NBER WORKING PAPER SERIES

THE IMPACT OF PRICE TRANSPARENCY IN OUTPATIENT PROVIDER MARKETS

Kayleigh Barnes
Sherry A. Glied
Benjamin R. Handel
Grace Kim

Working Paper 32580
http://www.nber.org/papers/w32580

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2024

We acknowledge and thank the New York State Health Foundation for funding under grant number #17-04903. We are very grateful to FAIRHealth for their partnership in the development and execution of this randomized trial and for making the evaluation data available to us. We especially thank Hunt Allcott for his invaluable advice and input. We thank seminar participants at the Electronic Health Economics Colloquium, APPAM Fall Conference, NYU Wagner seminar for comments, including Christopher Whaley, Kathy O'Regan, and Leanna Stiefel. The study was deemed by the New York University Institutional Review Board (IRB-FY2017-582) not to constitute human subjects research. Replication files are available from Grace Kim. The results and opinions in this study are those of the authors alone and may not reflect the views of the New York State Health Foundation, FAIRHealth, or those thanked in these acknowledgements. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Impact of Price Transparency in Outpatient Provider Markets
Kayleigh Barnes, Sherry A. Glied, Benjamin R. Handel, and Grace Kim
NBER Working Paper No. 32580
June 2024
JEL No. I11,L1

ABSTRACT

Medical provider price transparency is often touted as a way to lower health care spending. But the impact of price transparency is theoretically ambiguous: it could lower health care spending via increased consumer price shopping or improved insurer bargaining but could also raise health care prices via improved provider bargaining or provider collusion. We conduct a randomized-controlled trial to examine the impact of a state-wide medical charge transparency tool in outpatient provider markets in New York State. In the experiment, individual providers' billed charges (list prices) were released randomly at the level of the procedure and three-digit zipcode. We use a comprehensive commercial claims database to assess the impact of this intervention and find that it leads to a small increase in overall billed charges (+0.75%). This effect is concentrated among low-priced providers in markets with low out-of-network spending, suggesting that the transparency tool improves provider pricing information. We find no evidence of quantity effects. Results do not vary consistently across specialty groups, market concentration, frequency of service use, or frequency of website use. These results are consistent with price transparency having a minimal effect on consumer shopping while slightly improving provider information about competitors' charges.

Kayleigh Barnes
Federal Reserve Board of Governors
Constitution Ave NW & 20th St. NW
Washington, DC 20551
kayleigh.n.barnes@frb.gov

Sherry A. Glied
Robert F. Wagner Graduate School
of Public Service
New York University
105 East 17th Street
New York, NY 10003
and NBER
sherry.glied@nyu.edu

Benjamin R. Handel
Department of Economics
University of California, Berkeley
508-1 Evans Hall #3880
Berkeley, CA 94720
and NBER
handel@berkeley.edu

Grace Kim
New York University
gk1512@nyu.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/5853

1 Introduction

Health care policymakers, insurers, and private companies have frequently discussed the transparency of health care pricing information as a way to reign in rising health care spending (Reinhardt 2006; Sinaiko, Kakani, and Rosenthal 2019; Volpp 2016). Prices in health care markets are notoriously variable, opaque, and confusing, and price transparency has the potential to reduce search costs and information asymmetries for consumers seeking less expensive, high value health care. Most consumers purchasing services, especially those who intentionally or inadvertently purchase services outside their insurer networks, have no easy way to ascertain or compare prices charged by different providers. Recently, the federal government has sought to use regulatory levers to make healthcare prices more transparent, including upholding a CMS 2019 Final Rule that mandates hospitals release comprehensive information regarding negotiated rates (Wilensky 2019; Kullgren and Fendrick 2021; Glied 2021; UnitedHealthcare 2023).

Proponents of price transparency note the potential benefits in making price information available and easily accessible to consumers, reducing consumer search costs. They also note the potential benefits for provider pricing: if consumers are more price elastic then providers may lower prices in this more competitive landscape.

Skeptics are concerned that the release of pricing information might lead to higher prices either via explicit collusion among providers or via providers gaining a competitive advantage from realizing that they are systematically under-pricing relative to similar-quality peers (tacit collusion). If price elasticity at the provider level is low, providers may raise prices and charge more without losing much business. Albaek, Mollgaard and Overgaard (1997) document exactly this phenomenon in the context of the Danish concrete industry, where the publishing of transaction prices produced a 15-20% increase in concrete prices, as producers stopped offering confidential discounts to purchasers. Finally, if consumers equate price with quality, providers might raise prices to signal quality. With these kinds of issues in mind, some policymakers and economists have urged caution in promoting price transparency (see, e.g., Cutler and Dafny 2011).

In this study, we examine price transparency for providers across the state of New York, using a randomized control trial embedded in the information provision platform run by FAIR Health, a non-profit organization dedicated to promoting price transparency in health care markets. Prior to our intervention, FAIR Health provided market-level information on typical prices for a given procedure, but did not provide information for specific individual providers. We partnered with FAIR Health to implement a statewide randomized intervention providing individual-level provider billed charge information on their platform. We randomized whether this individual provider-level information would be provided for providers at the procedure-geozip level. For a given kind of procedure in a given market, all providers of that procedure are either randomized into our individual-level price information treatment or to usual treatment (market-level price information only).[1]

The design was set up specifically to capture market-level pricing and demand effects. Our intervention applies to 107 procedures and all geozips in New York and was in place for two years. The data captured all commercial claims in the FAIR Health data warehouse related to these procedures and geozips and encompassed over 110 million claims and over 205,000 providers. We present a series of descriptive statistics showing (i) that our intervention is well-balanced across treatment and control arms (ii) that there is meaningful heterogeneity across procedure-geozip markets in market power, out-of-network claims, and ex ante price dispersion and (iii) that usage of the information tool we study averages just fewer than 10,000 uses per month.

Our randomized intervention allows us to cleanly identify the net price and quantity effects of our information-provision tool while also allowing us to study heterogeneous effects related to (i) initial absolute procedure prices (ii) initial prices relative to peers (iii) specific types of medical procedures and (iv) specific kinds of market structures. Importantly, since our information provision focuses on individual provider billed charges, rather than negotiated rates between insurers and providers, our analysis is especially relevant for the out-of-network services that these charges are germane to. However, since billed charges also feed into negotiated rates with insurers, and are meaningfully correlated with them (Batty and Ippolito 2017). we also study services that are shoppable but typically obtained in-network. To our knowledge, this is the first randomized intervention of a price transparency tool that is specifically designed to address market-level effects as well as the effects on consumers and specific providers.[2]

Our randomized intervention directly guides our econometric approach to estimating our effects of interest. We use a difference-in-differences approach with different configurations of the fixed effects that compares key price and quantity outcomes for providers in treated procedure-location pairs to the same outcomes for providers in control procedure-location pairs. In our primary difference-in-differences specification, we control for procedure, geozip, and trimester fixed effects and also time fixed effects interacted with the procedure and geozip indicators, which leverage any procedure or geozip specific time-series variation spanning the periods pre- and post-intervention. We conduct a number of robustness analyses to our main difference-in-differences specification including, e.g., a difference-in-differences version without time fixed effects, and find similar results for these alternatives.

In our primary difference-in-differences specification, we find that, across all procedures and locations,

[1] The trial was submitted to the AEA RCT registry.

[2] Several prior major studies in health economics rely on randomized controlled trials as a "gold standard" for identification. See, for example, the 1974 RAND Health Insurance Experiment (HIE) studying the price elasticity of demand for health care and the Oregon Health Insurance Experiment studying the effects of expanding access to public health insurance (Manning et al. 1987; Newhouse 1996; Finkelstein et al. 2012).

providing individual-level provider charge information increases prices by 0.75% and has no statistically significant impact on quantity. We assess these impacts separately for providers whose prices were initially above or below the median for a given procedure in their geozip and find modest but larger increases in prices for providers who were initially below the median. In addition, we find no quantity impacts for providers who were initially high-priced as opposed to low-priced, suggesting that our intervention had no meaningful impact on the extent of consumer price shopping. We also find that providers located in procedure markets that are above median market concentration, measured with procedure-geozip HHI, have slightly larger price increases than those below median, with no statistically significant quantity differences.

Given that our intervention provides information on billed charges, rather than insurer-contracted prices, we focus especially on out-of-network claims, for which billed charges are relevant.[3] We observe essentially no price change for procedures with a high proportion of out-of-network claims in response to our intervention but a 2.5% price increase for procedures with a low proportion of out-of-network claims.

We also investigate the effects of our intervention for specific procedure categories. We utilized the sharpened False Discovery Rate (FDR) q-values to adjust for testing multiple hypotheses. After adjusting, we do not find significant differences across categories.

Taken together, these results are consistent with our intervention having (i) a minimal effect on consumer price shopping (as evidenced by no quantity effect) and (ii) a small positive effect on prices (0.75%-1.25% increase depending on the specification). These price increases are consistent with tacit collusion, or conscious parallelism, between providers who realized that they are under-charging relative to their peers.

Our results should be viewed with a number of caveats. Our results are specific to the information provision tool provided by FAIR Health and how many providers and consumers use the tool. While website utilization average around 10,000 people per month during our sample period, we cannot distinguish the type of user (consumer or provider).

Finally, our intervention applies to billed charges, which, though relevant for out-of-network claims and potentially relevant via what they signal about negotiated rates, ultimately are a noisy signal of insurer-provider negotiated rates. Despite these potential difficulties our results shed light on important market-level issues related to price transparency that prior studies (who also share some of these difficulties) are not able to address and we are able to do so using a gold standard randomized design.

Relevant prior work in the literature on price transparency has mostly focused on the impact of insurer-

[3] Consumers going out-of-network could pay the entire billed charge as given in our intervention or some reduced version of that billed charge if negotiated down between themselves and the provider or their insurer and the provider. The latter case is relevant for consumers whose insurance covers some portion of out-of-network claims, which is more typical in generous health plans, such as preferred provider organization (PPO) plans. See, e.g., Bai and G. F. Anderson 2016; Bai and G. F. Anderson 2017; Bai and G. F. Anderson 2015; Bai and G. F. Anderson 2018 for an extended discussion of consumer payments when accessing out-of-network providers.

provided price transparency tools on consumer price shopping, with equivocal results (Mehrotra, Brannen, and Sinaiko 2014). Robinson and Brown (2013) and Robinson and MacPherson (2012) show that information provision about prices has a meaningful impact on the providers patients choose in the context of a reference-pricing payments model implemented in California where consumers have a lot of money at stake in a context where the potential for price differences is quite salient. Several studies focused on homogeneous services (e.g. lab services and MRIs) find some evidence of price shopping behavior in some populations (Christensen, Floyd, and Maffett 2017; J. C. Robinson, T. Brown, and Whaley 2015; Sinaiko, Joynt, and Rosenthal 2016). There is also evidence that consumers respond to price information in the context of tiered networks (Prager 2020). However, most studies find both relatively low use of price shopping tools by health plan members and, even when accessing such information, little impact on price shopping behavior (S. Desai et al. 2016; Mehrotra, Brannen, and Sinaiko 2014; Sinaiko and Rosenthal 2016; Brot-Goldberg et al. 2017; M. Chernew et al. 2018; Cooper, Scott Morton, and Shekita 2020).

While there have been quite a few studies on short-run consumer responses to price transparency tools there have been only a few studies that examine provider responses to price transparency. Robinson and Brown (2013) and Wu et al. (2014) both find some evidence that providers lower prices after the introduction of reference pricing/price shipping tools. Both of these studies focused on enrollees in specific-insurance plans (Whaley, T. Brown, and J. Robinson 2019) exploit the staggered deployment of an online transparency tool to a large pool of insured consumers and finds that robust consumer use of the tool can drive providers to reduce prices for homogeneous services but not for differentiated services. There are two prior studies on market-wide deployment of price transparency tools in New Hampshire (S. M. Desai, Shambhu, and Mehrotra 2021; Z. Y. Brown 2019).

These studies find that price transparency led to more aggressive bargaining by insurers that had medium-run impacts lowering the prices of high-priced hospitals. To our knowledge, there are no prior papers studying a market-wide deployment of a price transparency tool focused on individual provider prices and certainly none where price transparency information was implemented in a randomized control trial together with researchers. The introduction of the New York Healthcare Online Shopping Tool (NY HOST) offers a unique opportunity to conduct such a trial and systematically and rigorously examine the effects of charge transparency on consumers and providers.

The paper proceeds as follows. We provide an overview of the background and setting for the experiment in Section 2. Section 2.2 describes the experimental design and randomization. We discuss the the mechanism by which the information provided by the tool might change the shopping behavior by consumers and price-setting by providers in Section 2.3. Section 3 provides an overview of the datasets utilized and construction of the datasets for analyses. Section 4 describes the empirical strategy. Section 5 provides the results of the empirical tests of the impact of the tool on providers' charges at both the individual provider level and aggregated market level. Section 6 examines the mechanisms behind the results and concludes.

2 Background

2.1 About FAIRHealth

FAIR Health is an independent non-profit organization that was established in 2009 as a replacement for Ingenix, a database owned by the insurance giant United Healthcare. FAIR Health maintains the nation's largest data repository of privately billed health insurance claims.[4] Its principal purpose is to provide insurers with an unbiased source of information on usual, customary, and reasonable rates to support the adjudication of out-of-network claims. The database contains claims information from insurers covering approximately 75% of the privately insured population of New York State, including information on both fully-insured claims and claims administered by insurers on behalf of self-insured plans.

In early 2011, FAIR Health created a consumer website that displayed educational information on providers and services but did not contain price information. In April 2011 this website was transformed into an independent, publicly-accessible consumer price transparency tool that displayed aggregate estimates of the charge and insurer allowed amount for a given procedure in each geozip across the country. Charges are posted by providers as the list price for services, typically the actual price to be paid by uninsured or out-of-network patients. The insurer allowed amount reflects the reimbursement rate that is negotiated between a provider and a health plan.

On September 12, 2017, FAIR Health re-launched a revamped version of its website for New York State as the New York Healthcare Online Shopping Tool (NYHOST).[5] The rollout was accompanied by an extensive, multi-pronged marketing effort to raise awareness of and draw people to the new consumer facing website. A statewide advertising campaign was estimated to have reached over 6 million consumers in New York State through traditional media, online advertising, and social media channels. The traditional media campaign included several components. In New York City and Albany, large billboards displayed ads in prominent places, including Times Square. Public service ads (featuring well-known personalities Larry King, Mandy Patinkin, and Nancy Grace) ran in New York City taxicabs; paid advertisements were featured in health clubs and shopping malls throughout the state; and magazine ads were featured in nearly two dozen national magazines (e.g. Harper's Bazaar, InStyle, Fortune, Food Network). The distribution of paid print advertising in malls and health clubs was focused on the most highly populated areas of the State, including New York

[4]Information on FAIRHealth can be accessed via https://www.fairhealth.org.
[5]The consumer tool is publicly accessible on the FAIR Health website via https://www.youcanplanforthis.org.

City and Albany. FAIR Health distributed press releases about the launch to websites with heavy internet traffic, such as Crain's New York and PR Newswire. A Facebook advertising campaign was estimated to have reached nearly two million people from September 2017 to July 2018, and Facebook clicks accounted for over a quarter of website hits. A digital banner campaign from September 2017 to January 2018 generated approximately an additional five percent of website hits. According to FAIR Health's provided data, direct searches (such as users typing the url), which may have been generated by the advertising and online media campaigns, generated the remaining website hits. FAIR Health also maintains an active social media presence, with accounts on Facebook and Twitter, and frequent updates featuring its events, services and publications (Kim and Glied 2021).

2.2 Experimental Design: New York Healthcare Online Shopping Tool (NYHOST)

In conjunction with the 2017 update, a randomized experiment was embedded within the website design of NYHOST that varied the level of charge information available to users across procedures and location. Based on data from FAIR Health, we identified 100 frequently performed procedures for professional outpatient medical services in New York State, spanning 30 different categories. Due to Current Procedural Terminology (CPT®) code changes during the 2017 calendar year, another 7 CPT codes were added for a total of 107 procedure codes included in the experiment. This set of categories and procedures were selected because they were both common and had a high rate of out-of-network use.[6] Working with the FAIR Health web development team, 19 categories spanning 50 procedures were assigned to have specific provider-level charge information featured in all the 3-digit zipcodes (referred henceforth as "geozip") in New York State, 31 in total, and those procedure codes were excluded from the randomization. For the other 57 procedures in the 11 remaining categories, specific provider-level charge information was released for a randomized set of geozip-procedure pairs. This randomization occurred across a set of 1767 procedure-geozip pairs with 948 procedure-geozips included in the treatment arm and 819 procedure-geozips included in the control arm. Each geozip was randomly allocated a set of procedures where provider-level charge information was displayed, with a range of 25 to 37 procedures in each geozip, for an average of 31 procedures treated procedures per geozip. The randomization process occurred as follows: first, the randomization algorithm assigned a random number to each procedure category and, within each geozip, assigned categories with progressively larger random numbers until the treatment group has 25 or more procedures. Next, the randomization algorithm

[6] Because of CPT codes that were discontinued in 2017, specifically mammogram codes, we restricted our analysis to a "balanced panel" of procedures, which included the set of 104 procedure codes that were actively billed during the time period examined, CY 2016 through the second quarter of 2019.

chooses the maximum over 1000 trials of minimum p-values in t-tests of equality of mean covariates between the treatment and control groups. A timeline describing the rollout and outcome of the experiment can be found in Figure 1. The experiment ran from September 12, 2017 through August 30, 2019, and during this time period, provider-level charge information was only available on the randomized geozip-procedure pairs.[7]

For procedure-geozips in the treatment group, the website featured provider-level charge information. Specifically, each provider listed on the website was given a range that their average billed charges for a procedure fell into. See Figure 3 for a screenshot of a sample search in the treatment group. In the treated procedure-geozips, the website displays only providers with above median volume based on the claims data, updated twice a year based on a rolling 12 months of data.[8] The provider charge ranges were created around actual charges to make sure the website was providing an estimate as opposed to an exact charge. FAIR Health chose this particular methodology to ensure the end user did not assume that the charge seen on the website would be what the provider would charge in exact amounts. The exact method of the construction of the charge ranges was not made public, so providers would be unlikely to be able to game the system (i.e. adjusting their charges to remain within a certain range). The range of the charges varied by the charge posted.[9]

The remaining procedure-geozip combinations were randomized to the control group. Searches for price information in the procedure-geozip combinations assigned to the control group would yield only the aggregated median charge information posted on the website - provider-specific charges were not available during the study period. See Figure 2 for a screenshot of a sample search in the control group. The aggregated charge information featured in the procedure-geozip combinations in the control group was not subject to any minimum volume requirements.

[7]The experiment ended on August 2019, and for select procedures, provider-level price information was released in all geozips in New York State.

[8]In the "unbalanced panel dataset", the median volume for each provider and procedure-by-geozip combination, each trimester was 5 claims, which aligns with our decision to truncate our"balanced panel dataset to those providers who rendered at least 5 services for each procedure-by-geozip pair in each trimester. After restricting the observations to the balanced panel (as referenced in A.4), the median volume posted for each provider and procedure-by-geozip combination, each trimester was 33 claims.

[9]The logic behind the construction of the charge bins was as follows: (1) Charge less than or equal to 25, set range 1-25; (2) charge greater than 25 but less than or equal to 50, set range to 26-50; (3) charge greater than 50 but less than or equal to 75, set range to 51-75; (4) charge greater than 75 but less than or equal to 100, set range to 76-100; (5) charge greater than 100 but less than or equal to 500, set range to (amount minus 50 and round down to nearest 10 to amount plus 50 round up to nearest 10); (6) charge greater than 500 but less than or equal to 2000, set range to (amount minus 100 and round down to nearest 100 to amount plus 100 round up to nearest 100); (7) charge greater than 2000 but less than or equal to 10000, set range to (amount minus 200 and round down to nearest 100 to amount plus 200 round up to nearest 100); (8) charge greater than 10000 but less than or equal to 20000, set range to (amount minus 500 and round down to nearest 100 to amount plus 500 round up to nearest 100); (9) charge greater than 20000 but less than or equal to 50000, set range to (amount minus 1000 and round down to nearest 100 to amount plus 1000 round up to nearest 100); (10) charge greater than 50000, set range to (amount minus 5000 and round down to nearest 100 to amount plus 5000 round up to nearest 100); (11) If price does not match any of above, set bottom range by taking current charge minus 100 and round down to nearest 100, set top charge to current charge plus 100 and round up to nearest 100.

2.3 How price transparency might affect prices and quantity at the provider and market levels

There is considerable variation in billed charges ("list prices") by each specialty within geographical markets. Prior to the introduction of price transparency tools, most consumers might have, at best, a sense of what average charges for a service in their market might be (and even this information would typically be difficult to find). While providers are aware of their own billed charges and the in-network rates that are offered to them by insurers, except in small, highly-concentrated markets, they typically lack information on their competitors' billed charges or negotiated rates. In response, a large trade literature and an army of consultants in the industry exists to advise doctors on how to set their prices. Revealing billed charges publicly provides new, much more accurate information on charges at the provider level both to consumers and to providers themselves.

The availability of new information on prices at the provider level would ordinarily be expected to shift demand toward lower-priced providers and to encourage providers to reduce their prices to attract additional demand, but that might not occur in this context. Three features of the out-of-network health care market may lead to unexpected results. First, information on quality in health care markets is inadequate. Consumers may perceive higher priced providers to be of higher quality (as they have in some other contexts where information on quality is poor) (Rao and Monroe 1989; Leavitt 1954). Providers, aware of this perception, might then choose to raise their own prices when they observe those of their peers. Second, billed charges are only relevant to insured consumers if they seek care out-of-network. For care provided in-network, consumers will pay cost-sharing based on prices negotiated between insurers and providers, which are generally well below list prices, and any such out-of-pocket payments will draw down remaining deductible and out-of-pocket maximums. A consumer choosing to use services out-of-network, then, will generally do so only because of a strong non-financial preference for a specific provider (due to perceived quality, convenience, referrals) or in cases of an emergency, consistent with evidence that consumers even in high-deductible plans choose high cost options for MRIs that were recommended by the referring provider, despite nearby cheaper options (M. Chernew et al. 2018). This suggests that the demand for out-of-network care from any specific monopolistically-competitive provider is relatively inelastic in price (completely inelastic to price in a surprise billing situation). Revealing competitors' price information to providers facing relatively inelastic demand could lead them to raise their prices. Finally, the availability of information on competitors' prices can facilitate collusion (Albæk, Møllgaard, and Overgaard 1997; Edlin 1997; Vaska 1985). Such collusion may be particularly valuable because market-level average billed charges have historically been used as a starting point for price negotiations with insurers. If providers within a market can maintain high levels of billed charges for out-of-network care, they may be able to command both higher out-of-network rates, and, perhaps higher negotiated rates.

3 Data

Our analyses utilize the FAIR Health database, which is comprised of medical claims in New York State with dates of services between January 1, 2016 to June 30, 2019, totalling over 110 million claims (110,422,511 claims in total). Our data covers about 75% of private insurance claims in New York. The data includes National Provider Identifier (NPI), 3-digit zipcode ("geozip"), date of service, procedure code (Current Procedural Terminology (CPT)), billed charge, place of service code, the patient's gender, and the patient's age group.

We matched NPI data to the CMS Physician Compare and the National Plan and Provider Enumeration System (NPPES) files to obtain information on provider characteristics. There were 205,023 unique NPIs in the FAIR Health data extract. We linked the provider NPIs represented in the FAIR Health data to the information in the 2017 CMS Physician Compare Downloadable File in order to access provider-level information, including gender, years in practice, medical school, group size, and hospital affiliation. The 2017 CMS Physician Compare File contains information on the providers who are participating in the CMS quality program, which encompasses all eligible providers (EPs) that qualify or participate in the program.[10] Since providers were able to be credentialed in multiple specialties and practice in several different locations, both the specialty and location was chosen as the first that appeared when sorted in alphabetical order. We also utilized U.S. Census Data to access population and geographic information.

We define our time periods at the trimester level, or a third of a year, with four months in each trimester. Defining the time period at the trimester level enables our analysis to align with the timing of the experiment, which went into effect in September 2017, two-thirds of the way into the calendar year. The post randomization period spanned from September 2017 through June 2019 period. Because we had incomplete data for the last trimester, spanning as May 2019 through August 2019, we restricted our final analyses to the 10 trimesters for which we had complete claims data. Thus, the time period examined in our study dates from January 2016 through April 2019, with five trimesters in the pre-randomization period (from January 2016 through August 2017), and five trimesters in the post-randomization period (from September 2017 through April 2019).

[10]Table A.1 shows the procedure categories that were included for the study. Table A.2 shows the datasets utilized and time periods encompassed. Table A.3 shows the match between the FAIR Health dataset and the CMS Physician Compare dataset. Approximately 58% of all of the provider NPIs in our FAIR Health data extract were captured in the 2017 CMS Physician Compare file.

For the provider level analysis, we collapsed the original claims data to the provider-geozip-procedure-trimester level, which encompassed 3,598,866 observations. Our billed charge outcome variable was constructed as the modal charge reported by each provider for each procedure in each trimester of the period studied. Given potential billing errors, we utilized the modal charge as our primary outcome variable to capture the most frequently billed unit charge for a given procedure for each trimester as the list price for that provider. As robustness checks, we also include the median, 95th percentile and 5th percentile of the billed charges for each provider in a given trimester, procedure, and geozip. To account for outliers and billing errors (since several claims had billed charges that ranged as high as $70,000), we winsorized the provider-level panel dataset at the 95th percentile, conditional upon the procedure code.[11]

To create a balanced panel of providers, we restricted our analysis to providers with at least five claims in each trimester. We also restricted to only physicians, dropping providers with credentials identifying them as a physicians assistant or a nurse practitioner, since these practitioners often submit separate claims for procedures and services they may have rendered supporting services for rather than as the primary provider. We also removed any CPT codes that were added or discontinued throughout the study period keeping only procedures for which charges were posted continuously across all the trimesters included in our study period.[12]

To assess the overall market impact of the tool, we created a panel dataset at the procedure and geozip level for each trimester, and constructed the aggregate volume-weighted charge and total volume for a given procedure and geozip in each time period. The market level dataset is based on all claims in the FAIR Health claims database, not only those included in the construction of the provider level data set. Our main outcome variables for the market level dataset include total volume of procedures, average billed charge and the coefficient of variation of charges within a market.

We present summary statistics for our provider and market level datasets in Table 1. We examine a variety of procedures, ranging from lower cost psychotherapy and physical therapy services to higher cost orthopedic and radiology services, resulting in substantial heterogeneity in the billed charge. Each provider had an average of 102 claims each trimester, with an average billed charge of $420. Providers' billed charges ranged from $2 to $59,000, with a standard deviation of $1,279. The volume of services rendered by providers ranged from 6 to 22,329, with a standard deviation of 198 claims.[13]

[11] Table A.4 shows the steps involved in the construction of the balanced panel.

[12] The codes were removed from the balanced panel because they were either discontinued or added to the CPT® (Current Procedural Terminology) list between 2016 and 2019. Those codes were 76641, 76642, 77052, 77056, 77065, 77066, 97161, 97162, and 97163.

[13] The billed charge for a given procedure varies widely across providers, and even for the same provider, the billed charge can vary over time since providers can update their chargemasters (list of billed charges) at will. There is significant dispersion in the distribution of the charge updates (presented as the change in the log of the price in Figure A.8). Approximately 15% of providers updated their charges in the first trimester of 2017 and 2018, but providers continue to update their charges later in a given calendar year (Figure A.9). This underlying price heterogeneity points to meaningful scope for price changes.

Figure 5 plots the histogram of normalized charge dispersion, the ratio of a provider's modal charge to the average charge in each procedure-by-geozip-by-trimester market, and documents heterogeneity in prices within a given market. This heterogeneity suggests that there is scope for the information treatment to affect equilibrium prices. At the market level, there was significant underlying price heterogeneity with the average interquartile range of prices within a market of $735. In our sample of procedures, there was an average of 5,956 claims rendered in a given procedure and geozip market in a single trimester, and of those claims, approximately 20% were rendered by a provider who was out-of-network with a given insurance product.[14]

We conducted balance checks to validate the randomization process of the experiment and compare market characteristics between the treatment and control groups, including measures of price, volume, and market concentration based on the claims with dates of service in 2016, the baseline period prior to the launch of the tool (Table 2). There was no statistically significant difference between the treatment and control groups on aggregate charges, the interquartile range of charges, volume of claims, volume out-of-network claims, insurer market concentration, and population density. Although most of the market characteristics were comparable, the control group had more concentrated provider markets at baseline (provider HHI of 860 in the control group compared to provider HHI of 732 in the treatment group), and the treatment group had slightly higher within market charge dispersion, defined as the standard deviation of charges in a given procedure-geozip market (0.09 in the treatment group compared to 0.08 in the control group), and slightly higher charge dispersion at the 90th quantile (1.10 in the treatment group compared to the 1.09 in the control group).

We compare market characteristics of procedures included in the randomization to other procedures for which FAIR Health receives claims. The subset of procedures that were selected for the experiment had a relatively high percentage of out-of-network claims; although most procedures were performed out-of-network less than 10% of the time, some procedures have an out-of-network percentage that ranged as high as 40 percent (Figure 6). Figure 7 demonstrates the market concentration of providers for each procedure and geozip market. The provider market concentration for each procedure and geozip was calculated using the Herfindahl-Hirschman Index (HHI), with the market share for each provider represented in the dataset by the National Provider Identification number (NPI). Prior research suggests that concentration of provider markets is associated with higher charges (Roberts, M. E. Chernew, and McWilliams 2017), and the variation in the provider market HHI demonstrates substantial heterogeneity in provider market concentration and corresponding price dispersion.

[14] We also calculated the insurer HHI as the sum of the square of the market share for each insurer within each each procedure and geozip market. Over 30 insurance companies in NYS represented in the claims data, with each insurer represented in the data by a FAIR Health "key" that kept the identity of each insurer confidential. The distribution of insurance market HHI shows that most markets are highly concentrated, and for most of the procedures in our sample, the insurer HHI was well above 4000 (Figure A.11).

4 Empirical Strategy

To capture the effect of the FAIR Health price transparency tool on both charges and volume of services provided, we estimate a difference-in-differences approach at the provider and market level. The provider-level analysis examines effects on the log of the modal charge that a provider bills over a trimester as well as the log of the total volume of a procedure performed by the provider. We choose to define the "price" as the modal charge because the mode is least insensitive to outliers and represents the price most likely to be faced by any given patient. As a sensitivity analysis, we confirmed that our results are consistent when the price is defined as the median or average charge. Providers in New York State are required to have one charge for a given procedure at a point in time (e.g. charge discrimination is not allowed), and our use of the modal charge reflects that requirement. Our market level analysis is conducted at the geozip-procedure level and captures volume-weighted effects on charges as well as charge dispersion. We estimate effects on the log of the average billed charge of a procedure in a given geozip in a trimester and on the total volume of services in a market.

4.1 Difference-in-differences specification

Our preferred specification is a difference-in-differences specification with fixed effects that account for time trends in procedure and geozip effects (Berck and Villas-Boas 2016). Because the randomization was conducted at the procedure-geozip level and the randomization units are not equal in size, geographic and procedure changes over time could bias our results because of compositional effects. The difference-in-differences specification with additional fixed effects reduces potential bias due to geozip or procedure-specific changes over time. Because of these considerations, we utilize a difference-in-differences specification that includes the "year by trimester" time variable interacted with the treatment effect to assess the impact over time. The provider-level difference-in-differences specification is as follows:

$$Y_{igpt} = \beta \cdot T_{gp} \cdot Post_t + \lambda_t \cdot \mathbb{1}(yrtri = t) + \gamma_{gp} \cdot \mathbb{1}(gp) + \pi_i \cdot \mathbb{1}(i) + \kappa_{pt} + \alpha_{gt} + \pi_i \cdot \mathbb{1}(i) + \varepsilon_{igpt} \tag{1}$$

Where Y_{igpt} is the outcome variable for provider i for procedure p in geozip g in trimester t, and is either $ln(P)_{igpt}$, the log of the modal charge for a provider, or $ln(Q)_{igpt}$, the log of total volume for a provider. In the difference-in-differences estimate, β specifies the treatment effect, T_{gp} is the treatment indicator (equal to 1 for the randomized procedure and geozips) which is interacted with $Post_t$, the indicator for the post-period (equal to 1 for the trimesters encompassing the period after September 2017). The model includes controls for time fixed effects λ_t, procedure-geozip fixed effects γ_{gp}, procedure dummy variables interacted with the time dummy variables κ_{pt}, the geozip dummy variables interacted with the time dummy variables α_{gt}, and provider fixed effects π_i. The errors ε_{igpt} are robust and clustered at the category-geozip level.

In addition to our preferred difference-in-differences specification with interacted time fixed effects (Equation 1), we also estimate a standard difference-in-differences framework with the treatment and post indicators, time fixed effects λ_t, procedure-geozip fixed effects γ_{gp}, and an error term, ε_{igpt}. (Equation 2)

$$Y_{igpt} = \beta \cdot T_{gp} \cdot Post_t + \lambda_t \cdot \mathbb{1}(yrtri = t) + \gamma_{gp} \cdot \mathbb{1}(gp) + \varepsilon_{igpt} \tag{2}$$

Due to volume differences between providers, a market-level model enables capturing effects on the volume-weighted price. Thus, we also estimate the difference-in-differences analyses at the market-level, with procedure-geozip-time as the unit of observation.

$$Y_{gpt} = \beta \cdot T_{gp} \cdot Post_t + \lambda_t \cdot \mathbb{1}(yrtri = t) + \gamma_{gp} \cdot \mathbb{1}(gp) + \kappa_{pt} + \alpha_{gt} + \varepsilon_{gpt} \tag{3}$$

In Equation 3, the dependent variable Y_{gpt} is either $ln(P)_{gpt}$ the log of the average volume-weighted actual billed charge in the market for procedure p in geozip g in time period t, $ln(Q)_{gpt}$ the log of total volume for a market, or CoV_{gpt} the coefficient of variation of billed charges within a market. The difference-in-differences estimate β specifies the treatment effect of interest, T_{gp} interacted with $Post_t$, and estimates the treatment effect on the market average charge for procedure p in geozip g after the launch of the tool. The treatment variable T_{gp} and time variable $Post_t$ has the same construct as in the provider-level model. The market-level difference-in-differences specification includes the time and procedure-geozip fixed effects separately and interacted with the time dummy variable, similar to the provider-level regressions in Equation 1. The model includes the error term, ε_{gpt} and robust standard errors are clustered at the category-geozip level.

We generate event study graphs that correspond to equations 1 and 3. These graphs plot the difference in outcome variable between treatment and control units over time, controlling for procedure-geozip, time, geozip-time and procedure-time fixed effects. The event studies allow us to visually interpret the treatment effect of NYHOST as well as compare pre-trends across our specifications.

4.2 Heterogeneity Tests

To complement and better understand our main results, we examine heterogeneity in our results along a variety of economically meaningful dimensions. Specifically, we examine how our results differ for providers that are above or below median for billed charges, number of services provided out-of-network and a provider's Herfindahl-Hirschman index (HHI) for a given procedure. Our analysis for above and below median price providers allows us to decompose overall price effects into high priced providers potentially lowering their prices or low price providers raising prices to match the market rate.

Additionally, we test heterogeneity by type of procedure. We group procedures into those that receive continuous care (defined as procedures where a patient typically have repeat visits with a provider at regular intervals, such as physical therapy) and non-continuous care (defined as procedures where a patient visit typically occurs once or infrequently, such as emergent surgical procedures or radiology services) as well grouping CPT codes into broader categories like psychological services or MRI scans. We chose these provider characteristic dimensions for heterogeneity because they correspond to competitive forces that may increase the impact of the change in information availability. Our analysis for above and below median price providers allows us to decompose overall price effects into high priced providers potentially lowering their prices or low price providers raising prices to match the market rate. The information provided by the tool likely has the most relevance for procedures rendered out-of-network or for providers that provide more out-of-network care because billed charges represent actual prices paid by patients. Analyzing heterogeneity by HHI examines the role that market competition plays; if there are relatively more providers to compare, the information treatment may have a stronger effect. For these provider characteristics we rerun our difference-in-differences specification both with and without additional fixed effects (Equation 1; Equation 2).

CPT codes themselves vary widely in out-of-network usage, market prevalence and frequency of usage which may contribute to heterogeneous treatment effects. We group CPT codes into continuous use codes (codes for psychological, physical therapy or chiropractic services) and non-continuous use codes (MRIs, CT scans and radiology). Across this cut we run both the difference in differences specifications with different configurations of the fixed effects. We then run a series of analyses examining each category of CPT codes, of which there are 10 included in our experiment. Because of the implementation of the experiment, for these analyses stratified by category, we are only able to fit our difference in differences specification.

5 Results

5.1 Website Usage

This analysis assesses utilization of the price transparency tool prior to and after the launch of the revamped tool in September 2017. Figure A.6 demonstrates the distribution of the utilization of the website, measured as the total number of searches, over years and months. Figure 4 depicts the distribution of website utilization across the treatment and control groups by month between January 2016 and June 2019. The persistently higher number of searches in the control group can be attributed to the number of searches that occurred in New York City for the procedures in that geozip that were assigned to the control group. When we exclude the 3-digit geozip that corresponds to the borough of Manhattan (geozip = 100), the number of searches between the treatment and control groups is more comparable (see Appendix). Because we randomized 57 different procedures across 31 geozips to the treatment and control groups, and the zipcodes are of different sizes in population density, random assignment to a highly populated zipcode can lead to a higher number of searches.

Web utilization of the tool by consumers appears to have been relatively consistent over time and low compared to the New York State overall population of over 19 million residents, with fewer than 15 searches on average for each procedure in a given month (Kim and Glied 2021). This suggests that demand side forces are unlikely to be a large contributor to any observed price or volume changes after the roll out of the experiment. Low numbers of searches relative to the population would be consistent with our theory that supply side forces played a larger role in use of the tool, where providers search to gain information on competitors' prices.

5.2 Provider-level Outcomes

We present the results of our analysis specified in equations 1 and 2 in Table 3. Our preferred specification results are presented in columns (3) and (6). We find that treatment is associated with a 0.75% increase in prices and no significant changes to the quantity of services provided.[15] Our difference-in-differences estimates presented in columns (1) and (3) are not significant. Figure 8 presents the event study plot associated with Table 3 column (2). Overall, our results suggest that meaningful changes to market prices or quantity of procedures are small and any changes that we do see are likely attributable to supply side market forces.

Our main results show the impact of the tool in aggregate but may mask heterogeneity in the effect on price and quantity. We test for heterogeneity across provider and procedure characteristics. Table 4

[15] This estimate increases to 1.2% without the provider fixed effects (Table A.13).

presents our results for heterogeneity across potentially meaningful cuts of the data. Specifically we analyze how our results vary for above versus below median charge providers (with the median determined at the procedure-by-geozip level), continuous versus non-continuous procedures, above versus below median out of network visits, above versus below median HHI and above versus below median website usage. Panel A presents the results for charges while Panel B presents the results for quantity treatment effects. We find that our overall results are primarily driven by increases in prices for below median price providers, consistent with providers raising their prices to match their peers. Additionally, we find that price increases are higher for non-continuous procedures (procedures that are not repeated for a patient) and for procedures that are less frequently provided out-of-network. For out-of-network claims, we find that procedures with a high proportion of out-of-network claims have essentially no price change while procedures with a low proportion of out-of-network claims have a 2.5% price increase as a result of our intervention. We also assessed the impact of the tool for below and above median price providers by these market characteristics (high versus low out-of-network use, market concentration, and website use). We find significant charge increases for low volume out-of-network procedures across the course of the experiment, with price increases driven entirely by "low price" providers, defined as those with billed charges below the median charge during the baseline period encompassing the pre-randomization trimesters from January 2016 through August 2017, with the median charge defined at the procedure-by-geozip level. Figure 10 depicts this upward trend in the modal charge posted by "low price" providers for the low out-of-network volume procedures. Figures 9 and A.15 feature event study graphs that highlight these trends over the course of the study period.

Table 4 Panel B presents our heterogeneity analysis for quantity effects. While overall we find null quantity effects, we do find some evidence that would support consumers price shopping. We find that volume of services decrease by about 3.4% for markets with above median website usage. Additionally, we find that the volume of services decreases by about 3.6% for continuous procedures.

We also investigate the effects of our intervention for specific procedure categories and present the findings in Table 5. Although we find variation in price changes across procedures, once we adjust for multiple hypothesis testing, none of these effects are statistically significant (M. L. Anderson 2008). A comparison of the calculated q-values with the p-values from the heterogeneity tests, by procedure category, are presented in Table A.16 for price effects and Table A.17 for quantity effects. The q-values adjusting for false positives suggest that none of the effects by category are significant.

The provider-level models utilize the balanced panel of providers constructed from providers with a minimum of 5 claims in a given trimester. As a robustness check, we present the results generated from the market-level dataset constructed at the geozip X procedure X trimester level. The market-level panel dataset utilizes the entirety of the dataset and thus captures the volume-weighted charge and aggregate market level effects. Table A.18 shows the results from the market-level specifications (Equation 3), which captures the aggregate charge and quantity effect. We find that the overall price effect is similar to the provider-level models, with no significant market level effects on overall volume. To test for heterogeneity in our market-level outcomes, we stratified the market-level dataset upon dimensions of market concentration, procedures with high vs. low out-of-network use, coefficient of variation, website utilization and for continuous vs. non-continuous services (Table A.19). Just as we did for the provider level results, we test for heterogeneity along continuity of care for continuous and non-continuous services. Similar to the provider-level results, our most significant result is that in procedure markets with low out-of-network claims at baseline, there is a 2.9% price increase. Table A.20 presents the effects of our intervention for specific procedure categories, and find a market-level price increase for MRI services (+8%) but no significant effects for all other categories. For our market level analysis, our heterogeneity tests involved separate analyses of results for markets with high or low price dispersion, measured as above median or below median coefficient of variation on billed charges.

6 Discussion

Our findings support the hypothesis that effects of the NYHOST price shopping tool were dominated by provider responses and price adjustments rather than consumer price shopping. Overall, provider-level prices and aggregated market prices increased more in the treated markets than in untreated markets. Our findings that providers with a lower percentage of their services rendered out of network were more likely to raise their charges in the post-randomization period suggests that the tool yielded useful information for providers with limited out-of-network experience. The providers who were already rendering a larger proportion of their services out-of-network may already have been aware of their competitors' charges or had set their charges optimally. For providers with limited out-of-network experience, the presence of the tool enabled them to see the charge information posted by their competitors in a given market and increase their charges accordingly.

Overall, these results are consistent with our intervention having (i) a minimal effect on consumer price shopping and (ii) a meaningful but modest effect on provider price increases, especially for less elective services that are almost always covered by insurance. These price increases are consistent with both tacit collusion between providers in an environment with greater provider-specific price information or with reduced information asymmetries that generally push providers towards realizing that they are under-charging relative to their peers.

Although price transparency is a laudable goal in a healthcare market dominated by information asymmetries, there may be perverse price effects due to supply constraints, the inelastic nature of the demand for healthcare services and opportunity for providers to engage in price-setting. Our results suggest caution about price transparency if physicians are more likely to leverage that information than consumers to set prices. Our results should be viewed with a number of caveats, including the limited time window during which we can study the effects of the intervention, and the application to billed charges, which, though relevant for out-of-network claims and potentially relevant via what they signal about negotiated rates, are not indicative of insurer-provider negotiated rates. The NY HOST price transparency tool is not representative of all price transparency interventions, and this is just one limited example. Further research is necessary to more fully assess how different kinds of interventions impact prices, quantities, and welfare. Recent policy actions, including the CMS price transparency rule mandating that insurers reveal provider-negotiated rates, necessitates future work investigating the impact of that rule.

References

Albæk, Svend, Peter Møllgaard, and Per B. Overgaard (1997). "Government-Assisted Oligopoly Coordination? A Concrete Case". In: *The Journal of Industrial Economics* 45.4, pp. 429–443.

Anderson, Michael L. (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". In: *Journal of the American Statistical Association* 103.484, pp. 1481–1495. DOI: 10.1198/016214508000000841.

Bai, Ge and Gerard F. Anderson (2015). "Extreme Markup: The Fifty US Hospitals With The Highest Charge-To-Cost Ratios". In: *Health Affairs* 34.6, pp. 922–928. DOI: 10.1377/hlthaff.2014.1414.

— (2016). "US Hospitals Are Still Using Chargemaster Markups To Maximize Revenues". In: *Health Affairs* 35.9, pp. 1658–1664. DOI: 10.1377/hlthaff.2016.0093.

— (2017). "Variation in the Ratio of Physician Charges to Medicare Payments by Specialty and Region". In: *JAMA* 317.3, pp. 315–318. DOI: 10.1001/jama.2016.16230.

— (2018). "Market Power: Price Variation Among Commercial Insurers For Hospital Services". In: *Health Affairs* 37.10, pp. 1615–1622. DOI: 10.1377/hlthaff.2018.0567.

Batty, Michael and Benedic Ippolito (2017). "Mystery Of The Chargemaster: Examining The Role Of Hospital List Prices In What Patients Actually Pay". In: *Health Affairs* 36.4, pp. 689–696. ISSN: 0278-2715. DOI: 10.1377/hlthaff.2016.0986.

Berck, P. and S. B. Villas-Boas (2016). "A note on the triple difference in economic models". In: *Applied Economics Letters* 23.4, pp. 239–242. DOI: 10.1080/13504851.2015.1068912.

Brot-Goldberg, Zarek C. et al. (2017). "What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics*". In: *The Quarterly Journal of Economics* 132.3, pp. 1261–1318. DOI: 10.1093/qje/qjx013.

Brown, Zach Y. (2019). "Equilibrium Effects of Health Care Price Information". In: *The Review of Economics and Statistics* 101.4, pp. 699–712. DOI: 10.1162/rest_a_00765.

Chernew, Michael et al. (July 2018). *Are Health Care Services Shoppable? Evidence from the Consumption of Lower-Limb MRI Scans*. DOI: 10.3386/w24869.

Christensen, Hans, Eric Floyd, and Mark Maffett (2017). "The Effects Of Charge-Price Transparency Regulation On Prices In The Healthcare Industry". In.

Cooper, Zack, Fiona Scott Morton, and Nathan Shekita (2020). "Surprise! Out-of-Network Billing for Emergency Care in the United States". In: *Journal of Political Economy* 128.9, pp. 3626–3677. DOI: 10.1086/708819.

Cutler, David and Leemore Dafny (2011). "Designing Transparency Systems for Medical Care Prices". In: *New England Journal of Medicine* 364.10, pp. 894–895. DOI: 10.1056/NEJMp1100540.

Desai, Sunita et al. (2016). "Association Between Availability of a Price Transparency Tool and Outpatient Spending". In: *JAMA* 315.17, pp. 1874–1881. DOI: 10.1001/jama.2016.4288.

Desai, Sunita M., Sonali Shambhu, and Ateev Mehrotra (2021). "Online Advertising Increased New Hampshire Residents' Use Of Provider Price Tool But Not Use Of Lower-Price Providers". In: *Health Affairs* 40.3, pp. 521–528. DOI: 10.1377/hlthaff.2020.01039.

Edlin, Aaron S. (1997). "Do Guaranteed-Low-Price Policies Guarantee High Prices, and Can Antitrust Rise to the Challenge?" In: *Harvard Law Review* 111.2, pp. 528–575. DOI: 10.2307/1342058.

Finkelstein, Amy et al. (2012). "The Oregon Health Insurance Experiment: Evidence from the First Year". In: *The Quarterly Journal of Economics* 127.3, pp. 1057–1106. DOI: 10.1093/qje/qjs020.

Glied, Sherry (2021). "Price Transparency—Promise and Peril". In: *JAMA* 325.15, pp. 1496–1497. ISSN: 0098-7484. DOI: 10.1001/jama.2021.4640.

Kim, Grace and Sherry Glied (2021). "Assessing Utilization of a Marketwide Price Transparency Tool". In: 27.

Kullgren, Jeffrey T. and A. Mark Fendrick (2021). "The Price Will Be Right—How to Help Patients and Providers Benefit from the New CMS Transparency Rule". In: *JAMA Health Forum* 2.2, e210102. DOI: 10.1001/jamahealthforum.2021.0102.

Leavitt, Harold J. (1954). "A Note on Some Experimental Findings About the Meanings of Price". In: *The Journal of Business* 27.3, pp. 205–210.

Manning, Willard G. et al. (1987). "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment". In: *The American Economic Review* 77.3, pp. 251–277.

Mehrotra, Ateev, Tyler Brannen, and Anna D. Sinaiko (2014). "Use Patterns of a State Health Care Price Transparency Web Site: What Do Patients Shop For?" In: *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* 51. DOI: 10.1177/0046958014561496.

Newhouse, Joseph P. (1996). "Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection". In: *Journal of Economic Literature* 34.3, pp. 1236–1263.

Prager, Elena (2020). "Healthcare Demand under Simple Prices: Evidence from Tiered Hospital Networks". In: *American Economic Journal: Applied Economics* 12.4, pp. 196–223. ISSN: 1945-7782. DOI: 10.1257/app.20180422.

Rao, Akshay R. and Kent B. Monroe (1989). “The Effect of Price, Brand Name, and Store Name on Buyers’ Perceptions of Product Quality: An Integrative Review”. In: *Journal of Marketing Research* 26.3, pp. 351–357. DOI: 10.2307/3172907.

Reinhardt, Uwe E. (2006). “The Pricing Of U.S. Hospital Services: Chaos Behind A Veil Of Secrecy”. In: *Health Affairs* 25.1, pp. 57–69. ISSN: 0278-2715. DOI: 10.1377/hlthaff.25.1.57.

Roberts, Eric T., Michael E. Chernew, and J. Michael McWilliams (2017). “Market Share Matters: Evidence Of Insurer And Provider Bargaining Over Prices”. In: *Health Affairs* 36.1, pp. 141–148. DOI: 10.1377/hlthaff.2016.0479.

Robinson, James C., Timothy Brown, and Christopher Whaley (2015). “Reference-Based Benefit Design Changes Consumers’ Choices And Employers’ Payments For Ambulatory Surgery”. In: *Health Affairs* 34.3, pp. 415–422. DOI: 10.1377/hlthaff.2014.1198.

Robinson, James C. and Timothy T. Brown (2013). “Increases In Consumer Cost Sharing Redirect Patient Volumes And Reduce Hospital Prices For Orthopedic Surgery”. In: *Health Affairs* 32.8, pp. 1392–1397. DOI: 10.1377/hlthaff.2013.0188.

Robinson, James C. and Kimberly MacPherson (2012). “Payers Test Reference Pricing And Centers Of Excellence To Steer Patients To Low-Price And High-Quality Providers”. In: *Health Affairs* 31.9, pp. 2028–2036. DOI: 10.1377/hlthaff.2011.1313.

Sinaiko, Anna D., Karen E. Joynt, and Meredith B. Rosenthal (2016). “Association Between Viewing Health Care Price Information and Choice of Health Care Facility”. In: *JAMA Internal Medicine* 176.12, pp. 1868–1870. DOI: 10.1001/jamainternmed.2016.6622.

Sinaiko, Anna D., Pragya Kakani, and Meredith B. Rosenthal (2019). “Marketwide Price Transparency Suggests Significant Opportunities For Value-Based Purchasing”. In: *Health Affairs* 38.9, pp. 1514–1522. DOI: 10.1377/hlthaff.2018.05315. URL: https://doi.org/10.1377/hlthaff.2018.05315.

Sinaiko, Anna D. and Meredith B. Rosenthal (2016). “Examining A Health Care Price Transparency Tool: Who Uses It, And How They Shop For Care”. In: *Health Affairs* 35.4, pp. 662–670. DOI: 10.1377/hlthaff.2015.0746.

UnitedHealthcare (Mar. 23, 2023). *Transparency in Coverage*. URL: https://www.uhc.com/content/dam/uhcdotcom/en/HealthReform/PDF/Provisions/reform-external-transparancy-FAQs.pdf.

Vaska, Michael K. (1985). “Conscious parallelism and price Fixing: Defining the boundary”. In: *The University of Chicago Law Review* 52.2, pp. 508–535.

Volpp, K. G. (2016). “Price Transparency: Not A Panacea For High Health Care Costs”. In: *JAMA* 315.17, pp. 1842–1843. URL: http://dx.doi.org/10.1001/jama.2016.4325.

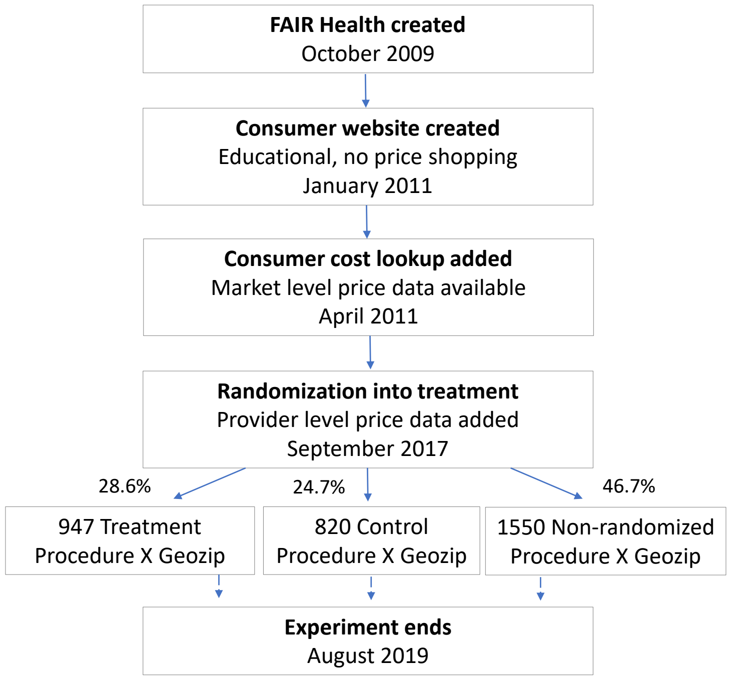
Whaley, Christopher, Timothy Brown, and James Robinson (2019). "Consumer Responses to Price Transparency Alone versus Price Transparency Combined with Reference Pricing". In: *American Journal of Health Economics* 5.2, pp. 227–249. DOI: 10.1162/ajhe_a_00118.

Wilensky, Gail (2019). "Federal Government Increases Focus on Price Transparency". In: *JAMA* 322.10, pp. 916–917. DOI: 10.1001/jama.2019.12912.

Wu, Sze-jung et al. (2014). "Price Transparency For MRIs Increased Use Of Less Costly Providers And Triggered Provider Competition". In: *Health Affairs* 33.8, pp. 1391–1398. DOI: 10.1377/hlthaff.2014.0168.

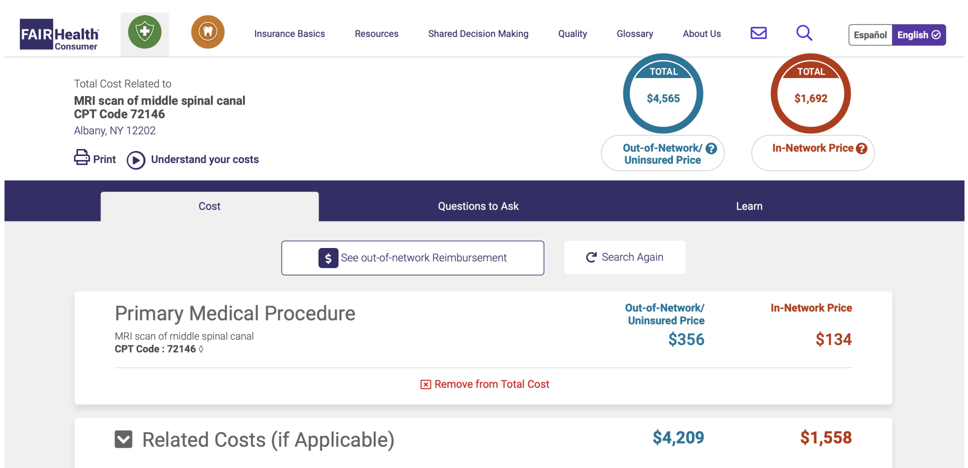
Tables and Figures

Figure 1: Timeline and Experimental Design



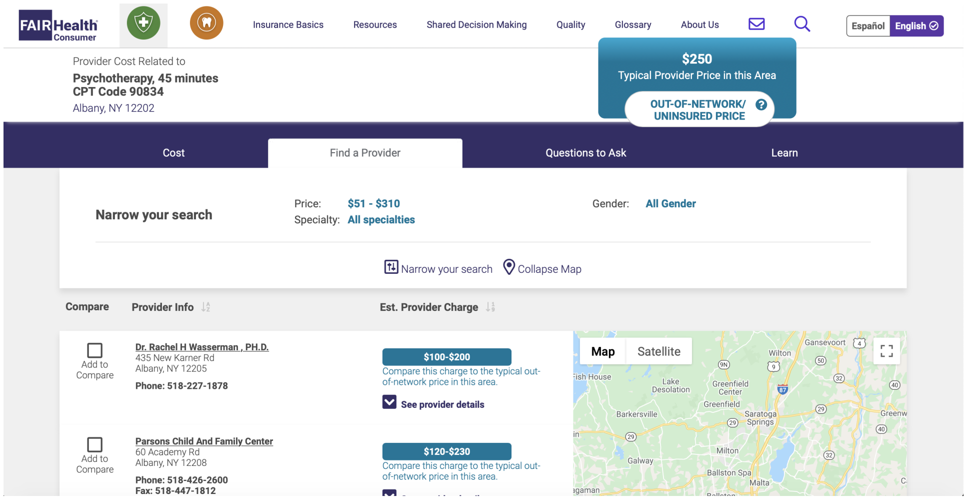
Notes: This shows the timeline of the implementation and randomization of the website.

Figure 2: NYHOST Website: Control Website Search Result



Notes: A snapshot of the FAIRHealth consumer shopping tool for a control procedure and geozip that does not contain provider-level information.

Figure 3: NYHOST Website: Treatment Website Search Result



Notes: A snapshot of the FAIRHealth consumer shopping tool for one of the randomized procedure and geozips that released provider-level price information.

Table 1: Provider and Market Summary Statistics

	Mean	SD	Min	Max
Panel A: Provider summary statistics				
Unit Charge	420	1,279	2	59,000
Volume	103	198	6	22,329
N	583,693			
Panel B: Market summary statistics				
Average charge	985	2,173	14	42,611
Average volume	4,562	19,349	6	350,776
Average volume out-of-network	688	4,446	0	96,661
Provider HHI	756	901	5	10,000
Website usage	2	11	0	565
N	13,141			

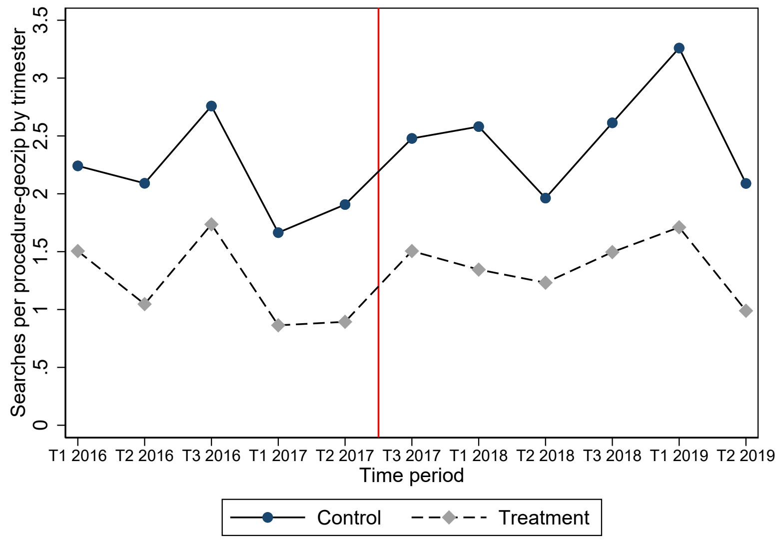
Notes: This table shows the summary statistics for charge and volume information for the provider-level dataset at the NPI-by-procedure-by-geozip-by-trimester level as well as market level summary statistics of charges, volume and market characteristics.
Data source: FAIRHealth.

Table 2: Balance Check

Variable	(1) Treatment Mean/SD	(2) Control Mean/SD	T-test P-value (1)-(2)
Average charge	1,025.13	918.30	0.58
	(7,045.04)	(5,282.47)	
Average volume	4,251.48	4,980.21	0.69
	(51,717.63)	(60,955.76)	
Average volume out-of-network	574.69	900.36	0.46
	(9,550.12)	(16,700.61)	
Provider HHI	716.19	812.44	0.18
	(2,061.27)	(2,399.04)	
Website usage	1.43	2.36	0.31
	(13.62)	(37.43)	
N	2175	1822	
Clusters	165	144	
F-test of joint significance (p-value)			0.01**
F-test, number of observations			3997

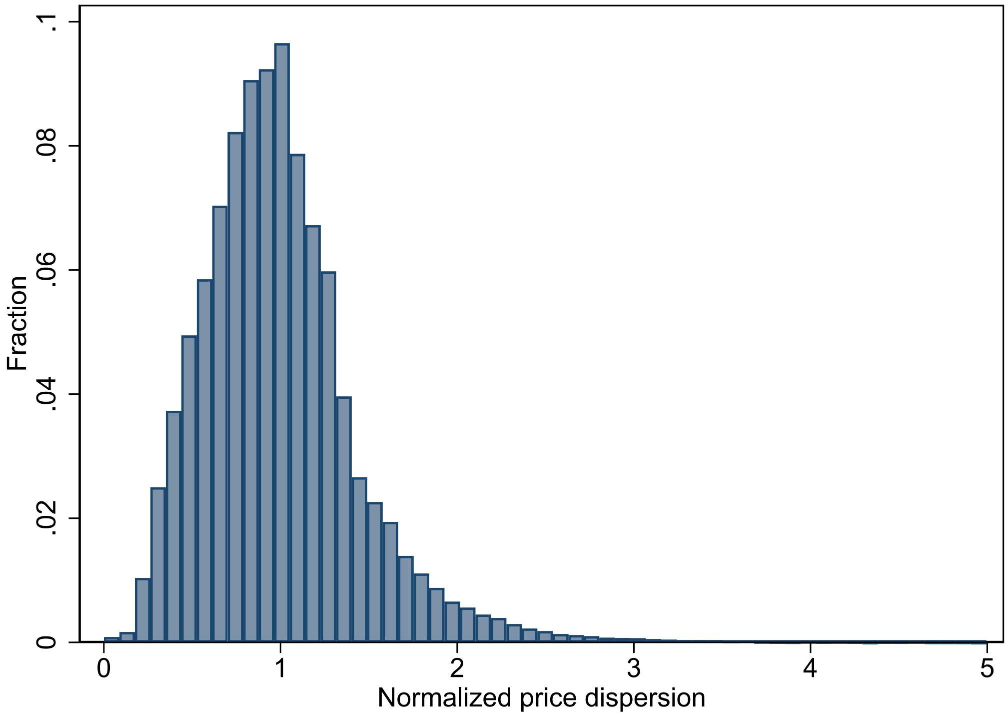
Note: This table checks for balance of market level summary stats in the pre-period for treatment and control markets.

Figure 4: Website Utilization Between Treatment and Control Groups



Notes: This figure plots the average monthly website utilization per procedure-by-geozip market (data on NYHOST web searches provided by FAIRHealth) for procedures in the treatment and control groups. The persistently higher searches in the control group can be attributed to the number of searches that occurred in New York City for the procedure-by-geozip combinations in the control group.

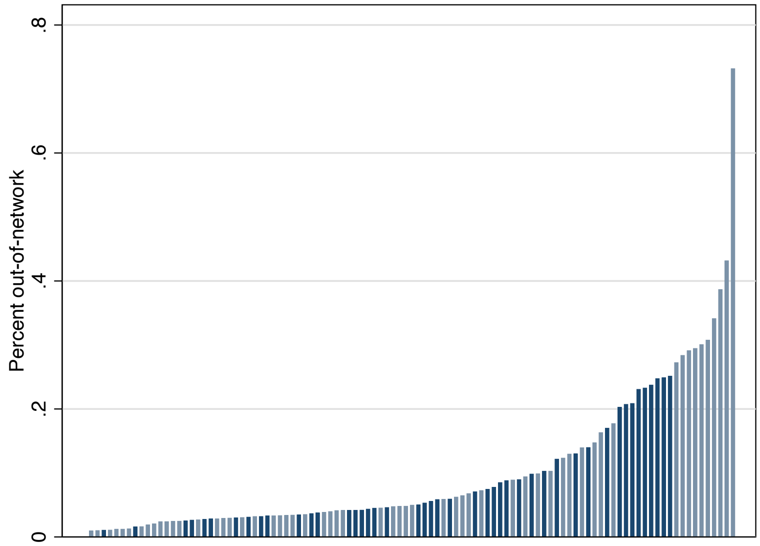
Figure 5: Normalized Charge Dispersion



,

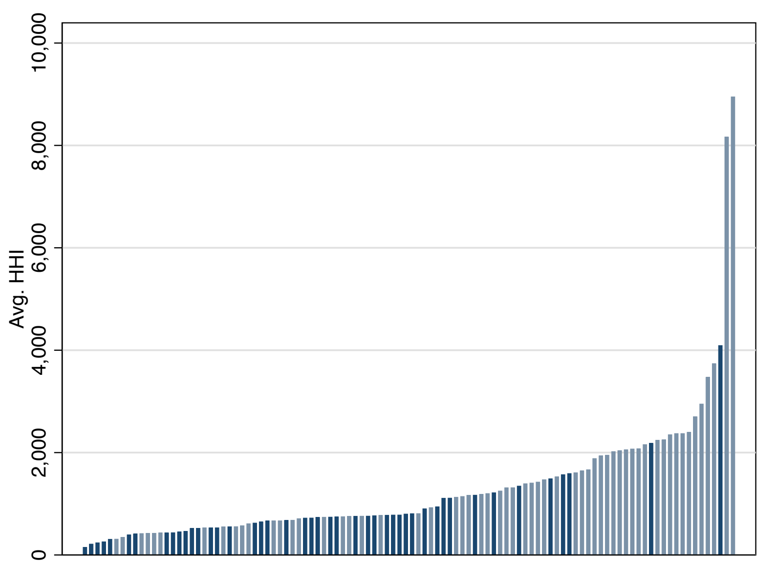
Notes: This graph presents the histogram of normalized modal charge dispersion using procedure-by-geozip-by-trimester data. Normalized modal charge is calculated as the ratio of a provider's modal charge to the market average, defined as the mean of the charge in each procedure-by-geozip-by-trimester market. The histogram is restricted to normalized charges below 5.

Figure 6: Percentage of Claims Out-of-Network, by Procedure Code



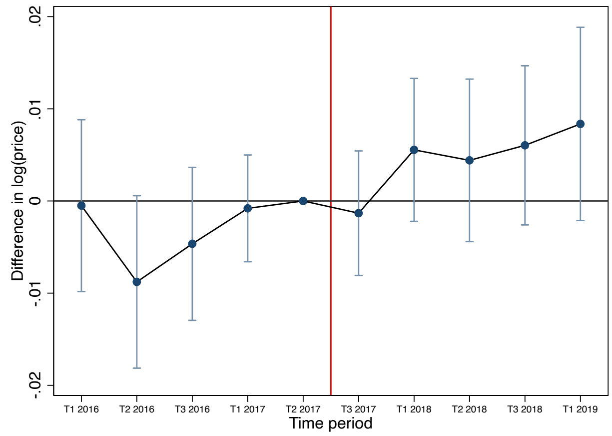
Notes: The graph represents the distribution of the percentage of out-of-network claims for procedure codes ordered lowest to highest. Dark blue bars indicate procedure codes used in the final analysis sample.

Figure 7: Physician Market Concentration



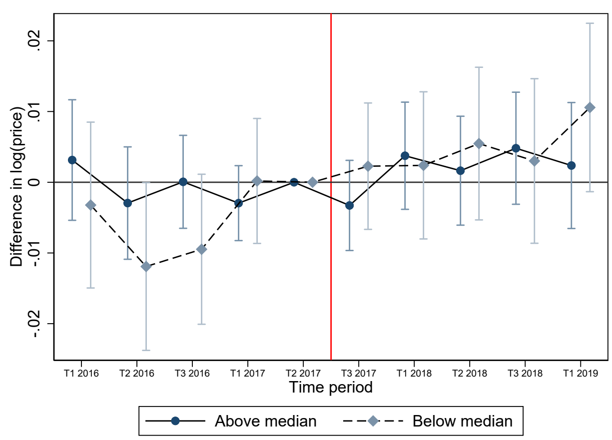
Notes: This figure represents the distribution of physician market concentration by procedure code in 2016. Provider market concentration was constructed as the HHI for each procedure and geozip, with the market share constructed as each provider, defined by the NPI. Dark blue bars indicate procedure codes used in the final analysis sample.

Figure 8: Event Study: Difference between Treatment and Control with Difference-in-Differences Specification



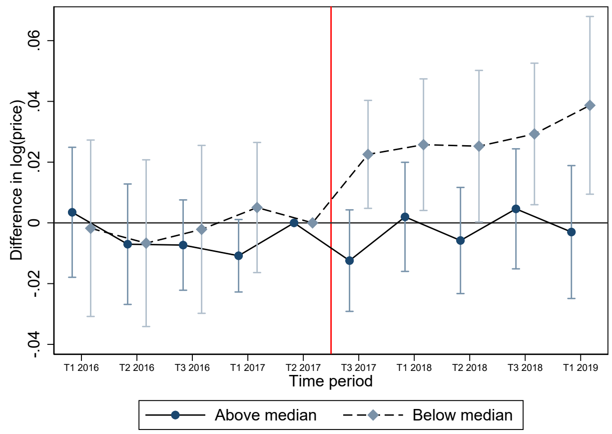
Note: This figure plots coefficients from a regression of log(price) on an interaction between treatment and trimester, with time (trimester-year), market (procedure-by-geozip), procedure-by-trimester, geozip-by-trimester, and provider fixed effects. Treatment began at the start of trimester 3 in 2017. This event study corresponds to a difference-in-differences regression model specification. Standard errors are clustered at the category-by-geozip level.

Figure 9: Event Study: Treatment Effect for High vs. Low Price Providers



Note: This figure plots coefficients from difference-in-differences regressions of log(price) on an interaction between treatment and trimester, with time (trimester-year), market (procedure-by-geozip), procedure-by-trimester, geozip-by-trimester, and provider fixed effects. The figure plots the trends for high vs. low price providers, with high price providers defined as providers with a billed charge in that trimester above the median charge for that procedure-by-geozip and trimester, and low price providers defined as providers with a billed charge below the median charge. Treatment began at the start of trimester 3 in 2017. This event study corresponds to a difference-in-differences regression model specification. Standard errors are clustered at the category-by-geozip level.

Figure 10: Event study: Treatment Effect for Low Out-of-Network Volume Procedures, High vs. Low Price Providers



Note: This event study graph plots the difference in log(price) between the treatment and control group for high price (above median charge) and low price (below median charge) providers by trimester, with a differences-in-differences specification. This figure plots coefficients from a regression of log(price) on an interaction between treatment and trimester, with time (trimester-year), market (procedure-by-geozip), procedure-by-trimester, geozip-by-trimester, and provider fixed effects. Treatment began at the start of trimester 3 in 2017. This event study corresponds to a difference-in-differences regression with interacted fixed effects. Standard errors are clustered at the category-by-geozip level.

Table 3: Provider-Level Regressions: Treatment Effect of NYHOST

	log(Price)			log(Quantity)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment effect	0.0013	0.0012	0.0075*	-0.0071	-0.0040	0.0009
	(0.0052)	(0.0046)	(0.0042)	(0.0113)	(0.0083)	(0.0071)
Observations	583469	583469	583469	583469	583469	583469
Adjusted R^2	0.945	0.946	0.946	0.601	0.601	0.601
ProcedureXGeozip FE	Yes	Yes	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes	Yes	Yes
ProcedureXTime FE		Yes	Yes		Yes	Yes
GeozipXTime FE			Yes			Yes
Provider FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table contains coefficients from heterogeneity tests with a difference-in-differences regression model testing for the impact of log(price) on an interaction between the treatment and a post indicator, with time (trimester-by-year), market (procedure-by-geozip), procedure-by-time, geozip-by-time and provider fixed effects. Treatment began at the start of trimester 3 in 2017. Standard errors are clustered at the category-by-geozip level.

Table 4: Provider-Level Regressions: Heterogeneity Tests for the Treatment Effect of NYHOST

	Baseline price		Continuous procedures		OON procedures		Market HHI		Website use	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	$> Median$	$< Median$	Non-continuous	Continuous	$> Median$	$< Median$	$> Median$	$< Median$	$> Median$	$< Median$
log(P) effect	0.0023	0.0095	0.0103	0.0027	0.0062	0.0128*	0.0065	0.0084*	0.0064	0.0114**
	(0.0040)	(0.0064)	(0.0079)	(0.0044)	(0.0043)	(0.0075)	(0.0083)	(0.0044)	(0.0052)	(0.0055)
Observations	313057	242724	156491	333587	407252	175987	86596	496188	293394	288071
Adjusted R^2	0.972	0.950	0.941	0.903	0.922	0.945	0.952	0.950	0.955	0.942
Panel B: Quantity treatment effects										
log(q) effect	0.0001	0.0125	0.0153	-0.0190*	-0.0066	0.0006	0.0174	-0.0000	-0.0112	-0.0004
	(0.0096)	(0.0099)	(0.0097)	(0.0113)	(0.0084)	(0.0104)	(0.0159)	(0.0080)	(0.0127)	(0.0081)
Observations	313057	242724	156491	333587	407252	175987	86596	496188	293394	288071
Adjusted R^2	0.657	0.684	0.435	0.540	0.578	0.564	0.685	0.609	0.604	0.612

Notes: This table contains coefficients from a regression of log(price) on an interaction between treatment and a post indicator with fixed effects for time (trimester-by-year), market (procedure-by-geozip), procedure-by-time and geozip-by-time corresponding to a difference in difference in difference regression testing for heterogeneity. Treatment began at the start of trimester 3 in 2017. Standard errors are clustered at the category-by-geozip level.

Table 5: Provider level regressions: Treatment effect of NYHOST by procedure category

	(1) CT	(2) MRI	(3) RAD	(4) GI	(5) EYE	(6) ORTHO	(7) OB	(8) PSYCH	(9) PTOT	(10) CHIRO
Panel A: Charge Treatment Effects										
log(P) effect	-0.03433	0.03651	0.03000	-0.01443	-0.00265	0.03838*	-0.02067	-0.01720**	0.00653	-0.00660
	(0.02346)	(0.02591)	(0.02154)	(0.02572)	(0.00808)	(0.02250)	(0.01539)	(0.00662)	(0.00919)	(0.00690)
Observations	33868	33405	55382	27315	65897	20298	13285	129105	159101	45266
Adjusted R^2	0.850	0.807	0.869	0.866	0.835	0.951	0.970	0.818	0.836	0.860
ProcXZip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Provider FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Quantity Treatment Effects										
log(Q) effect	0.01567	0.05155	0.06113**	0.01341	-0.00281	-0.00142	0.04034*	0.00346	-0.04116*	-0.00416
	(0.03661)	(0.03439)	(0.02358)	(0.03201)	(0.01858)	(0.02011)	(0.02363)	(0.01281)	(0.02069)	(0.02523)
Observations	33868	33405	55382	27315	65897	20298	13285	129105	159101	45266
Adjusted R^2	0.536	0.537	0.482	0.632	0.509	0.640	0.546	0.619	0.449	0.509
ProcXZip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Provider FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table contains coefficients from a regression of log(price) and log(quantity) on an interaction between treatment and a post indicator with fixed effects for time (trimester-year) and market (procedure-by-geozip) corresponding to a difference in difference regression for each procedure category. Treatment began at the start of trimester 3 in 2017. Standard errors are clustered at the category-by-geozip level.

A Appendix

A.1 Categories of service for procedures examined

Category	Randomized	Number of Procedures
Acupuncture	N	4
Allergy	N	4
Bone Density	N	1
Cardiology	N	2
Chemotherapy	N	1
Chiropractic	Y	4
CAT Scan (Radiology)	Y	4
Dermatology	N	4
ENT	N	5
Gastroenterology	Y	3
Infusion	N	1
Mammogram	Y	8
MRI	Y	5
Neuromuscular	N	1
Obstetrics & Gynecology	Y	4
Ophthalmology	Y	4
Orthopaedic	Y	6
Pain	N	5
Plastic Surgery	N	1
Psychotherapy	Y	5
Physical Therapy/Occupational Therapy	Y	8
Pulmonology	N	3
Radiology	Y	6
Sleep Medicine	N	2
Spine	N	4
Surgery	N	3
Urology	N	1
Ultrasound	N	4
Ultrasound-OB	N	2
Vascular Radiology	N	2

Notes: These categories were selected on the basis of encompassing procedures that were commonly serviced and non-emergent.

A.2 Description of the datasets utilized

Dataset	**Description**
FAIRHealth	2016-2019
FAIRHealth NYHOST Website data	2016-2019
CMS Physician Compare	2017
CMS National Plan and Provider Enumeration System (NPPES)	2017

Notes: These datasets were utilized to conduct the analyses of the impact of the NYHOST price transparency tool. The CMS Physician Compare file utilized is the most recent dataset that was able to be accessed on 2/20/2020.

A.3 Data on provider characteristics

Dataset	**# Distinct NPI in FAIRHealth Data**	**# Distinct NPI in CMS Compare (2017)**
Total number of providers in each dataset.	205,258	1,142,428
Providers in both FAIRHealth + CMS Physician Compare	119,583	119,583
# Distinct NPI in FAIRHealth Data and CMS Physician Compare, with specialties associated with the MD/DO credential.	78,509	78,509
Number of providers represented in the balanced panel (subset of the total number of providers)	21,601	14,146

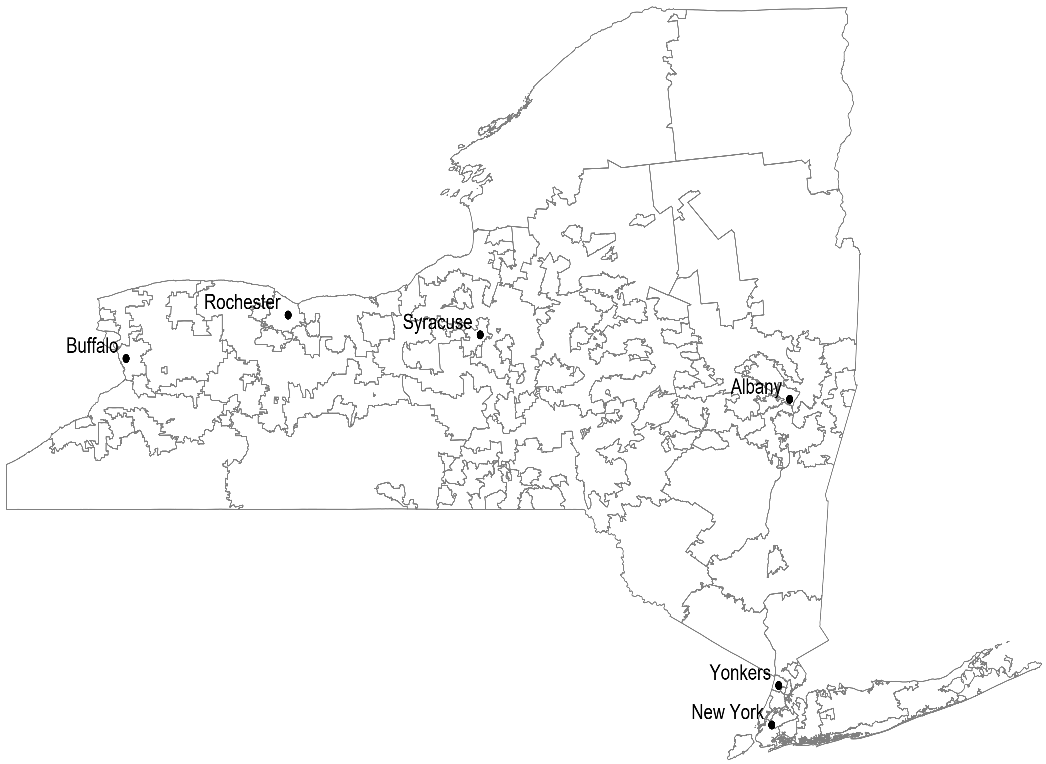
Notes: This table demonstrates the number of providers represented in the FAIR Health dataset, and the match of the NPIs in the FAIRHealth dataset with the 2017 CMS Physician Compare File. The CMS Physician Compare File includes information on providers, including credentials, medical school, gender, and affiliated hospitals.

A.4 Construction of the Balanced Panel

Steps to Create Balanced Panel	**# Observations**	**# Observations dropped**
1. Total number of observations for the provider-level dataset with the modal charge (NPI X geozip X procedure X year X trimester).	3,598,866	
2. Drop any observation with fewer than 5 claims in a trimester.	1,708,771	1,890,095
3. Create a balanced panel of providers by restricting the sample to providers with at least five claims in each trimester for the study period.	1,083,963	624,808
4. Create a balanced panel of procedures (drop any procedures that were added or discontinued during this time period).	1,028,785	55,178
5. Drop any procedures for which provider-level charges were released across the state.	667,014	361,771
6. Drop any observations where the volume is above the 99th percentile of the volume for a given procedure-by-geozip-by-trimester.	654,425	12,589
7. Drop any observations where the modal charge is above the 99th percentile of the charge for a given procedure-by-geozip-by-trimester.	628,428	25,997

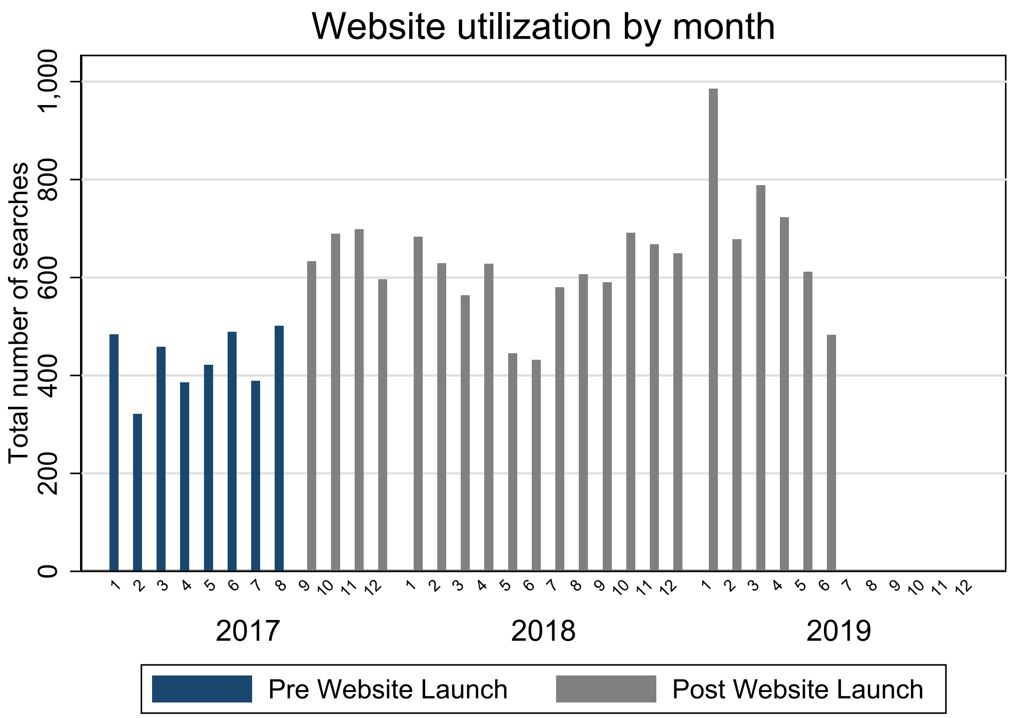
Notes: We constructed a balanced panel of providers and procedures in order to ensure that we were following the same panel of providers over time to assess the impact of the randomization on their billed charges over time after the release of the price transparency tool. The balanced panel was constructed by restricting the sample to the procedures that were randomized across geozips, to providers with at least five claims in each quarter, and procedures that were active throughout the study period.

A.5 Map of New York Geozips



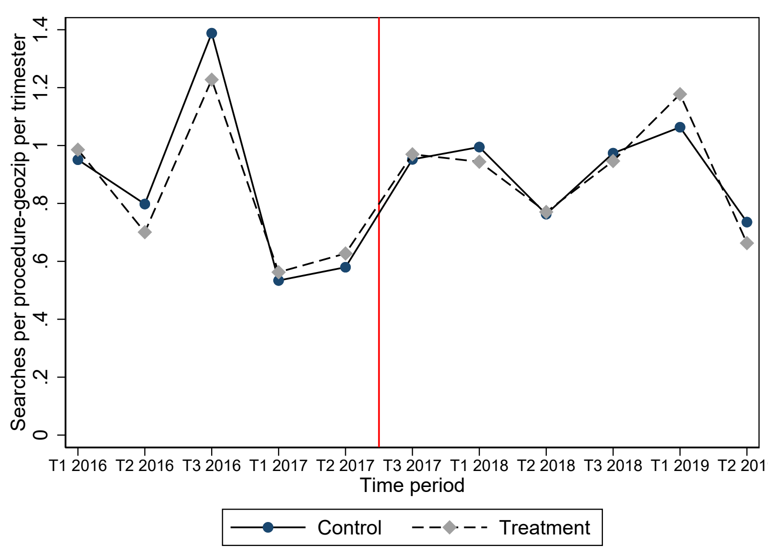
Notes: Map of New York State geozips, with select cities indicated.

A.6 Total Website Utilization by Month



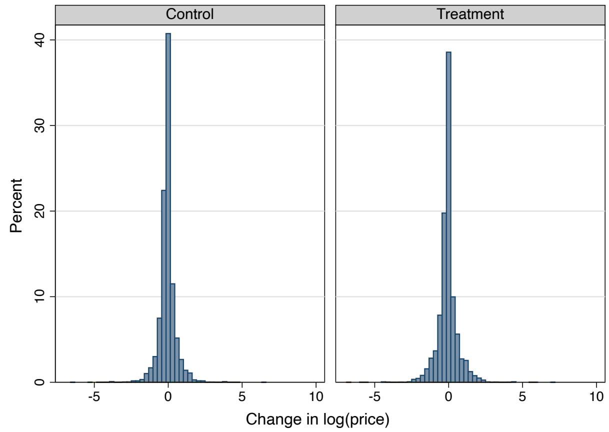
Note: This figure plots total number of monthly searches across the pre- and post- period. Large spikes in 2016 and 2019 are associated with major website overhauls and marketing changes.The experiment was in effect from 9/2017 through 6/2019.

A.7 Average Website Utilization Excluding Manhattan



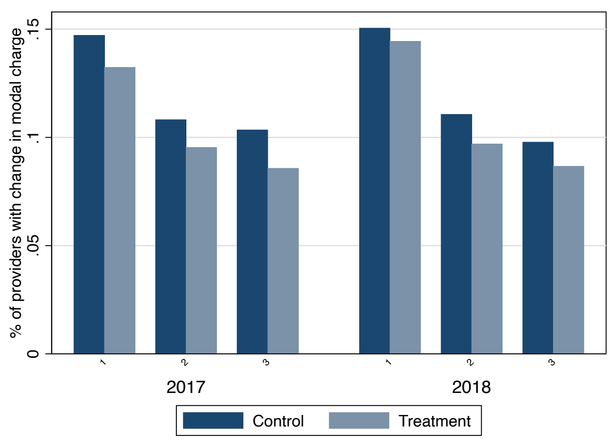
Notes: This figure plots the average website utilization (data on NYHOST web searches provided by FAIRHealth) per procedure-by-geozip combination for procedures in the treatment and control groups when the geozip corresponding to Manhattan (geozip = 100) is excluded.

A.8 Changes in log(Price) in Treatment and Control Groups



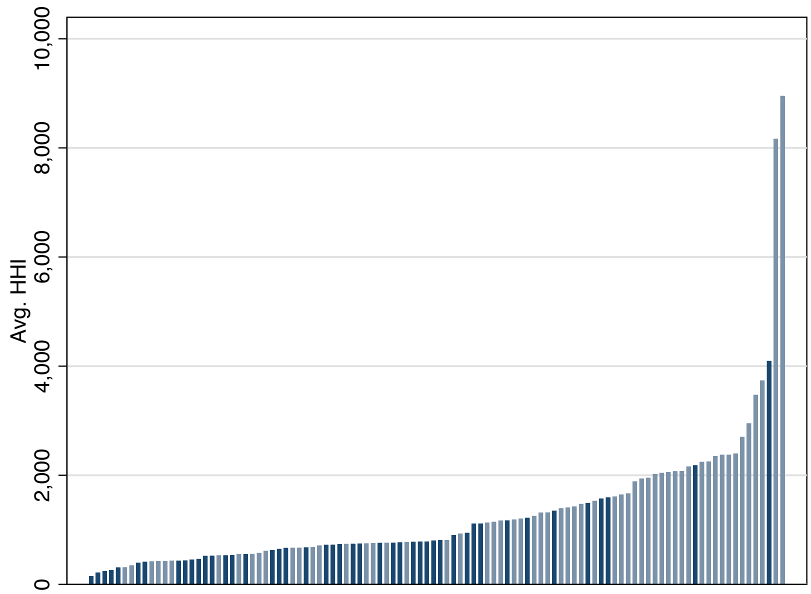
Note: This figure presents histograms of charge updates for providers in the treatment and control groups.

A.9 Changes in log(Price) over Time (2017-2018)



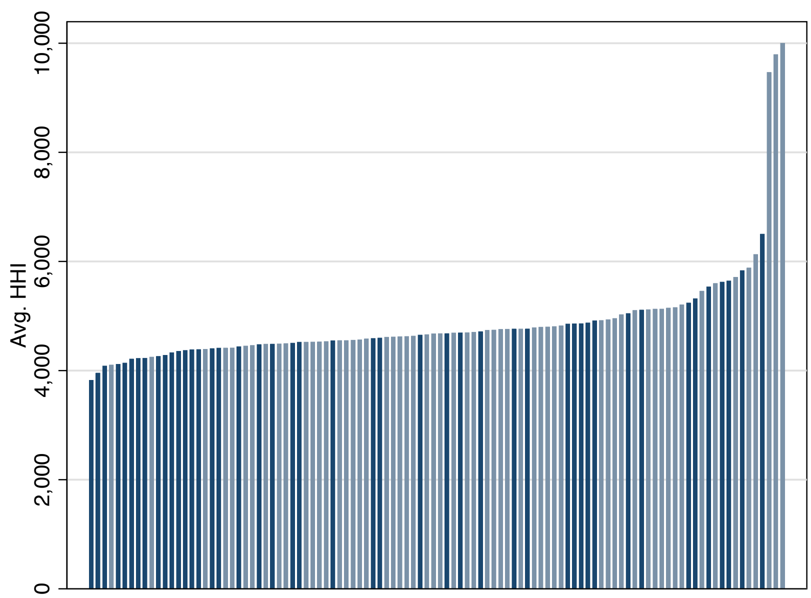
Data source: FAIRHealth.
Note: This figure presents histograms of charge updates for providers in the treatment and control groups. This displays the percentage of providers who update their charges each trimester.

A.10 Provider Market Concentration



Notes: Provider market concentration in 2016, by procedure code, with each provider identified by the National Provider Identifier (NPI).

A.11 Insurer Market Concentration



Notes: Insurer market concentration in 2016, by procedure code, with each insurer identified by a FAIR Health "key".

A.12 Provider-Level Results: Treatment Effect of NYHOST with Alternative Fixed Effects Specification

	log(Price)		log(Quantity)	
	(1) DiD	(2) Triple Diff	(3) DiD	(4) Triple Diff
Treatment effect	0.0013	0.0075*	-0.0071	0.0009
	(0.0052)	(0.0042)	(0.0113)	(0.0071)
Observations	583469	583469	583469	583469
Adjusted R^2	0.945	0.946	0.601	0.601
ProcedureXGeozip FE	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes
ProcedureXTime FE		Yes		Yes
GeozipXTime FE		Yes		Yes
Provider FE	Yes	Yes	Yes	Yes

Notes: This table contains coefficients from a difference-in-differences regression of log(price) on an interaction between the treatment variable and a post indicator with different fixed effects configurations, including time (trimester-year), market (procedure-by-geozip), procedure-by-time, and geozip-by-time. Treatment began at the start of trimester 3 in 2017. Standard errors are clustered at the category-by-geozip level.

A.13 Provider-Level Results: Robustness Test of the Treatment Effect on the Percentile of Providers' Charges

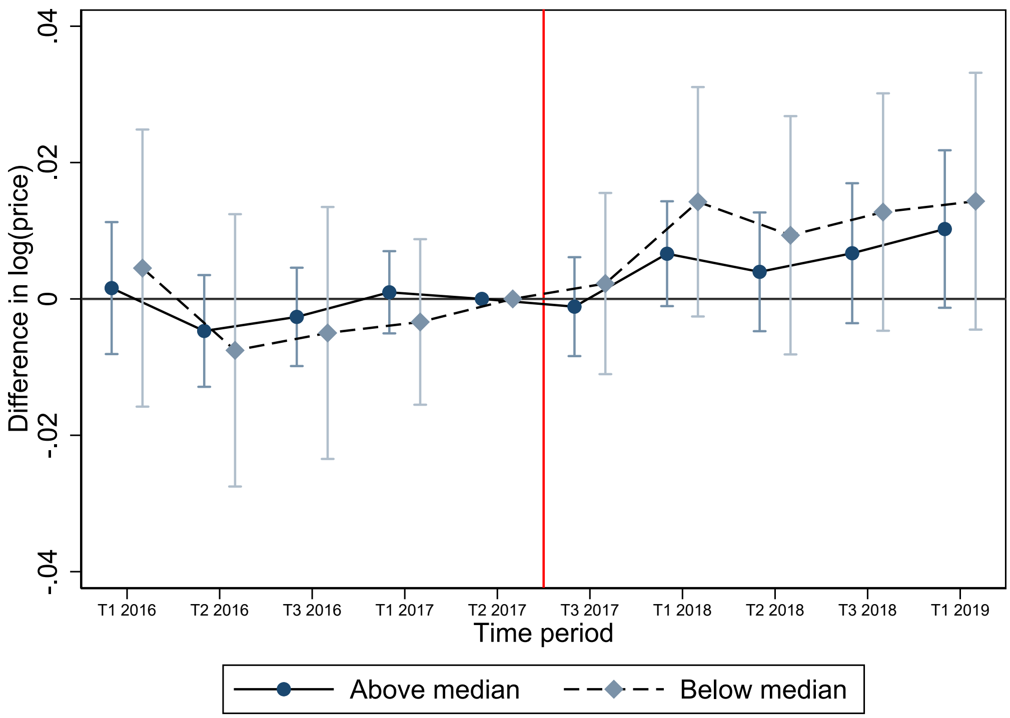
	(1) 95th	(2) 95th	(3) 50th	(4) 50th	(5) 5th	(6) 5th
Treat*Post	0.003	0.012***	0.003	0.012***	0.007	0.016***
	(0.006)	(0.004)	(0.006)	(0.004)	(0.006)	(0.004)
Constant	5.016***	5.013***	4.948***	4.945***	4.839***	4.837***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Observations	510815	510815	510815	510815	510815	510815
Adjusted R^2	0.848	0.848	0.846	0.846	0.815	0.816
ProcedureXGeozip Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Post	Yes	Yes	Yes	Yes	Yes	Yes
ProcedureXPost Dummies		Yes		Yes		Yes
GeozipXPost Dummies		Yes		Yes		Yes

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

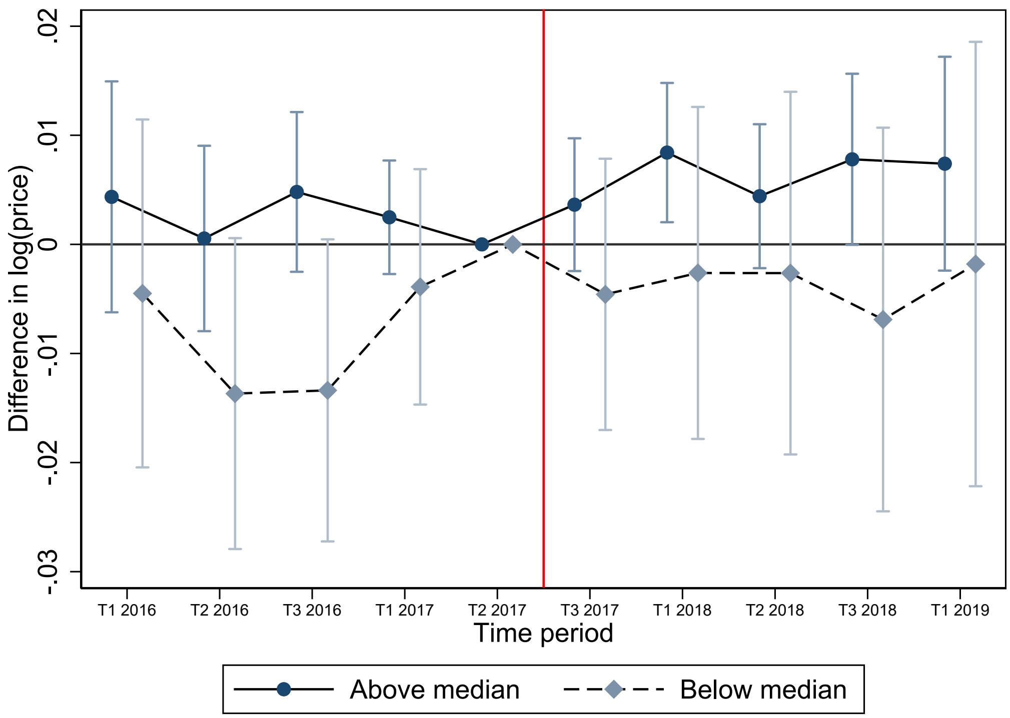
Notes: Fixed effects models with procedure-by-geozip and time fixed effects. The DD estimates demonstrate the overall impact of NY HOST on the log of the percentile of a provider's charge (5th, 50th, and 95th percentile) each trimester as the outcome variable. The time fixed effects are measured the by the"Post"dummy variable, signifying the time period after the experiment went into effect. The difference-in-differences specification includes post-by-procedure dummy variables and post-by-geozip dummy variables.

A.14 Event Study: Difference between Treatment and Control with Differences-in-Differences Specification, for High vs. Low OON Volume Procedures



Note: This figure plots coefficients from a regression of log(price) on an interaction between treatment and trimester, with time (trimester-year), market (procedure-by-geozip), procedure-by-trimester, geozip-by-trimester, and provider fixed effects, for above and below median OON procedures. Treatment began at the start of trimester 3 in 2017. This event study corresponds to a difference-in-differences regression model specification. Standard errors are clustered at the category-by-geozip level.

A.15 Event Study: Difference between Treatment and Control with Differences-in-Differences Specification, for High Out-of-Network Procedures



Note: This figure plots coefficients from a regression of log(price) on an interaction between treatment and trimester and fixed effects for time (trimester-year), market (procedure-by-geozip), procedure-by-trimester and geozip-by-trimester for high OON procedures across above and below median price providers. Treatment began at the start of trimester 3 in 2017. This event study corresponds to a difference regression. Standard errors are clustered at the category-by-geozip level.

A.16 Provider-level results: Multiple Hypothesis Testing for Heterogeneity Tests by Category, for Price Effects

	(1)									
	CHIRO	CT	GASTRO	MRI	OB	OPHTHO	ORTHO	PSYCH	PTOT	RAD
P-Values	0.347	0.154	0.579	0.169	0.189	0.745	0.0980	0.0140	0.483	0.174
Q-Values	0.608	0.516	0.804	0.516	0.516	0.865	0.516	0.177	0.674	0.516

Notes: This table presents the p-values from the heterogeneity tests by category, from difference-in-differences specification with log(price) as the dependent variable, and the sharpened False Discovery Rate (FDR) q-values for multiple hypothesis testing.

A.17 Provider-level results: Multiple Hypothesis Testing for Heterogeneity Tests by Category, for Quantity Effects

	(1)									
	CHIRO	CT	GASTRO	MRI	OB	OPHTHO	ORTHO	PSYCH	PTOT	RAD
P-Values	0.870	0.672	0.678	0.144	0.0980	0.881	0.944	0.789	0.0560	0.0150
Q-Values	0.865	0.825	0.825	0.516	0.516	0.865	0.894	0.865	0.507	0.177

Notes: This table presents the p-values from the heterogeneity tests by category, from difference-in-differences specification with log(quantity) as the dependent variable, and the sharpened False Discovery Rate (FDR) q-values for multiple hypothesis testing.

A.18 Market-Level Regressions: Treatment Effect of NYHOST on Outcomes

	log(Price)		log(Quantity)	
	(1)	(2)	(3)	(4)
	DiD	Triple Diff	DiD	Triple Diff
Treatment effect	0.0127**	0.0176***	-0.0111	0.0149
	(0.0063)	(0.0055)	(0.0272)	(0.0131)
Observations	15871	15864	15871	15864
Adjusted R^2	0.994	0.995	0.940	0.987
ProcedureXGeozip FE	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes
ProcedureXTime FE		Yes		Yes
GeozipXTime FE		Yes		Yes

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: This table contains coefficients from difference-in-differences regression of log(price) on an interaction between treatment and a post indicator with time (trimester-by-year), market (procedure-by-geozip), procedure-by-time, and geozip-by-time fixed effects. Treatment began at the start of trimester 3 in 2017. Standard errors are clustered at the category-by-geozip level.

A.19 Market-Level Regressions: Heterogeneity Tests for the Treatment Effect of NYHOST

	Coefficient of Variation		Continuous Procedures		OON Procedures		Market HHI		Website Use	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	>Median	<Median	Non-Continuous	Continuous	>Median	<Median	>Median	<Median	>Median	<Median
log(P) effect	0.0164*	0.0189***	0.0120	-0.0046	0.0083	0.0286***	0.0196**	0.0141**	0.0148**	0.0282**
	(0.0088)	(0.0069)	(0.0089)	(0.0088)	(0.0072)	(0.0083)	(0.0097)	(0.0058)	(0.0060)	(0.0121)
Observations	8045	7801	7653	4750	7827	8037	7922	7922	13406	2159
Adjusted R^2	0.994	0.996	0.995	0.984	0.996	0.993	0.993	0.998	0.995	0.997
ProcedureXGeozip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ProcedureXTime FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GeozipXTime FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
* $p < .10$, ** $p < .05$, *** $p < .01$										
Panel B: Quantity treatment effects										
log(Q) effect	0.0201	0.0183	0.0612***	-0.0506	-0.0213	0.0600***	0.0194	0.0086	0.0148	-0.0159
	(0.0154)	(0.0226)	(0.0171)	(0.0488)	(0.0219)	(0.0170)	(0.0270)	(0.0088)	(0.0151)	(0.0183)
Observations	8045	7801	7653	4750	7827	8037	7922	7922	13406	2159
Adjusted R^2	0.990	0.985	0.984	0.988	0.989	0.981	0.969	0.996	0.985	0.994
ProcedureXGeozip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ProcedureXTime FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GeozipXTime FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: This table contains coefficients from a regression of charges (constructed here as the log(price)) on an interaction between treatment and a post indicator with fixed effects for time (trimester-by-year), market (procedure-by-geozip), procedure-by-time and geozip-by-time corresponding to a difference in difference in difference regression testing for heterogeneity. Treatment began at the start of trimester 3 in 2017. Standard errors are clustered at the procedure-by-geozip level.

A.20 Market-Level Regressions: Treatment Effect of NYHOST by Category

	(1) CT	(2) MRI	(3) RAD	(4) GI	(5) EYE	(6) ORTHO	(7) OB	(8) PSYCH	(9) PTOT	(10) CHIRO
Panel A: Charge treatment effects										
log(P) effect	-0.01550	0.08102***	-0.02828	-0.00601	-0.00679	0.02856	0.00558	-0.00928	-0.00500	0.02081
	(0.02119)	(0.02502)	(0.02057)	(0.02529)	(0.01567)	(0.02107)	(0.01726)	(0.01009)	(0.01224)	(0.01465)
Observations	1240	1550	1766	930	1240	1858	1239	1549	1985	1219
Adjusted R^2	0.955	0.887	0.936	0.965	0.997	0.984	0.995	0.939	0.980	0.937
ProcXZip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
* $p < .10$, ** $p < .05$, *** $p < .01$										
Panel B: Quantity treatment effects										
log(Q) effect	0.06304*	0.08516***	-0.21411	-0.00588	0.06716**	-0.04532	0.11407**	0.01925	-0.03083	-0.09618
	(0.03725)	(0.02756)	(0.13690)	(0.02968)	(0.03316)	(0.03495)	(0.05126)	(0.03232)	(0.05479)	(0.09246)
Observations	1240	1550	1766	930	1240	1858	1239	1549	1985	1219
Adjusted R^2	0.980	0.974	0.613	0.989	0.982	0.977	0.968	0.990	0.965	0.970
ProcXZip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: This table presents the regression results when stratifying the procedures by category. This table contains coefficients from a regression of log(price) and log(quantity) on an interaction between treatment and a post indicator with fixed effects for time (trimester-year) and market (procedure-by-geozip) corresponding to a difference in difference regression for each procedure category. Treatment began at the start of trimester 3 in 2017. Standard errors are clustered at the category-by-geozip level.