79	0.16	0.03	40	783	61
Gall Bladder Removal								
HRR Coefficient of Variation	0.40	0.44	0.41	0.04	0.60	74	1729	119
Adjusted Baseline Price	7109.14	6483.42	7761.75	0.15	0.30	78	1763	125
Raw price (not risk adjusted)	6862.99	6606.37	8030.07	0.56	0.06	78	1763	125
Hip Replacement								
HRR Coefficient of Variation	0.28	0.25	0.27	0.33	0.91	23	465	36
Adjusted Baseline Price	28398.78	25755.67	29184.62	0.18	0.81	29	544	44
Raw price (not risk adjusted)	28629.63	25908.05	29205.54	0.17	0.86	29	544	44

Notes: Matched control group constructed using nearest neighbor matching on fraction of admissions and outpatient visits from HCCI contributors at baseline, bed total in baseline year, adjusted price in baseline year, HHI (by hospital) in baseline year, HHI (by system) in baseline year, and matched exactly on whether the facility is an academic medical center or belongs to a system. Baseline year is first year hospital is in data, either 2010 or 2011. HRR is Hospital Referral Region, which represent tertiary care referral regions and are constructed by The Dartmouth Atlas (dartmouthatlas.org). Adjusted price is constructed following previous literature to generate a ratio of observed price to the expected price based on patient characteristics, and multiplied by a scalar, the overall mean.

Table 3-3: Average percent of discharges from HCCI contributors in each state

State	Average Fraction of Claims from HCCI		State	Average Fraction of Claims from HCCI
AK	0.043		MT	0.012
AL	0.080		NC	0.039
AR	0.038		ND	0.019
AZ	0.104		NE	0.093
CA	0.046		NH	0.043
CO	0.114		NJ	0.071
CT	0.067		NM	0.044
DC	0.093		NV	0.068
DE	0.053		NY	0.046
FL	0.097		OH	0.078
GA	0.103		OK	0.086
IA	0.058		OR	0.035
ID	0.026		PA	0.065
IL	0.057		RI	0.130
IN	0.048		SC	0.056
KS	0.089		SD	0.013
KY	0.111		TN	0.053
LA	0.074		TX	0.126
MA	0.041		UT	0.051
ME	0.046		VA	0.070
MI	0.057		VT	0.010
MN	0.081		WA	0.041
MO	0.075		WI	0.110
MS	0.065		WY	0.032

Notes: Fraction of HCCI claims created by dividing total admissions and outpatient visits per year recorded in claims database from HCCI by total admissions and outpatient visits reported by hospital in AHA annual survey. State means are across all years in data (2010-14). Total observations=25,952. Hospitals top-coded at 0.9.

Table 3-4: Characteristics of hospitals used in matching analysis in baseline year (2010 or 2011)

	Treatment (CO, MA, OR) Hospitals	Control Hospitals (Full Sample)	Matched Control Hospitals	p-value (Difference between Treatment and full Control)	p-value (Difference between treatment and matched control)	Number hospitals treatment	Number of hospitals full control	Number of hospitals matched control
Total Beds (Mean)	256.40	261.50	248.06	0.83	0.74	88	1902	143
HHI (by system) (Mean)	3638.30	3499.66	3950.90	0.59	0.55	88	1902	143
HHI (by hospital) (Mean)	2542.11	2769.95	2924.55	0.39	0.46	88	1902	143
Academic Medical Center (share)	0.14	0.10	0.12	0.34	0.70	88	1902	143
System Membership (share)	0.67	0.68	0.68	0.81	0.90	88	1902	143
Average Share of Patients from HCCI contributors								
Normal Delivery	0.0735	0.0716	0.0637	0.85	0.34	75	1249	95
Cesarean Section Delivery	0.0802	0.0928	0.0811	0.41	0.95	46	708	59
Appendix Removal	0.0964	0.0926	0.0815	0.78	0.31	40	783	61
Gall Bladder Removal	0.0692	0.0707	0.0680	0.88	0.90	78	1763	125
Hip Replacement	0.1061	0.1016	0.0983	0.80	0.74	29	544	44
Adjusted Baseline Price								
Normal Delivery	5409.66	5092.63	5391.51	0.13	0.95	75	1249	95
Cesarean Section Delivery	9430.86	7356.25	8164.84	0.00	0.02	46	708	59
Appendix Removal	6473.69	6067.21	6578.58	0.51	0.88	40	783	61
Gall Bladder Removal	7109.14	6483.42	7761.75	0.15	0.30	78	1763	125
Hip Replacement	28398.78	25755.67	29184.62	0.18	0.81	29	544	44

Notes: Matched control group constructed using nearest neighbor matching on fraction of admissions and outpatient visits from HCCI contributors at baseline, bed total in baseline year, adjusted price in baseline year, HHI (by hospital) in baseline year, HHI (by system) in baseline year, and matched exactly on whether the facility is an academic medical center or belongs to a system. Baseline year is first year hospital is in data, either 2010 or 2011. Adjusted price is constructed following previous literature to generate a ratio of observed price to the expected price based on patient characteristics, and multiplied by a scalar, the overall mean. HHI is constructed by market, in which each market is defined as a Dartmouth Atlas HRR region (dartmouthatlas.org)

Table 3-5: Cross-sectional regression results, APCD states with public release compared to states without APCD or APCD without public release

Risk-Adjusted Price (2014)				Coefficient of Variation (2014)			
	Coefficient	Standard Error	P-value		Coefficient	Standard Error	P-value
APCD state	1281.61	205.71	0.000	**APCD state**	-0.028	0.008	0.001
Procedure (Childbirth)	ref			**Procedure (Childbirth)**	ref		
Procedure (Csec)	2826.50	139.32	0.000	**Procedure (Csec)**	-0.009	0.007	0.198
Procedure (Appendix removal)	523.32	134.68	0.000	**Procedure (Appendix removal)**	0.099	0.007	0.000
Procedure (Gall bladder removal)	1432.85	114.39	0.000	**Procedure (Gall bladder removal)**	0.122	0.005	0.000
Procedure (Hip replacement)	22849.81	440.27	0.000	**Procedure (Hip replacement)**	0.004	0.007	0.547
Fraction with HCCI contributor admissions	-2621.04	683.96	0.000	**HHI (facility based)**	0.000	0.000	0.000
HHI (facility based)	0.02	0.02	0.452	**Hospital Count**	0.000	0.000	0.208
System membership	904.99	146.90	0.000	**Constant**	0.270	0.009	0.000
Academic medical center	773.09	202.50	0.000				
Sole community provider	518.68	390.13	0.184				
Bed total	0.67	0.25	0.007				
Constant	10056.64						
Total Observations	5,315			**Total Observations**	4,985		

Notes: Risk-adjusted price modeled using a generalized linear model with a log link and gamma family. Marginal effects are reported. Coefficient of variation modeled using ordinary least squares regression. States with an APCD that released data in 2014 include CO, MA, OR, ME and NH. Standard errors are clustered at the hospital level for prices and HRR level for COV.

Table 3-6: Regression estimates of price changes

	Price in levels			Price in Logs		
	Coefficient	Standard Error	P-value	Coefficient	Standard Error	P-value
Post X Treatment	103.74	166.83	0.535	0.019	0.016	0.234
Year: 2010	ref			ref		
Year: 2011	722.61	102.01	0.000	0.082	0.009	0.000
Year: 2012	914.53	144.09	0.000	0.112	0.013	0.000
Year: 2013	1490.78	150.80	0.000	0.182	0.014	0.000
Year: 2014	1745.20	173.35	0.000	0.211	0.016	0.000
Percent of Claims in HCCI	-7.10	13.31	0.594	0.000	0.001	0.999
HHI (hospital)	-0.12	0.20	0.559	0.000	0.000	0.946
HHI (system)	0.08	0.21	0.697	0.000	0.000	0.898
System Membership	-4.34	368.68	0.991	0.028	0.034	0.417
Constant	9379.19	541.95	0.000	8.853	0.061	0.000
N (Number of observations)	3187			3187		
Number of groups	652			652		

Notes: Estimates are from a fixed effects model, with fixed effects for each hospital-procedure pair. The coefficient of interest is on the interaction of Post x Treatment, which measures changes in risk-adjusted price in hospitals the year after a state publicly releases price information through an APCD. Estimates based all hospitals in treatment states that met volume thresholds for each procedure, compared with hospitals in control states matched on fraction of admissions and outpatient visits from HCCI contributors at baseline, bed total in baseline year, adjusted price in baseline year, HHI (by hospital) in baseline year, HHI (by system) in baseline year, whether the facility is an academic medical center or belongs to a system. Standard errors clustered at the hospital level.

Table 3-7: Sensitivity check: price estimates using full sample as control

	Price in levels			Price in logs		
	Coefficient	Standard Error	P-value	Coefficient	Standard Error	P-value
Post X Treatment	38.67	132.25	0.770	0.0029	0.0133	0.829
Year: 2010	0.00			0.0000		
Year: 2011	477.97	33.81	0.000	0.0668	0.0033	0.000
Year: 2012	786.98	45.41	0.000	0.1108	0.0046	0.000
Year: 2013	1308.16	49.61	0.000	0.1779	0.0050	0.000
Year: 2014	1728.55	56.60	0.000	0.2293	0.0055	0.000
Percent of Claims in HCCI	-7.62	6.35	0.230	-0.0008	0.0006	0.171
HHI (hospital)	-0.22	0.09	0.020	0.0000	0.0000	0.791
HHI (system)	0.20	0.08	0.008	0.0000	0.0000	0.429
System Membership	-29.52	173.37	0.865	0.0038	0.0165	0.820
Constant	8214.23	225.04	0.000	8.7496	0.0208	0
N (Number of observations)	25,952			25,952		
Number of groups	5,315			5,315		

Notes: Estimates are from a fixed effects model, with fixed effects for each hospital-procedure pair. The coefficient of interest is (Post X Treatment), which measures changes in risk-adjusted price in hospitals the year after a state publicly releases price information through an APCD. Estimates based all hospitals in treatment states that met volume thresholds for each procedure, compared with all hospitals in control states. Standard errors clustered at the hospital level.

Table 3-8: Sensitivity check: difference-in-differences models with price in levels, logs and predicted via GLM

	Price in Levels			Price in Logs			GLM marginal effects		
	Coefficient	Standard Error	P-value	Coefficient	Standard Error	P-value	Coefficient	Standard Error	P-value
Post x Treatment	39.99	392.66	0.919	0.0032	0.03	0.915	134.52	294.47	0.648
Treatment group	188.07	560.52	0.738	0.02	0.05	0.739	16.33	184.21	0.929
Year: 2010	0.00						0.00		
Year: 2011	682.73	112.03	0.000	0.08	0.01	0.000	782.67	251.05	0.002
Year: 2012	869.13	141.15	0.000	0.11	0.01	0.000	961.92	234.97	0.000
Year: 2013	1352.11	167.02	0.000	0.17	0.02	0.000	1492.74	241.06	0.000
Year: 2014	1688.30	206.54	0.000	0.20	0.02	0.000	1834.81	273.90	0.000
Childbirth	ref			0.00			0.00		
Cesarean Section Childbirth	3693.71	351.62	0.000	0.47	0.03	0.000	3718.65	161.49	0.000
Appendix Removal	1152.51	326.48	0.001	0.12	0.04	0.005	1158.95	169.09	0.000
Gall Bladder Removal	1976.92	260.84	0.000	0.21	0.03	0.000	1913.51	147.34	0.000
Hip Replacement	24354.12	1309.22	0.000	1.56	0.04	0.000	23970.06	685.28	0.000
Percent of Claims in HCCI	9.55	78.89	0.904	-0.0037	0.00	0.257	-31.60	17.09	0.064
HHI (hospital)	0.43	0.29	0.129	-0.00001	0.00	0.745	0.02	0.08	0.772
HHI (system)	-0.15	0.27	0.564	0.00002	0.00	0.275	0.20	0.09	0.034
System Membership	989.04	761.17	0.195	0.16	0.06	0.005	1348.60	233.21	0.000
Constant	3773.85	606.18	0.000	8.38	0.06	0.000			
N (Number of observations)	3187			3187			3187		

Notes: Each model uses a standard difference-in-differences framework where the treated group comprises hospitals in CO, MA, and OR and the control group is a sample of hospitals in other states matched to treatment group hospitals on fraction of admissions and outpatient visits from HCCI contributors at baseline, bed total in baseline year, adjusted price in baseline year, HHI (by hospital) in baseline year, HHI (by system) in baseline year, whether the facility is an academic medical center or belongs to a system. Standard errors clustered at the hospital level.

Table 3-9: Regression estimates of change in price levels stratified by market concentration

	Price in Levels			Price in Logs		
	Coefficient	Standard Error	P-value	Coefficient	Standard Error	P-value
HHI < 2700						
Post x Treatment	57.66	180.67	0.750	0.022	0.023	0.360
Year: 2010	0.00			0.000		
Year: 2011	460.34	135.29	0.001	0.067	0.016	0.000
Year: 2012	725.66	175.03	0.000	0.125	0.021	0.000
Year: 2013	1279.06	192.81	0.000	0.209	0.023	0.000
Year: 2014	1474.88	176.73	0.000	0.236	0.026	0.000
Percent of Claims in HCCI	3.04	18.69	0.871	0.000	0.002	0.943
HHI (hospital)	1.03	1.06	0.331	0.000	0.000	0.164
HHI (system)	0.35	0.41	0.401	0.000	0.000	0.300
System Membership	-18.75	398.07	0.963	-0.019	0.045	0.672
Constant	4971.99	1429.53	0.001	8.619	0.127	0.000
N (Number of observations)	1287			1287		
Number of groups	266			266		
HHI = 2700 - 5400						
Post x Treatment	326.59	281.74	0.248	0.036	0.024	0.145
Year: 2010	0.00			0.000		
Year: 2011	810.90	152.32	0.000	0.096	0.013	0.000
Year: 2012	1124.94	212.35	0.000	0.122	0.018	0.000
Year: 2013	1568.18	226.07	0.000	0.179	0.020	0.000
Year: 2014	2042.29	323.39	0.000	0.212	0.026	0.000
Percent of Claims in HCCI	-0.65	9.52	0.946	0.000	0.001	0.779
HHI (hospital)	-0.26	0.35	0.466	0.000	0.000	0.683
HHI (system)	0.21	0.30	0.471	0.000	0.000	0.267
System Membership	-394.99	758.35	0.603	0.028	0.059	0.635
Constant	9840.08	839.73	0.000	8.835	0.071	0.000
N (Number of observations)	1205			1205		
Number of groups	247			247		
HHI > 5400						
Post x Treatment	-179.43	370.02	0.629	-0.019	0.023	0.414
Year: 2010	0.00			0.000		
Year: 2011	1018.47	284.77	0.001	0.088	0.020	0.000
Year: 2012	792.54	388.95	0.046	0.077	0.025	0.003
Year: 2013	1606.63	375.50	0.000	0.148	0.028	0.000
Year: 2014	1607.47	411.22	0.000	0.179	0.027	0.000

Percent of Claims in HCCI	-37.57	55.88	0.504	-0.001	0.002	0.535
HHI (hospital)	0.28	0.48	0.558	0.000	0.000	0.548
HHI (system)	-0.54	0.53	0.312	0.000	0.000	0.277
System Membership	162.20	1376.09	0.907	0.117	0.040	0.005
Constant	15298.15	2310.84	0.000	9.288	0.235	0.000
N (Number of observations)	695			695		
Number of groups	144			144		

Notes: Estimates are from a fixed effects model, with fixed effects for each hospital-procedure pair. The coefficient of interest is on the interaction of Post x Treatment, which measures changes in risk-adjusted price in hospitals the year after a state publicly releases price information through an APCD. Estimates based all hospitals in treatment states that met volume thresholds for each procedure, compared with hospitals in control states matched on fraction of admissions and outpatient visits from HCCI contributors at baseline, bed total in baseline year, adjusted price in baseline year, HHI (by hospital) in baseline year, HHI (by system) in baseline year, whether the facility is an academic medical center or belongs to a system. Standard errors clustered at the hospital level. Stratification of HHI based on bottom quartile of distribution of treatment hospitals, middle two quartiles (25-75%) and top quartile. Standard errors clustered at facility level.

Table 3-10: Regression estimates of changes in price levels stratified by shoppability

	Coefficient (logged price)	Standard Error	P-value
Treatment Group (Colorado hospitals)	0.091	0.084	0.280
Shoppable services (normal childbirth, Csection childbirth, hip replacement)	0.383	0.077	0.000
Treatment x Shoppable	-0.094	0.116	0.420
Treatment x Post	0.043	0.050	0.391
Treatment x Post x Shoppable	-0.054	0.029	0.067
Year: 2010	ref		
Year: 2011	0.104	0.017	0.000
Year: 2012	0.156	0.024	0.000
Year: 2013	0.257	0.025	0.000
Year: 2014	0.315	0.048	0.000
Fraction of Claims in HCCI	0.011	0.004	0.013
HHI (facility-based)	-0.00004	0.00005	0.390
HHI (system-based)	0.00016	0.00006	0.008
System Membership	-0.038	0.092	0.680
Constant	7.965	0.132	0.000
N (Number of observations)	1115		
Notes: Estimates from a standard difference-in-differences model with one pre and one post period on hospitals in Colorado, which implemented a publicly released APCD and, additionally, a consumer facing website that reported provider-specific prices for certain procedures. Standard errors clustered at the facility level. Marginal change in shoppable services in post-period (-0.011); marginal change in non-shoppable services in post-period (0.043); p-value from Wald test of differences (0.067).			

Table 3-11: Quantile regression estimates of change in prices at percentiles of price distribution

Percentile of distribution	All hospitals		HHI < 2700		HHI 2700 – 5400		HHI > 5400	
	Marginal Effect	**p-value**	**Marginal Effect**	**p-value**	**Marginal Effect**	**p-value**	**Marginal Effect**	**p-value**
10th	0.014	0.085	0.008	0.839	-0.030	0.003	0.044	0.056
15th	0.017	0.032	0.034	0.030	-0.021	0.310	0.031	0.280
25th	0.014	0.063	0.033	0.065	0.005	0.644	-0.006	0.825
50th	0.016	0.100	0.012	0.562	0.023	0.150	0.005	0.867
75th	0.002	0.788	0.017	0.483	0.006	0.722	-0.008	0.623
85th	0.005	0.533	0.003	.847	0.023	0.021	-0.00056	0.979
90th	0.006	0.606	0.007	0.797	0.027	0.071	-0.014	0.550
Number of Observations	3187		3178		3178		3178	

Notes: Estimates are from a fixed effects model, with fixed effects for procedures and facilities. Table reports marginal effects on the log of risk-adjusted price in a year after a treatment state is treated at each specified quantile, with interactions used for analyses stratified by market concentration. Estimates based on all hospitals in treatment states that met volume thresholds for each procedure, compared with hospitals in control states matched on fraction of admissions and outpatient visits from HCCI contributors at baseline, bed total in baseline year, adjusted price in baseline year, HHI (by hospital) in baseline year, HHI (by system) in baseline year, whether the facility is an academic medical center or belongs to a system. Stratification of HHI based on bottom quartile of distribution of treatment hospitals, middle two quartiles (25-75%) and top quartile. Standard errors are robust and p-value taken from the marginal effect calculations to represent the probability of obtaining this estimate for all hospitals or each strata if the null (effect size = 0) is true.

Table 3-12: Sensitivity check: quantile regression estimates with placebo cutoff points

	HHI < 2700		HHI 2700-5400		HHI >5400	
Post Period in 2011	**Marginal Effect**	**p-value**	**Marginal Effect**	**p-value**	**Marginal Effect**	**p-value**
Quantile: .25	0.0074	0.757	0.038	0.001	0.032	0.281
Quantile: .50	-0.002	0.923	0.026	0.254	0.055	0.362
Quantile: .75	-0.017	0.521	0.018	0.373	0.018	0.52
Post Period in 2012						
Quantile: .25	0.003	0.825	0.028	0.004	0.018	0.422
Quantile: .50	0.016	0.331	0.033	0.061	0.003	0.944
Quantile: .75	0.017	0.364	0.036	0.001	0.02	0.228

Notes: Estimates are from a quantile regression of log price at the specified quantile on treatment states in the years after the placebo cutoff (2011 or 2012), category of concentration, an interaction of market concentration on treatment in post period, year, interaction between year and market concentration, procedure fixed effects, facility fixed effects, system membership, HHI at the facility level, and percent of discharges from the HCCI sample. Table reports marginal effects on the log of risk-adjusted price in a year after a treatment state is treated at each specified quantile and, for the stratification by market concentration, the marginal effect for that specific HHI level. Standard errors are robust.

Table 3-13: Regression estimates on coefficient of variation

	Coefficient	Standard Error	P-value
Treatment x Post	-0.034	0.020	0.087
Year: 2010			
Year: 2011	-0.015	0.010	0.152
Year: 2012	-0.044	0.016	0.010
Year: 2013	-0.018	0.018	0.308
Year: 2014	-0.014	0.021	0.501
HHI	0.000	0.000	0.001
System membership	0.020	0.049	0.682
Percent HCCI contributors	-0.011	0.005	0.029
Hospital count	0.021	0.016	0.186
Percent HCCI *Hospital count	0.002	0.001	0.040
Constant	0.286	0.108	0.011
N (Number of observations)	444		
Number of groups	89		
Notes: Estimates from a linear regression with hospital referral region (HRR)-procedure specific fixed effects. HRR were matched on number of academic medical center, total beds in HRR, percent of hospitals affiliated with a system, and baseline HHI at the facility and system level. Regression specification was determined using specification tests including AIC/BIC and a Wald test. Standard errors clustered at the HRR level.			

Table 3-14: Sensitivity checks on regression estimates on coefficient of variation: full sample as control and difference-in-differences estimates

	Full Sample as Control			Matched Sample, Diff-in-Diff Specification		
	Coefficient	Standard Error	p-value	Coefficient	Standard Error	P-value
Treatment x Post	-0.008	0.018	0.667	-0.019	0.024	0.440
Treatment group				-0.031	0.017	0.070
Year_2010	0			ref		
Year_2011	0.001	0.004	0.876	-0.013	0.023	0.566
Year_2012	-0.017	0.006	0.004	-0.042	0.023	0.063
Year_2013	-0.026	0.006	0.000	-0.019	0.023	0.414
Year-2014	-0.032	0.007	0.000	-0.025	0.029	0.394
Procedure 1				ref		
Procedure_2				0.034	0.019	0.076
Procedure_3				0.151	0.026	0.000
Procedure_4				0.123	0.018	0.000
Procedure_5				0.063	0.023	0.006
HHI	-0.0000044	0.000	0.459	-0.000030	0.000	0.000
System membership	0.035	0.027	0.198	-0.044	0.039	0.262
Percent HCCI contributors	-0.003	0.001	0.011	-0.010	0.003	0.003
Hospital count	0.007	0.005	0.143	0.004	0.004	0.309
Percent HCCI * Hospital count	0.000	0.00048	0.017	0.00023	0.000	0.549
Constant	0.254	0.046	0.000	0.425	0.059	0.000
N (Number of observations)	4184			444		

Notes: Columns 1-3 represent estimates from a linear regression with hospital referral region (HRR)-procedure specific fixed effects using the full non-treatment sample of hospitals. Columns 4-6, specifies the framework as a standard diff-in-diff where the treated group comprises hospitals in CO, MA, and OR and the control group is a sample of hospitals in other states matched to treatment group hospitals on fraction of admissions and outpatient visits from HCCI contributors at baseline, bed total in baseline year, adjusted price in baseline year, HHI (by hospital) in baseline year, HHI (by system) in baseline year, whether the facility is an academic medical center or belongs to a system. Standard errors clustered at the HRR level.

Table 3-15: Sensitivity checks on regression estimates on coefficient of variation: placebo cutoffs

| All HRRs	2011			2012		
	Coefficient	Standard Error	P-value	Coefficient	Standard Error	P-value
Treatment x Post	-0.051	0.025	0.043	-0.055	0.025	0.030
Year: 2010	0.000			0.000		
Year: 2011	0.007	0.015	0.626	-0.014	0.010	0.175
Year: 2012	-0.022	0.016	0.157	-0.021	0.015	0.173
Year: 2013	-0.002	0.018	0.925	0.000	0.019	0.989
Year: 2014	-0.007	0.022	0.745	-0.004	0.022	0.851
HHI	0.000	0.000	0.057	0.000	0.000	0.049
System membership	0.020	0.058	0.731	0.022	0.056	0.693
Percent HCCI contributors	-0.011	0.005	0.034	-0.010	0.005	0.041
Hospital count	0.020	0.016	0.216	0.017	0.016	0.267
Percent HCCI *Hospital count	0.002	0.001	0.088	0.002	0.001	0.077
Constant	0.293	0.154	0.061	0.312	0.152	0.043
N (Number of observations)	444			444		
Number of groups	89			89		
Notes: Estimates from a linear regression with hospital referral region (HRR)-procedure specific fixed effects. HRR were matched on number of academic medical center, total beds in HRR, percent of hospitals affiliated with a system, and baseline HHI at the facility and system level. Cutoffs set at points before the actual intervention to test differential trends in the outcome. Standard errors clustered at the HRR level.						

Table 3-16: Regression estimates of coefficient of variation stratified by market concentration

	HHI < 2700			HHI 2700 – 5400			HHI > 5400		
	Coefficient	Standard Error	P-value	Coefficient	Standard Error	P-value	Coefficient	Standard Error	P-value
Treatment x Post	-0.071	0.035	0.068	-0.015	0.025	0.557	-0.025	0.054	0.623
Year: 2010									
Year: 2011	-0.038	0.020	0.074	-0.016	0.016	0.317	0.023	0.015	0.150
Year: 2012	-0.055	0.035	0.128	-0.040	0.015	0.010	-0.031	0.047	0.520
Year: 2013	-0.024	0.034	0.484	-0.039	0.022	0.077	0.026	0.045	0.565
Year: 2014	0.057	0.043	0.201	-0.043	0.030	0.167	0.035	0.046	0.451
HHI	0.000	0.000	0.559	0.000	0.000	0.002	-0.000003	0.000	0.939
System membership	-0.402	0.157	0.018	0.049	0.065	0.458	-0.082	0.291	0.782
Percent HCCI contributors	0.027	0.026	0.307	-0.001	0.008	0.887	-0.001	0.010	0.955
Hospital count	0.064	0.034	0.072	0.013	0.014	0.351	-0.019	0.068	0.780
Percent HCCI *Hospital count	0.002	0.002	0.312	-0.00021	0.001	0.835	-0.001	0.002	0.605
Constant	-0.702	0.506	0.180	0.474	0.125	0.000	0.392	0.294	0.195
N (Number of observations)	105			219			120		
Number of groups	21			44			24		

Notes: Estimates from a linear regression with hospital referral region (HRR)-procedure specific fixed effects. HRR were matched on number of academic medical centers, total beds in HRR, percent of hospitals affiliated with a system, and baseline HHI at the facility and system level. Regression specification was determined using specification tests including AIC/BIC and a Wald test. Standard errors clustered at the HRR level.

Figure 3-1: State APCD implementation as of January 2019

APCD with price release
APCD with no price release
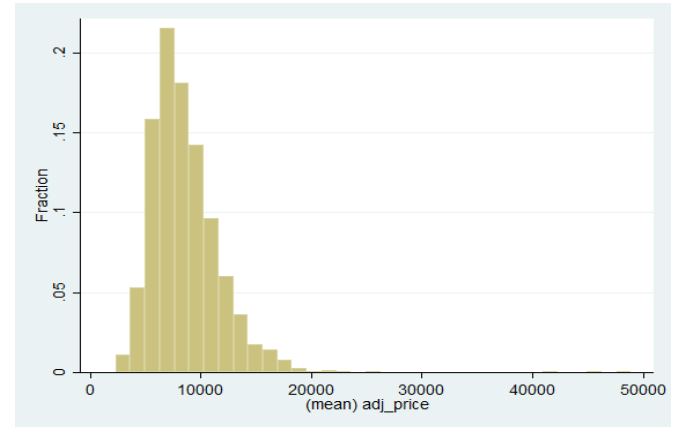APCD in implementation
No APCD
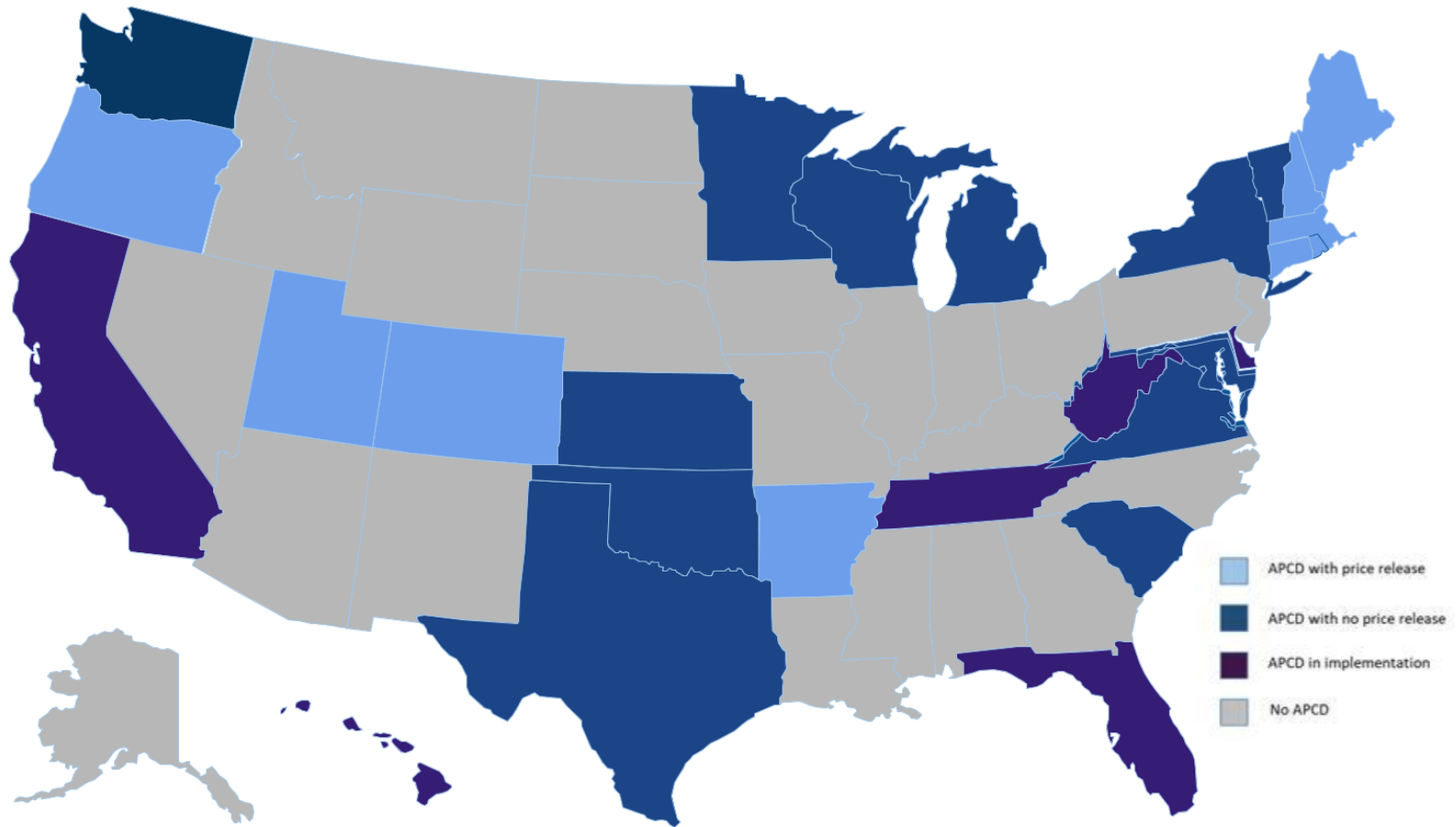
Notes: Per a state website accessed Aug 2018, NY is developing procedures to allow data to be released

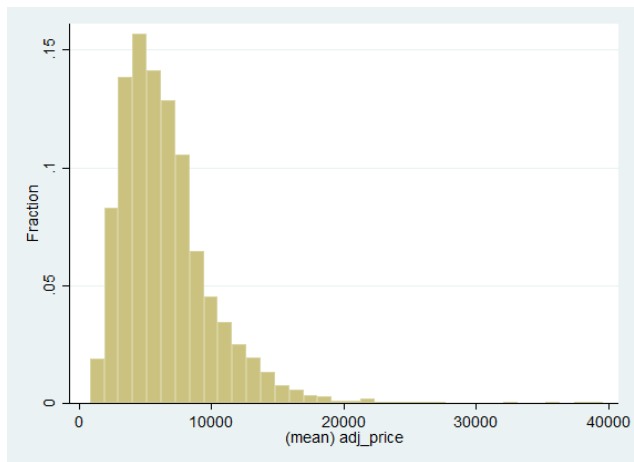Source: APCD Council (www.apcdcouncil.org), state APCD websites

Figure 3-2: Histograms of mean hospital prices, by procedure
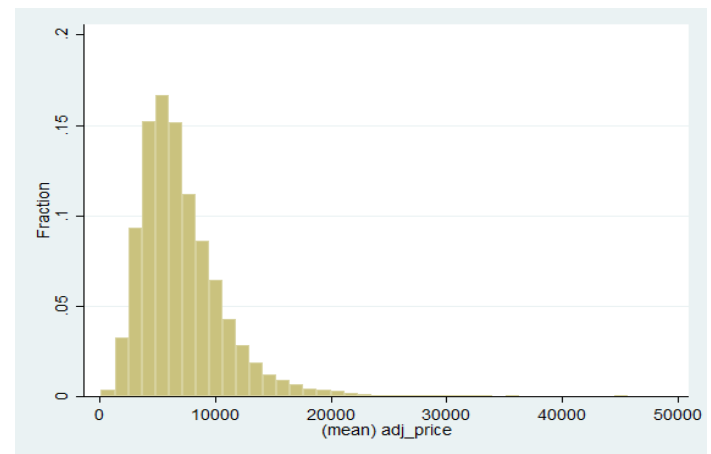*a: Normal Delivery Prices*
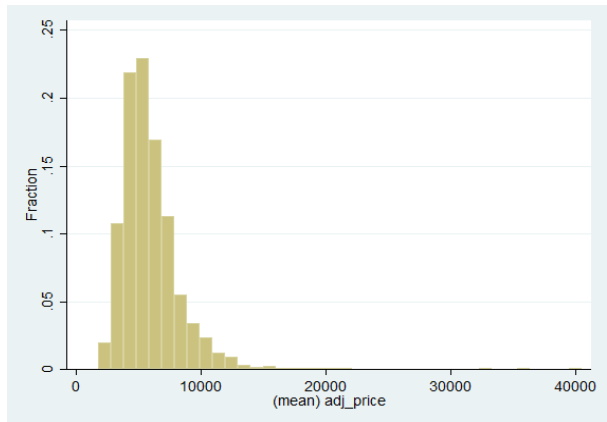
b: Cesarean Section Delivery Prices

c: Appendectomy Prices

d: Gall Bladder removal prices
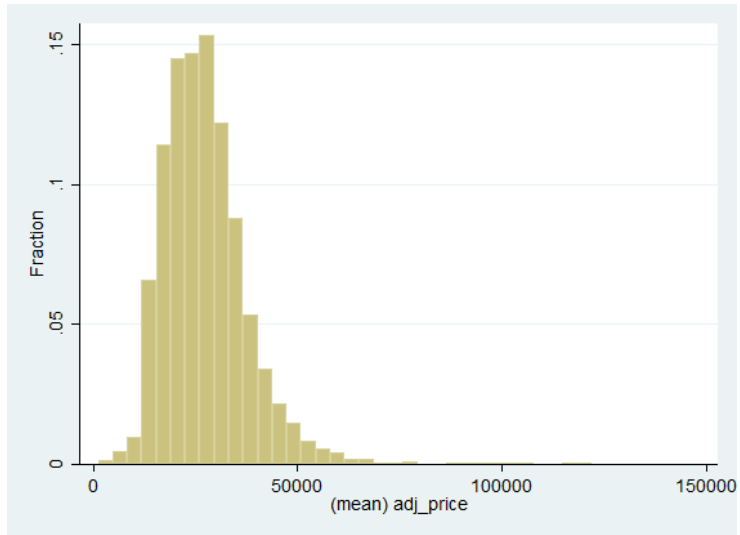
e: Hip Replacement Prices

Figure 3-3: Patient risk-adjusted prices over time, price ratios
Prices are shown as the ratio of the actual compared to the expected price. Ratio of 1 means that the transaction price is exactly the same as the price expected based on average prices for a given set of patient characteristics. Deviations from average occur based on transaction price variations not attributable to patient characteristics. Vertical dashed lines represent beginning of treatment for Massachusetts and Oregon (2012) and Colorado (2013).

Figure 3-4: Stratified price ratios

Prices are shown as the ratio of the actual compared to the expected price. Ratio of 1 means that the transaction price is exactly the same as the price expected based on average prices for a given set of patient characteristics. Deviations from average occur based on transaction price variations not attributable to patient characteristics. Vertical dashed lines represent beginning of treatment for Massachusetts and Oregon (2012) and Colorado (2013).
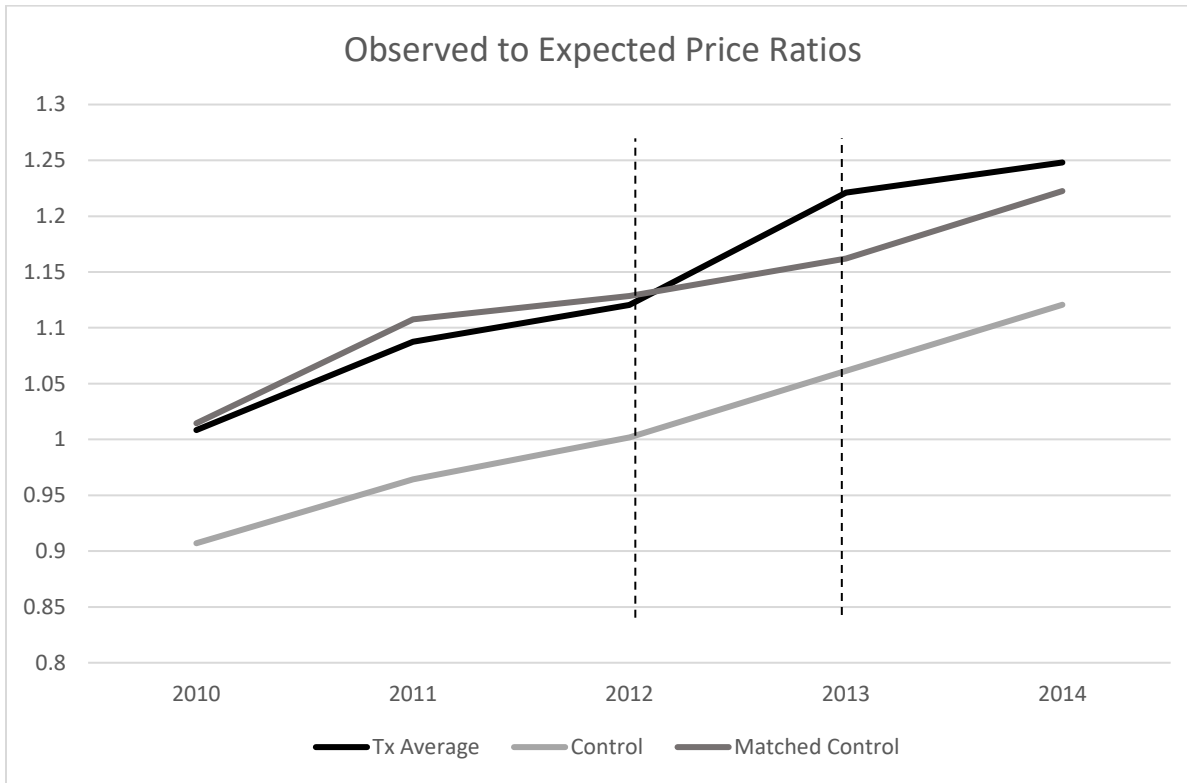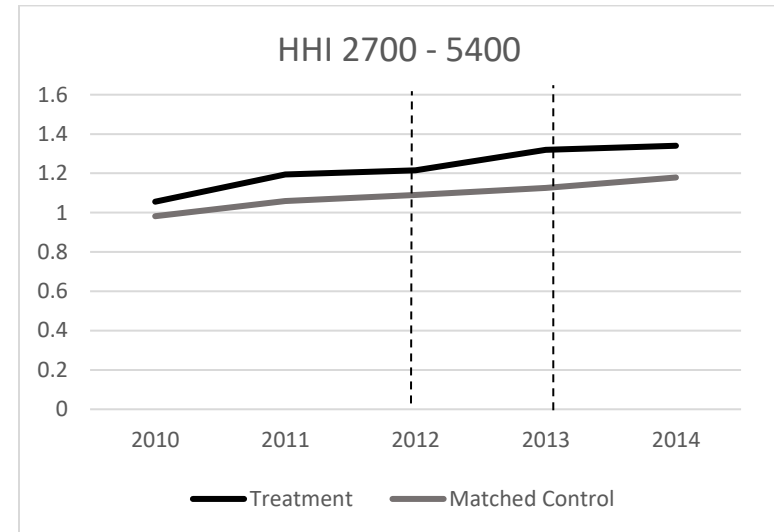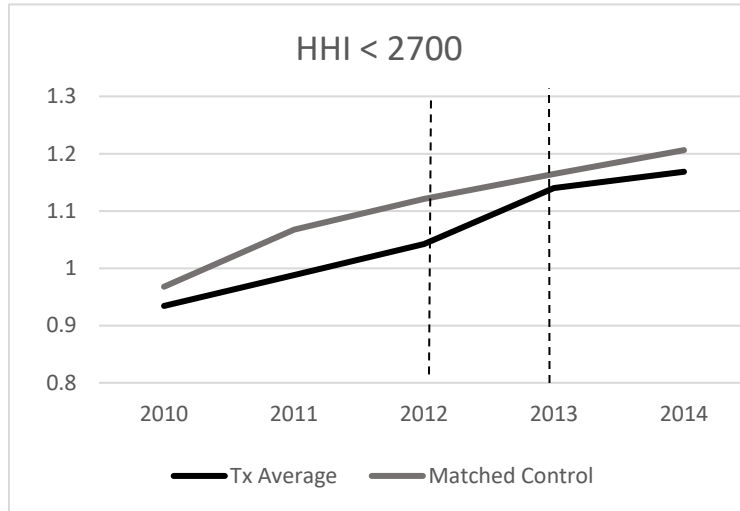
HHI > 5400

118
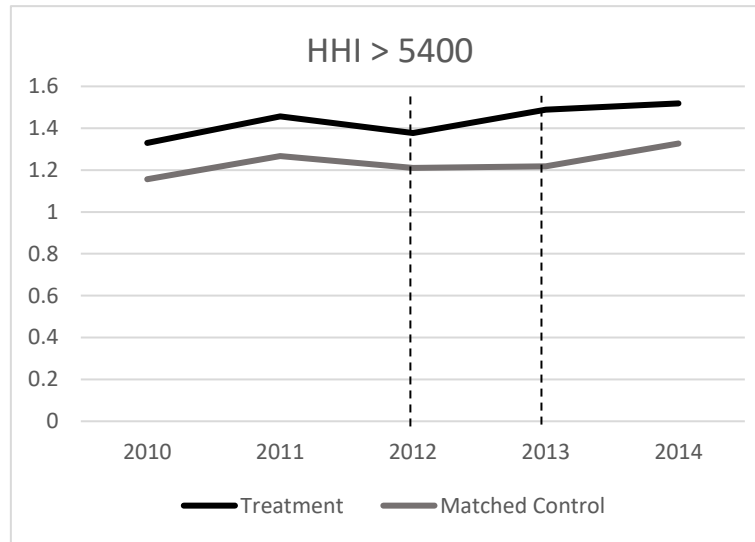
Figure 3-5: Average risk-adjusted prices, by year and procedure

Hip Replacement

	Treatment	Matched Control
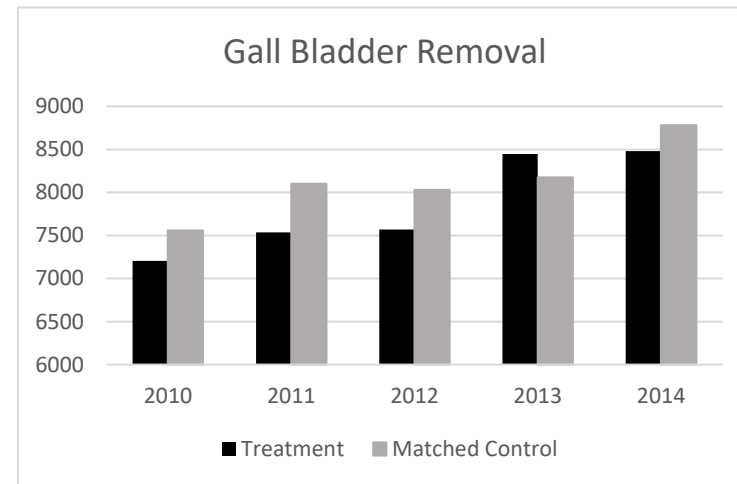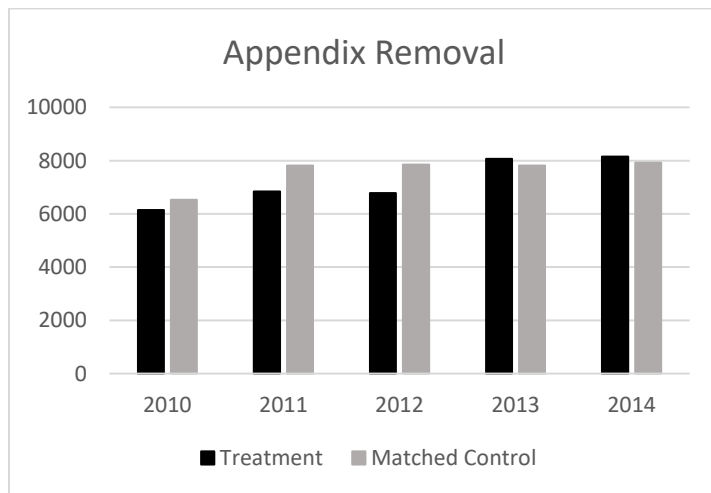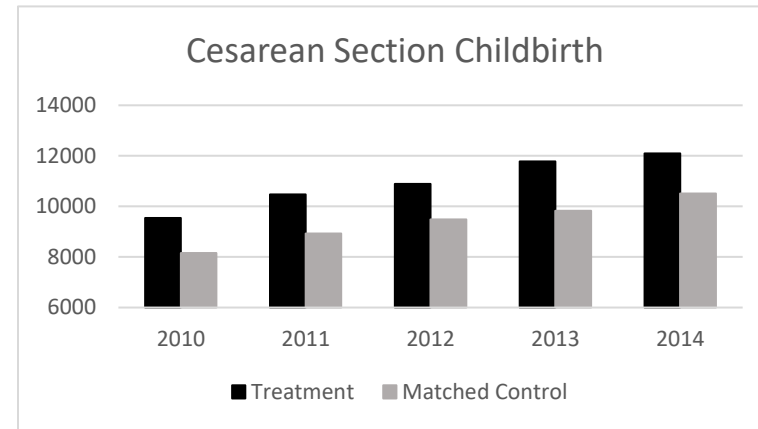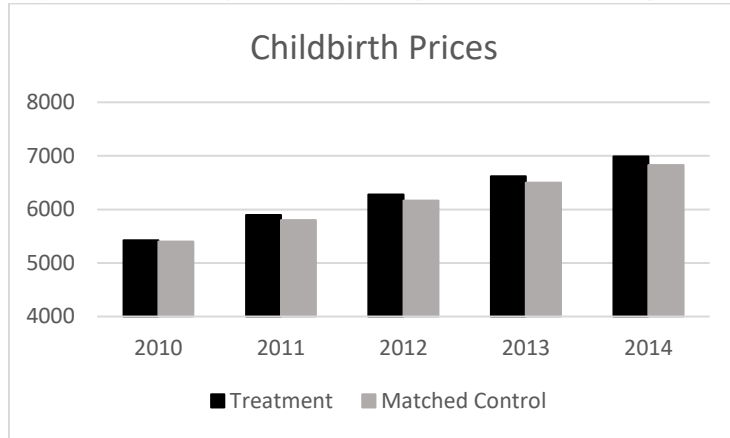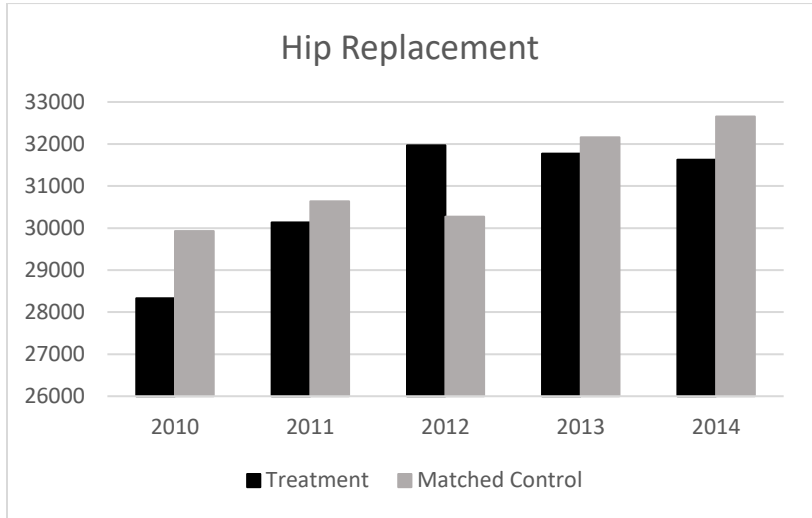
Figure 3-6: Estimates from quantile regressions, all hospitals and stratified by HHI

Estimates represented as points on the graph are from a fixed effects model, with fixed effects for procedures and facilities. Bands show 95% confidence intervals for each estimates. More information about the model used can be found in Table 3-11.

a) All hospitals
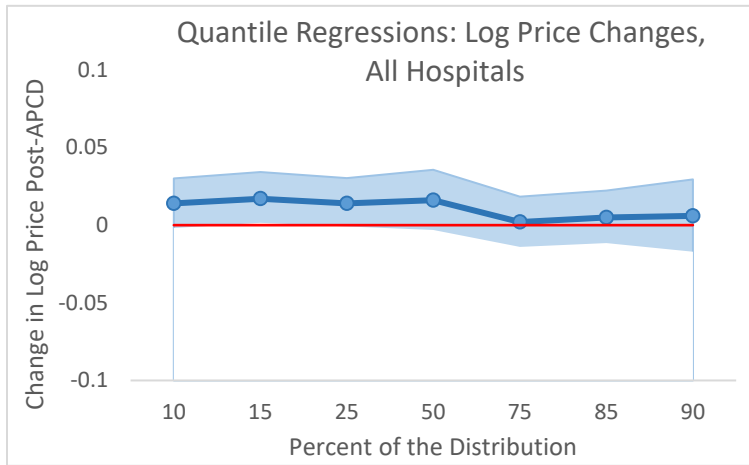
b) Least concentrated
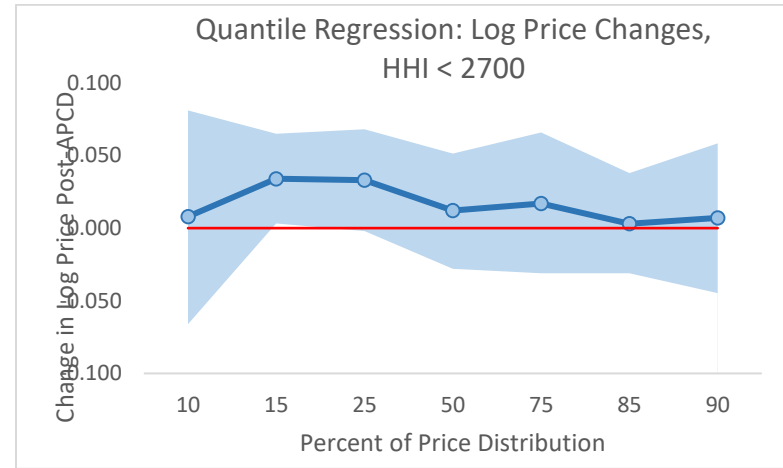
c) Middle quantiles of concentration
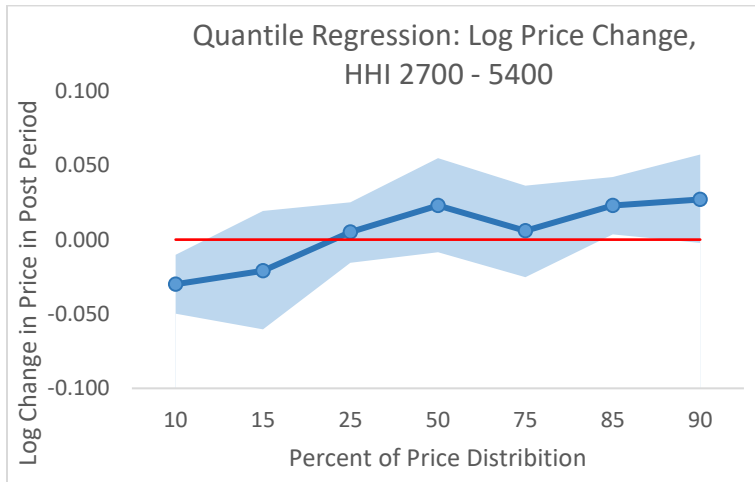
d) Most concentrated
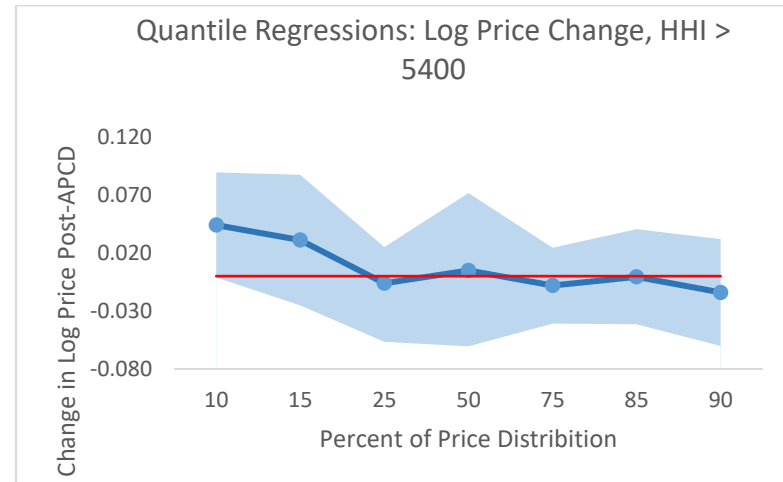
Figure 3-7: Coefficient of variation by procedure and year

Hip Replacement
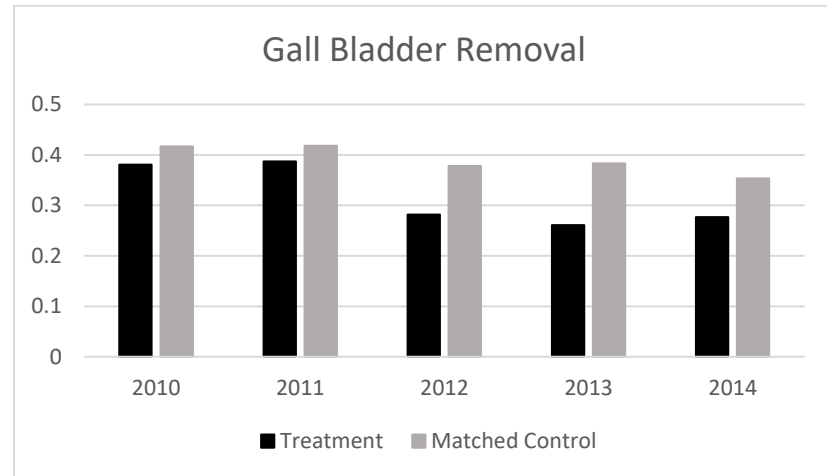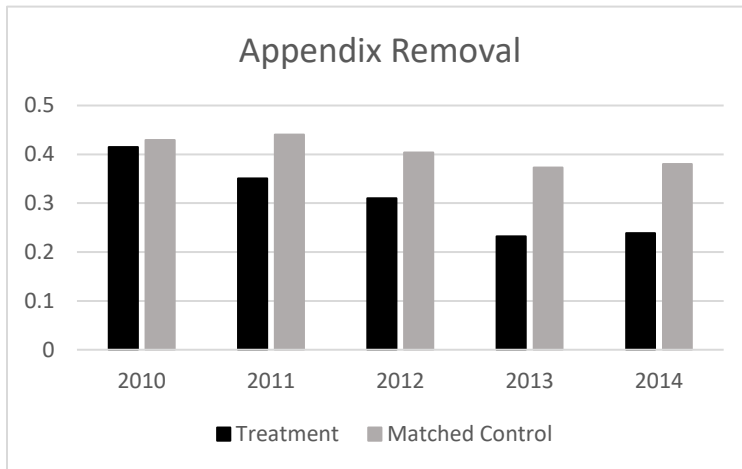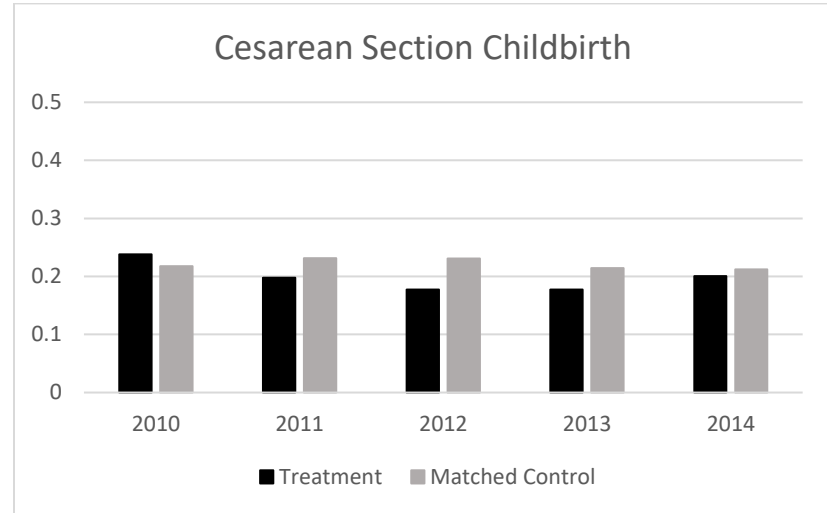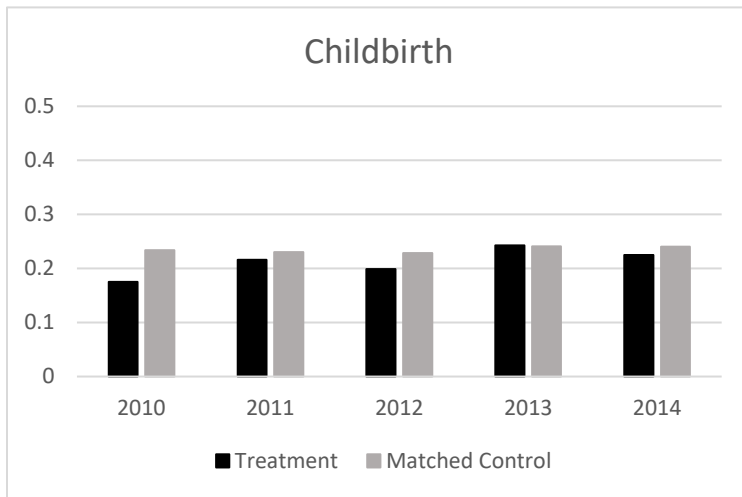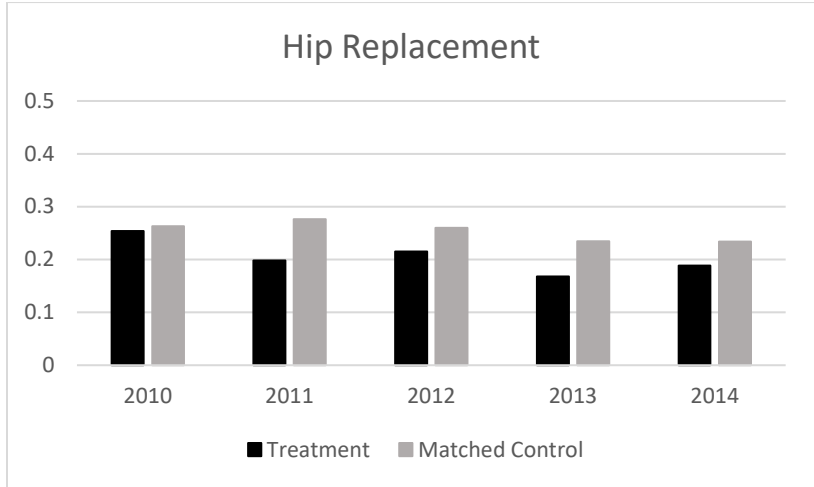
Figure 3-8: Coefficient of variation over time, all procedures

Coefficient of Variation
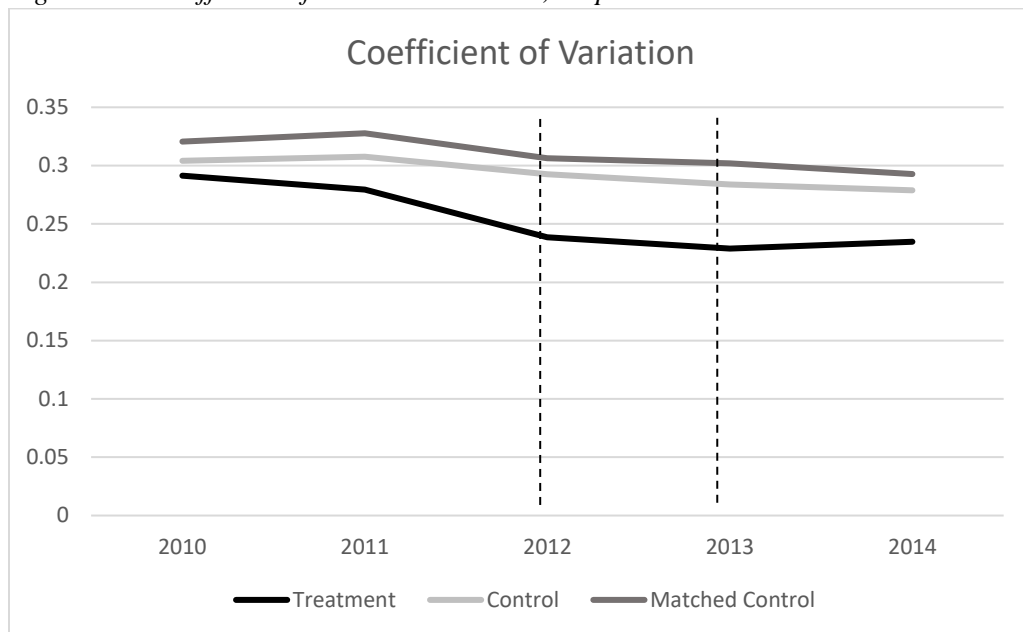
Figure 3-9: Coefficient of variation stratified by market concentration
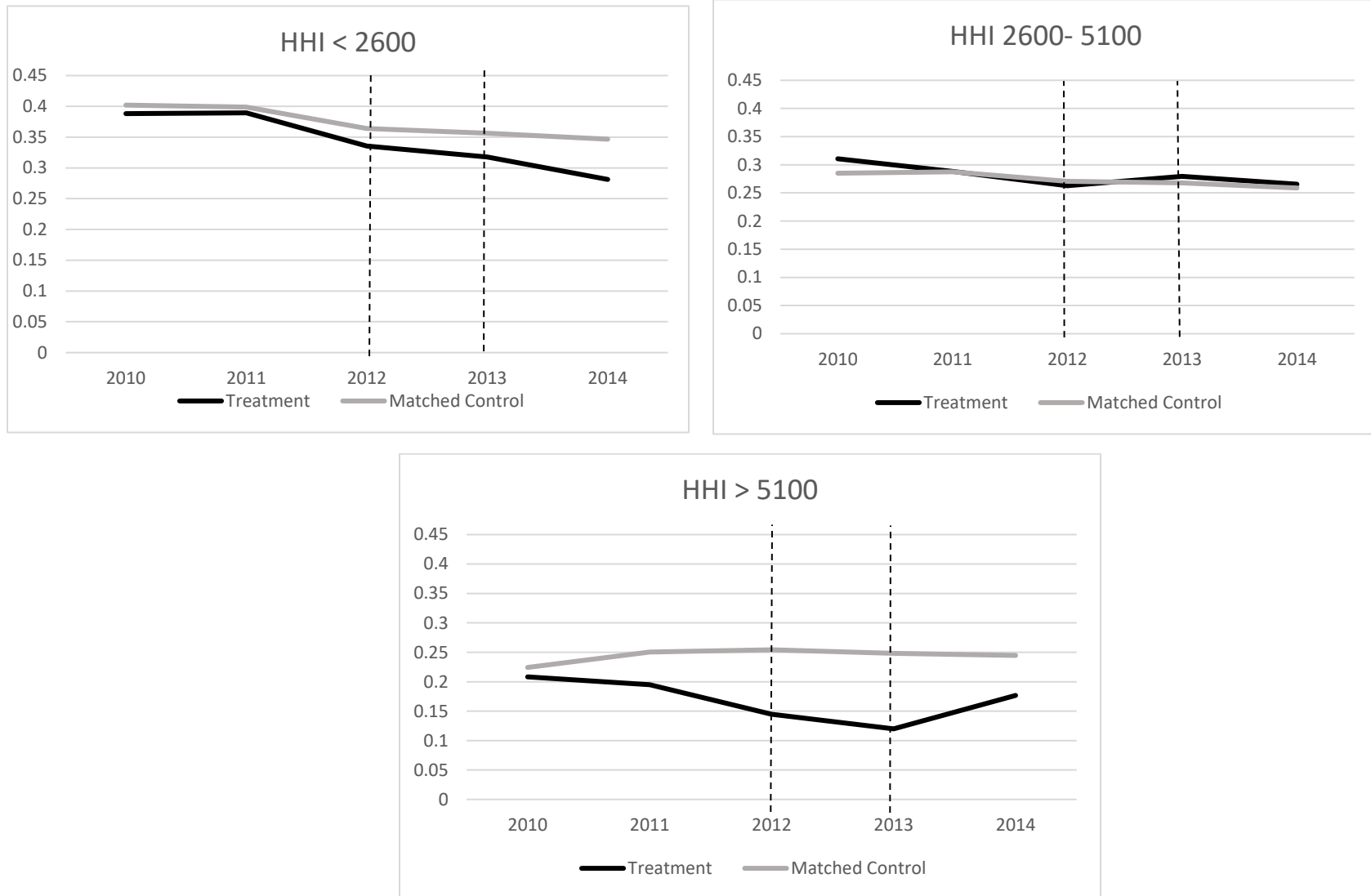
Chapter 4 : Shopping for Lower-Priced Care:
Do High-Deductible Health Plans Affect Price Paid for Childbirth?

INTRODUCTION

Using enrollee cost sharing in health insurance benefit design to mitigate spending has a rich

theoretical and empirical history in health economics (Arrow 1963b; Manning et al. 1987; M. V.

Pauly 1968).   In recent years, largely due to an effort by employers to lessen their health care

expenses, the use of cost sharing has accelerated. In 2006, just 7 percent of employers offered

health plans with high deductibles; that number increased to 29 percent of employers in 2017

(Claxton et al. 2018). High deductible health plans (HDHPs) use high upfront cost sharing to

increase enrollee price sensitivity, with the idea it will mitigate moral hazard and result in

consumers making economically efficient medical decisions (Buchmueller 2009; Bundorf 2012).

Evidence shows enrollees in HDHPs spend less than enrollees in other types of health insurance

plans, which persists over time and across service settings (Beeuwkes Buntin et al. 2011;

Fronstin 2013; Reddy et al. 2014; Brot-Goldberg et al. 2017a; Agarwal, Mazurenko, and

Menachemi 2017). Yet, it is not entirely clear how these spending reductions are achieved.

There are several mechanisms through which spending in HDHPs may decrease. Each

mechanism implies different conclusions about both the ability to sustain spending reductions

across time and populations, and the potential impacts of those spending reductions on health.

First, there is good evidence showing that consumers with lower health risk self-select into

HDHPs, accounting for some of the comparative savings over more generous benefits (Fronstin

2012; Kullgren, Volpp, and Polsky 2013; Lave et al. 2011). As these plans are used across a

wider variety of populations, however, some of these savings could disappear. Second, early

evidence did not show reduced spending as a result of the plan structure itself (Buchmueller

2009),but more recent studies have found that the HDHP structure discourages spending,

primarily by leading to reductions in the quantity of services used (Beeuwkes Buntin et al. 2011;

Brot-Goldberg et al. 2017b). Reductions in quantity, however, are potentially problematic as

consumers cut back on services important to maintaining good health such as high-value

preventive services. Indeed, available evidence suggests that HDHP enrollees do reduce their

use of services important for prevention and early treatment of disease, leading to concerns about

the impact of HDHP enrollment on health (Reddy et al. 2014; Wharam et al. 2008, 2019)(Reddy

et al. 2014; Wharam et al. 2008, 2019)(Reddy *et al.*, 2014; Wharam *et al.*, 2008, 2019).

A third potential channel for spending reductions is shopping on price. Spending reductions on

this dimension have been particularly sought, as they are much less likely to lead to adverse

impacts on health than reductions in the quantity of medical services used. (In theory, price

reductions could harm health if price were positively correlated with quality, but evidence

contradicts that idea (Cooper et al. 2018; White, Reschovsky, and Bond 2014)). However,

evidence is scarce that HDHP enrollees engage in this type of behavior. Survey results suggest

very little price shopping in the general population, with mixed evidence of higher rates of

comparison among people in HDHPs (Kullgren et al. 2018; Mehrotra et al. 2017; Sinaiko AD,

Mehrotra A, and Sood N 2016). These results also suggest a general lack of willingness to price

shop, as most HDHP enrollees do not view it as helpful for finding care or saving money (Cliff

et al. 2019).

Several studies have used administrative data to look at whether actual prices paid differ between those in HDHPs and other plans. One study, a cross-sectional analysis, found very little evidence of lower prices among HDHP enrollees in large employer plans (Sood et al. 2013). However, this analysis could not control for selection. A more precisely identified study examined one large employer that rolled over nearly all employees to an HDHP at the same time (Brot-Goldberg et al. 2017b). That study found no evidence of price shopping in two years after the switch to HDHP, though its generalizability is limited due to the specificity of the setting. Additionally, that paper and others have not characterized whether there are differences in HDHP enrollee behavior by market. More consolidated health care markets present less opportunity to choose among providers, and may be less amenable to price shopping.

This essay seeks to understand whether previous findings are generalizable and whether there are differences by market in the effect of HDHP enrollment on price. I use a large panel dataset with health insurance claims from 2010 to 2014 from all 50 US states to examine whether prices for one commonly shopped for service, normal childbirth, are lower after enrollment in an HDHP. In this sample, 47% of births covered by an HDHP are subject to a deductible, and 80% are subject to some cost sharing.

I identify enrollees who switched from a non-HDHP to an HDHP due to a full employer rollover, mitigating individual selection issues and allowing me to identify effects based on the plan design. For the full sample, I find transaction price for childbirth virtually unchanged after switching to an HDHP, and cannot rule out no change or increases in price. Market stratification, however, makes a difference. In markets with more hospitals, price decreases from an average of $5702 pre-HDHP to $5551 after the switch to HDHP, a 3 percent drop. Yet when provider fixed effects are added to the equation, the results change very little. This result

suggests that the majority of the observed change in spending is due to a change in provider prices after rollover to HDHP and not to enrollee shopping behavior. Provider prices could change for a number of reasons including because of a re-negotiation of payer-provider contracts or because the employer, in addition to switching to an HDHP, shopped around for an insurer with lower provider prices.

These main findings are supported by other results from this study, showing very small changes in charge prices after rollover to HDHP. While charge prices have little relevance to transaction prices, finding significant differences here could imply that women are *trying* to shop, and are doing so on prices that are widely available but unrelated to the transaction price. That does not seem to be the case. Additionally, I find small impacts on the use of a provider that is below-median price for its market. Estimates are positive in my preferred specification—indicating HDHP enrollees use lower priced providers in a market—but are imprecise and require strong assumptions to be considered causal.

These results are consistent with previous literature that has found very little effect of enrollment in HDHPs on shopping behavior. While this study does identify spending reductions, which is consistent with previous literature, it is likely these reductions are due to providers lowering prices rather than consumer shopping behavior. It is possible that increased price sensitivity brought on by greater consumer cost exposure increases insurer bargaining power, though that was not directly tested in this essay. It is also possible that the HDHP rollover was part of larger re-examination of insurance benefits by the employer who found a plan with lower negotiated service rates. Regardless, this paper adds to literature showing that simply increasing exposure to health care prices is not enough to induce shopping behavior. For policy, the result implies two distinct paths. If consumer incentives are to be used, HDHP enrollment and price shopping

tools are not enough. Other efforts to enhance consumers' use of price information are needed.

However, it's unclear that these types of incentives will ever result in meaningful change in

health care spending without impacting health. A second policy path is to pursue other efforts to

reduce spending growth, through provider incentives or other policy changes that more directly

impact prices.

BACKGROUND: SHOPPING IN HIGH-DEDUCTIBLE HEALTH PLANS

Theoretical Framework

Using cost sharing in health insurance benefits to control spending has long held appeal, likely

because it is relatively easy to implement and because of its straightforward relationship to

economic theory. The utilization and price paid for health care services can be understood

within a classic microeconomics framework that models the trade-off between price and quantity

demanded (Arrow 1963b; M. V. Pauly 1968; R.

Zeckhauser 1970). Line D_0 in the graph models

the relationship between the price of a medical

service and demand for that service; a robust

literature empirically confirms the prediction from

this model that changing prices changes the

quantity of medical care demanded (Finkelstein *et*

al., 2012; Manning *et al.*, 1987; McGuire, 2011;
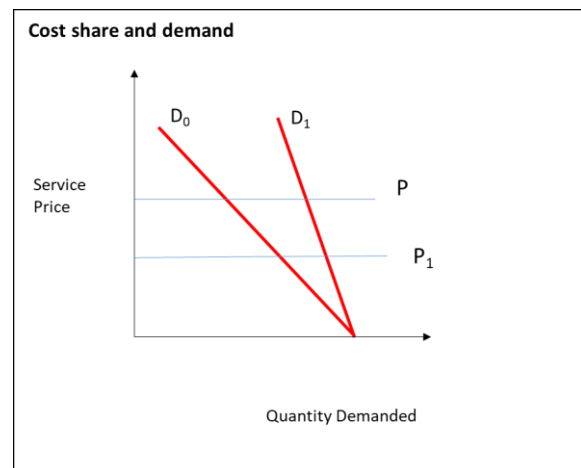
Newhouse, 1993).   When a person gains protection from financial liability for health care costs

through insurance, the demand curve rotates clockwise, as shown in the figure in the movement

from D_0 to D_1. That has a number of implications and, relevant here, lessens sensitivity to

changes in price, so the change in quantity demanded when a price drops from P to P_1 is greater for curve D_0 than D_1. High-deductible health plans remove this financial protection, which should make medical care consumers more price sensitive. In addition to a decrease in quantity, economic theory predicts that increased sensitivity to price (coupled with transparency about prices) will lead to an increase in shopping behavior. Indeed, the simplicity of this prediction and its relationship to the most basic tenets of microeconomics have led some to question, in the absence of evidence of this effect, whether health care ever functions as a traditional market (M. Chernew et al. 2018).

Price Shopping in Real Life

The simplicity of the theoretical prediction belies the complexity of navigating the U.S. health care system and, in this context, finding reliable information about the pricing of health care procedures. About one-third of all medical services are considered shoppable, that is they are not urgent and are discrete (Frost and Newman 2016). However, not all shoppable services are equally likely to be shopped for. For example, some services may be bundled with others—such as a blood test done at a primary care appointment. Particularly when these services are inexpensive relative to other medical care, the time cost of shopping and returning to a medical office may outweigh monetary gains from cheaper prices. In other cases, services are potentially so expensive that a lower price may be moot to the enrollee. Hip replacements are typically non-urgent and discrete, but the average cost--around $26,000 in a national sample of commercially insured--is so high that almost any insured individual will reach the out-of-pocket maximum on the procedure price alone. Finally, services that are done as direct referrals might be less amenable to shopping than other services. Physicians often refer to a specific center, which

patients may take as a proxy for quality or as a directive that they are reluctant to ignore. That said, there is evidence that some people shop for imaging services, a common referral service, and save money doing so (Wu et al. 2014).

Childbirth is one of the most commonly shopped for services in this population. A study by Sinaiko and co-authors (Sinaiko AD, Mehrotra A, and Sood N 2016) found that childbirth was the third most commonly searched for services in people with Aetna health insurance, who are included in this study. The mean price for vaginal delivery is about $5000 in the 2010, and available evidence suggests people would pay a significant portion of that. Average deductibles in an HDHP in 2010, the first year of our data were about $2100 for a family, and in that year about one-third of people in employer-sponsored coverage had out-of-pocket maximums of at least $3000 (Claxton et al. 2010). In HDHPs that year, families could have an out-of-pocket maximum of up to $11,900 under Internal Revenue Service regulations. Prices also vary considerably; previous work with this same dataset found the regional COV for normal childbirth was about 0.23.

Barriers to Price Shopping

For the subset of services that are amenable to shopping, it is not well known why more HDHP enrollees do not shop. Price transparency tools are becoming widespread and within market variation in price is substantial (Dunn, Shapiro, and Liebman 2013; Mehrotra et al. 2017; White, Reschovsky, and Bond 2014). Despite the tools and variation, however, survey research suggests that enrollees do not understand the variation in prices, or consider using price transparency tools when seeking medical services (Kullgren et al. 2019). Recent surveys suggest that provider relationships and referrals from providers are a significant driver of patient choice

of medical facility, and that prices matter less (M. Chernew et al. 2018). Patients also may have limited numbers of provider choices (Kullgren et al. 2019; Mehrotra et al. 2017). Indeed, another health policy trend, narrow networks of providers, could undermine the push toward price shopping by limiting the number of providers to choose from.

Another barrier, particularly for an expensive services such as childbirth, could be an enrollees' rational reaction to the non-linear aspect of the insurance contract. Indeed, if at the beginning of the year, an enrollee knows she is likely to spend higher than her deductible or out-of-pocket maximum regardless of the provider she uses, she may not be sensitive to price differences between providers. However, this type of forward-looking price behavior has not been shown in HDHPs or other healthcare contexts (Brot-Goldberg et al. 2017b; Einav, Finkelstein, and Schrimpf 2015).

One often overlooked additional barrier is the quality of information on price that patients receive. While most large insurers, including those represented in this study, have price transparency tools, few patients use them (Desai et al. 2016). The tools can be hard to find and patients may not know to look for them. Patients may be inclined to search for prices in the way that they search for other goods—through Google or another search engine. While there is scant evidence in the scientific literature about how people shop for medical prices online, there is some information in gray literature. For example, a survey commissioned by a firm that creates a price transparency tool, HealthSparq, found that 35% of consumers who had gone online to look for prices used Google, 23% had used an online tool not associated with their health plan and 65% had used a website associated with their health plan "or other online portal" (Hanover Research 2017). (Respondents could check more than one choice.) For consumers who use sites not associated with their health plan, the relevance of results to the price a person will actually

133

pay is questionable. For example, search results from the top 2 pages of a Google or Bing search for price information about specific nonemergency medical procedures only yielded geographically-relevant price information 17% of the time (Kratka et al. 2018). That same study found these estimates varied widely: upper GI endoscopy ranged from $230 to $1,950. Survey research finds that more than half of people who have previously price shopped for medical services agree it is difficult (Cliff et al. 2019).

Despite these barriers, enrollees in HDHPs increasingly have tools to shop and often have incentives to do so. This essay asks whether, given the incentives, a situation in which price shopping is likely helpful, and probable access to information, there is any evidence of price shopping among those newly enrolled in an HDHP.

DATA AND ESTIMATION

Data

This essay uses childbirth claims from enrollees in a commercial insurance plan in the large or small group market from one of three insurers: Aetna, Humana, or UnitedHealthcare. The data come from the Health Care Cost Institute (HCCI) and contain claims for about 50 million individuals, or 25% of the population in the employer-sponsored market. The medical claims data contain diagnostic and procedural codes, as well as information about the medical encounter, including length of stay and service setting. Importantly, HCCI data are one of the only large-scale sources of actual negotiated transaction prices for medical services. I use these transaction prices, along with both the charge amounts (the "sticker price" of a service set by the hospital) and out-of-pocket amounts also in the data.

The HCCI enrollment data have age in bands of approximately 10 years, gender, and enrollment at the month level. There is also limited information about insurance plan structure including the market in which it is sold (large group, small group or individual), the type of structure it has (point of service, preferred provider organization, health maintenance organization), and whether or not it is considered a consumer-directed health plan, which identifies a member as enrolled in a high-deductible health plan. Finally, the data contain a group ID, the purpose of which is somewhat variable but, roughly, it corresponds with an employer-plan grouping.

These data also include encrypted National Provider Identifiers (NPI), which can be used to identify hospitals. Through a proprietary file containing NPI numbers encrypted to match the HCCI database, I merge hospital attributes in the American Hospital Association Survey Dataset (AHA survey) for all years with the HCCI claims data. That allows me to a calculate hospital-bed based Hefindahl-Hirschman Index (HHI) to measure market concentration for each market. Finally, I merge in several variables from the U.S. Census Bureau Area Health Resource file. From that file, I am able to adjust for population and the median income at the core-based statistical area level.

Identifying HDHP Rollovers

The key identifying assumption in this study is that the impetus for switching from a traditional (low deductible) health insurance plan to a high-deductible plan is an employer, not individual, decision. Employer decisions are less likely to be made for reasons correlated to the outcome of lower prices, including expected medical spending and a preference for price shopping. I cannot directly observe employer decisions in these data so I infer it when I see an entire group of

enrollees switch group numbers and switch from a low-deductible plan to an HDHP. To do that, I leverage the enormity of the data to find specific plans for which I can be certain that the employer switched an entire group of employees to an HDHP, jettisoning those observations that do not meet strict criteria. I identify enrollees in groups of at least 10 in which all members have a do not have an HDHP at the beginning of one year and, in a subsequent year, all members have an HDHP and the group is 95 to 105 percent of its original size. I follow each enrollee as long as he or she remains in that specific HDHP, up to 4 years. Additionally, I keep only those enrollees who do not move between hospital referral regions (HRRs) during the period in which they are observed in this study. More details about the approach as well as the number of potentially eligible enrollees and final sample are in Table 4-1. The final sample includes 572,231 observations of 184,156 individuals in 8,096 groups.

I form a control group by taking the entire universe of enrollees who remain in a non-HDHP for the years of observations, 2010-14. I exclude enrollees under 18 or over 64 and those who move between HRRs. The resulting control group comprises 51.5 million observations of 17.5 million individuals, which is too large to analyze efficiently with my existing computer power, and not necessary for estimation of effects in the treatment group. So, I randomly cut 75% of the observations to create a more manageable control group with 13.2 million observations of 4.7 million individuals. The demographic makeup of the full control and randomly selected control group are almost identical (Table 4-2). I will refer to the randomly selected control group as simply the control group for the remainder of this paper.

Outcome Measures: Hospital Prices

The HCCI dataset contains medical claims with diagnostic and procedure codes, out-of-pocket spending amounts, negotiated transaction prices, charge prices and encrypted provider codes. I focus on price shopping for childbirth. To measure a homogenous set of procedures, I restrict my analysis to only vaginal deliveries without complications (DRG 775) and in which the primary procedural codes are either manual assistance with delivery or repair of obstetrical tear (CPT 7359 or 7569), implying that nothing more serious was performed. I further limited the analysis those done in a hospital inpatient setting and for which length of stay was less than 5 days.

The main outcome measure was the transaction price for the medical service: the price paid by the combination of the insurer and patient to the hospital. I also look at the out-of-pocket cost to the enrollee for the procedure, and the charge price of the hospital. While charge price is not representative or highly correlated of the true price paid for the procedure (0.44 in these data), it is easier to find than the negotiated price. For example, most state hospital association websites include a list of childbirth charge prices with hospital names on their websites (Kullgren, Duey, and Werner 2013). Because these sites may be more common or easier to find than websites with negotiated prices, women searching for prices may inadvertently shop for price on the wrong price.

One issue with using the negotiated price as a proxy for price shopping is that, given that the move to HDHP represents a plan change, there may be changes in the negotiated prices or changes in the network of providers. Thus, enrollees in these plans may receive lower prices simply by virtue of insurer negotiations or via network restriction. To mitigate this concern, I construct an additional outcome variable to measure price shopping: the proportion of enrollees who choose a low-price provider in a market. I create for each service a group-area-year choice

137

set of prices defined as the set of average service prices across all unique billing entities visited by at least 1 enrollee in an HRR in each employer group. I additionally create the same measures, looking at any provider in the HSA and measuring the likelihood of choosing one with a below-median price. While this has the advantage of better measuring choices than an estimate of prices, it does require an assumption that the price to an enrollee of a non-chosen option is consistent across plans in an HSA in observed rank relative to the chosen option. This assumption could be plausible within the same health plan, but it is a much stronger assumption when the choice set is defined through observations taken from any birth in an HSA.

Independent Variables

The primary independent variable is the employer-mandated switch to an HDHP, the construction of which is described above. Additionally, to measure one of the potential modifiers of price shopping, the presence of choices in a market, I analyze the differences in switchers' behavior in markets with few choices compared to markets with some or many hospital choices. I define a market using the Dartmouth Atlas hospital service areas, which are a collection of zip codes that form local markets for common hospital services (dartmouthatlas.org).

The ideal measure would be to look at the number of effective choices that each enrollee has. That depends on her insurance plan structure, the geography of her hospital service area, and perhaps other things, such as differences in hospital quality or obstetrician admitting privileges. These data do not include variables that allow direct observation of these factors; instead we use the total number of hospitals in an HSA as both a ceiling and a proxy for the relative number of choices for each enrollee. Thus, enrollees in the bottom quartile of the number of hospitals in an

HSA are considered to have few or no choices; those in the top 3 quartiles are considered to have at least some, and perhaps many choices. I split the sample into those in the bottom quartile and those in the top three quartiles with the rationale that enrollees with at least some choices were more likely to price shop than those with few (or no) choices.

I measure market structure in two ways. First, I perform a simple count of the number of providers for the service. Then, I use HHI based on bed size. For both, I split the sample based on the 25% of enrollees who live in the markets with the fewest hospital choices or the highest hospital concentration, 3 or fewer providers or an HHI ≥ 3937. The 75% of enrollees in the markets with more provider choices or in less concentrated markets make up the second group.

Analysis and Estimation

I leverage the panel nature of the dataset and breadth of national coverage to compare prices and spending outcomes for enrollees whose employers roll over to an HDHP to those with stable benefit plans. I analyze the data at an individual-year level with the following empirical specification. For the main analysis, I use the following equation:

$$Y_{it} \text{ or } \Pr(Y_{it} = 1) = f(\alpha + \beta_1 Switchers_{it} + \beta_2 SwitchersPost_{it} + \beta_6 X_{it} + \gamma_t + \delta_i + \epsilon_{kit})$$

The outcome is one of three continuous spending outcomes (member out-of-pocket, transaction price, charged amount) or the probability that a woman chooses a lower priced provider for person, i, at time, t. The primary coefficient of interest for the main analysis is β_2, which picks up the difference in the spending outcome for those who switch to an HDHP after they have

switched.   In the equation, the *Switchers* variable controls for the pre-period differences in the

treatment group; γ_t are year fixed effects, δ_i are regional fixed effects (northeast, Midwest,

south, west); and **X** is a vector of individual and area-level covariates: age (in bands), plan type

(PPO, HMO, POS, Other) mean area income in 2010 (by core-based statistical area (CBSA)),

and area population in 2010.   In robustness checks, I use CBSA fixed effects in place of the

regional fixed effects and find it makes no difference in the significance or magnitude of the

outcomes.

To check the sensitivity of the main analysis, I additionally use a matching model in which I

match each treatment observation to its nearest neighbor in the control sample. I match exactly

on plan type, CBSA residence, year, and whether an enrollee is age 35 or older.   I compare

spending outcomes among enrollees with an HDHP rollover to matched individuals with stable

benefit plans and assume standard errors are independent and identically distributed. I

additionally perform matching on providers and the same set of demographic covariates to test

whether prices change among individual providers after rollover to HDHP. Finally, to mitigate

issues of changes in networks, I include only the subset of enrollees who have a point-of-service

(POS) plan for the entirety of their observation in the sample.

To determine the exact functional form, I examined the structure of the primary dependent

variables of interest: negotiated price, out-of-pocket spending, and charge prices. All three are

highly skewed (Figure 4-1); charge price is especially highly skewed. Generalized linear models

(GLM) allow more flexibility in the expectation of y, and thus are better at correctly modeling

highly skewed data (Deb, Norton, and Manning 2017). To determine the correct distribution

family for the GLM equation, I used a Modified Park Test on each of the three outcome

variables. Following suggestions in the econometric literature, I used a log link function in all

models (Deb, Norton, and Manning 2017). I use the margins command in Stata (version 14.2) to transform the difference-in-difference estimates from the models into real dollar amounts.

To model the probability of choosing a lower-priced provider, I use a linear probability model and probit model as a sensitivity check. I use the same structure as above, with covariate adjustment for age, insurance plan type, region of residence and year.

To test whether there is a differential effect when an HDHP has more choices in her market, I employ I triple-differences strategy, stratifying the effect by market structure with the following equation:

$$
\begin{aligned}
Y_{it} = f(\alpha &+ \beta_1 Switchers_{it} + \beta_2 SwitchersPost_{it} + \beta_3 MoreChoices_{it} + (\beta_4 MoreChoices_{it} \\
&* Switchers_{it}) + (\beta_5 MoreChoices_{it} * SwitchersPost_{it}) + \beta_6 X_{it} + \gamma_t + (\gamma_t \\
&* MoreChoices_{it}) + \delta_i + \epsilon_{kit})
\end{aligned}
$$

As with the main equation, the outcome is measured for person, i, at time, t, and the outcomes are total enrollee spending, transaction price, or charge amount. The primary coefficients of interest is β_5, which picks up the difference in the spending outcome after switch to an HDHP in markets with more provider choices, relative to the difference in markets with fewer choices. All other covariates are the same as those in the main equation. This equation also includes the interaction of year and number of choices in a market, as the concentration of providers may not only affect initial levels of prices but also the trend in price growth. Once I obtain these results, I additionally add a provider fixed effect to this equation to test whether prices drop conditional on using the same provider. As noted above, the hypothesis tested here is that markets with more provider choices (or less provider concentration) will be more amenable to shopping and, thus,

prices will be lower relative post-HDHP switch relative to markets with fewer choices (more provider concentration).

Pre-Trends and Selection into Birth

Two potential threats to internal validity in this analysis are whether pre-period trends in prices are parallel between the treatment and control group, and whether the change in benefit design changes who is giving birth, i.e. whether there is selection into births after rollover to HDHP. In the first case, the parallel trends assumption is a key identifying assumption in any difference-in-difference analysis. If pre-period trends are not parallel, it calls into question whether there are unobserved differences between the treatment and control group that may bias the estimated effect of the treatment (Ryan, Burgess Jr, and Dimick 2015). Transaction price means from a GLM model in which a dummy signifying treatment prior to HDHP switch is interacted with a linear time trend are shown in Figure 4-2. The graph shows roughly parallel lines, with a slightly steeper slope for the comparison group; Table 4-3 shows the estimated slope from this model, which includes covariates, is about 1% and not statistically different from 0. These all provide assurance that transaction price trends are approximately parallel in the pre-period. As with transaction prices, charge prices in the treatment and control group are roughly parallel with a difference in trends indistinguishable from 0 (Figure 4-2; Table 4-3). Out-of-pocket spending, however, does not show parallel pre-period trends. As both the graph and table show, out-of-pocket spending is higher for the treatment group in the pre-period but growing much more quickly for the control group (Figure 4-2; Table 4-3). This differential trend here would bias the treatment estimates downward.

The second potential threat is that after a switch to an HDHP, there will be selection into medical treatment. Previous studies have found quantity reductions in medical service use after HDHP, which begs the question of whether, when services are used, the population using them is significantly different (Brot-Goldberg et al. 2017b; Reddy et al. 2014; Wharam et al. 2019). Childbirth is somewhat of a different type of medical service than many others for a number of reasons, not least of which is that planned pregnancies are typically a lifestyle rather than medical choice and, unplanned pregnancies are, by nature, accidents. Table 4-4 shows that the percent of each group giving birth remains at between 0.5% and 0.6% in both the treatment and control group, offering little suggestion of reduction in births based on HDHP enrollment.

Descriptive Statistics

Employees of firms who rollover to an HDHP are more likely to be in the 35-54 age range than are those in stable benefit plans, more likely to live in the Midwest and in an area with lower population (Table 4-5). The analytic sample is the subset of the full sample that gives birth between 2010 and 2014. That sample includes 71,381 observations of 67,099 individuals, 2,805 in the rollover group and 64,294 in the control group. As shown in Table 4-6, the comparison of the demographic profile mirrors the full sample (but for gender, not shown). Within the subsample, a higher percent of the treatment group is between 35 and 44 years old and is more likely to live in the Midwest or the West. Here, as in the full treatment and control samples, POS network structures are more common among the treatment group.

Overall, transaction prices increase over the study period from about $5000 per birth for the treatment and control groups to nearly $7000 in 2014 (Figure 4-3). However, it's evident that there are no more than small differences in non-adjusted price between the two groups. Out-of-

pocket spending, by contrast, shows a sharp jump in the post period (Figure 4-4).   The graph

shows somewhat higher spending in the treatment group and (as also shown in the parallel trends

analysis) a steeper slope in the control sample up to 2013. Yet the post-treatment group (the

group that switches to an HDHP) shows a distinct jump in levels of out-of-pocket spending such

that by 2014 the mean is close to $1,400 per year.   Charge amount increases in both groups

between 2010 and 2014 (Figure 4-5). The graph shows that average charges begin at about

$10,000 per birth in 2010—about twice the average actual transaction price—and rise to close to

$14,000 by 2014.

RESULTS

Changes in Out-of-Pocket Price

As expected, a woman who gives birth after switching to an HDHP spends more out-of-pocket

after switching. The GLM model adjusted for demographic covariates shows the total member

out-of-pocket spending is an average of $227 more post-switch than it would be had she not

switched plan types (Table 4-7). The parallel trends analysis suggested that this estimate may be

biased downward; it is possible the actual amount is higher. Alternate specifications are

consistent with the main results (Table 4-7), including with an individual fixed effect model

using the subset of women with more than one observed birth (Table 4-8).

Changes in Charge Prices

Changes in charge prices may be significant if they reflect attempts to shop on price. However,

because they are not strongly correlated with transaction prices, drops in charge prices are

unlikely to affect spending. From the full sample, there is an estimated drop of $365 in charge price paid, about 3 percent from baseline, though it is not statistically significant at conventional levels (p=0.078; Table 4-7). The estimates is robust to alternative specifications (Table 4-7).

Changes in Transaction Prices

In the analysis of the full sample, I am unable to detect differences in transaction prices after a woman switches to HDHP. In a GLM model adjusted for demographic covariates (Table 4-7), I find the average price drops by $26, though the confidence intervals are relatively wide (CI:-$191; $138). These estimates imply that, given a baseline price of approximately $5800 in these models, transaction price drops by as much as 3.3 percent and could rise by 2.4 percent. Using alternate specifications, I find estimates consistent in magnitude with the confidence intervals of the main specification and, in some cases, a negative effect on prices that is statistically different from 0. Both GLM and OLS models without covariates show a statistically significant drop in transaction prices post-rollover, implying that demographic differences are important in these models (Table 4-7). Matching on demographic characteristics and area of residence (CBSA), I find a decrease of $142 per birth, or about 2 percent of the baseline price (Table 4-7). Results are also consistent limiting the sample to women who have a POS plan only, both before and after HDHP switch (Table 4-7). With an individual fixed effect model, using a subset of women who have >1 birth in the data, I find a $100 decrease in price, though it is not statistically significant (Table 4-8). Finally, matching on provider and demographic characteristics, I find a statistically significant decrease in price of $157 post-rollover (Table 4-7), showing prices drop for HDHP enrollees even conditional on using the same providers.

Taken together, these estimates suggest that there are small, if any, decreases in childbirth price after HDHP rollover, and that those decreases are due to provider price changes.

Unlike the full sample, women who have more choices in their health service area, or who live in an area with lower concentration, pay lower prices for childbirth after switching to HDHP (Table 4-9). These women pay an average of $512 less relative to switchers in markets with fewer provider choices post-switch. For women with fewer provider choices, the marginal effect is positive ($343), signifying an increase in average price, though not statistically significant at the 95% level (p=0.078). However, as with the main findings, much of the effect can be attributed to changes in prices at the provider level. In a model with a provider fixed effect, prices decreased $446 after HDHP rollover for enrollees in markets with more choices relative to lower choice markets post-switch (Table 4-9, model 8). Comparing this estimate to the most similar model without provider fixed effects, which uses OLS rather than GLM as a functional form and estimates a decrease of $482 post-switch (Table 4-9, model 3), provider price changes explain 93% of the drop in price after HDHP rollover in markets with a higher number of providers. Running models with other functional forms, without covariates, or using the subsample of women with a POS plan type before and after the HDHP switch, returns generally consistent estimates. When the sample is limited to those who had a POS plan through the entire study, the estimated decrease in transaction price is not statistically significant and the magnitude is slightly smaller. However, the conclusion drawn from the full sample—that decreases are entirely accounted for by provider fixed effects—holds for this subset of the sample as well. (Table 4-9, models 9-11).

To get a better sense of the how these estimates might affect actual prices, Figure 4-6 shows the predicted prices for three groups: the control sample, treatment sample pre-rollover, and

treatment sample post-rollover. As the graph shows, in markets with more choices, women pay

an average of $5551 after HDHP rollover compared to $5702 prior to rollover, a 3 percent

relative drop.

I additionally examined transaction prices for women without cost sharing, meaning they had

both used their deductible and reached their maximum out-of-pocket amount prior to childbirth.

The idea is that these women should not be price sensitive, so any observed decrease would be

due to provider price changes and not shopping behavior. I find decreases in price of

approximately $200 post-rollover (Table 4-10), suggesting some provider effects but leaving

room for the main estimates to include shopping behavior. However, these results are so

imprecise and the subsample so small (N=8,088), it is hard to draw conclusions from this

estimate.

Changes in Provider Choices

Another measure of whether HDHP enrollees are actively shopping for lower prices is whether

they are choosing lower priced providers in their area. Results for this measure were mixed,

depending on how the choice set of providers was defined.

When the choice set was defined as providers available in the same group health plan, using any

provider used by a member of a group in a year, estimates are negative, suggesting enrollees in

HDHPs are less likely to choose lower-priced providers, but are small and statistically

insignificant (Table 4-11). However, with this definition of choice set, 74% of enrollees have

only 1 provider choice. It is possible that the number of provider choices is endogenous to the

rollover into HDHP; if enrollees all use lower-priced provider post-enrollment it would affect the

observed choice set. To test for selection into fewer providers, I regressed the number of choices

on a variable that took on values of 1-3 for control sample, treatment in the pre-period and treatment in the post-period. While the treatment group did have significantly fewer providers in their choice set, there was no difference in the pre to post period (Table 4-12). There could also be endogeneity if enrollees are choosing the same number but lower-priced providers, such that the entire distribution shifts down. While I do find evidence that the median provider price is lower post-rollover (Table 4-12), and that result is marginally significant (p=0.079), this measure is quite imperfect. With nearly three-fourths of the sample having only one choice, the median price for most of the observations in the regression is a direct function of the enrollees' choices. Without changing the definition of the choice set it is hard to test this particular source of endogeneity in this case.

As a second measure, I defined the choice set as any hospital in an HSA. Under this definition, the median number of choices is 6 and three-fourths of the population has 3 or more hospitals in her HSA. The disadvantage, clearly, is that I cannot measure network structure and so am unsure about whether each of these hospitals is included in a particular plan network. Under this definition, women are 3 to 6 percentage points more likely to choose a lower-priced provider, though these estimates are not statistically significant under most specifications (Table 4-11). Also, as noted above, this measure relies on strong assumptions about the rank order of prices in a choice set relative to an enrollees' actual choice set.

DISCUSSION

This study found evidence consistent with the idea that, for at least one common shoppable service, childbirth, HDHP enrollees pay lower prices due to enrollment in the plan. The finding is limited enrollees in areas in which there are more choices of providers and the effects appear due to changes in provider prices rather than through active choices made by enrollees.

However, the decrease in prices, about 3 percent relative to the pre-period price in these markets, could represent substantial savings, especially given that childbirth is a common service in the employer-sponsored insurance population.

These results are consistent with evidence presented, though not highlighted, by several recent studies. Chernew and co-authors (2018) found that for another procedure, lower limb MRI, people with full exposure to prices (as in HDHPs) pay less than those with partial or no exposure. These authors highlight that the plan design explained a small part of the variance in price, and that physician referrals explain most of it. They are also consistent with Brot-Golberg et al (2017) who find a 15 percent decrease in provider-specific childbirth prices in a market with multiple choices, and no evidence of price shopping, after a full firm rollover to HDHP.

To be clear, this project had a number of limitations. The identification of an employer's choice to switch to an HDHP is not directly observable. Nevertheless, this strategy of implied rollover has been used previously to identify the effects of moving to an HDHP design in similar datasets (Wharam et al. 2017) and, in large administrative claims datasets, it is typically not possible to observe detailed decisions of employers. Second, in constructing the choice sets of enrollees for measuring the distribution of prices, I only observe a price conditional on at least some observations in the treatment or control sample visiting that provider. As noted in the methodology, the observed choice set is some cases a subset of the full choice set, and here, is potentially endogenous to the plan change if substantially all enrollees respond by choosing differently-priced providers. Using the full range of providers in an HSA, agnostic to observed choices in a plan, requires strong assumptions about the rank order of actual prices faced by an enrollee. Third, this study only looked at prices in the first 4 years, at most, after rollover to HDHP; long run effects may be different. Finally, the identification strategy does not allow me

to directly separate different contracting or network effects, such as changes in network size or structure, from consumer shopping. The results with provider fixed effects and matching on provider are strongly suggestive that contracting plays a role in observed decreases in prices in competitive markets. Nevertheless, this study was not able to test that directly.

In sum, this study finds impacts of enrollment in HDHP on the prices in more competitive markets and suggests these changes are due to changes in provider prices rather than choices made by consumers. Evidence presented in this paper, on the whole, is not consistent with consumers shopping among providers on price. As with other studies, I find little evidence that HDHPs increase the use of lower-priced providers or encourage women to pay less for a common, shoppable service.

While the intention of HDHPs is to put encourage pricing pressure through consumer shopping behavior, it is possible that prices decrease for other reasons including an increase in insurer bargaining power due to greater consumer price elasticity or to shopping behavior on the part of the employer. This study was not set up to test these possibilities, and future work should look more directly at these effects. Whether this finding generalizes to other procedures and holds up when more directly tested will show whether HDHPs live up to their potential to control spending without negative impacts on health.

Table 4-1: HDHP sample creation

Step	Observations/Individuals/Groups Dropped	Notes/Limitations Introduced
1. Begin with enrollee file that contains enrollees grouped by group number and year (at least 1 observation for each year an enrollee is in the data; more if change group numbers). Enrollees included who have a) small or large group market; b) > 1 observation; c) a mix of HDHP and non-HDHP observable	-19,278,691 observations - 4,752,383 individuals - 175,624 groups	
2. Drop if fewer than 10 individuals in a group in a year (to mitigate very small employers and selection on individual biases)	-798,545 observations -330,139 individuals - 145,065 groups)	
3. Drop when employer groups have both HDHP and non-HDHP plans in the same year because we cannot know that this is a mandated choice and, thus, switching may be subject to individual biases.	-14,425,129 observations -4,227,412 individuals -10,680 groups	This creates a situation where people may have non-consecutive observable years of enrollment despite actual consecutive years of enrollment. For example, one individual has 5 years of enrollment. Year 1: the entire group has a non-HDHP, year 2/3 split HDHP/non-HDHP, year 4/5: all HDHP. Group ID Is the same all 5 years. We keep years 1, 4,5.
4. Drop when there is reverse switching from HDHP to traditional plans; including HDHP in first year or non-HDHP in last year	-432,274 observations -408,815 individuals -11,241 groups	
5. Drop people who have more than 2 plan switches in a year	-4,667 observations -879 individuals -953 groups	
6. Drop when 2 plans in a year if not associated with a full group switch (>=85% of group must have 2 group plan numbers in the year)	- 104,518 observations -56,226 individuals -8,783 groups	
7. To ensure no individual has more than one observation per year I drop the observations for individuals with 2 observations/year that represents fewer plan months	-22,792 observations -0 individuals	
8. Clean up after these steps. Dropping when there is no switching between	-2,081,362 observations -1,343,585 individuals -20,301 groups	

HDHP and non-HDHP, or only one year in data drops 2,081,362 (1,343,585 individuals in 20,301 groups)		
9) Drop people who move between HRRs	-126, 742 observations -53,331 individuals -12,036 groups	
10. Drop if not full group rollover, defined as between 95 and 105% of group switch from non-HDHP to HDHP from one year to a subsequent year	-429,917 observations -137,301 individuals -13,933 groups	This does not have to be in consecutive years.
Drop kids(0-17), or adults > 64 years old	-227,032 observations -64,387 individuals	
Drop if CBSA is missing or redacted (drops 61,599 observations	-61,559 observations	
Clean up: Drop if only 1 observation or no HDHP switch	-18,848 observations	
	Final HDHP switcher sample: 572,231 observations of 184,156 individuals in 8,096 groups Group includes 3065 observations on 2,805 individuals with a childbirth claim	-

Table 4-2: Full control vs. random sample

	Random sample	Excluded from random sample	T-statistic	p-value
Percent of Individuals:				
Female	51.6%	51.7%	3.554	<0.001
Ages 18-24	17.9%	17.5%	-23.462	<0.001
Ages 25-34	23.0%	22.4%	-24.535	<0.001
Ages 35-44	20.9%	21.2%	14.322	<0.001
Ages 45-54	20.3%	21.0%	31.565	<0.001
Ages 55-64	14.5%	14.6%	3.716	<0.001
Insurance Product: EPO	4.4%	4.3%	-10.020	<0.001
Insurance Product: HMO	18.6%	18.4%	-7.861	<0.001
Insurance Product: Indemnity	0.4%	0.4%	8.027	<0.001
Insurance Product: POS	56.7%	57.2%	18.190	<0.001
Insurance Product: PPO	19.9%	19.6%	-11.163	<0.001
Residence: Northeast	20.5%	20.8%	12.308	<0.001
Residence: South	42.8%	42.9%	7.292	<0.001
Residence: Midwest	15.5%	15.4%	-4.124	<0.001
Residence: West	19.8%	19.4%	-17.166	<0.001
Number individuals	4,713,262	12,743,044		

Notes: Within insurance product, EPO= exclusive provider organization; HMO=health maintenance organization; IND=indemnity plan; POS=point of service; PPO=preferred provider organization.   Metro population is defined as population within a core-based statistical area (CBSA), a Census-Bureau definition roughly corresponding to the cluster of population around an urban center.   All values are from the baseline year, the first year a person appears in the data.   P-values are taken from a ttest of sample equivalence.

Table 4-3: Pre-trends in spending outcomes

	Covariates	Model	Estimate (Tx * time)	p-value
Out-of-Pocket Spending				
[1]		GLM	0.020	0.019
[2]	X	GLM	0.012	0.165
Transaction Price				
[1]		GLM	-0.071	0.000
[2]	X	GLM	-0.081	<0.001
Charge Amount				
[1]		GLM	0.026	0.017
[2]	X	GLM	0.014	0.180
Notes: Estimates are from a GLM model with a log link, so interpretation is in semi-elasticities (e.g. a 1 year increase in time leads to an X% increase in outcome). Covariate adjustment for age, plan type, region of residence, median area income and area population.				

Table 4-4: Number of births

	Year				
	2010	2011	2012	2013	2014
Control					
Number of births	13,714	16,994	15,146	12,945	9,517
Percent of population	0.52%	0.50%	0.51%	0.50%	0.58%
N(population)	2,617,277	3,376,077	2,957,829	2,592,854	1,632,008
Treatment Pre					
Number of births	600	330	402	387	
Percent of population	0.54%	0.48%	0.57%	0.66%	
N(population)	110,661	68,596	70,003	58,477	
Treatment Post					
Number of births		206	241	345	554
Percent of population		0.57%	0.52%	0.49%	0.50%
N(population)		35,844	46,395	70,549	111,706

Table 4-5: Sample demographic characteristics: treatment vs. control

	Treatment	Full Control	p-value
Percent female	50.8%	51.6%	<0.001
Age Band			
Percent 18-24	12.0%	17.9%	<0.001
Percent 25-34	20.0%	23.0%	<0.001
Percent 35-44	24.4%	20.9%	<0.001
Percent 45-54	27.0%	20.3%	<0.001
Percent 55-64	13.7%	14.5%	<0.001
Insurance Product			
EPO	3.0%	4.4%	<0.001
HMO	14.3%	18.6%	<0.001
IND	0.0%	0.4%	<0.001
POS	69.6%	56.7%	<0.001
PPO	13.1%	19.9%	<0.001
Region of Residence			
Northeast	23.1%	20.5%	<0.001
South	37.2%	42.8%	<0.001
Midwest	24.0%	17.3%	<0.001
West	16.0%	19.8%	<0.001
Median CBSA Income ($)	56845.40	56332.25	<0.001
Metro Population (in thousands)	4416.6	5092.6	<0.001
N (individuals)	184,156	4,713,262	

Notes: Within insurance product, EPO= exclusive provider organization; HMO=health maintenance organization; IND=indemnity plan; POS=point of service; PPO=preferred provider organization.   Metro population is defined as population within a core-based statistical area (CBSA), a Census-Bureau definition roughly corresponding to the cluster of population around an urban center.   All values are from the baseline year, the first year a person appears in the data.   P-values are taken from a ttest of sample equivalence.

Table 4-6: Descriptive statistics among women who give birth

	Treatment	Full Control	p-value
Age Band			
Percent 18-24	12.73%	17.47%	0.000
Percent 25-34	67.09%	65.84%	0.171
Percent 35-44	19.82%	16.01%	0.000
Percent 45-54	0.07%	0.10%	0.595
Percent 55-64	0.00%	0.01%	0.676
Insurance Product			
EPO	1.71%	5.70%	0.000
HMO	8.09%	9.34%	0.026
IND	0.04%	0.05%	0.791
POS	75.40%	68.85%	0.000
PPO	14.76%	16.06%	0.065
Region of Residence			
Northeast	20.18%	17.34%	0.000
South	35.58%	45.27%	0.000
Midwest	24.28%	18.66%	0.000
West	20.36%	19.01%	0.077
Median CBSA Income ($)	56789.07	56109.20	0.000
Metro Population (in thousands)	4312.9	4737.9	0.000
N(individuals)	2,805	64,294	

Notes: Within insurance product, EPO= exclusive provider organization; HMO=health maintenance organization; IND=indemnity plan; POS=point of service; PPO=preferred provider organization.   Metro population is defined as population within a core-based statistical area (CBSA), a Census-Bureau definition roughly corresponding to the cluster of population around an urban center.   All values are from the baseline year, the first year a person appears in the data.   P-values are taken from a ttest of sample equivalence.

Table 4-7: Regressions on spending outcomes: full sample

Outcome and Model Number	Covariates	Model Specification	Estimate: Treatment Group (Pre-period)	Estimate: Post x Treatment (difference-in-differences estimate)	Constant
Out-of-Pocket Spending					
[Main Estimate]	X	GLM	146.69***	227.37***	790.10
[1]		OLS	146.08***	340.59***	639.69
[2]	X	OLS	152.99***	348.29***	901.41
[3]		GLM	140.87***	217.54***	790.10
[4]		Matching		484.41***	
Transaction Price					
[Main Estimate]	X	GLM	-81.87	-26.18	5821.81
[1]		OLS	33.78	-198.61**	5053.81
[2]	X	OLS	-113.06**	-71.75	3213.15
[3]		GLM	10.87	-167.67**	5828.86
[4]		Matching (CBSA)		-142.22**	
[5]		Matching (provider)		-157.13**	
[6]	X	OLS (subgroup with POS only)	-162.49*	-169.30*	3411.76
[7]	X	GLM (subgroup with POS only)	-142.65**	-97.90	5923.48
Charge Amount					
[Main Estimate]	X	GLM	-701.13***	-365.23*	11545.12
[1]		OLS	-447.89***	-804.84***	10086.71
[2]	X	OLS	-724.63***	-365.75*	9337.64
[3]		GLM	-499.85***	-740.24**	11546.53
[4]		Matching		-545.07***	

Notes: Estimates are from a regression model that uses a basic difference-in-differences framework to control for baseline differences between women who switch to an HDHP and those who remain in a stable benefit plan.   Covariates used in some models: Age, insurance type (e.g.   PPO, HMO), region of residence, median income at core-based statistical area (CBSA) level, CBSA population.   Year fixed effects in all models.   For matching estimates, each treatment individual is matched exactly to a person in the control group on plan type, CBSA or provider, year and whether an enrollee is 35 or older.   Each regression includes 71,381 observations on 67,099 individuals except rows 6-7 under "Transaction Price" which have N=36,320.   Standard errors are robust.
*p≤0.1 **p≤0.05 ***p≤0.001

Table 4-8: Spending outcome estimates from individual fixed effect regressions

	Covariates	Specification	Estimate: Treatment Post Period	Constant
Out-of-Pocket Spending				
[1]		OLS	244.53**	616.27
[2]	X	OLS	228.75**	27.32
Transaction Price				
[3]		OLS	-105.96	5253.68
[4]	X	OLS	-100.86	4966.24
Charge Amount				
[5]		OLS	78.59	10802.68
[6]	X	OLS	48.13	10246.05
Notes: Covariates in some models include insurance type. Regressions use 8,457 observations of 4,175 individuals with >1 observed. All models include year and individual fixed effects. Standard errors are robust. *p≤0.1 **p≤0.05 ***p≤0.001				

Table 4-9: Changes in transaction price by market

	Covariates	Model	Market Definition	Estimate: Treatment Group	Estimate: Market	Estimate: Market x Treatment Group	Estimate: Post x Treatment (diff-in-diff estimate)	Estimate: Post x Treatment x Market (triple diff estimate)	Constant
[Main Estimate]	X	GLM	> 3 total HSA providers	-392.13**	-424.12***	404.64**	343.14*	-511.94**	5806.11
[1]		GLM	> 3 total HSA providers	-271.54**	-423.52***	373.14**	109.45	-398.70*	5810.47
[2]		OLS	> 3 total HSA providers	-250.26*	-426.30***	381.57**	53.69	-379.07*	5323.29
[3]	X	OLS	> 3 total HSA providers	-387.05***	-483.52***	375.68**	262.47	-482.47**	3569.05
[4]		GLM	< median concentration	-195.63	-305.85***	251.38	17.01	-249.23	5810.47
[5]	X	GLM	< median concentration	-332.86**	-416.47***	297.52*	282.22	-404.00*	5806.25
[6]		OLS	< median concentration	-172.06	-307.84***	256.15	47.04	-217.74	5248.62
[7]	X	OLS	< median concentration	-333.19**	-497.71***	277.36*	162.78	-313.97	3555.85
[8]	X + fixed effect for providers	OLS	> 3 total HSA providers	-238.85**	-69.77	261.91	253.65*	-445.81**	4925.02
[9]	X	OLS (POS	>3 total HSA providers	-468.28***	-563.39***	390.62**	30.62	-253.12	3358.58

161

		plan type only)							
[10]	X	GLM (POS plan type only)	>3 total HSA providers	-502.86***	-506.85***	440.42**	119.66	-263.85	5895.85
[11]	X + fixed effect for providers	OLS (POS plan type only)	>3 total HSA providers	-279.47**	-121.21**	262.90**	109.71	-325.36	5138.47

Notes: Estimates are from generalized linear (GLM) or ordinary least squares (OLS) regression models that use a difference-in-differences framework to control for baseline differences between women who switch to an HDHP and those who remain in a stable benefit plan.   Estimates above show the main effects of belonging to the treatment group and of market type (more choices or less concentration), the relative difference that market definition makes for those in the treatment group relative to the control, the relative difference HDHP rollover (post-period) makes relative to control, and the relative difference that being in a market with more choices (or less concentration) makes relative to markets with fewer choices and compared with the control group.   The number of providers was determined by counting up the total number of hospitals that provided childbirth services in each HSA.   Market concentration was determined by using the Herfindahl-Hirschman Index (HHI).   Standard errors are robust.   All models include a year and year-market size fixed effects.   Number of observations=52,885 in all regressions except rows 9-11 which have N=36,320..
*p≤0.1 **p≤0.05 ***p≤0.001

Table 4-10: Estimates for women with no cost sharing for childbirth

	GLM with market interactions	Standard error	p-value	Linear Regression with Market Interactions	Standard error	p-value
Treatment Estimates	166.96	0.06	0.92	138.33	360.97	0.70
Treatment X Post	-168.84	0.08	0.85	-153.48	504.00	0.76
Treatment x Post x More_Choices	-196.81	0.10	0.85	-163.66	615.70	0.79
N (observations)	8,088			8,088		

Notes: The specifications used to generate these estimates mirror those in Table 4-9 but are limited to the subset of women who have no cost-sharing responsibilities. For the treatment group, that means they have spent through their deductible. Regressions include main effects for treatment, number of providers in market, post period, and year, and interactions on treatment x number of providers, post-period treatment and number of providers, and number of providers and year. Regressions are adjusted for age (in bands), type of insurance product, region of residence, median area income, and area population.

Table 4-11: Changes in likelihood of choosing a below-median price provider

	Covariates	Model	# Provider Choices	Estimate: Treatment	Estimate: Treatment X Post	N (observations)
Provider Choices Specific to Observed Group						
[Main Estimate]	X	LPM	≥ 1	0.033	-0.026	13,947
[1]		LPM	≥ 1	0.050	-0.055	13,947
[2]	X	Probit	≥ 1	0.033	-0.026	13,947
[3]		Probit	≥ 1	0.049	-0.054	13,947
Provider Choices Are Any Provider in HSA						
[Main Estimate]	X	LPM	≥ 1	0.00012	0.033	49,492
[1]		LPM	≥ 1	-0.003	0.034	49,492
[2]	X	Probit	≥ 1	0.00011	0.033	49,492
[3]		Probit	≥ 1	-0.0031	0.034	49,492
[4]	X	LPM	≥ 3	-0.022	0.061**	37,912
[5]	X	LPM	≥ 6	-0.026	0.047	24,051

Notes: Estimates are from a linear probability model (LPM) or probit model in which the transaction price is the outcome. Estimates reported are coefficients in LPM and marginal effects in probit. In models adjusted with covaraties, adjustments are made for age (in bands) insurance plan type, region of residence, median area income and area population. Standard errors are robust.
*$p \leq 0.1$ **$p \leq 0.05$ ***$p \leq 0.001$

Table 4-12: Testing for selection on number of provider choices and price of median provider

|	Mean Number of Observed Choices in Each Plan Group)	Linear Regression Estimate	p-value	Average of median-priced providers in plan-group choice set	Linear Regression Estimate	p-value
Control	1.58	[base]		5870.45	[base]	
Treatment (Pre-Period)	1.38	-0.18	0.053	5938.33	41.49	0.69
Treatment (Post-Period)	1.42	-0.22	0.043	5686.96	-164.54	0.15
Linear Combination (Treatment_Pre - Treatment_Post)		0.040	0.66		206.02	0.079
Notes: Provider choices are observed provider choices for each identifiable group, which is either an employer group or an employer-plan group. Regression is OLS regression on number of observed choices (minimum, 1; maximum 12) or median provider price with individuals split into groups (control, treatment pre and treatment-post). Fixed effects for year included in model. The linear combination represents the difference between the treatment in the pre-period and treatment group in the post period. Standard errors are robust.						

Figure 4-1: Histograms of transaction prices, charge prices, OOP spending

1a. Negotiated Transaction Price
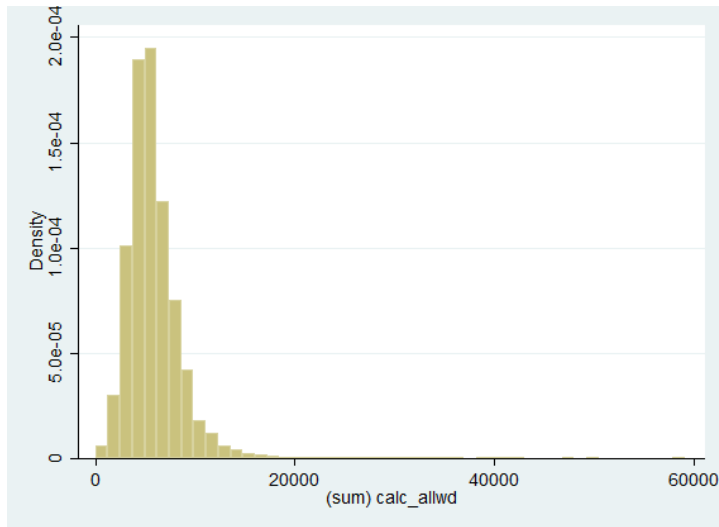
1b. Out-of-Pocket Spending
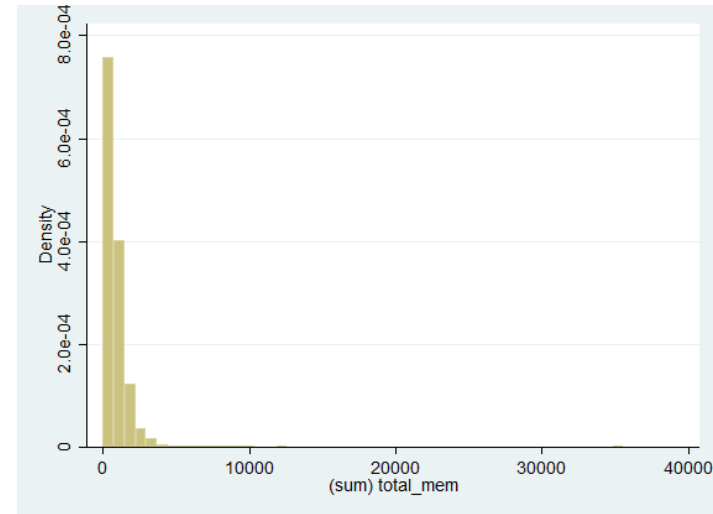
5c. Charge Price
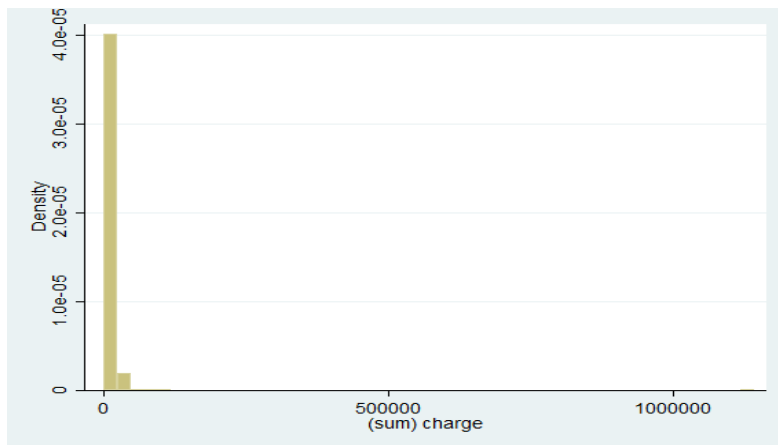
Figure 4-2: Pre-trends in transaction price, charge price and OOP spending

Notes: These graphs show predicted spending amounts from a GLM models with linear time trends and an interaction on dummy variable for women who later switch to an HDHP.   Estimates are adjusted for the following demographic covariates: age (in bands), type of insurance plan, region of residence, median area income, area population.

2a: Pre-Trends in Transaction Price
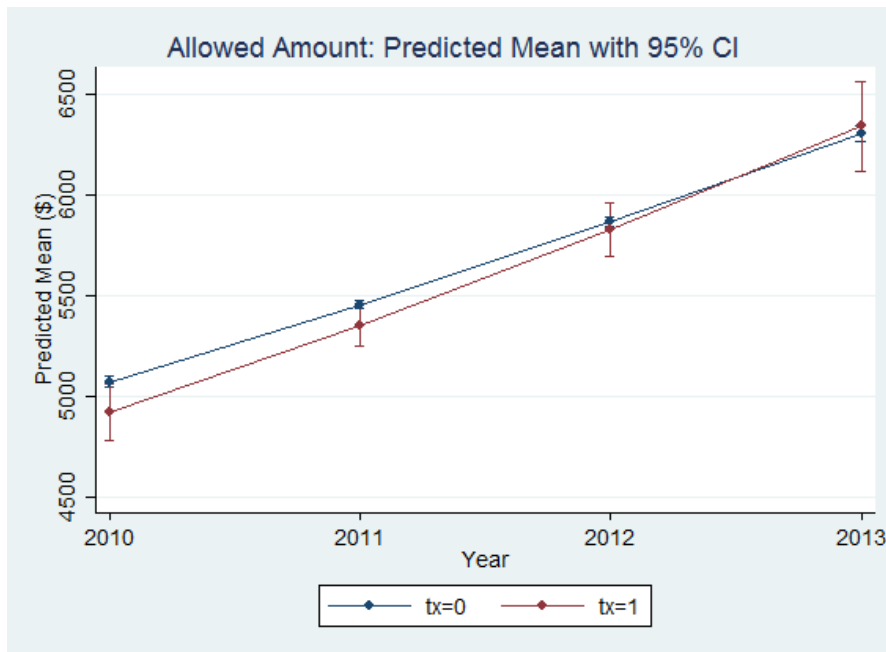
2b: Pre-Trends in Charge Prices

ChargePrice: Predicted Mean with 95% CI
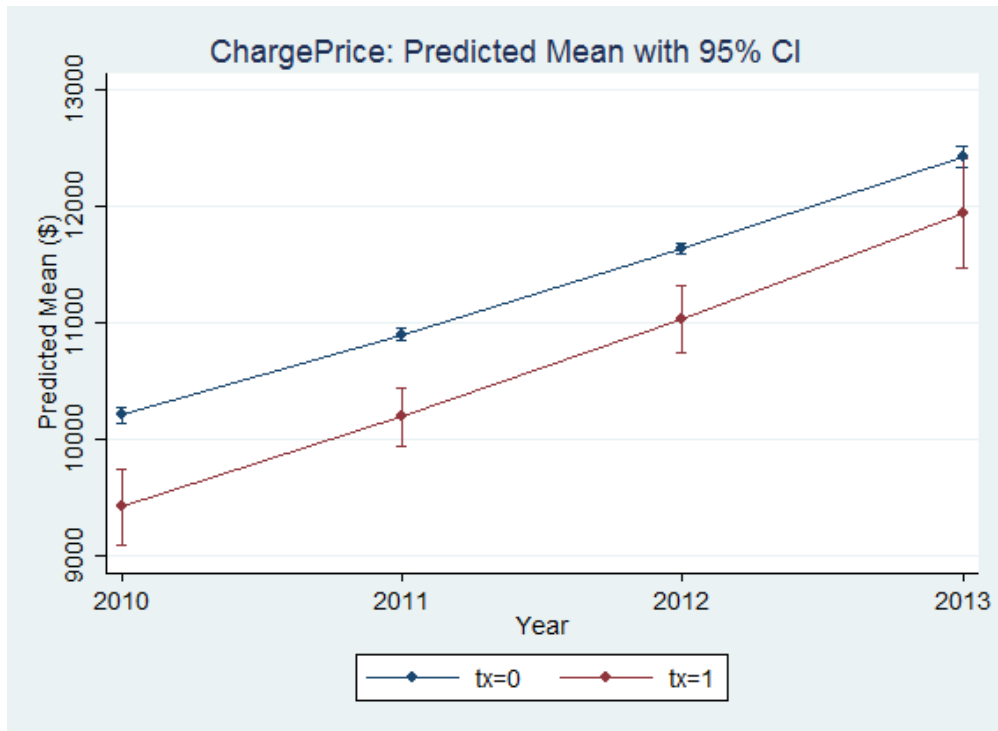
2c: Pre-Trends in Out-of-Pocket Spending

OOP Spending: Predicted Mean with 95% CI
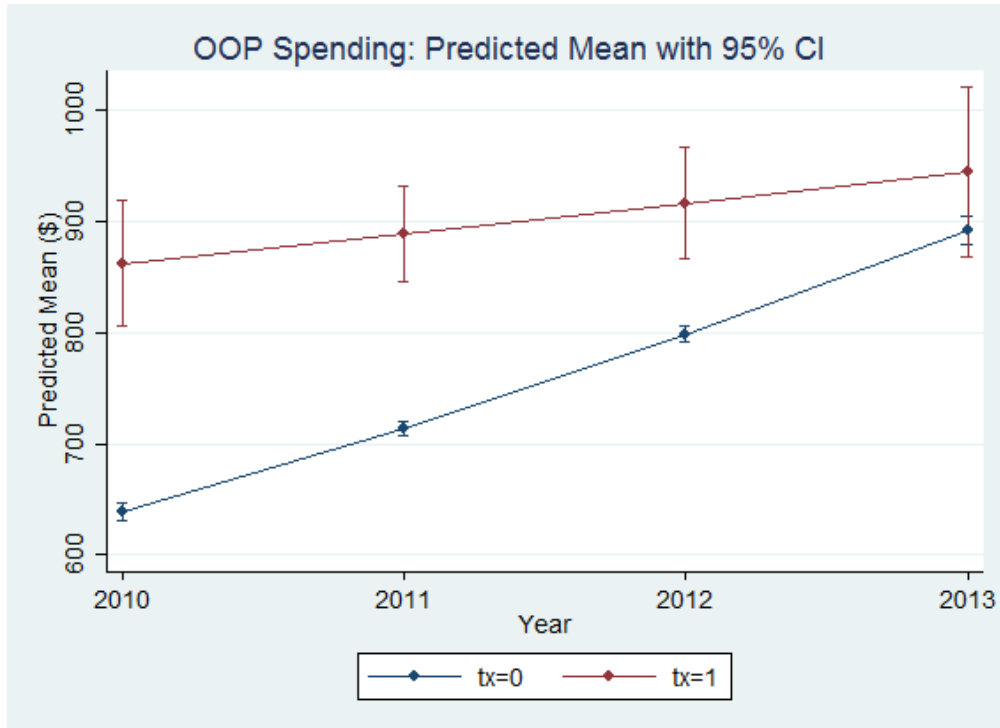
Figure 4-3: Mean annual transaction prices: treatment and control

Mean Negotiated Price

Legend: —— Control ••••• Treatment: Pre —— Treatment: Post

Y-axis: Dollars ($), ranging from 0 to 8000 in increments of 1000.

X-axis: 2010, 2011, 2012, 2013, 2014

Figure 4-4: Mean annual charge prices: treatment and control
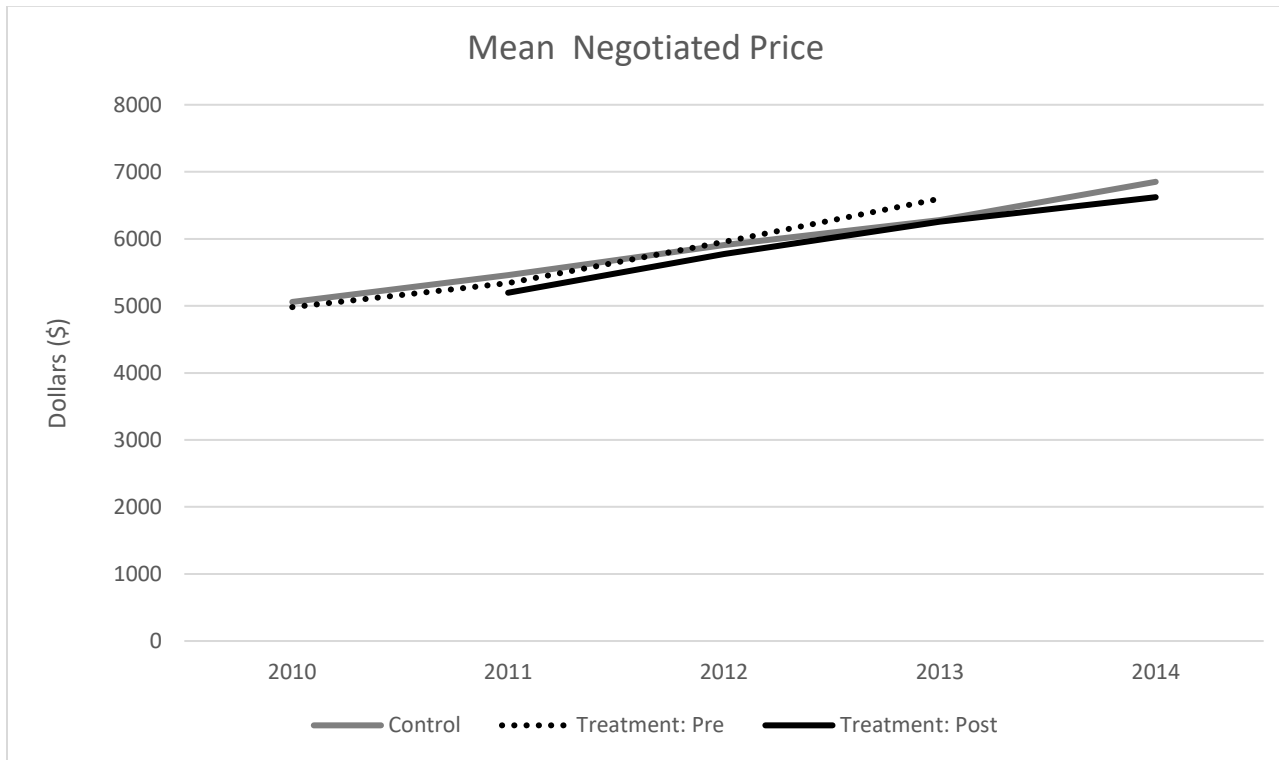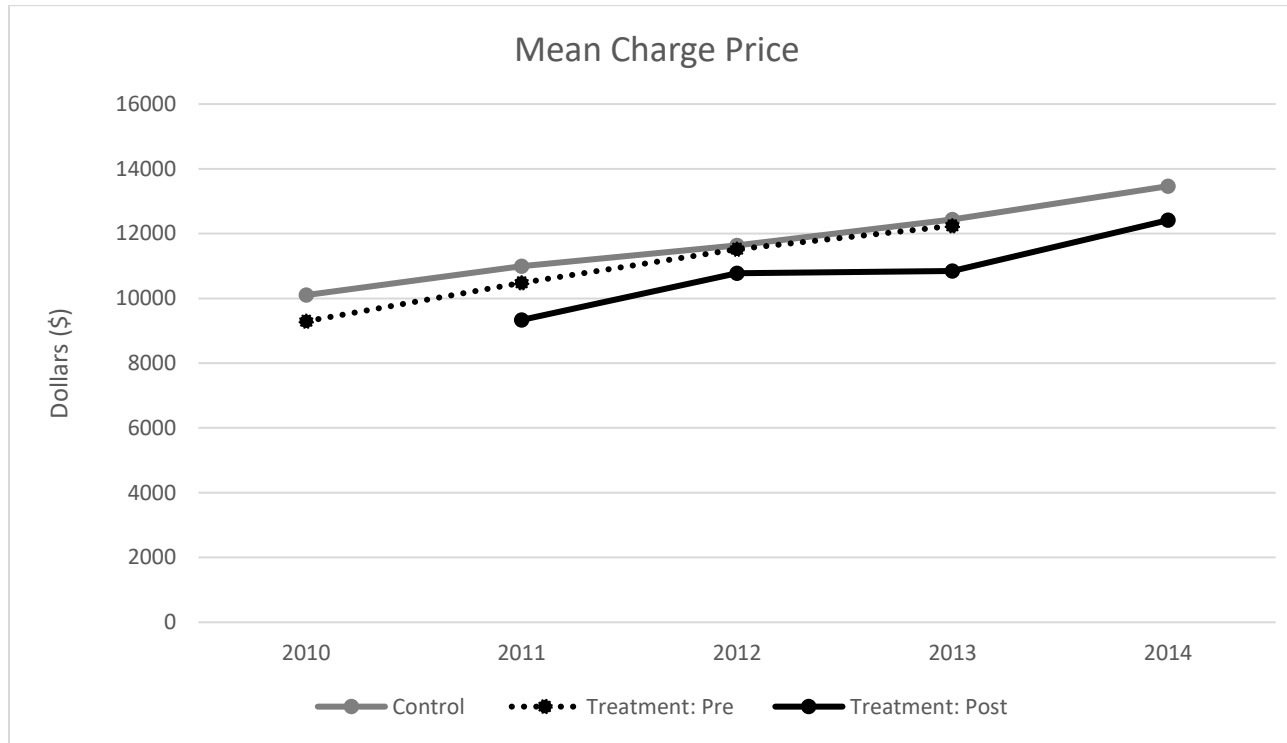
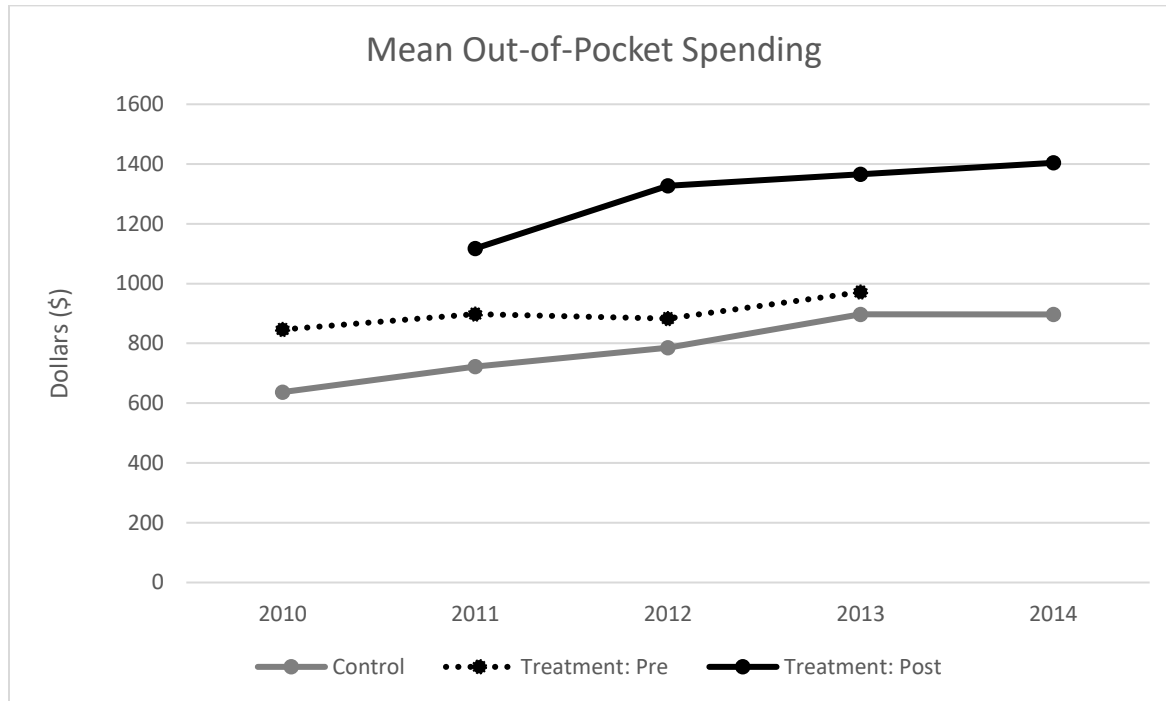*Figure 4-5: Mean annual OOP spending: treatment and control*

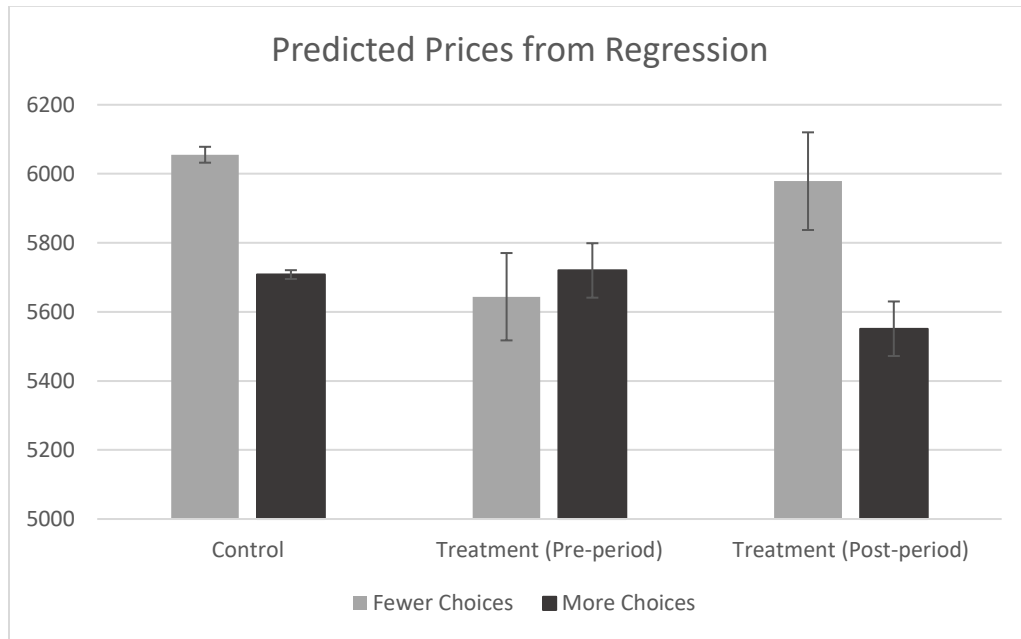*Figure 4-6: Predicted prices from GLM regression*

Predicted Prices from Regression

- Fewer Choices
- More Choices

Categories: Control, Treatment (Pre-period), Treatment (Post-period)

Chapter 5 : Conclusion

This dissertation explored three consumer-based interventions meant to encourage individuals to discern value, either on price or clinical benefit, and thus create more efficient spending patterns. However, based on the results presented in this dissertation, there are serious shortcomings of each of these strategies as they are currently deployed. If consumer-based incentives are to be used to mitigate spending growth, or at least to buy more health for each dollar spent, each may need to be augmented to see results.

In the first essay, I examine an intervention that combined value-based insurance design with mandated receipt of primary preventive care to stay in a cost-advantaged benefit plan. Specifically, the intervention mandated preventive office visits at regular intervals as well as high-value preventive screenings, and eliminated out-of-pocket cost sharing for both. Using claims from one year before and two years after the intervention for the population eligible for treatment compared with claims from a control sample that did not change benefit designs, I find that the intervention did achieve its aim of increasing receipt of high-value service. But, it also increased the use of low-value services, for which cost sharing and incentives were unchanged. In fact, the cross-price elasticity of low-value services was nearly identical to the own-price elasticity of high-value services, showing that consumers or their physicians did not discern between high- and low-value care. Using decomposition, I found that nearly 60 percent of the

total increase in low-value care was attributable to increases in preventive office visits, suggesting such visits result in a bundle of preventive services of both high- and low-value. The increase in low-value care represents an unintended spillover and undermines efficiencies gain through greater high-value preventive service use.

In the second essay, I looked at how price transparency initiatives may impact hospital-insurer bargaining. Price transparency is typically considered a consumer-based intervention, though some types of price release may allow hospitals or insurers to gather information about competitors. I examined the release of provider-specific price information via public databases in three states--Massachusetts, Oregon and Colorado-- through all-payer claims databases (APCDs) and measured whether exposure to new information changed hospital prices. To observe price information from these states and others used as a control, I obtained claims data from 2010-2014 for 3 large national insurers who sell private health insurance in the employer and individual market in all 50 states. I find no change in mean price, though some suggestion that prices at the lower tail of the distribution increase, particularly in markets that are less concentrated. Dispersion also drops in states that adopt APCDs, though its unclear if the drop is because of or coincident with the introduction of the APCD. These findings, while suggestive and not conclusive, align with theoretical work postulating that the widespread release of price information could have anti-competitive effects. It shows that less expensive hospitals may use additional information to raise prices, but that expensive hospitals are not similarly affected. Perversely, the release of pricing information in this way could make it harder—not easier—for consumers to ultimately find low-priced providers. This chapter calls into question the wisdom of releasing price information widely through APCDs, particularly given scarce evidence of consumer attention to such release.

In my third essay, I analyze the impacts on price paid for a common, shoppable service, childbirth, when people switch into a high-deductible health plans (HDHP). HDHPs aim to engage consumers in discerning value among medical services and providers by making them responsible for a greater portion of upfront medical costs. I use claims data that includes an indicator for HDHP plans and, in order to mitigate individual selection issues, compare prices from employers who rolled over an entire group of employees from non-HDHP to an HDHP plan against employers with stable benefit plans. Consistent with other studies, I do find reductions—by about 3 percent--in the prices paid for childbirth once enrollees are in an HDHP in markets with more choice for childbirth providers. However, these reductions remain even with provider fixed effects, implying that enrollees in HDHPs are paying slightly lower prices at the same providers compared with those in non-HDHP plans. This essay adds to previous work showing scant evidence of shopping behavior, even for services considered shoppable and are often shopped for.

In sum, these papers suggest that consumer-based efforts on their own are not currently sufficient to affect more efficient health spending patterns. These findings provide an indictment of several primary consumer-directed health care strategies, at least in their current form, and suggest several potential paths forward. One path would be to abandon or greatly diminish the role of consumerism in health care. Indeed, in the debate over whether people who use the health care system should be treated more as patients or consumers, the frustration of additional responsibilities without additional support often takes center stage. On this path, concentrating incentives and decisions within providers may then prove more fruitful. A second path is to continue to employ consumer-based incentives, and work to strengthen their usefulness to consumers and mitigate negative effects. There are good reasons to do so. In many spheres,

including health care, consumers value choice. And, although health care is a complex market, consumers have the most to gain by wringing value from it. Educating consumers to make better decisions may be easier than tweaking incentives to prohibit gaming by profit-maximizing providers.

Given both consumer preference for choice and the current policy environment, increased consumerism in health care is unlikely to abate. Yet as evidence mounts that giving consumers skin in the game is insufficient to engender efficient spending, more must be done. The ultimate solution likely lies somewhere between these outlined policy paths; we must move to support individual decision-making and better align provider incentives with value.

Bibliography

Agarwal, Rajender, Ashutosh Gupta, and A. Mark Fendrick. 2018. "Value-Based Insurance Design Improves Medication Adherence Without An Increase In Total Health Care Spending." *Health Affairs* 37 (7): 1057–64. https://doi.org/10.1377/hlthaff.2017.1633.

Agarwal, Rajender, Olena Mazurenko, and Nir Menachemi. 2017. "High-Deductible Health Plans Reduce Health Care Cost And Utilization, Including Use Of Needed Preventive Services." *Health Affairs* 36 (10): 1762–68. https://doi.org/10.1377/hlthaff.2017.0610.

Albæk, Svend, Peter Møllgaard, and Per B. Overgaard. 1997. "Government-assisted Oligopoly Coordination? A Concrete Case." *The Journal of Industrial Economics* 45 (4): 429–43.

Altman, Stuart H., David Shactman, and Efrat Eilat. 2006. "Could US Hospitals Go the Way of US Airlines?" *Health Affairs* 25 (1): 11–21.

Armstrong, Natalie. 2018. "Overdiagnosis and Overtreatment as a Quality Problem: Insights from Healthcare Improvement Research." *BMJ Qual Saf* 27 (7): 571–75.

Arrow, Kenneth J. 1963a. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review* 53 (5): 941–73.

Arrow, Kenneth J. 1963b. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review* 53 (5): 941–73.

Austin, D. Andrew, and Jane G. Gravelle. 2007. *Does Price Transparency Improve Market Efficiency? Implications of Empirical Evidence in Other Markets for the Health Sector*.

Baicker, Katherine, and Helen Levy. 2015. "Cost Sharing as a Tool to Drive Higher-Value Care." *JAMA Internal Medicine* 175 (3): 399–400. https://doi.org/10.1001/jamainternmed.2014.7595.

Baicker, Katherine, Joshua Schwartzstein, and Sendhil Mullainathan. 2015. "Behavioral Hazard in Health Insurance." *National Bureau of Economic Research Bulletin on Aging and Health*.

Batty, Michael, and Benedic Ippolito. 2017. "Mystery of the Chargemaster: Examining the Role of Hospital List Prices in What Patients Actually Pay." *Health Affairs* 36 (4): 689–96.

Becker's Hospital CFO Report. 2011. "The Ins and Outs of Successful Hospital Insurance Negotiations and Service Pricing," September 21, 2011. https://www.beckershospitalreview.com/finance/the-ins-and-outs-of-successful-hospital-insurance-negotiations-and-service-pricing.html.

Beeuwkes Buntin, Melinda, Amelia M Haviland, Roland McDevitt, and Neeraj Sood. 2011. "Healthcare Spending and Preventive Care in High-Deductible and Consumer-Directed Health Plans." *The American Journal of Managed Care* 17 (3): 222–30.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky. 1986. "The Nash Bargaining Solution in Economic Modelling." *The RAND Journal of Economics*, 176–88.

Brot-Goldberg, Zarek C, Amitabh Chandra, Benjamin R Handel, and Jonathan T Kolstad. 2017a. "What Does a Deductible Do? {The} Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *The Quarterly Journal of Economics* 132 (3): 1261–1318.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. 2017b. "What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *The Quarterly Journal of Economics* 132 (3): 1261–1318.

Brown, Zach Y. 2016. "An Empirical Model of Price Transparency and Markups in Health Care." Working Paper, Columbia University.

———. 2018. "Equilibrium Effects of Health Care Price Information." Working Paper.

Buchmueller, Thomas C. 2009. "Consumer-Oriented Health Care Reform Strategies: A Review of the Evidence on Managed Competition and Consumer-Directed Health Insurance." *The Milbank Quarterly* 87 (4): 820–41.

Bump, Suzanne M. 2017. *Evaluation of the 2012 Health Care Cost Containment Law in Massachusetts*. [Massachusetts. Department of the State Auditor].

Bundorf, M. Kate. 2012. "Consumer-Directed Health Plans: Do They Deliver?" http://www.rwjf.org/en/library/research/2012/10/consumer-directed-health-plans.html.

Cady, John F. 1976. "An Estimate of the Price Effects of Restrictions on Drug Price Advertising." *Economic Inquiry* 14 (4): 493–510.

Capps, Cory, David Dranove, and Mark Satterthwaite. 2003. "Competition and Market Power in Option Demand Markets." *RAND Journal of Economics*, 737–63.

Charlesworth, Christina J., Thomas H. A. Meath, Aaron L. Schwartz, and K. John McConnell. 2016. "Comparison of Low-Value Care in Medicaid vs Commercially Insured Populations." *JAMA Internal Medicine* 176 (7): 998. https://doi.org/10.1001/jamainternmed.2016.2086.

Chernew, Michael, Zack Cooper, Eugene Larsen-Hallock, and Fiona Scott Morton. 2018. "Are Health Care Services Shoppable? Evidence from the Consumption of Lower-Limb MRI Scans." National Bureau of Economic Research.

Chernew, Michael E., Allison B. Rosen, and A. Mark Fendrick. 2007. "Value-Based Insurance Design." *Health Affairs (Project Hope)* 26 (2): w195-203. https://doi.org/10.1377/hlthaff.26.2.w195.

Chernew, Michael E., Mayur R. Shah, Arnold Wegh, Stephen N. Rosenberg, Iver a. Juster, Allison B. Rosen, Michael C. Sokol, Kristina Yu-Isenberg, and a. Mark Fendrick. 2008. "Impact of Decreasing Copayments on Medication Adherence within a Disease Management Environment." *Health Affairs* 27 (1): 103–12. https://doi.org/10.1377/hlthaff.27.1.103.

Christensen, Hans, Eric Floyd, and Mark Maffett. 2017. "The Effects of Charge-Price Transparency Regulation on Prices in the Healthcare Industry."

Claxton, Gary, Bianca DiJulio, Benjamin Finder, Janet Lundy, Megan McHugh, Awo Osei-Anto, Heidi Whitmore, Jeremy Pickreign, and Jon Gabel. 2010. "Employer Health Benefits 2010 Annual Survey." Kaiser Family Foundation/Health Research & Educational Trust. https://kaiserfamilyfoundation.files.wordpress.com/2013/04/8085.pdf.

Claxton, Gary, Matthew Rae, Michelle Long, Anthony Damico, and Heidi Whitmore. 2018. "Health Benefits In 2018: Modest Growth In Premiums, Higher Worker Contributions At Firms With More Low-Wage Workers." *Health Affairs* 37 (11): 1892–1900.

Claxton, Gary, Matthew Rae, Michelle Long, Anthony Damico, Heidi Whitmore, and Gregory Foster. 2017. "Health Benefits In 2017: Stable Coverage, Workers Faced Considerable Variation In Costs." *Health Affairs* 36 (10): 1838–47. https://doi.org/10.1377/hlthaff.2017.0919.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2013. *Bargaining in the Shadow of a Giant: Medicare's Influence on Private Payment Systems*. National Bureau of Economic Research Washington, DC.

Cliff, Betsy Q., Christopher Krenz, Brady T. West, Helen Levy, A. Mark Fendrick, Tyler NA Winkelman, Jonathan So, Angela Fagerlin, and Jeffrey T. Kullgren. 2019. "Attitudes About Consumer Strategies Among Americans in High-Deductible Health Plans." *Medical Care*.

Coakley, Martha. 2011. "Examination of Health Care Cost Trends and Cost Drivers. Report for Annual Public Hearing." *Massachusetts Attorney General's Office. Boston, Massachusetts*.

Colla, Carrie H., Nancy E. Morden, Thomas D. Sequist, William L. Schpero, and Meredith B. Rosenthal. 2015. "Choosing Wisely: Prevalence and Correlates of Low-Value Health Care Services in the United States." *Journal of General Internal Medicine* 30 (2): 221–28. https://doi.org/10.1007/s11606-014-3070-z.

Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen. 2018. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." National Bureau of Economic Research.

Cutler, David, and Leemore Dafny. 2011. "Designing Transparency Systems for Medical Care Prices." *New England Journal of Medicine* 364 (10): 894–95. https://doi.org/10.1056/NEJMp1100540.

Cutler, David M, and Richard J Zeckhauser. 2000. "THE ANATOMY OF HEALTH INSURANCE." In *Handbook of Health Economics, Volume 1*, 1:563–643.

Dafny, Leemore, Kate Ho, and Robin S. Lee. 2016. "The Price Effects of Cross-Market Hospital Mergers." National Bureau of Economic Research.

Deb, Partha, Edward C. Norton, and Willard Graham Manning. 2017. *Health Econometrics Using Stata*. Stata Press College Station, TX.

Desai, Sunita, Laura A. Hatfield, Andrew L. Hicks, Michael E. Chernew, and Ateev Mehrotra. 2016. "Association between Availability of a Price Transparency Tool and Outpatient Spending." *Jama* 315 (17): 1874–81.

Devers, Kelly J., Lawrence P. Casalino, Liza S. Rudell, Jeffrey J. Stoddard, Linda R. Brewster, and Timothy K. Lake. 2003. "Hospitals' Negotiating Leverage with Health Plans: How and Why Has It Changed?" *Health Services Research* 38 (1p2): 419–46.

Dunn, Abe, Adam Hale Shapiro, and Eli Liebman. 2013. "Geographic Variation in Commercial Medical-Care Expenditures: A Framework for Decomposing Price and Utilization." *Journal of Health Economics* 32 (6): 1153–65.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2015. "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D." *The Quarterly Journal of Economics* 130 (2): 841–99.

Ellis, Randall P., and Willard G. Manning. 2007. "Optimal Health Insurance for Prevention and Treatment." *Journal of Health Economics* 26 (6): 1128–50. https://doi.org/10.1016/j.jhealeco.2007.09.002.

Fendrick, A. Mark, and Michael E. Chernew. 2006. "Value-Based Insurance Design: Aligning Incentives to Bridge the Divide between Quality Improvement and Cost Containment." *American Journal of Managed Care* 12 (SPEC. ISS.): 5–10.

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

Fronstin, Paul. 2012. "Findings from the 2012 EBRI/MGA Consumer Engagement in Health Care Survey." *EBRI Issue Brief / Employee Benefit Research Institute*, no. 379 (December): 1–27.

———. 2013. "Findings from the 2013 EBRI/Greenwald & Associates Consumer Engagement in Health Care Survey." *EBRI Issue Brief / Employee Benefit Research Institute*, no. 393 (December): 4–25.

Frost, Amanda, and David Newman. 2016. "Spending on Shoppable Services in Health Care." *Health Care Cost Institute Issue Brief*, no. 11.

Gaynor, Martin, Kate Ho, and Robert J. Town. 2015. "The Industrial Organization of Health-Care Markets." *Journal of Economic Literature* 53 (2): 235–84.

Ginsburg, P. B. 2007. "Shopping For Price In Medical Care." *Health Affairs* 26 (2): w208–16. https://doi.org/10.1377/hlthaff.26.2.w208.

Goldman, Dana, and Tomas J. Philipson. 2007. "Integrated Insurance Design in the Presence of Multiple Medical Technologies." *American Economic Review* 97 (2): 427–32. https://doi.org/10.1257/aer.97.2.427.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town. 2015. "Mergers When Prices Are Negotiated: Evidence from the Hospital Industry." *American Economic Review* 105 (1): 172–203.

Grennan, Matthew, and Ashley Swanson. 2016. "Transparency and Negotiated Prices: The Value of Information in Hospital-Supplier Bargaining." National Bureau of Economic Research.

Gruber, Jonathan, Johanna Catherine Maclean, Bill Wright, Eric Wilkinson, and Kevin Volpp. 2016. "The Impact of Increased Cost-Sharing on Utilization of Low Value Services: Evidence from the State of Oregon." Cambridge, MA. https://doi.org/10.3386/w22875.

Hanover Research. 2017. "Awareness and Usage of Healthcare Tools Prepared for HealthSparq, Inc." December.

Hirth, Richard A., Elizabeth Q. Cliff, Teresa B. Gibson, M. Richard McKellar, and A. Mark Fendrick. 2016. "Connecticut's Value-Based Insurance Plan Increased The Use Of Targeted Services And Medication Adherence." *Health Affairs* 35 (4): 637–46. https://doi.org/10.1377/hlthaff.2015.1371.

Ho, Kate, and Robin S. Lee. 2013. *Insurer Competition and Negotiated Hospital Prices*. National Bureau of Economic Research Cambridge (MA).

Ho, Katherine. 2009. "Insurer-Provider Networks in the Medical Care Market." *American Economic Review* 99 (1): 393–430.

Hong, Arthur S., Dennis Ross-Degnan, Fang Zhang, and J. Frank Wharam. 2017. "Clinician-Level Predictors for Ordering Low-Value Imaging." *JAMA Internal Medicine* 177 (11): 1577. https://doi.org/10.1001/jamainternmed.2017.4888.

Horn, Henrick, and Asher Wolinsky. 1988. "Bilateral Monopolies and Incentives for Merger." *The RAND Journal of Economics*, 408–19.

Kaestner, Robert, and Anthony T. Lo Sasso. 2015. "Does Seeing the Doctor More Often Keep You out of the Hospital?" *Journal of Health Economics* 39 (January): 259–72. https://doi.org/10.1016/J.JHEALECO.2014.08.003.

Kratka, Allison, Charlene A Wong, Riley Herrmann, Kathryn Hong, Aleena Karediya, Iris Yang, and Peter A Ubel. 2018. "Finding Health Care Prices Online—How Difficult Is It to Be an Informed Health-Care Consumer?" *JAMA Internal Medicine* 178 (3): 423–424.

Kullgren, Jeffrey T., Betsy Q. Cliff, Chris D. Krenz, Helen Levy, Brady West, A. Mark Fendrick, Jonathan So, and Angela Fagerlin. 2019. "A Survey Of Americans With High-Deductible Health Plans Identifies Opportunities To Enhance Consumer Behaviors." *Health Affairs* 38 (3): 416–24.

Kullgren, Jeffrey T., Elizabeth Q. Cliff, Christopher Krenz, Brady T. West, Helen Levy, A. Mark Fendrick, and Angela Fagerlin. 2018. "Consumer Behaviors Among Individuals Enrolled in High-Deductible Health Plans in the United States." *JAMA Internal Medicine* 178 (3): 424–26.

Kullgren, Jeffrey T, Katia A Duey, and Rachel M Werner. 2013. "A Census of State Health Care Price Transparency Websites." *JAMA: The Journal of the American Medical Association* 309 (23): 2437–38. https://doi.org/10.1001/jama.2013.6557.

Kullgren, Jeffrey T., Kevin G. Volpp, and Daniel Polsky. 2013. "Are the Healthy Behaviors of US High-Deductible Health Plan Enrollees Driven by People Who Chose These Plans? Smoking as a Case Study." *PLoS ONE* 8 (2): e56154. https://doi.org/10.1371/journal.pone.0056154.

Kwoka, John E. 1984. "Advertising and the Price and Quality of Optometric Services." *The American Economic Review* 74 (1): 211–16.

Lave, Judith R., Aiju Men, Brian T. Day, Wei Wang, and Yuting Zhang. 2011. "Employee Choice of a High-Deductible Health Plan across Multiple Employers." *Health Services Research* 46 (1 Pt 1): 138–54. https://doi.org/10.1111/j.1475-6773.2010.01167.x.

Leuven, Edwin, and Barbara Sianesi. 2003. "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing"."

Lewis, Matthew S., and Kevin E. Pflum. 2015. "Diagnosing Hospital System Bargaining Power in Managed Care Networks." *American Economic Journal: Economic Policy* 7 (1): 243–74.

Maciejewski, Matthew L., Daryl Wansink, Jennifer H. Lindquist, John C. Parker, and Joel F. Farley. 2014. "Value-Based Insurance Design Program in North Carolina Increased Medication Adherence but Was Not Cost Neutral." *Health Affairs* 33 (2): 300–308. https://doi.org/10.1377/hlthaff.2013.0260.

Manning, Willard G, Joseph P Newhouse, Naihua Duan, Emmett B Keeler, and Arleen Leibowitz. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *The American Economic Review*, 251–77.

McGuire, Thomas G. 2011. "Demand for Health Insurance." In *Handbook of Health Economics*, 2:317–96. Elsevier.

McWilliams, J Michael, and Aaron L Schwartz. 2017. "Focusing on High-Cost Patients - The Key to Addressing High Costs?" *The New England Journal of Medicine* 376 (9): 807–9. https://doi.org/10.1056/NEJMp1612779.

Mehrotra, Ateev, Katie M. Dean, Anna D. Sinaiko, and Neeraj Sood. 2017. "Americans Support Price Shopping for Health Care, but Few Actually Seek out Price Information." *Health Affairs* 36 (8): 1392–1400.

Morden, Nancy E., Carrie H. Colla, Thomas D. Sequist, and Meredith B. Rosenthal. 2014. "Choosing Wisely—the Politics and Economics of Labeling Low-Value Services." *New England Journal of Medicine* 370 (7): 589–92.

Newhouse, J. P., and The Insurance Experiment Group. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge, Massachusetts: Harvard University Press.

Newhouse, Joseph P. 2006. "Reconsidering the Moral Hazard-Risk Avoidance Tradeoff." *Journal of Health Economics* 25 (5): 1005–14. https://doi.org/10.1016/j.jhealeco.2006.06.003.

Norton, Edward C. 2016. "Applied Econometrics in Health Services Research: HMP 826 Class Notes." University of Michigan Department of Health Management and Policy.

"Oregon All Payer All Claims Database (APAC) Use Case Document." 2017. Oregon Health Authority. https://www.oregon.gov/oha/HPA/ANALYTICS/APAC%20Page%20Docs/APAC-Use-Cases.pdf.

Osborne, Nicholas H., Lauren H. Nicholas, Andrew M. Ryan, Jyothi R. Thumma, and Justin B. Dimick. 2015. "Association of Hospital Participation in a Quality Reporting Program with Surgical Outcomes and Expenditures for Medicare Beneficiaries." *Jama* 313 (5): 496–504.

"Overview of the Massachusetts All-Payer Claims Database." 2014. Center for Health Information and Analysis. https://www.apcdcouncil.org/sites/apcdcouncil.org/files/media/state/ma-apcd-overview-2014.pdf.

Pauly, Mark. 1968. "The Economics of Moral Hazard." *The American Economic Review* 58 (3): 531–37. https://doi.org/10.2307/1813785.

Pauly, Mark V. 1968. "The Economics of Moral Hazard: Comment." *American Economic Review* 58 (3): 531–37.

Pauly, Mark V., and Fredric E. Blavin. 2008. "Moral Hazard in Insurance, Value-Based Cost Sharing, and the Benefits of Blissful Ignorance." *Journal of Health Economics* 27 (6): 1407–17. https://doi.org/10.1016/j.jhealeco.2008.07.003.

Reddy, Sheila R, Dennis Ross-Degnan, Alan M Zaslavsky, Stephen B Soumerai, and James F Wharam. 2014. "Impact of a High-Deductible Health Plan on Outpatient Visits and Associated Diagnostic Tests." *Medical Care* 52 (1): 86–92. https://doi.org/10.1097/MLR.0000000000000008.

Reinhardt, Uwe E. 2006. "The Pricing of U.S. Hospital Services: Chaos behind a Veil of Secrecy." *Health Affairs (Project Hope)* 25 (1): 57–69. https://doi.org/10.1377/hlthaff.25.1.57.

Ryan, Andrew M., James F. Burgess, and Justin B. Dimick. 2015. "Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences." *Health Services Research* 50 (4): 1211–35. https://doi.org/10.1111/1475-6773.12270.

Ryan, Andrew M., James F. Burgess Jr, and Justin B. Dimick. 2015. "Why We Should Not Be Indifferent to Specification Choices for Difference-in-differences." *Health Services Research* 50 (4): 1211–35.

Schwartz, Aaron L., Michael E. Chernew, Bruce E. Landon, and J. Michael McWilliams. 2015. "Changes in Low-Value Services in Year 1 of the Medicare Pioneer Accountable Care Organization Program." *JAMA Internal Medicine* 175 (11): 1815. https://doi.org/10.1001/jamainternmed.2015.4525.

Schwartz, Aaron L, Bruce E Landon, Adam G Elshaug, Michael E Chernew, and J Michael McWilliams. 2014. "Measuring Low-Value Care in Medicare." *JAMA Internal Medicine* 02115 (7): 1–9. https://doi.org/10.1001/jamainternmed.2014.1541.

Sinaiko AD, Mehrotra A, and Sood N. 2016. "Cost-Sharing Obligations, High-Deductible Health Plan Growth, and Shopping for Health Care: Enrollees with Skin in the Game." *JAMA Internal Medicine* 176 (3): 395–97. https://doi.org/10.1001/jamainternmed.2015.7554.

Sinaiko, Anna D., and Meredith B. Rosenthal. 2011. "Increased Price Transparency in Health Care—Challenges and Potential Effects." *New England Journal of Medicine* 364 (10): 891–94.

Slightham, Toni. 2016. "Question about Oregon APAC Database," April 25, 2016.

Smith, Robert A., Kimberly S. Andrews, Durado Brooks, Stacey A. Fedewa, Deana Manassaram-Baptiste, Debbie Saslow, Otis W. Brawley, and Richard C. Wender. 2018. "Cancer Screening in the United States, 2018: A Review of Current American Cancer Society Guidelines and Current Issues in Cancer Screening." *CA: A Cancer Journal for Clinicians* 68 (4): 297–316.

Sood, Neeraj, Zachary Wagner, Peter Huckfeldt, and Amelia M. Haviland. 2013. "Price Shopping in Consumer-Directed Health Plans." In *Forum for Health Economics and Policy*, 16:35–53. De Gruyter.

Town, Robert, and Gregory Vistnes. 2001. "Hospital Competition in HMO Networks." *Journal of Health Economics* 20 (5): 733–53.

Tu, Ha, and Rebecca Gourevitch. 2014. "Moving Markets: Lessons from New Hampshire's Health Care Price Transparency Experiment." Mathematica Policy Research.

Tu, Ha T, and Johanna R Lauer. 2008. "Word of Mouth and Physician Referrals Still Drive Health Care Provider Choice." *Research Brief*, no. 9 (December): 1–8.

US Preventive Services Task Force. 2010. "The Guide to Clinical Preventive Services 2010-2011: Recommendations of the US Preventive Services Task Force."

Weiss, Jeff. 2012. "A Better Way to Negotiate Healthcare Costs." *U.S. News & World Report*, April 6, 2012. https://www.usnews.com/opinion/blogs/economic-intelligence/2012/04/06/a-better-way-to-negotiate-healthcare-costs.

Whaley, Christopher. 2015. "Provider Responses to Online Price Transparency."

Wharam, J. Frank, Alison A. Galbraith, Ken P. Kleinman, Stephen B. Soumerai, Dennis Ross-Degnan, and Bruce E. Landon. 2008. "Cancer Screening before and after Switching to a High-Deductible Health Plan." *Annals of Internal Medicine* 148 (9): 647–55.

Wharam, J. Frank, Fang Zhang, Emma M. Eggleston, Christine Y. Lu, Stephen Soumerai, and Dennis Ross-Degnan. 2017. "Diabetes Outpatient Care and Acute Complications before and after High-Deductible Insurance Enrollment: A Natural Experiment for Translation in Diabetes (NEXT-D) Study." *JAMA Internal Medicine* 177 (3): 358–68.

Wharam, J. Frank, Fang Zhang, Jamie Wallace, Christine Lu, Craig Earle, Stephen B. Soumerai, Larissa Nekhlyudov, and Dennis Ross-Degnan. 2019. "Vulnerable And Less Vulnerable Women In High-Deductible Health Plans Experienced Delayed Breast Cancer Care." *Health Affairs* 38 (3): 408–15.

White, Chapin, James D. Reschovsky, and Amelia M. Bond. 2014. "Understanding Differences between High- and Low-Price Hospitals: Implications for Efforts to Rein in Costs." *Health Affairs* 33 (2): 324–31.

Wu, Sze-jung, Gosia Sylwestrzak, Christiane Shah, and Andrea DeVries. 2014. "Price Transparency for MRIs Increased Use of Less Costly Providers and Triggered Provider Competition." *Health Affairs* 33 (8): 1391–98.

Zeckhauser, Richard. 1970. "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory* 2 (1): 10–26.

Zeckhauser, Richard J. 1970. "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Jounral of Economic Theory* 2: 10–26.

Zheng, Sarah, Zhong Justin Ren, Janelle Heineke, and Kimberley H. Geissler. 2016. "Reductions in Diagnostic Imaging With High Deductible Health Plans." *Medical Care* 54 (2): 110–17. https://doi.org/10.1097/MLR.0000000000000472.