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

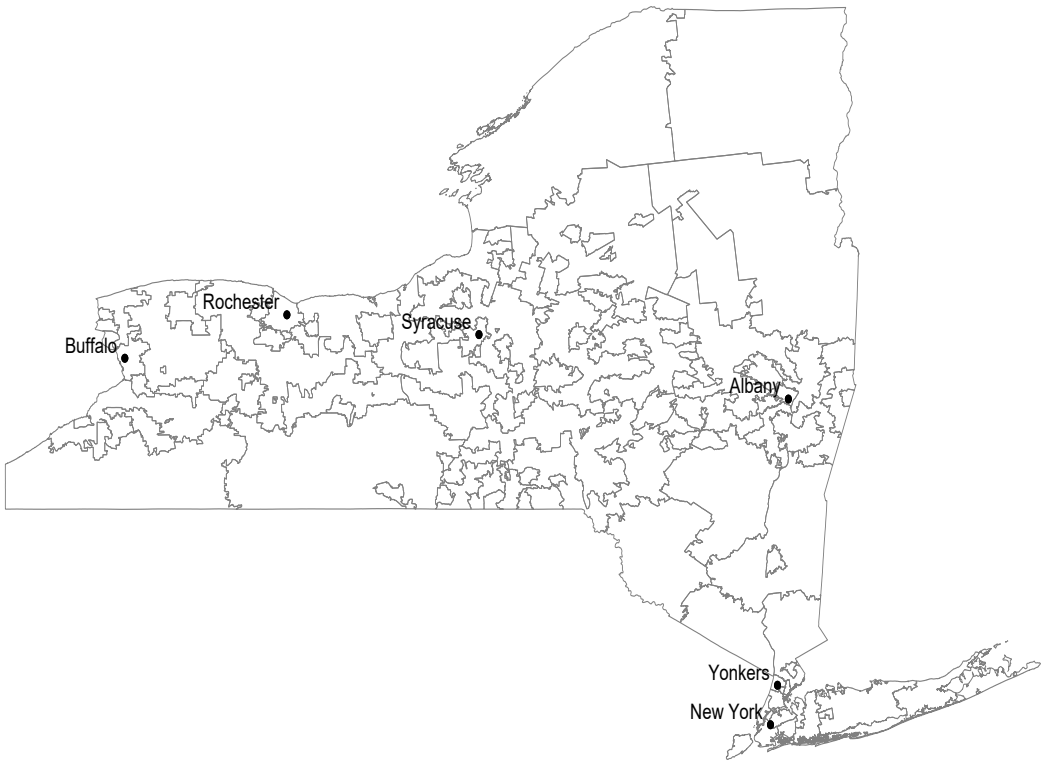
Notes: This table demonstrates the number of providers represented in the FAIR Health dataset, and the match of the NPIs in the FAIRHealth dataset with the 2017 CMS Physician Compare File. The CMS Physician Compare File includes information on providers, including credentials, medical school, gender,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

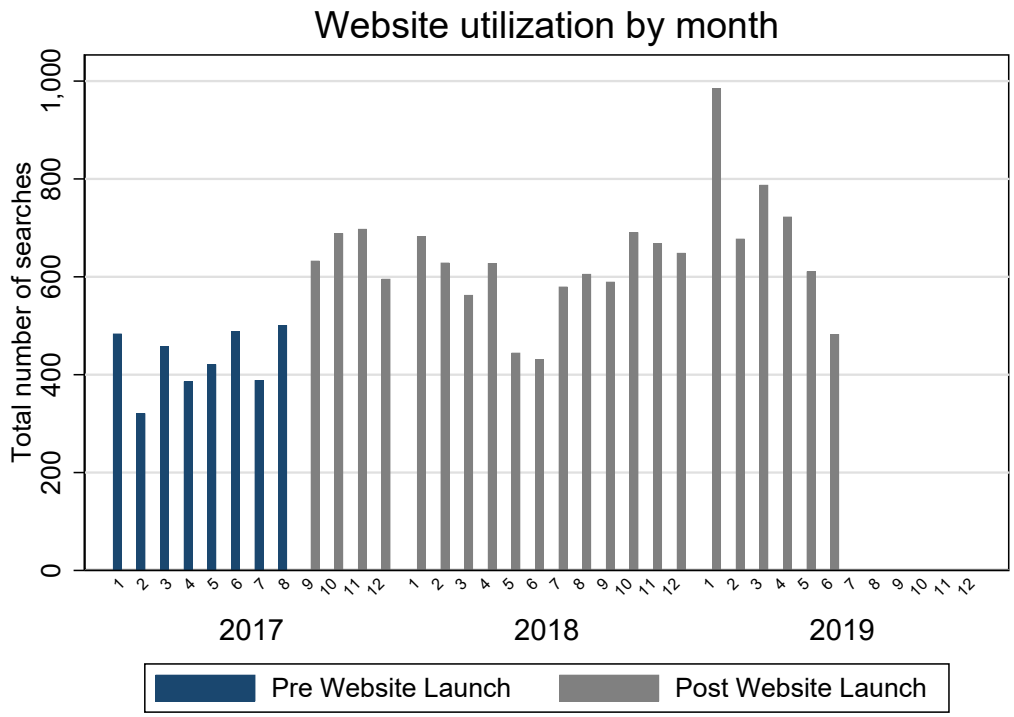
Notes: We constructed a balanced panel of providers and procedures in order to ensure that we were following the same panel of providers over time to assess the impact of the randomization on their billed charges over time after the release of the price transparency tool. The balanced panel was constructed 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



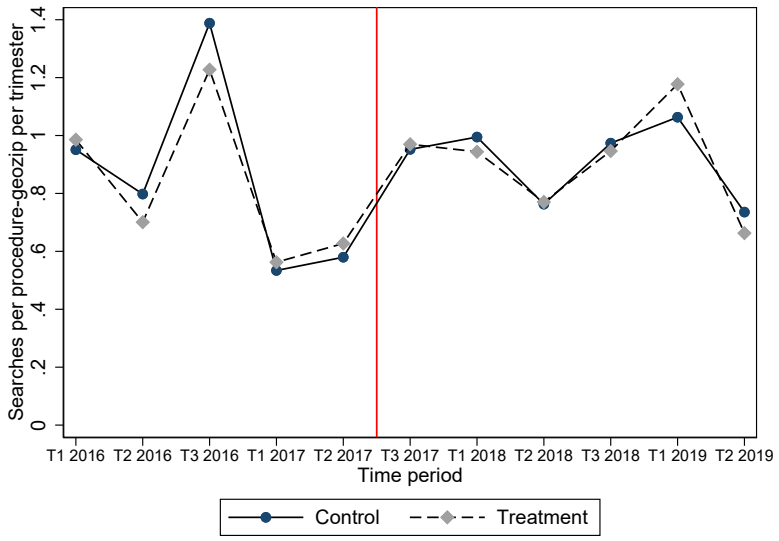
Notes: Map of New York State geozips, with select cities indicated.

A.6 Total Website Utilization by Month



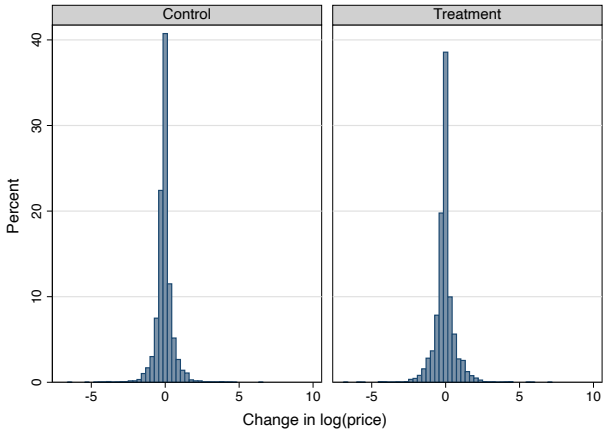
Note: This figure plots total number of monthly searches across the pre- and post- period. Large spikes in 2016 and 2019 are associated with major website overhauls and marketing changes.The experiment was in effect from 9/2017 through 6/2019.

A.7 Average Website Utilization Excluding Manhattan



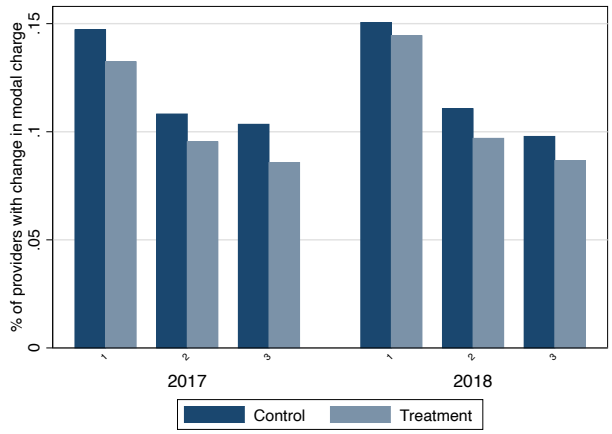
Notes: This figure plots the average website utilization (data on NYHOST web searches provided by FAIRHealth) per procedure-by-geozip combination for procedures in the treatment and control groups when the geozip corresponding to Manhattan (geozip = 100) is excluded.

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



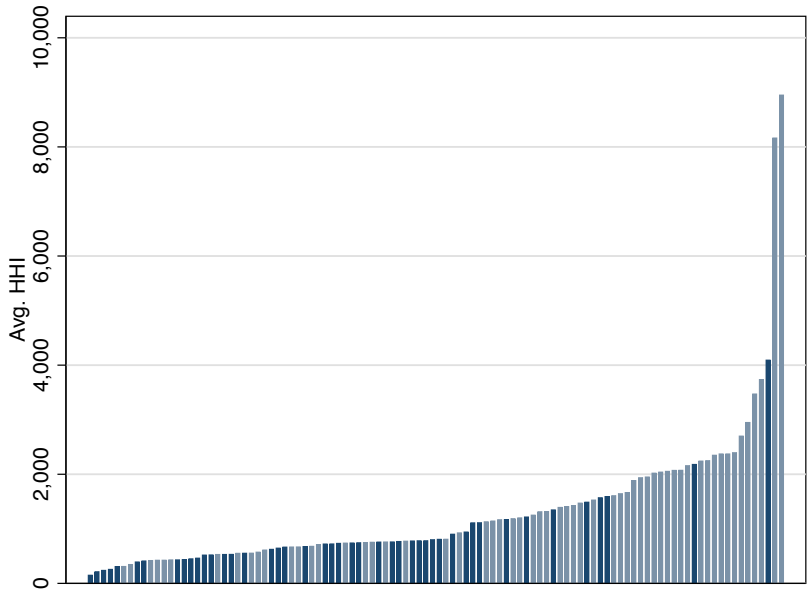
Note: This figure presents histograms of charge updates for providers in the treatment and control groups.

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



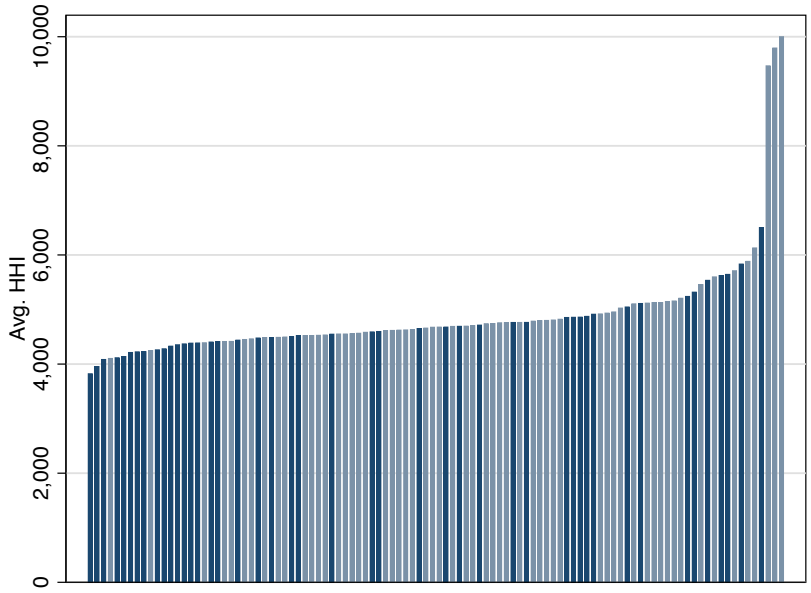
Data source: FAIRHealth.
Note: This figure presents histograms of charge updates for providers in the treatment and control groups. This displays the percentage of providers who update their charges each trimester.

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

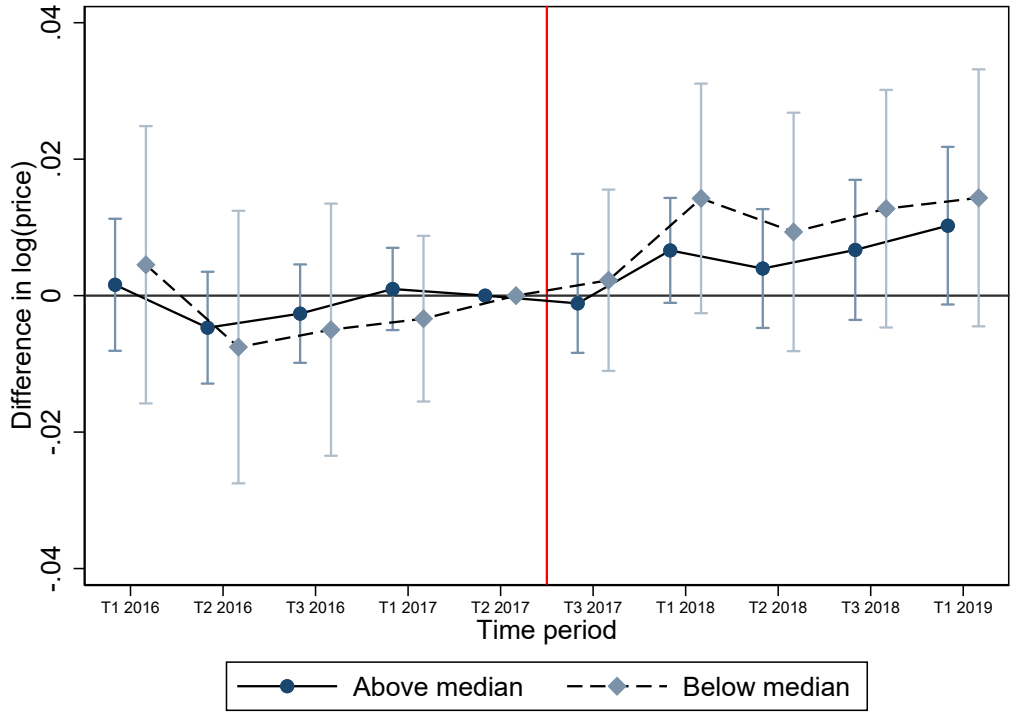
	(1) 95th	(2) 95th	(3) 50th	(4) 50th	(5) 5th	(6) 5th
Treat*Post	0.003	0.012***	0.003	0.012***	0.007	0.016***
	(0.006)	(0.004)	(0.006)	(0.004)	(0.006)	(0.004)
Constant	5.016***	5.013***	4.948***	4.945***	4.839***	4.837***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Observations	510815	510815	510815	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

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

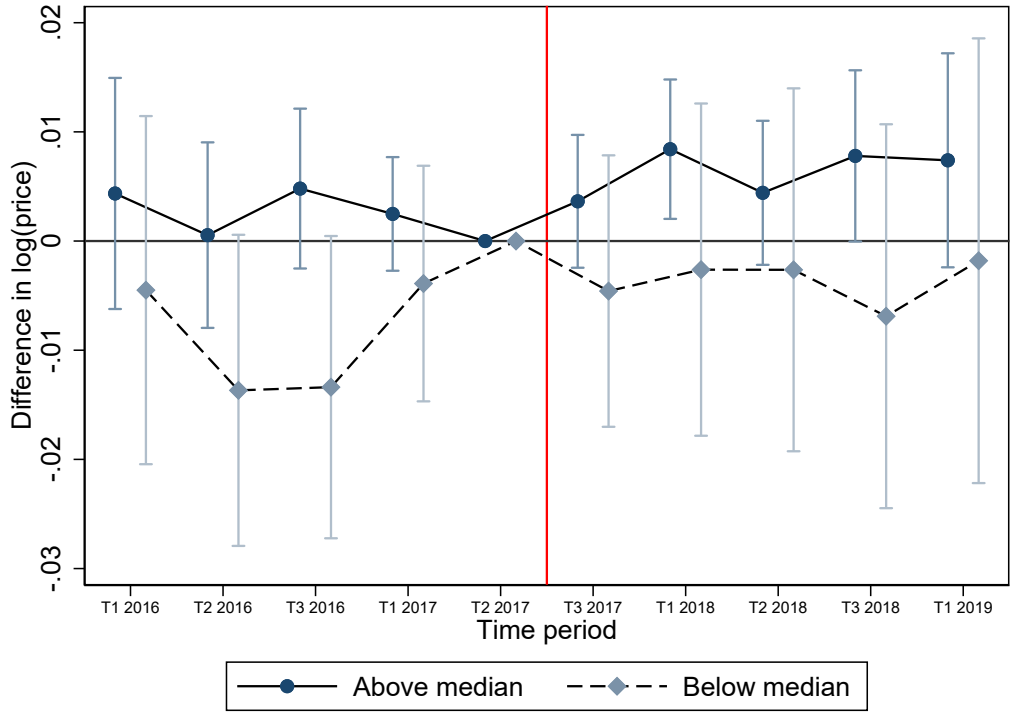
Notes: Fixed effects models with procedure-by-geozip and time fixed effects. The DD estimates demonstrate the overall impact of NY HOST on the log of the percentile of a provider's charge (5th, 50th, and 95th percentile) each trimester as the outcome variable. The time fixed effects are measured the 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) >Median	(2) <Median	(3) Non-Continuous	(4) Continuous	(5) >Median	(6) <Median	(7) >Median	(8) <Median	(9) >Median	(10) <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.