Diagnosing Price Dispersion:

Demand, Bargaining, and Search in Hospital-Supplier Contracting

Matthew Grennan*
University of Pennsylvania, The Wharton School & NBER
grennan@wharton.upenn.edu

Ashley Swanson
University of Pennsylvania, The Wharton School & NBER
aswans@wharton.upenn.edu

February 26, 2019

Abstract

Using detailed purchase order data for a large sample of US hospitals 2009-15, we document large price dispersion across hospital buyers for identical brands in a variety of important medical supply categories. Hospitals also vary dramatically in the size and composition of the set of suppliers they contract with, and on average contract with an order of magnitude fewer suppliers than are available in the market. We develop a model and identification strategy to determine the extent to which this dispersion is determined by brand preferences, search/contracting costs, and bargaining abilities. Estimates suggest that markups are primarily driven by lack of price sensitivity among health care providers in their usage decisions. Hospital administrator bargaining ability varies widely across hospitals, driving most of the observed price dispersion. Reducing search/contracting costs does reduce prices and price dispersion, but mostly impacts hospital surplus through putting higher value brands in the choice set. These effects vary dramatically across device categories, hospitals, and brands. Hospitals with previously small and low quality choice sets gain enormously from lower search costs, while others gain more modestly. Similarly, some high value, high search cost brands go from infrequent to heavy use across the market, while other brands' profits decrease.

*We would like to thank Sebastian Fleitas, Kate Ho, Robin Lee, Aviv Nevo, Tobias Salz, Ali Yurukoglu, and audiences at Carnegie Mellon, McGill, Penn, Stanford, and the Federal Trade Commission for helpful feedback. The data used in this paper were generously provided, in part, by the ECRI Institute (www.ecri.org). We gratefully acknowledge financial support from the Wharton Public Policy Initiative and NSF Award SES-1559485. Stuart Craig, Gi Heung Kim, Changhwa Lee, Jiaxuan (Grace) Nie, and Donato Onorato provided excellent research assistance. Any errors are our own.

1 Introduction

Business-to-business markets, in which buyer firms contract with a set of suppliers to fulfill their input needs, comprise a large part of the economy. The Law of One Price is frequently violated in such markets, with a large degree of price dispersion across buyers for similar or even identical brands. The underlying economic mechanisms and frictions driving price dispersion in business-to-business settings include: heterogenous demand for differentiated products and bargaining abilities in television content (Chipty and Snyder 1999; Crawford and Yurukoglu 2012), medical devices (Grennan 2013, 2014), and hospital services (Ho 2009; Ho and Lee 2017); search/contracting costs for buyers seeking to add a supplier in coal (Stigler 1961) and sanitation (Salz 2017); and strategic exclusion of suppliers in video rentals (Ho et al. 2012), hospital services (Sorensen 2003; Ho and Lee 2018), and pharmaceuticals (Starc and Swanson 2018). Price dispersion in such markets is important because it is typically an indicator of market power among suppliers and/or buyers (Hemphill and Rose 2018), with implications for static and dynamic efficiency. These prices impact buyer firm profitability in the short run, and, due to their intermediate place in vertical supply chains, downstream market efficiency in the long run.

In this paper, we quantify the economic mechanisms underlying price dispersion for a wide variety of consumable hospital supply markets. These markets exhibit large price dispersion, with coefficients of variation across hospitals for the exact same brand-month in the range 0.05-0.33 (Grennan and Swanson 2019).[1] Consumable hospital supplies are also important inputs in hospital care, representing 24 percent of hospital operating costs.[2] Thus, the variation in these input prices is meaningful in terms of its potential impact on both hospital margins and downstream markets for hospital services.[3]

We focus on three mechanisms that we expect to be particularly relevant in the hospital supply context: (1) heterogeneity in users' preferences; (2) heterogeneity in relative bargaining weights; and (3) contracting frictions – encompassing the process of searching for suppliers and acquiring the information and cooperation needed to develop a new buyer-supplier relationship.[4] At the point of use, medical supplies are selected by individuals, often

[1]To put these numbers in perspective, note that the average input price variation documented in Grennan and Swanson (2019) is approximately half the coefficient of variation for common procedure prices charged by hospitals in different markets (Cooper et al. 2018). It is near the top of the range of coefficients of variation found in consumer goods markets (Scholten and Smith 2002).

[2]Consumable supplies are the second largest component of hospital operating costs, after labor (Craig et al. 2018).

[3]According to the American Hospital Association 2018 Trendbook, the average hospital operating margin in 1995-2016 was 4.4 percent (https://www.aha.org/system/files/2018-05/2018-chartbook-table-4-1.pdf).

[4]Throughout this paper, we use the terms "search" or "contracting" costs to refer to any costs of adding a brand to a hospital's consideration set.

physicians, whose usage is price-insensitive and may be brand-loyal.[5] However, hospital purchasing administrators negotiate the contracts that add brands to the consideration set, and the hospital pays the price of the brands contracted and used. Hospitals and physicians often disagree about tradeoffs regarding price and perceived quality across competing brands, and agency conflicts emerge: physicians would like purchasing administrators to establish contracts for their preferred brands, hospital managers would like purchasing agents to seek and negotiate low prices for high-value brands, and hospital managers would like physicians to choose high-value brands within the contracted set. These agency conflicts have been a source of tension within hospitals, and past efforts by hospitals to standardize surgical materials and devices have encountered significant resistance from surgeons (Nugent et al. 1999). However, hospitals vary dramatically along several dimensions which are likely to mediate these conflicts: management practices (McConnell et al. 2013; Bloom et al. 2014), alignment with physicians (Ketcham et al. 2009; Baker et al. 2014; Swanson 2019), and reliance on purchasing and information intermediaries (Schneller 2009; Grennan and Swanson 2019). Industry participants and policymakers are eager to understand which, if any, of these features improve purchasing outcomes. Thus, it is an important and difficult task to understand the relative importance of the demand, bargaining, and search/contracting mechanisms in generating the observed variation.[6]

We present and estimate a structural empirical model that allows for equilibrium supplier networks, prices, and quantities to be determined by the interaction of search/contracting costs, demand preferences, and bargaining, any of which may vary across buyer-supplier pairs. We estimate the model using detailed data on hospital purchase orders for a sample of US hospitals, including monthly prices and quantities at the vendor, manufacturer, and stock keeping unit (SKU) level. We estimate the model separately across 19 different product categories that are important in terms of total hospital expenditure. The breadth of these product categories – including "physician preference items" (PPIs) (e.g. knee implants), commodities (e.g. surgical gloves), and other medical/surgical products (e.g. catheters) – allow us to analyze a fairly representative set of hospital supply purchases, including categories that likely vary in importance of the underlying mechanisms. For example, comparing PPIs to commodities and other medical/surgical products, hospitals rely on purchasing intermediaries less for PPIs (Schneller 2009), physician preferences are expected to be stronger for PPIs, and the total set of brands available to search over is smaller for PPIs. Despite these

[5] At an extreme, usage of devices known as "physician preference items" (PPIs) such as cardiac and orthopedic implants are associated with strong brand preferences.

[6] The latter space is evolving rapidly, as the increasing digitization of the hospital supply chain has led to the emergence of intermediaries, including Amazon.com Inc., seeking to use data and analytics to facilitate easier search and contracting across suppliers (Evans and Stevens 2018).

differences, we document similar price dispersion for PPIs and non-PPI product categories.

The primary challenge to estimating an empirical model that simultaneously allows for search costs, preference heterogeneity, and bargaining heterogeneity is to separately identify these different mechanisms. This issue is easily seen by considering each separate pair of factors individually. First, the presence of both search costs and preference heterogeneity introduces an identification problem similar to the familiar selection problem in the labor economics literature (beginning with Heckman 1979): the unobservable shocks in the demand equation may be correlated with the process that generated the *set of brands* under consideration. Second, the presence of both search costs and bargaining in a market may lead to bias in models that only account for one or the other. For example, high search costs can reduce consumers' bargaining leverage and mute the effect of supply-side concentration on prices (Allen et al. 2013, 2018). Finally, the effect of any source of market power, concentration or search, is further muted by the fact that hospitals exercise their monopsony power and negotiate prices that are even lower than competition and their outside options alone might suggest. A number of researchers have shown the importance of modeling the bargaining stage in models with differentiated products demand and negotiated prices.[7]

Our identification approach proceeds as follows. First, we contend with the endogeneity of the consideration set using a similar logic as that in Hausman (1996) and Nevo (2001) – rather than using prices of the same good in other markets as cost shifters, we use exposure variables capturing interactions with the same vendor in other, unrelated product categories as search cost shifters. Our exclusion restriction is that exposure to vendors in unrelated product categories impacts formation of the consideration set, but does not reflect correlated preferences over vendors across unrelated product categories, conditional on the consideration set and controls. The institutional details underlying this strategy are in Section 2.2.

Second, we jointly estimate differentiated products demand using observed brand shares within each hospital-year consideration set, and marginal costs and bargaining parameters in a Nash-in-Nash bargaining framework. Estimation of preferences relies on rich and plausibly exogenous variation in consideration sets over time, and estimation of price sensitivity relies on price shocks occurring when hospitals subscribed to a benchmarking database, as documented in Grennan and Swanson (2019). Relative bargaining abilities for each brand-hospital pair are identified by the extent to which a brand's price changes as the added value of the brand to the hospital changes.

Finally, we use observed consideration sets and demand and supply parameter estimates

[7]See, e.g., Grennan (2013) regarding medical devices, and also Gowrisankaran et al. (2015) regarding insurer-hospital negotiations, for cases where inelastic demand from end-users would imply negative marginal costs when prices are modeled as the outcome of a Bertrand pricing game, but the estimates from a bargaining model imply more reasonable marginal costs.

to infer search/contracting costs. Search costs rationalize which brands are included in vs. excluded from hospitals' consideration sets, given supply and demand parameters. We do not observe buyer search in the way that online studies observe it (e.g., Dinerstein et al. (2018) observe actual browsing data from eBay); however, we do leverage rich panel variation in demand realizations and consideration sets over time, which provides similar identifying information. Our moment inequalities approach identifies search costs using simple conditions: e.g., that if a brand is in the consideration set, it must have been optimal to search *at some point* (when brands are substitutes, the highest possible added value is relative to the outside good); and that if a brand is not in the consideration set, it must not have been worth searching (again when brands are substitutes, the lowest possible added value is relative to the full consideration set).[8]

The demand and bargaining model estimates illustrate several interesting and intuitive patterns. First, while demand for all products is fairly price-insensitive, demand for physician preference items is more than an order of magnitude less price-sensitive than demand for more commoditized, non-PPI medical/surgical products. Second, supplier firms tend to capture a greater portion of the total surplus generated by hospital supply negotiations for PPIs than for other products. These patterns are consistent with the widely-held belief that PPIs are overpriced due to misalignment between the objectives of hospitals (i.e., the purchasers) and physicians (i.e., the users) (Robinson 2008).

With the estimated demand and bargaining models, we are able to make progress in decomposing the price variation across hospitals, and also understanding the sources of market power that make this variation possible. We do this by computing a series of counterfactual equilibria where we alternately shut down heterogeneity across hospitals in demand and bargaining. Holding choice sets fixed as observed in the data, we estimate that price variation across hospitals is driven primarily by bargaining (especially in commodities); but demand heterogeneity also accounts for substantial price variation. This is important for several reasons. First, while research suggests that hospitals have made substantial improvements in purchasing of commodities since the introduction of the prospective payment system, these results indicate that there is still considerable heterogeneity driven by relative bargaining power. This bargaining power variation could in turn capture heterogeneity in management

[8]This approach is complementary to recent work on bargaining with strategic exclusion, in which buyers' ability to replace contracted suppliers with substitutes outside the consideration set enables them to elicit greater price discounts (see e.g., Ho and Lee (2018)). Strategic exclusion may play a role in our setting, but a model with both search and strategic exclusion is computationally infeasible, given the large sets of potential suppliers in our setting. Our interpretation of qualitative interviews with purchasing executives and quantitative evidence in the data (e.g., the brands excluded from hospitals' consideration sets are difficult to rationalize using strategic exclusion alone) is that search/contracting costs are a dominant factor that limits consideration set sizes in the hospital supply setting, and thus we choose to focus on them in our analysis.

(Bloom et al. 2014), information (Grennan and Swanson 2019), or something else (Lewis and Pflum 2015). Second, the results confirm that preference heterogeneity generates significant market power for PPIs – the welfare impact of this preference heterogeneity will depend on whether it is driven by true product differentiation in quality vs. physician-specific brand preferences generated by marketing.

The search cost estimates suggest that contracting frictions are on the order of 5 percent of price, on average. In order to better understand the impact of these costs on the market, we consider counterfactuals where we lower search costs by one half, allow hospitals to optimally add additional brands to their choice sets, and compute equilibrium prices and quantities for these new choice sets. This sheds light on the role of search/contracting costs in generating markups (see, e.g., Hortacsu and Syverson (2004); Allen et al. (2018); Galenianos and Gavazza (2018)), and the interactions of search with demand and bargaining. As choice set sizes increase, average prices decrease, but relatively little. The largest gains from lowering contracting frictions appear to be the expected consumer surplus gains from increasing choice, not increasing price competition.

This paper contributes to the growing literature on empirical models of negotiated price markets based on "Nash-in-Nash" bargaining (Crawford and Yurukoglu 2012; Grennan 2013, 2014; Gowrisankaran et al. 2015; Ho and Lee 2017). Whereas prior studies have typically taken the set of buyer-supplier relationships as given, we consider the role of frictions in the search/contracting process and how it interacts with demand and bargaining to generate markups and price dispersion across buyers. In this aspect, our study is most closely related to Allen et al. (2018)'s study of search and negotiation of mortgage quotes and Salz (2017)'s study of waste management contracts, though our modeling differs substantially due to the different data and institutional contexts. Our approach to search, using moment inequalities based on stability conditions, is also related in spirit to Ghili (2018), but our "weaker" stability conditions are built to be consistent with a wide variety of processes that have been considered the the industrial organization search literature (Sorensen 2003; Hortacsu and Syverson 2004; Hong and Shum 2006; Honka 2014). While limiting the counterfactuals that can be considered, our approach provides one path forward for cases where buyers contract with a set of differentiated substitute suppliers for needs that are realized over time, a common scenario in business-to-business markets.

A final noteworthy contribution of this paper is the large number and variety of product markets we are able to analyze. In this way, our study is a bridge between the literature on price dispersion across markets (Scholten and Smith 2002; Kaplan and Menzio 2015) and the in-depth empirical case studies in bargaining and search above.

2 Data and Background on Hospital Purchasing

Health care in the hospital setting has high fixed capital costs in the form of facilities and equipment, but it also has high variable costs in the form of skilled labor, pharmaceuticals, and consumable supplies such as implantable medical devices. The price dispersion we document in this setting is particularly notable because, in the short run, hospitals are typically reimbursed a fixed amount by private or public insurers for the services they provide. Thus, variations in supply prices have a direct impact on hospitals' bottom lines.[9] In this Section, we provide some background on how consumable medical supplies are used and purchased, and we describe the unique data set and research setting that allow us to analyze the determinants of price dispersion.

2.1 Hospital Purchase Order Data

The primary data used in this study come from a unique database of consumable supply purchases made by a large number of US hospitals during the period 2009-2015. The data are from the PriceGuide™ benchmarking service (hereafter, "PriceGuide data") offered by the ECRI Institute, a non-profit health care research organization. For each transaction, we observe price, quantity, transaction month, and supplier for a wide range of product categories.

The reported price and quantity data are of high quality because they are typically transmitted as a direct extract from a hospital's materials management database. The PriceGuide™ benchmarking service compares each hospital's submitted data to that of others in the database and generates several analyses of the hospital's savings opportunities; thus, the hospitals have strong incentives to report prices accurately. Related to its materials management origins, the data is at the stock-keeping-unit (SKU) level, requiring us to use machine learning algorithms to group SKUs that belong to the same manufacturer-brand.[10] For stents and surgical staplers, we also validate our algorithms against data collected from manufacturer catalogs and find that our machine learning algorithm performs well in identifying brands. See Appendix A for details.

[9] The supplies in our database comprise 24 percent of hospital operating costs (Craig et al. 2018).

[10] The goal of the machine learning procedure is to identify the level of product at which hospital-supplier contracts are negotiated. E.g., for stents, prices are negotiated separately for each brand, and each brand subsumes a large number of SKUs. We use the RE-EM tree package to flexibly group SKUs (defined by a set of dummies for all potential alphanumeric characters in each SKU position) into brands based on observed price variation within each manufacturer-vendor combination. In Appendix C, we present results for alternative aggregations of SKUs and find our results qualitatively unchanged.

2.1.1 Representativeness of the benchmarking database sample

The hospitals in the purchase order data voluntarily joined a subscription service that allows them to benchmark their own prices and quantities to those of other hospitals in the database and thus may not be a random sample of US hospitals. In particular, subscription is costly, so we expect hospitals with greater concerns about supply costs to be overrepresented in the database. In a survey of database members, "cost reduction on PPIs" and "cost reduction on commodities" were the first and second (and nearly tied) most commonly cited reasons for joining. As discussed in detail in Grennan and Swanson (2019), we do not find evidence that hospitals differentially join the database during times where prices are trending up or down relative to market price trends. However, we do note that the PriceGuide members are overrepresented in the American west and underrepresented in the south, and that they are larger than the average US hospital.

2.2 Product Categories

Each transaction in our data includes a product category identifier from the ECRI Institute's Universal Medical Device Nomenclature System (UMDNS).[11] We chose the products for this analysis as those in the top 100 UMDNS codes by total spending; we then removed products that were overly broad (e.g., "office supplies"), those that had missing or inconsistent quantity data, those whose files were too large for us to estimate supply and demand with a reasonable bound on computing resources, and those for which the identification strategy we use to address endogenous consideration set formation is not sufficiently powerful.[12] After applying these filters, 19 top product categories remain. In many of our analyses, we separately consider product categories by class according to the Food and Drug Administration's classification system. Specifically, we distinguish FDA risk class III from FDA risk classes I-II; for brevity of notation, we refer to class III supplies as physician preference items (PPIs) and to classes I-II as non-PPIs.[13]

[11] A note on terminology: throughout this draft, we use "product category" to refer to the UMDNS grouping included in the transaction files. The UMDNS system is employed to classify any device or supply based on its intended purpose, with some distinctions for mechanism of action. It covers all medical devices and supplies, clinical laboratory equipment and reagents, and selected hospital furniture, among other items. For example, drug-eluting coronary stents have UMDNS code 20383. For a finer level of distinction, we use the term "brand" to refer to the "product" level at which prices are negotiated – e.g., Medtronic Resolute Integrity drug-eluting coronary stent. The use of "brand" is not meant to connote any particular marketing strategy. Finally, for a coarser level of distinction, we use "product class" to refer to broad groupings of product categories: physician preference items (PPIs) such as drug-eluting coronary stents, and non-PPIs such as surgical gloves.

[12] See Appendix A for the lists of product categories removed at each step.

[13] Class I devices, such as gloves, are deemed to be low risk and are therefore subject to the least regulatory controls. Class II devices, such as catheters, are higher risk devices with greater regulatory controls to provide

The top panel of Table 1 summarizes our data for non-PPIs; product categories are listed in decreasing order of average yearly spend.[14] Some of these product categories are roughly commodities, such as surgical gloves: products that can be used in a hospital setting by staff members with a variety of roles and scopes of practice. Conditional on a few characteristics, such as material, we do not expect particular manufacturers or brands to be strongly preferred.[15] Non-PPIs also include other medical and surgical items that may be important inputs in moderately invasive procedures, and that may or may not be associated with strong brand preferences. For example, electrophysiologists may have strong preferences over the ablation/mapping catheters they employ in ablation procedures. Non-PPIs are a quite heterogeneous category: as shown in Table 1, these product categories vary in popularity, and spending per hospital-year varies from \$19 thousand to \$457 thousand. Similarly, price per unit varies from less than a dollar (isolation gowns) to \$2,691 (bone grafts). Non-PPIs are purchased by 494 sample hospitals on average.[16] The variation in negotiated price across hospitals depends on the UMDNS code under consideration: for surgical gloves, the coefficient of variation across hospitals, within brand-time, is only 0.07, while for trocars, the CV is 0.21. There are many brands to choose from, but hospitals may have limited awareness of the total set of brands (and corresponding prices) available: the average hospital only purchases $|\mathcal{J}_h| = 7$ of the 132 unique brands available for the average non-PPI UMDNS code in a given year. Even more notably, it is usually the case that the most popular brand j^* is not purchased by the majority of hospitals: on average, the probability that j^* is in hospital h's consideration set $\mathcal{J}_h$ is only 0.38. For non-PPI product categories, particularly the more commodity-like categories, this is unlikely to be driven by provider preference heterogeneity; it seems much more likely that supply factors, such as contracting frictions, drive this

reasonable assurance of the devices' safety and effectiveness. Class III devices, such as replacement heart valves and coronary stents, are the highest risk devices and must typically be approved by FDA before they are marketed.

[14]Table 1 summarizes data for final analytic sample. See Appendix A.1 for details regarding data cleaning.

[15]We excluded many commodities from our set of focal product categories due to inconsistencies across hospitals in how quantities are reported. Exam gloves, one of the most popular product categories in our data, are excluded for this reason.

[16]The full dataset contains 1,228 hospitals, but we restrict analysis in each product category to hospitals purchasing the given product category in significant volumes, those for which we observe the date they joined the benchmarking service, and those we were able to match to external hospital characteristics. To perform the analysis in the current study, we obtained permission to contract a trusted third-party to match facilities in the PriceGuide data to outside data on hospital characteristics from the American Hospital Association (AHA) annual surveys. The third-party then provided us with access to the merged data for analysis, with hospital-identifiable information removed. See Appendix A.1 for details.

variation.[17]

PPIs are summarized in the bottom panel of Table 1. We examine four important PPIs: drug-eluting coronary stents, acetabular hip prostheses, humeral shoulder prostheses, and patellar knee prostheses. For physician preference items, usage is driven by strong brand preferences of physicians, often surgeons, choosing which product to use to treat a given patient. These strong PPI brand preferences are frequently noted in policy discussions. PPIs tend to be expensive, high-tech products used in specific, advanced procedures – often cardiac and orthopedic procedures – and are thus not necessarily purchased by all hospitals: only 414 sample hospitals purchased the average PPI. At the extreme, only 321 sample hospitals purchased humeral shoulder prostheses. The average hospital purchasing PPIs spends \$398 thousand per year on each PPI UMDNS code. The coefficients of variation for PPIs are in a similar range as those observed for non-PPIs, but given how expensive PPIs are (\$1,303 per unit on average), the dollar values of the price variations observed across hospitals are more extreme. These product categories can often only be purchased directly from one of a few manufacturers, and manufacturers exert substantial effort in marketing their brands to hospitals with relevant patients. It is thus not surprising that we observe relatively more comprehensive consideration sets: the average hospital's consideration set contains $|\mathcal{J}_h| = 11$ brands, of the $|\mathcal{J}| = 79$ available, and the modal brand is in most hospitals' consideration sets ($\overline{\Pr[j^* \in \mathcal{J}_h]} = 0.66$).

2.3 Institutions: Hospital Contracting Environment

Hospitals purchase thousands of product categories, and prices for most of these product categories are determined in negotiation. In determining which brands to contract, purchasers within a hospital must factor in clinical value, safety, cost, and other conditions of sale and service (e.g., for capital equipment, which is not in our data, maintenance is an important consideration). Negotiation can take place directly between a hospital administrator and a representative of the brand's manufacturer, or hospitals may rely on group purchasing organizations or other contracting coalitions to negotiate their contracts. GPO prices are often used as a starting point for direct hospital-manufacturer negotiations for physician preference items and capital equipment (Schneller 2009).[18]

[17] In Appendix D, we ultimately use our detailed demand and costs estimates to explore the extent to which these choice sets could be rationalized by a model of strategic exclusion with no search/contracting frictions, and we find that hospital-brand specific preferences would have to be extremely negative, relative to our estimated distributions, across *all* brands in $\mathcal{J} \setminus \mathcal{J}_h$, to rationalize such a model.

[18] A GAO report from 2003 noted that there are hundreds of GPOs, some of which operate regionally; however, at the time of the study, seven national GPOs with purchasing volumes over \$1 billion accounted for more than 85 percent of all hospital purchases nationwide made through GPO contracts.

Table 1: Summary of Purchasing Categories

	N_h	Annual Spend \$1000s	p		$\lvert\mathcal{J}\rvert$	$\lvert\mathcal{J}_h\rvert$		$Pr[j^* \in \mathcal{J}_h]$	$Pr[j^* = j_h^*]$
			μ	$\frac{\sigma}{\mu}$		μ	$\frac{\sigma}{\mu}$		
Other Medical/Surgical Supplies (Non-PPIs)									
Linen Underpads	602	$32	$0.30	0.08	84	2	0.58	0.16	0.11
Isolation Gowns	501	$19	$0.45	0.08	56	2	0.55	0.19	0.13
Surgical Gloves	758	$86	$0.86	0.07	220	7	0.62	0.25	0.09
Pulse Oximeter Probes	366	$146	$10	0.18	101	5	0.59	0.15	0.10
Liquid Adhesives	696	$59	$17	0.12	70	2	0.54	0.35	0.24
Pneumatic Compression Cuffs	351	$102	$19	0.11	55	3	0.58	0.30	0.26
Trocars	669	$64	$35	0.21	218	7	0.53	0.35	0.12
GI Staples	609	$125	$133	0.18	172	6	0.56	0.30	0.16
Linear Staplers	583	$84	$142	0.14	134	6	0.58	0.19	0.10
Orthopedic Fixation Systems	441	$127	$396	0.20	134	17	0.58	0.76	0.36
Hemostatic Media	290	$120	$447	0.06	19	3	0.45	0.64	0.34
Electrosurgical Forceps	453	$168	$523	0.20	88	8	0.56	0.64	0.40
Ablation/Mapping Cath.	324	$423	$880	0.14	131	18	0.48	0.49	0.20
Allografts	369	$226	$956	0.14	364	15	0.59	0.24	0.07
Bone Grafts	393	$457	$2,691	0.13	128	9	0.48	0.72	0.20
Average(15)	494	$149	$417	0.14	132	7	0.55	0.38	0.19
Physician Preference Items (PPIs)									
Patellar Knee Prosth.	470	$100	$414	0.24	31	5	0.54	0.61	0.23
Acetabular Hip Prosth.	516	$276	$1,152	0.23	157	18	0.59	0.75	0.34
Drug Eluting Stents	351	$995	$1,471	0.06	10	4	0.37	0.84	0.40
Humeral Shoulder Prosth.	321	$222	$2,173	0.21	116	16	0.42	0.45	0.12
Average(4)	414	$398	$1,303	0.18	79	11	0.48	0.66	0.27

Providers – who are the end users who ultimately decide when and what brands to use, conditional on patient needs – can also play a role in choosing which suppliers are contracted with, particularly for medical devices like PPIs. For this reason, suppliers and their representatives work closely with physicians and hospital support staff in order to promote their brands, provide training for new brands, and even provide on-site technical assistance in the operating suite (Montgomery and Schneller 2007). This implies that suppliers' representatives have highly specialized knowledge about their brands and physician users. As noted above, hospitals typically purchase a small share of the brands available in a given product category. This may be due to heterogeneity in preferences: we would expect a hospital to be more likely to contract a given brand and thus include it in its consideration set for use in procedures if it is particularly preferred by the hospital's affiliated physicians. This phenomenon may also be driven by factors that vary across hospital-vendor pairs, such as distribution costs and search/contracting frictions.

In our demand analysis, we use this logic to motivate an identification strategy based on factors that impact search/contracting frictions, but not end-user preferences, across multiple dissimilar product categories. For example, a supply chain administrator who has found Medtronic easy to work with in procuring coronary stents may also be more likely to consider Medtronic as a supplier of tracheal tubes, despite the fact that these product categories have different staff users, therapeutic indications, and levels of technological sophistication. We argue that, conditional on appropriate controls, search/contracting frictions such as these are likely to generate correlations across dissimilar product categories in terms of which brands enter the *consideration set* (i.e., which brands are in the storeroom), but that the siloed and specialized nature of medical device sales and training implies that *user* (i.e., physician) preferences over particular brands are likely uncorrelated across sufficiently dissimilar categories, conditional on consideration sets.[19]

We define "sufficiently dissimilar categories" by starting with the UMDNS system, which imposes a hierarchy over product categories based on their intended use. For example, Figure 1 below displays part of this hierarchy as a tree structure with up to seven splits leading to coronary drug-eluting stents. Coronary drug-eluting stents (in pink) are directly under the parent code for coronary balloon-expandable stents, less directly under the parent code for non-active implantable devices. Polymeric mesh and bone grafts (in green) are also under the parent code for non-active implantable devices, while cardiac pacemakers (in yellow) are instead under the parent code for active implantable devices. The most distant product category from coronary drug-eluting stents in this example is tracheal tubes (in yellow), which are not under the parent code for implants and prostheses. In our empirical analysis, we will assume that, conditional on having contracts for both Medtronic tracheal tubes and Medtronic stents, users in those product categories' verticals do not have correlated preferences over Medtronic brands at the point of usage.[20]

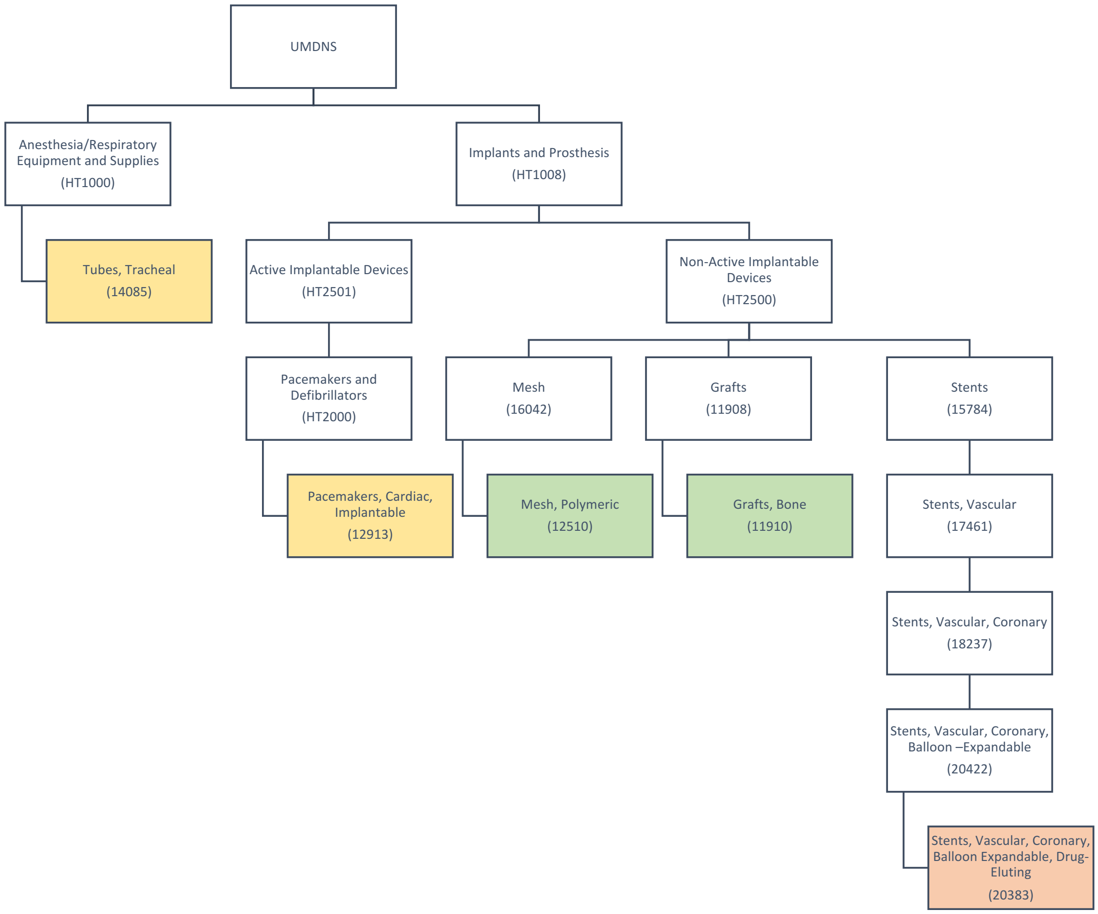
We use the above hierarchy to flag each pair of UMDNS codes (among the top 200 in overall spending) as likely similar ("near") or dissimilar ("far"). A "near" pair is one with more than one shared parent in the UMDNS tree structure. In the above example, tracheal tubes and cardiac pacemakers are classified as "far" from coronary drug-eluting stents, while polymeric mesh and bone grafts are classified as "near." Starting with this set of candidate "far" pairs, we then identified each pair with significant vendor overlap – UMDNS code A and UMDNS code B have "significant vendor overlap" if at least 70 percent of spending in A is contributed by vendors that sell products in B. Finally, among the pairs with significant vendor overlap, we checked whether each product category would typically be used in the same or similar procedures; for example, balloon catheters and coronary stents are used in the same procedures; cardiac valve prostheses and annuloplasty rings are used in different procedures for heart valve abnormalities.[21] Our final list of dissimilar pairs of UMDNS codes

[19]Our exclusion restriction is discussed more precisely in Section 3.1.1 below.

[20]In the remainder of the paper, we use the term "vertical" to denote unique purchasing entities within hospitals.

[21]To perform the check for whether a given pair of product categories is used in the same or similar procedures, we engaged a nursing student to create a database of common procedure codes and descriptions

Figure 1: UMDNS Code Hierarchy



are those that are "far" according to the UMDNS tree hierarchy, and that either *do not* have significant vendor overlap, or *do* have significant vendor overlap but are not used in the same or similar procedures.

In the identification strategy laid out in Section 3.1.1 below, we use exposure of a given hospital-vendor pair across dissimilar product categories by this definition to address a form of selection bias. Selection bias is potentially introduced by the fact that we only observe demand realizations for the brands in a given hospital's contracted consideration set, which are likely those that are particularly preferred by the hospital's physicians.

for each product category. She then cross-referenced the lists of codes and descriptions across product categories in each pair with significant vendor overlap.

2.3.1 Observed determinants of price

In this business-to-business bargaining setting, we expect prices, quantities, and the size of consideration sets in a given product category to be a function of users' preferences, relative bargaining power, search and contracting frictions, and the gradient between consideration set size and price. In this Section, we provide preliminary evidence that a large dispersion in price across hospitals persists, even after taking into account each of the features that conventional wisdom predicts would have large effects: 1) purchaser size, measured by hospital (and affiliated system) bed size and purchase quantity; 2) strategic exclusion in the form of standardization; and 3) contracting the services of a group purchasing organization. The basic framework for this analysis is the following equation:

$$ln(p_{ujht}) = \theta_{ujt} + X_{ht} * \beta + \epsilon_{ujht}$$

where the dependent variable $ln(p_{jht})$ is the price that hospital h pays for brand j of UMDNS code u in year t; X_{ht} are characteristics of hospital h in year t; ϵ_{ujht} are residuals. In the main text, we report and discuss the pooled coefficients $\beta^{Non-PPI}$ and β^{PPI} from a regression of all observations in the Non-PPI and PPI classes, weighting each UMDNS code-specific observation by that UMDNS code's aggregate yearly spending across all included hospitals in the analytic sample. All continuous covariates in X_{ht} are standardized within each product category, and ujt fixed effects are included in all specifications. Thus, a coefficient of $\hat{\beta}^c_k = 0.1$ would suggest that, on average, a standard deviation increase in X_{htk} leads to a 10 percent increase in price for the categories in class c.

The left panel of Figure 2 shows the pooled effects of various purchaser characteristics on prices in the Non-PPI class. The analogous results for PPIs are in the right panel.[22]

First, consider the role of purchaser size. Conventional wisdom holds that large buyers will obtain better prices, and hospital leaders use this logic to argue before antitrust authorities that horizontal mergers will reduce costs (Craig et al. 2018; Noether and May 2017). That said, a large theoretical literature notes that the effect of size on prices depends on several factors, including the supplier competition and the curvature of the bargaining-surplus function.[23] The top four points in each panel of Figure 2 show the correlations of price with several measures of purchaser size – hospital beds, total beds in hospital's affiliated system, total units purchased in the hospital-year (q_{ht}), and total units purchased in the system-year ($\sum_{h' \in s(h)} q_{h't}$). For non-PPIs, each estimate indicates that larger purchasers obtain slightly

[22]In each panel, point estimates and 95% confidence intervals are plotted, with standard errors clustered by UMDNS code-hospital. Appendix Table A3 contains the same results for each individual product category.

[23]See, e.g., Horn and Wolinsky (1988); Stole and Zwiebel (1996); Chipty and Snyder (1999); Inderst and Wey (2007); Snyder (1996, 1998); Dana (2012); Gans and King (2002); Marvel and Yang (2008).

lower prices – the largest effect indicates that a standard deviation increase in total purchasing at the system level is associated with a 0.8 percent decrease in prices for non-PPI product categories. We see somewhat different patterns for PPIs – hospital and system bed size have small *positive* effects on PPI prices, whereas a standard deviation increase in hospital (system) purchase volume is associated with a 3.2 (3.7) percent decrease in PPI prices.[24] In sum, purchaser size is associated with a small, but significant, decrease in prices, which is larger for PPIs than for non-PPIs.

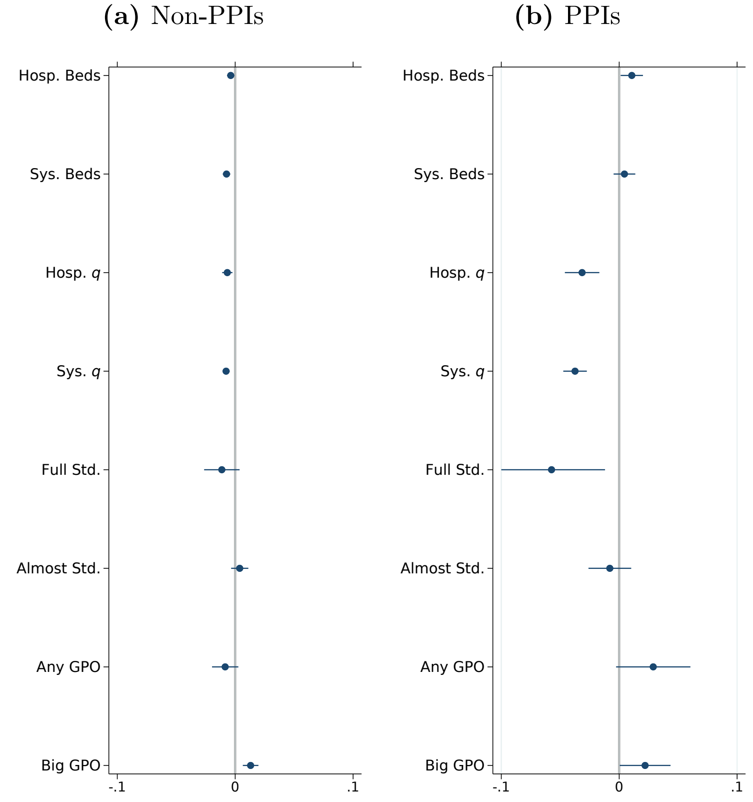
Next, we examine the effect of consideration set size on prices. The classic "Nash-in-Nash" model used in empirical bargaining studies generates a clear prediction that, if brands are substitutes, the addition of a brand to the set a hospital contracts with ($\mathcal{J}_h$) will weakly decrease the negotiated price of the other brands in the set, as an additional substitute reduces the gains from trade for inframarginal brands. A more nuanced set of predictions can be obtained from models in the recent empirical bargaining literature that allow firms to strategically employ small consideration sets in order to extract larger discounts from included vendors (e.g., Ho and Lee 2018; Ghili 2018; Liebman 2018). This practice is known as "standardization" in the hospital industry, and is thought to be a useful source of savings (Noether and May 2017). Models with this type of exclusion, such as the Nash-in-Nash with Threat of Replacement model proposed by Ho and Lee (2018), may predict that, for a given set of potential suppliers $\mathcal{J}^{potential}$, prices will be *increasing* in the size of the set the hospital contracts with $\mathcal{J}_h$ (at the margin, if exclusion is optimally applied).[25] There is compelling empirical evidence from other contexts that restrictive networks of health care providers (Ho and Lee 2018; Gruber and McKnight 2016; Sorensen 2003), restrictive drug formularies (Duggan and Scott Morton 2010), and restrictive pharmacy networks (Starc and Swanson 2018) can lead to lower costs for insurers. However, hard evidence on the ubiquity or usefulness of standardization in supply procurement is scarce, and Grennan and Swanson (2019) found no evidence of either for coronary stents. The next two points in each panel of Figure 2 show the association between price and two measures of standardization – a dummy for the hospital-category-year being fully standardized with one vendor, and a dummy for being "almost" standardized with 80 percent share going to one vendor.[26] As shown in Appendix Table A4, standardization is quite rare in our data: 5 (1) percent of non-PPIs (PPIs) are

[24]This is consistent with the finding in Grennan and Swanson (2019) that there is a steeper gradient in coronary stent prices with respect to purchase volume than there is with respect to hospital bed size; indeed, small specialty hospitals may purchase PPIs in greater quantities than similarly-sized acute care hospitals.

[25]In related theoretical work, Dana (2012) posits that buyer groups' primary advantage results from their commitment to purchase from a single supplier in differentiated product markets.

[26]Standardization on a single vendor is clearly a more extreme form of exclusion than the "narrow networks" contemplated in the insurer-hospital bargaining literature (Ho and Lee 2018; Ghili 2018; Liebman 2018); however, it is widely advocated as a source of supply chain savings to hospitals.

Figure 2: Correlation between p_{jht} and X_{ht}



	ujtFE	+Beds Vars	$+q$ Vars	+Std Vars	+GPO Vars	$+h$FE	$+uh$FE
Non PPIs							
R^2	0	0	0	0	0	.0002	.0012
PERCENTILES							
10TH	2.97	2.97	2.97	2.97	2.97	2.96	2.99
25TH	4.91	4.91	4.91	4.91	4.91	4.92	4.93
50TH	6.03	6.03	6.03	6.03	6.03	6.01	5.98
75TH	6.74	6.74	6.74	6.74	6.74	6.74	6.74
90TH	7.48	7.48	7.48	7.48	7.48	7.46	7.46
PPIs							
R^2	0	.0008	.0109	.0109	.0118	.0511	.0807
PERCENTILES							
10TH	6.77	6.77	6.78	6.78	6.78	6.82	6.83
25TH	7.02	7.02	7.02	7.02	7.02	7.04	7.04
50TH	7.25	7.25	7.25	7.25	7.25	7.23	7.25
75TH	7.39	7.39	7.39	7.39	7.39	7.38	7.350
90TH	7.60	7.60	7.60	7.60	7.60	7.58	7.57

Notes: Selected percentiles of the distribution of residual (log) prices after controlling for brand-year fixed effects and various covariate sets X_{ht}; R^2 of the regression for that covariate set (net of the R^2 for the specification with only brand-year fixed effects) also reported. Column (1) controls for brand-year effects only; column (2) adds hospital and system bed size to the specification; column (3) adds hospital and system purchase volume; column (4) adds standardization measures; column (5) adds GPO variables; column (6) replaces all of the previous hospital covariates with hospital fixed effects; and the final column (7) includes hospital-category fixed effects.

fully standardized, while 19 (14) percent of non-PPIs (PPIs) are "almost" standardized.[27]

[27] There is dramatic variation across product categories in the number of available brands and vendors; accordingly, some categories have much higher standardization rates than others. Isolation gowns have

For non-PPIs, there is no statistically significant association between standardization and price; for PPIs, full standardization is associated with a statistically significant 5.7 percent decrease in price. In Appendix Table A3, we see that this is entirely driven by a large negative association between price and full standardization for patellar knee prostheses. On balance, we find limited evidence that restrictive choice sets are associated with lower prices; associations are negative and significant in two categories (patellar knees and trocars) and positive and significant in two categories (allografts and ablation/mapping catheters).

Finally, the results on hospital and system size may obscure other opportunities for hospitals to bargain collectively. Conditional on hospital/system size, hospitals may vary in their use of group purchasing organization services or their membership in other purchasing coalitions. We have one proxy for this behavior in the form of the GPO membership variable reported in the AHA surveys. As shown in Appendix Table A4, about 94 percent of sample hospitals report some GPO membership. It is unclear whether the remaining 6 percent are truly not using GPO services, or whether this is simply measurement error. Given this ambiguity, we focus on two variables – a dummy for any GPO reported, and a dummy for reporting use of one of the top GPOs (MedAssets, Novation, Premier, UHC, VHA, and Vizient).[28] Reporting any GPO membership is associated with an insignificant -0.8 percent decrease in price for non-PPIs, an insignificant 2.9 percent increase in price for PPIs. However, conditional on reporting any GPO membership, membership in a large GPO is associated with 1-2 percent *higher* prices for both non-PPIs and PPIs.[29] These results reinforce our finding of limited benefits of purchaser size, contrary to conventional wisdom.

To put these results in perspective, the table at the bottom of Figure 2 reports selected percentiles of the distribution of residual (log) prices after controlling for brand-year fixed effects and various covariate sets X_{ht}; it also reports the R^2 of the regression for that covariate set (net of the R^2 for the specification with only brand-year fixed effects). Briefly, although we document above that there are several factors that have significant effects on medical supply prices, they have little explanatory power. The interquartile and interdecile ranges of the residuals are remarkably stable across specifications (1)-(5). Similarly, there is no detectable effect of hospital covariates on the R^2 of price for non-PPIs (precision shown

our highest standardization rate among non-PPIs, at 52 percent full standardization and 83 percent almost standardization. Within PPIs, patellar knees are 7 percent fully standardized and 30 percent almost standardized.

[28]Prior to 2015, VHA and UHC jointly owned Novation; in 2015, they merged and rebranded under the name Vizient.

[29]For individual product categories, Appendix Table A3 shows evidence of statistically significantly higher prices for large GPO members in liquid adhesives, isolation gowns, ablation/mapping catheters, and patellar knees; and of statistically significantly lower prices for surgical gloves and linen underpads.

to the fourth decimal point); all hospital covariates increase R^2 by only 0.012 for PPIs.[30] Moving from hospital covariates to hospital fixed effects decreases the residual dispersion slightly, and increases the (net) adjusted R^2 to 0.0002 for non-PPIs, 0.0511 for PPIs. We see the most dramatic effects when we add hospital-category fixed effects: comparing column (7) to column (5), the interquartile range decreases by 1 percent for non-PPIs and by 16 percent for PPIs. These patterns support the (somewhat obvious) notion that there are many unobserved factors that vary across hospitals and that impact prices – e.g., physician training, relationships with the device industry, practice style, and management, to name just a few. It is somewhat more striking that the variation within hospital across product category is approximately as important as the variation across hospitals; that is, given the siloed nature of hospital purchasing, those unobserved factors do not bear out equally across purchasing entities within the hospital.

In sum, we document that hospital characteristics account for little of the observed variation in prices of medical devices; moreover, hospitals' success in bargaining varies both across product categories and across brands within product categories. We propose that a model that incorporates heterogeneity in preferences and bargaining power across hospital-brand pairs will do a better job of explaining the variation in our data given observed consideration sets; we then go further to explore the roles of search and contracting frictions. For the sake of expediency, we rely on the predictions of Nash-in-Nash bargaining to estimate supply and demand parameters, as detailed in the following Section. However, we note that this model will not incorporate any potential benefits of strategic exclusion (Ho and Lee 2018).[31] We consider the role of standardization to be an important topic for further research.

3 Model

We model hospital consideration set formation and demand decisions in a framework with negotiated prices. We assume hospital demand is derived from the preferences of its staff and the needs of its patient population. First, those involved in hospital purchasing add brands to the consideration set given their beliefs about brand quality and bargaining parameters, and

[30] Note that non-PPIs are a very heterogeneous category, so the model with (UMDNS-specific) brand-year fixed effects has an adjusted R^2 of 0.99. The analogous R^2 is only 0.7 for the smaller set of less heterogeneous PPIs.

[31] In Appendix D, we more directly consider the consideration set stability predictions of the Nash-in-Nash with Threat of Replacement model, using our estimated supply and demand parameters and alternative assumptions about hospitals' quality expectations and the level of product at which search takes place. This exercise provides evidence that it would be difficult to rationalize small observed consideration sets solely by strategic exclusion.

given the search and contracting frictions associated with adding brands to the consideration set. Upon termination of search, the hospital and vendor negotiate prices for each brand in the (endogenous) consideration set. Finally, conditional on the consideration set and negotiated prices, the hospital purchases a brand for each use case from that set according to its demand function. The timing and notation are as follows:

1. Hospital h has ex ante beliefs regarding brand j and time t defined by (all parameters known unless specified):

   Preferences $\theta_h + \theta_{jt} + \theta_{v(j),hrr(h)} + \xi^o_{jht} + \xi^u_{jht}$ (unknown ξ^u_{jht})

   Costs of production/distribution $C_j(q_{jh}; \gamma)$

   Bargaining $\frac{\beta_{jt}}{\beta_h}\eta_{jht}$ (unknown η_{jht})

   Search/contracting costs $sc_{jht} = X^{sc}_{jht}\psi + \nu_{jht}$

2. Hospital consideration set $\mathcal{J}_{ht}$ determined and $\{\xi^u_{jht}\}, \{\eta_{jht}\}$ learned via search/contracting process.

3. Contract prices $p_{jht}(mc_j, \mathcal{J}_{ht}, \theta_{jht}, \beta_{jht})$ determined.

4. Quantities $q_{jht}(\mathcal{J}_{ht}, p_{jht}, \theta_{jht})$ demanded.

Below, we describe each step of the search-bargaining-demand model, in reverse order of model timing.

3.1 Demand model

The utility of brand $j \in \mathcal{J} = \{1, ..., J\}$ for use case i (often a doctor/patient combination, for implantable medical devices) at hospital h is

$$u_{ijht} = \delta_{jht} + \varepsilon_{ijht}. \tag{1}$$

The use-specific i.i.d. unobservable $\varepsilon_{ijht} = \epsilon_{ight} + (1 - \lambda_g)\epsilon_{ijht}$ is the random coefficients representation (from Cardell 1997) of the nested logit model where ϵ_{ight} is a random component common to all goods in group g; and ϵ_{ijht} is the standard type I extreme value error term (with scale normalized to one). As a nesting parameter $\lambda_g \in [0, 1]$ approaches 1, there is more substitution among brands within group g than to the outside good and other brands outside g.

The mean utility across use cases is specified as

$$\delta_{jht} = \theta_h + \theta_{jt} + \theta_{v(j),hrr(h)} - \theta^p p_{jht} + \xi^o_{jht} + \xi^u_{jht}, \tag{2}$$

where $\theta_h + \theta_{jt}$ are hospital and brand-time specific dummy variables and their utility weights, $\theta_{v(j),hrr(h)}$ is a dummy for each vendor-region pair and its utility weight (to account for regional variation in vendors' sales and marketing activity), $\theta^p p_{jht}$ is the negotiated price for pair hj and its utility weight, and $\xi^o_{jht} + \xi^u_{jht}$ is a hospital-brand specific unobservable preference heterogeneity term. The non-standard element here is that this model admits the possibility that the hospital observes a nonzero ξ^o_{jht} *before* the consideration set is formed. This induces the selection problem discussed intuitively in the industry detail section. We provide the details of our method for addressing it in the identification and estimation section below.

Given contracts for a set of brands $\mathcal{J}_{ht}$ and flow of choice opportunities M_{ht}, we assume the hospital chooses the brand in the consideration set that maximizes utility for each use opportunity, so that quantities demanded are given by:

$$q_{jht} = M_{ht} Pr[u_{ijht} > u_{ikht}, \forall k \in \mathcal{J}_{ht}] = M_{ht} \frac{e^{\frac{\delta_{jht}}{1-\lambda_g}}}{\sum_{k\in g} e^{\frac{\delta_{kht}}{1-\lambda_g}}} \frac{\left(\sum_{k\in g} e^{\frac{\delta_{kht}}{1-\lambda_g}}\right)^{1-\lambda_g}}{1 + \sum_g \left(\sum_{k\in g} e^{\frac{\delta_{kht}}{1-\lambda_g}}\right)^{1-\lambda_g}} \ , \tag{3}$$

and hospital surplus across all contracted brands is given by:

$$\pi_h(\mathcal{J}_h) = M_{ht} \frac{1}{\theta^p} \ln \left(1 + \sum_g \left(\sum_{k\in g} e^{\frac{\delta_{kht}}{1-\lambda_g}}\right)^{1-\lambda_g}\right) \ . \tag{4}$$

3.1.1 Demand identification and estimation

We follow the procedure in Berry (1994), setting choice probabilities implied by the demand model equal to market shares observed in the data, and inverting the system to yield a linear correspondence between a function of market shares and the mean utility for each brand:

$$\ln(s_{jht}/s_{0ht}) - \lambda_g \ln(s_{jht}/s_{ght}) = \delta_{jht} = \theta_h + \theta_{jt} + \theta_{v(j),hrr(h)} - \theta^p p_{jht} + \xi^o_{jht} + \xi^u_{jht}, \tag{5}$$

leading to the linear estimation problem $\ln(s_{jht}/s_{0ht}) = \lambda_g \ln(s_{jht}/s_{ght}) + \theta_h + \theta_{jt} + \theta_{v(j),hrr(h)} - \theta^p p_{jht} + \xi^o_{jht} + \xi^u_{jht}$.

Estimating this equation faces two well-known challenges in that theory suggests both

$\ln(s_{jht}/s_{ght})$ and p_{jht} are correlated with the unobservable term $\xi^o_{jht} + \xi^u_{jht}$. We take an instrumental variables approach to solving this identification problem. For $\ln(s_{jht}/s_{ght})$, we follow much of the literature (e.g., Berry et al. 1995; Berry and Waldfogel 1999) in using choice set size as instruments, which leverages the fact that more variety will on average affect substitution independent of the individual brand's unobservable. Due to our concern about endogeneity of the choice set construction process, we use predicted choice set size from our first stage regression of choice set composition on controls and instruments to form the instruments $(\widehat{|\mathcal{J}_{ght}|}, \widehat{|\mathcal{J}_{ght}|}^2, \ln(\widehat{|\mathcal{J}_{ght}|}))$.

For p_{jht}, we build on the results in Grennan and Swanson (2019) that show how access to benchmarking information generates a price shock that varies across brand-hospital combinations based on their place in the pre-information price and quantity distributions.[32] Specifically, we instrument for price with the full set of interactions between variables indicating: whether benchmarking information is available for that brand at that hospital at that time $1_{\{post_{jht}\}}$; the hospital-brand's quartile of quantity purchased, relative to all other hospital-brands in the year before information is available (e.g., $1_{\{q_{jht} > p75(q)\}}$); and the hospital-brand's quintile of price relative to all other hospitals purchasing the relevant brand in the year before information is available (e.g., $1_{\{p_{jht} > p80(p)\}}$). We also follow Grennan (2013) in instrumenting for the price of each hospital-brand-year using the price of the same hospital-brand in the previous year.[33]

We also wish to correct for a specific kind of sample selection problem that could be introduced if the consideration set formation process is somehow correlated with demand unobservables. One version of such a problem would be if the hospital observes a nonzero ξ^o_{jht} *before* the consideration set is formed, and the search and contracting processes are directed by this information. In this case, the brands in the consideration set may specifically be brands that are preferred (in expectation) by the hospital. This would tend to introduce a positive bias in the estimated fixed effects $\theta_h + \theta_{jt}$.

We address this problem by introducing a control function for $E[\xi^o_{jht} | j \in \mathcal{J}_{ht}]$ as in Petrin and Train (2009) and Attanasio et al. (2008). Specifically, suppose that the search process can be approximated by the following reduced form:

$$\mathbb{1}(j \in \mathcal{J}_{ht}) := \mathbb{1}(\phi_h + \phi_{jt} + \phi_{v(j),hrr(h)} + Z^s_{jht}\phi^Z + \varepsilon^s_{jht} > 0) \tag{6}$$

where Z^s_{jht} is a set of instruments that impact search but not demand, and ε^s is a shock to

[32]Grennan and Swanson (2019) found evidence that access to benchmarking leads to price decreases for a variety of product categories, particularly for hospitals and brands that involved high prices and large purchase quantities prior to benchmarking access.

[33]In the demand equation, we also control for whether the brand was included in the hospital's consideration set in the previous year.

the search process which may in general be correlated with the demand unobservable. Our approach then takes the following steps: (1) we estimate the reduced form search model as a linear probability model, regressing $\mathbb{1}(j \in \mathcal{J}_{ht})$ on the instruments Z^s_{jht} and controls; and (2) we include the control function $f(\widehat{Pr(j \in \mathcal{J}_{ht})})$ as a regressor in the demand model. In our preferred specification, we set $f(.)$ as a cubic polynomial in $(1 - \widehat{Pr(j \in \mathcal{J}_{ht})})$.[34] This specification of the control function has the intuitive property that it is small when the instruments push a brand into the consideration set with probability one. Under the story of search directed based on knowledge of the unobservable ξ^o_{jht}, this generates a clear prediction that the coefficient on the control function will be zero if no selection is present, and it will be positive and correct the fixed effect distribution downward if there is a positive correlation between ξ^o_{jht} and ε^s.[35]

As discussed in Section 2.3, the unique data on *all* consumable supply purchases across numerous product categories in each hospital, paired with the phenomenon of many vendors supplying product categories in disparate "verticals" across the hospital, provides potential instruments Z^s_{jht} based on the exposure of the hospital to the vendor of brand j in verticals which are unrelated from a demand perspective. The set of instruments we currently employ for this purpose are *vendor-hospital* exposure – the total spend observed for the given hospital on the current brand's vendor *in dissimilar categories* in the same year, relative to *all spend* by that hospital in those dissimilar categories. We define "dissimilar" categories based on the UMDNS code hierarchy previously discussed. Letting $\mathcal{P}_u$ be the set of UMDNS code u's "parents," a category u' is in the "dissimilar" set $\mathcal{D}_u$ if u and u' share at most one parent: $|\mathcal{P}_u \cap \mathcal{P}_{u'}| \leq 1$. In the example hierarchy in Figure 1, the green highlighted categories are "similar" to coronary drug-eluting stents (e.g., bone grafts) and the yellow highlighted categories are "dissimilar" (e.g., tracheal tubes).

Given the dissimilar set $\mathcal{D}_u$, we construct the exposure variable $\tilde{Z}^s_{ujht}$ as:

$$\tilde{Z}^s_{ujht} = \frac{\sum_{u' \in \mathcal{D}_u} E_{u'hv(j)t}}{\sum_{u' \in \mathcal{D}_u} \sum_{v'} E_{u'hv't}}$$

where E_{uvht} is total expenditure by hospital h on vendor v's brands in UMDNS code u at time t. The denominator is the hospital's total purchasing in the dissimilar categories.

To illustrate the power of this identification strategy, Figure 3 shows the results of a

[34] Results are similar with linear and quadratic control functions.

[35] See Appendix B for further discussion and Monte Carlo simulation results on the performance of this estimator.

regression within each top UMDNS code u:

$$\mathbb{1}(j \in \mathcal{J}_{uht}) = \theta_{ujt} + \theta_{uh} + \theta_{v(j),hrr(h)} + \beta_{near} 1_{\{\tilde{Z}^{s,near}_{ujht} > p50(\tilde{Z}^{s,near})\}} + \beta_{far} 1_{\{\tilde{Z}^{s,far}_{ujht} > p50(\tilde{Z}^{s,far})\}}.$$

In this specification, $\tilde{Z}^{s,far}_{ujht}$ is our focal instrument: the exposure of hospital h to brand j's vendor at time t in *dissimilar* categories. We control for θ_{ujt} in order to focus on variation in exposure *within* brand-time. We also control for $\tilde{Z}^{s,near}_{ujht}$: exposure of h to j's vendor in *similar* categories; we argue that exposure to vendors in similar categories is more likely than exposure to vendors in dissimilar categories to be driven by correlated preferences across verticals. For each $\tilde{Z}^s$ variable, we examine the effect of above-median exposure (as a rough proxy for "high" exposure) of the hospital-vendor pair in other categories. Thus, our coefficient of interest β_{far} captures the effect of above-median exposure to a given brand's vendor in dissimilar categories on a hospital's tendency to include that brand in its consideration set. The solid markers show the results of the specification when we control only for brand-time fixed effects. The hollow markers show the results of our preferred specification, in which we also include hospital fixed effects and vendor-HRR fixed effects. The former remove variation driven by a given hospital's overall preference for variety; the latter absorb variation driven by regional sales and marketing activity of specific vendors.

Figure 3 plots the estimated coefficients $\hat{\beta}_{far}$ and corresponding 95 percent confidence intervals, with and without hospital and vendor-HRR fixed effects, for each of our top UMDNS codes.[36] In each panel, the top estimates (blue circles) are for non-PPIs; the bottom (red triangles) for PPIs. In panel (a), we show results for the 70 (of the original 100) top spending product categories that were not deemed overly broad (e.g., "office supplies") and did not have missing or inconsistent quantity data (see Appendix A). In panel (b), we show results for the final 19 product categories in our analytic sample; these categories are required to have a powerful exposure first stage in our preferred specification and must also be sufficiently small as to not run up against a memory constraint in estimation.

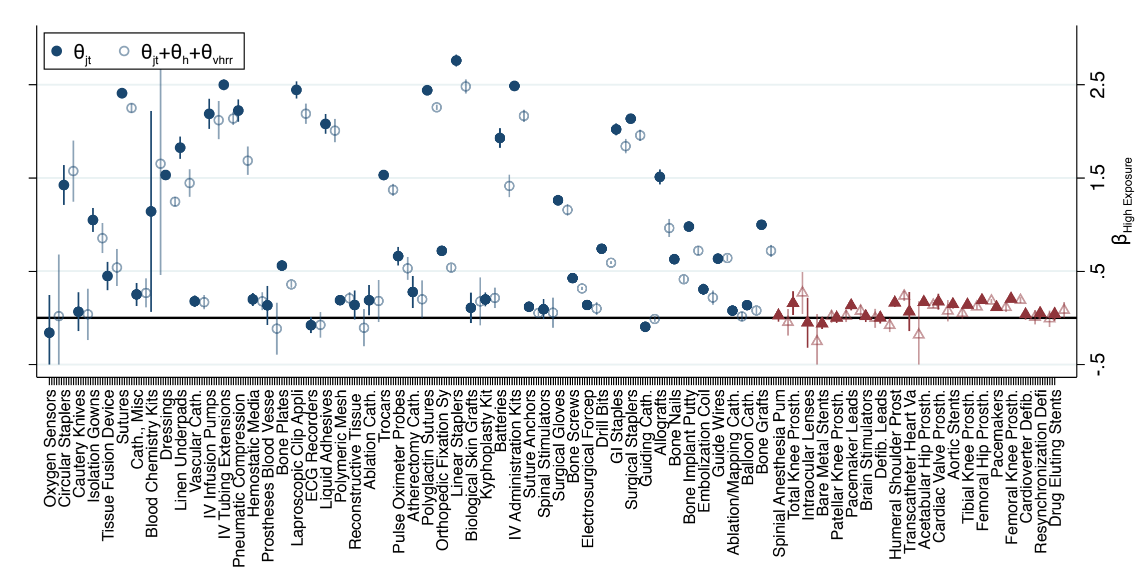
Within both product classes, we see positive estimated effects of high exposure on brand adoption for the vast majority of product categories. In many cases, these effects are also statistically significant within category in the preferred fixed effects specification (hollow markers).[37] The effect of including hospital and HRR-vendor fixed effects is usually to attenuate the estimate of β_{far} slightly. The fact that they are not dramatically different

[36]In order to facilitate comparison across widely varying UMDNS codes with different choice set sizes, we normalize each coefficient by dividing through by the mean probability of consideration set inclusion across hospital-brand-years in the UMDNS code.

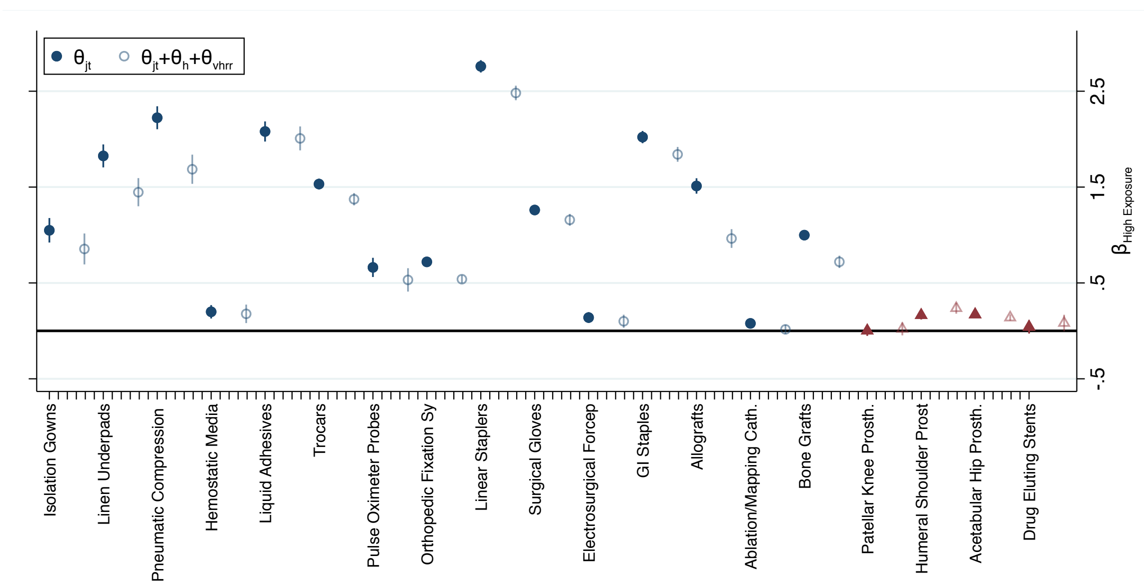
[37]Note that the few product categories with insignificant estimates of $\hat{\beta}_{far}$ in panel (b) do by definition have a powerful first stage in the preferred first stage selection equation described below, which leverages richer variation in $\tilde{Z}^{s,far}$ than shown here.

Figure 3: Reduced Form Evidence of Exposure and Consideration Set Inclusion

(a) All UMDNS Codes



(b) Final UMDNS Codes



suggests that most of our identifying variation is coming from changes over time as cross-market mergers, new product introductions, etc. move around $\tilde{Z}^{s,far}$.

Taken together, these results suggest that exposure to vendors in dissimilar categories has a large positive effect on inclusion of those vendors' brands in the consideration set for the focal category. This identification strategy leads us to the demand model we take to the data:

$$\ln(s_{jht}/s_{0ht}) = \lambda \ln(s_{jht}/s_{ght}) + \theta_h + \theta_{jt} + \theta_{v(j)hrr(h)} - \theta^p p_{jht} + \theta^f f(Z^s_{jht}\hat{\phi^Z}) + \xi^u_{jht}, \quad (7)$$

which is a linear instrumental variables specification, where we instrument for price and the nested logit term as described above, and $\hat{\phi^Z}$ is obtained from a first stage regression of $\mathbb{1}(j \in \mathcal{J}_{ht})$ on excluded instruments and controls. Our preferred specification lets our excluded instruments in the "exposure" first stage Z^s_{jht} be a series of dummy variables for quintiles of $\tilde{Z}^{s,far}$, and we also control for quintiles of $\tilde{Z}^{s,near}$ in both the exposure first stage and in the demand model.[38] Finally, $f(.)$ is a cubic polynomial in $\widehat{Pr(j \in \mathcal{J}_{ht})}$.

3.2 Supply model of negotiated prices

In the business-to-business market for a given hospital supply, the price for a given brand is buyer-specific. We assume that prices are determined between the hospital and the set of vendors with which it contracts as a Nash Equilibrium of simultaneous bilateral Nash Bargaining problems. Each price maximizes the bilateral Nash product, taking other prices as given (t subscripts omitted for brevity):

$$\begin{aligned} p_{hj} &= \arg\max \left(q_{hj}p_{hj} - C_j(q_{hj})\right)^{b_j(h)} \left(\pi_h(\mathcal{J}_h) - \pi_h(\mathcal{J}_h \setminus j)\right)^{b_h(j)} \\ &= \frac{C_j(q_{hj})}{q_{hj}} + \frac{b_j(h)}{b_j(h) + b_h(j)} \left[\left(1 + \frac{\partial q_{hj}}{\partial p_{hj}} \frac{p_{hj} - \frac{C_j(q_{hj})}{q_{hj}}}{q_{hj}} \right) \frac{\pi_h(\mathcal{J}_h) - \pi_h(\mathcal{J}_h \setminus j)}{q_{hj}} + p_{hj} - \frac{C_j(q_{hj})}{q_{hj}} \right] \end{aligned} \quad (8)$$

where $C_j(q_{hj})$ is a function capturing the costs of manufacturing and distributing quantity q_{hj} of brand j to hospital h. The terms $b_j(h)$ and $b_h(j)$ are relative bargaining ability weights that capture the extent to which the optimal price depends on vendor profits vs. the expected additional hospital surplus in the case that a contract is agreed to for brand j: $AV^{CS} \equiv \frac{\pi_h(\mathcal{J}_h) - \pi_h(\mathcal{J}_h \setminus j)}{q_{hj}}$. All the terms in this pricing equation are assumed to be known to all market participants at the time of bargaining.

3.2.1 Supply identification and estimation

We jointly estimate the above (linearized) demand model with the control function correction term and the supply model using a generalized method of moments approach. This enables us to simultaneously recover the demand parameters θ, marginal costs, and mean relative bargaining weights.

We model marginal costs as a fraction of the price charged to the hospital paying the min-

[38] We have also tried linear, quadratic, log, and alternative quantile functions of $\tilde{Z}^{s,far}$. We chose the quintiles specification as having the most powerful first stage.

imum price for each brand. This specification balances the challenges of estimating marginal costs in a market with large markups (Grennan 2013) and absent product characteristics data within our product categories, with the desire to allow for flexible marginal costs across brands within a category that may indeed be quite different from one another. We interpret the estimated parameter as the average minimum margin in each product category.[39,40]

$$mc_j := \gamma \min_{ht} p_{jht}. \tag{9}$$

This specification follows the approach in Grennan (2013, 2014) by assuming no unobservable term in marginal costs, and instead loads the supply side unobservable in the bargaining parameters. We prefer this specification as our prior is that, in our empirical setting, marginal costs are less likely to vary across hospitals and time in unobservable ways than bargaining outcomes. Intuitively, relative bargaining abilities for each brand-hospital pair are identified by the slope with which price changes as the added value of the brand changes, and marginal costs are identified as the intercepts in this relationship as added value approaches zero.

We think of the bargaining parameters themselves as representing characteristics of the hospital and brand that enter the negotiation, but are separate from the economic factors that define total surplus: cost, willingness-to-pay, and disagreement points. We model the relative ratio of the two bargaining parameters by:

$$\frac{b_{jt}(h)}{b_{ht}(j)} := e^{\beta_{jt} - \beta_h - 1_{\{Info_{jh}\}} X^{pq}_{jh} \beta^{Info,pq}_{jh} + \eta_{jht}}. \tag{10}$$

Here β_{jt} and β_h represent brand-year and hospital fixed effects, respectively. The term $1_{\{Info_{jh}\}} X^{pq}_{jh} \beta^{Info,pq}_{jh}$ represents whether the hospital has access to benchmarking information on the brand $1_{\{Info_{jh}\}}$ and where the hospital was in the price and quantity distributions relative to other hospitals prior to having that information X^{pq}_{jh}. Our inclusion of the benchmarking information regressors relates directly to our use of this variation in identifying demand, and is intended to capture the finding from Grennan and Swanson (2019) that benchmarking appears to solve an asymmetric information problem in which hospitals may

[39] In unreported results, we analyze robustness of this assumption. We have tried estimating models with less flexibility across brands, which tend to push marginal costs towards zero and which we believe overstate margins and understate true cost heterogeneity. Models using product characteristics $mc_j = X^{mc}_j \gamma$ seem to work better in the product categories for which they are available, but the potential richness of the relevant characteristics varies widely across categories, and combined with the size of the vendor and product spaces, collecting such data from manufacturer catalogs across all product categories has proven overwhelming. We provide estimates of such a model for selected categories where we have collected such data.

[40] In ongoing work, we are seeking to incorporate potential returns to scale in distribution at the product- and vendor-hospital level.

be uncertain about a brand's supplier's negotiator type.[41] Finally, η_{jht} is the econometric unobservable term for the supply side moments.

In order to recover the supply parameters, we rearrange the supply equation and take logs to obtain the following expression:

$$\begin{aligned}\ln(\eta_{jht}) &= -\beta_j + \beta_h + 1_{\{Info_{jh}\}} X_{jh}^{pq} \beta_{jh}^{Info,pq} + \ln(p_{jht} - mc_j) \\ &- \ln\left(\left(1 + \frac{\partial q_{jht}}{\partial p_{jht}} \frac{p_{jht} - mc_j}{q_{jht}}\right) \frac{\pi_h(\mathcal{J}_{ht}) - \pi_h(\mathcal{J}_{ht} \setminus j)}{q_{jht}}\right) .\end{aligned}$$

In this expression, prices p_{jht}, product characteristics $X_j^{mc}(= \min_{ht} p_{jht})$, and demand observables X_{jht}^d enter as data, and we condition out the bargaining regressors. That leaves only the marginal cost parameter γ to be recovered from this moment equation, which is identified under the assumption $E\left[\ln(\eta_{jht}) Z_{jht}^B\right] = 0$. For our supply side instrument we use the optimal instrument (Hansen 1982): $Z^B := \frac{\partial \ln(\eta)}{\partial \gamma}$.

We combine the supply and demand moments in a GMM estimator. We estimate demand and supply jointly, imposing supply optimality constraints:

$$mc_j \in [0, p_{jht}], \tag{11}$$

and

$$\frac{\partial s_{jht}}{\partial p_{jht}} \frac{p_{jht} - mc_j}{s_{jht}} \in [-1, 0]. \tag{12}$$

3.3 Search/contracting model

The demand and pricing models specified thus far are based upon a consideration set $\mathcal{J}_{ht}$ that has been determined by hospital h's search over the set of all possible brands available at a point in time $\mathcal{J}_t$. The search process we specify is intended to accommodate the following features: (1) Allowing for various sources of heterogeneity across brands in (beliefs about) preferences θ_{hj}; bargaining β_j, β_h; and search costs sc_{hj} seems important for fitting the data and intuitions about agent information and behavior. (2) Hospitals make repeated purchases from $\mathcal{J}_{ht}$, so the composition of $\mathcal{J}_{ht}$ matters (in a similar vein to optimal retailer assortment or portfolio choice problems). (3) Unless the full brand set $\mathcal{J}_t$ is small, the computability of an (optimal) search model relies on the *ordering* of brands not changing (too much) with the size/composition of the inframarginal consideration set $\mathcal{J}_{ht}$.

[41]This seems like the most natural way to map asymmetric information about bargaining parameters into the Nash-in-Nash framework. Providing noncooperative foundations of such a model, however, is not straightforward, and would presumably involve extending the ideas in Collard-Wexler et al. (2017) to asymmetric information bargaining as in Rubinstein (1985).

One cannot satisfy all of $[1-3]$ simultaneously and also meet the assumptions required for the algorithms commonly applied to simplify dynamic search problems (e.g. Chade and Smith 2006; Weitzman 1979). Our approach is to instead construct moment inequalities based on weak search assumptions. In this setting, we find that very weak assumptions are still informative. Moreover, adding inequalities based on stronger assumptions generates identical results to a full search model consistent with those assumptions (results varying the strength of the bounds assumptions available by request). Importantly, the approach is computationally tractable because our particular bounds can be computed prior to search cost estimation. Our bounds can also be constructed from inequalities consistent with both simultaneous and sequential search models.

3.3.1 Estimating the distribution of unobservables for counterfactual product-hospital combinations

In order to infer search costs (and for later counterfactuals), we need to consider demand for brand-hospital combinations that we do not observe in the data. Our control function selection correction allowed us to estimate the unconditional brand and hospital fixed effect distributions.[42] However, we still need an estimate of brand-hospital-specific taste shocks for these $k \in \mathcal{J}_t \setminus \mathcal{J}_{ht}$. The reasoning behind the selection correction was that the distribution of demand unobservables $\xi_{jht} = \xi^o_{jht} + \xi^u_{jht}$ is expected to be shifted upward for $j \in \mathcal{J}_{ht}$, relative to the unconditional distribution, so these cannot be used directly.

However, we observe a close approximation to the unconditional distribution of ξ_{jht} in the demand realizations jht with $\widehat{Pr(j \in \mathcal{J}_{ht})} \approx 1$. Intuitively, hospitals with $\widehat{Pr(j \in \mathcal{J}_{ht})} \approx 1$ were "forced" to purchase brand j due to plausibly exogenous variation in exposure to those brands' vendors. To leverage this fact, we use the following procedure:

1. Estimate a piecewise uniform distribution $g^{uc}(\xi)$ using the demand residuals from the sample $jht|\widehat{Pr(j \in \mathcal{J}_{ht})} > 0.8$. $\hat{g}^{uc}(\xi)$ is the estimated unconditional distribution of ξ.[43]

2. Using the same piecewise uniform functional form and interval cutoffs from (1), estimate $g^{c,in}(\xi)$ using the demand residuals for all $j \in \mathcal{J}_{ht}$. $\hat{g}^{c,in}(\xi)$ is the estimated conditional distribution of ξ for hospital-brand-years with positive purchase.

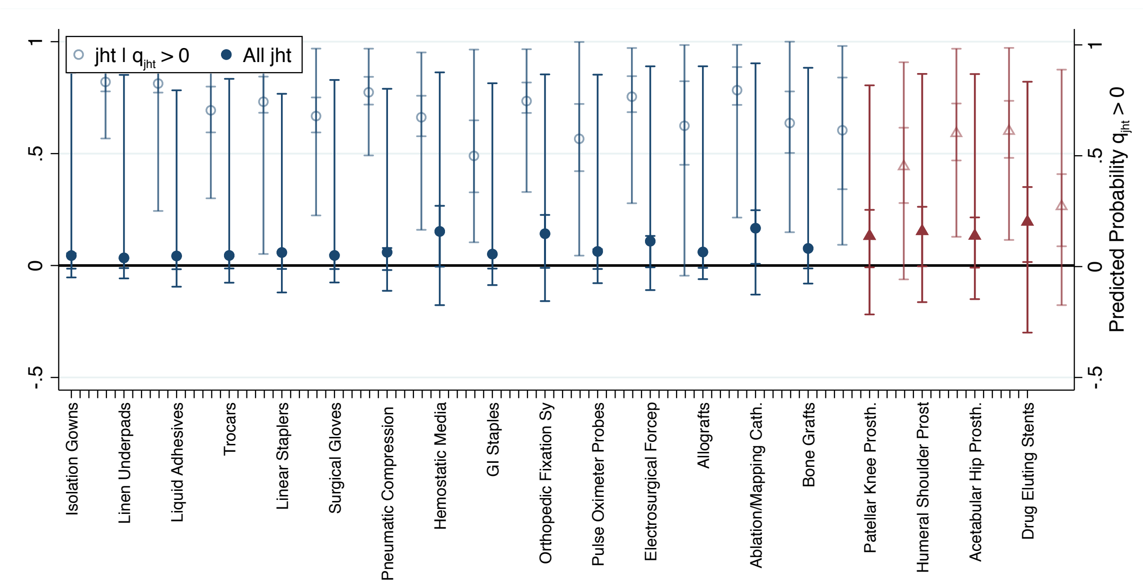
[42] Similar to work in teacher "value added", we apply Bayesian shrinkage procedures to the parameters $\theta_{jt}, \theta_h, \beta_{jt}, \beta_h$ to account for the fact that some are estimated from very few observations.

[43] In our current implementation, we divide the $\xi_{jht}|\widehat{Pr(j \in \mathcal{J}_{ht})} > 0.8$ into quintiles, and assume the functional form of a uniform distribution within each quintile, bounded at the ends by the min and max unobservables in the set. Assuming that the conditional and unconditional distributions of ξ share the same functional form, estimating the conditional distribution of ξ amounts to estimating the probability density for each of these same five intervals.

3. Infer the discrete distribution of ξ for hospital-brand-years with zero purchase $g^{c,out}(\xi)$ using the law of total probability: $g^{uc}(\xi) = (1 - Pr(j \in \mathcal{J}_{ht})) * g^{c,out}(\xi) + Pr(j \in \mathcal{J}_{ht}) * g^{c,in}(\xi)$.

To illustrate the power of our "forcing" variables Z^s, and to demonstrate how much support is available for estimating the unconditional distribution of ξ, Figure 4 plots the distributions of $\widehat{Pr(j \in \mathcal{J}_{ht})}$ for all jht and for $jht|\mathbb{1}(j \in \mathcal{J}_{ht})$. The main points to take away from this plot are (1) that the distribution of $\widehat{Pr(j \in \mathcal{J}_{ht})}$ is clearly shifted upward for realizations with $q_{jht} > 0$; and (2) that, in each included product category, there is a significant mass of observations with $\widehat{Pr(j \in \mathcal{J}_{ht})} \approx 1$ for which we observe demand realizations, enabling us to estimate the unconditional distributions of ξ.

Figure 4: Distributions of predicted consideration set inclusion propensities



Notes: Distributions of $\widehat{Pr(j \in \mathcal{J}_{ht})}$ for all jht such that $q_{jht} > 0$ (hollow markers) and for all possible jht (solid markers; given small observed consideration sets, the vast majority of these have $q_{jht} = 0$). Markers represent distribution means. Bars represent $1^{st} - 99^{th}$ percentiles. 25^{th} and 75^{th} percentile tick marks shown as well.

3.3.2 Search identification and estimation

We first consider how brands $j \in \mathcal{J}_{ht}$ provide upper bounds on search costs. Assuming that the set of firms purchased from in the data must be a subset of the firms searched, $\mathcal{J}_{ht} \subseteq \mathcal{J}_{ht}^{search}$, it follows that any $j \in \mathcal{J}_{ht}$ must have been worth paying search costs for at some stage during the search process. When brands are all substitutes for one another, the weakest such assumption comes from the value of j versus the outside good:

$$E_{\xi,\eta}[AV_{jht}(\theta,\beta,\gamma;\emptyset\cup j)] > sc_{jht}\ \forall j \in \mathcal{J}_{ht}. \tag{13}$$

This assumption is weak in that it provides a greater value than adding j to any other set, and also in that it is potentially consistent with both simultaneous and sequential models of search (and in the case of sequential, any order of search).

Analogously, brands $k \in \mathcal{J}_t \setminus \mathcal{J}_{ht}$ provide lower bounds on search costs. Assuming there is at least one brand k that a given hospital has not searched, $k \in \mathcal{J}_t \setminus \mathcal{J}_{ht}^{search}$, then it must *not* have been worth paying search costs for this brand at any time during the search process. For substitutes, the weakest such assumption comes from value of k as part of the full possible choice set:

$$E_{\xi,\eta}[AV_{kht}(\theta,\beta,\gamma;\mathcal{J}_t)] < sc_{kht}\ \forall k \in \mathcal{J}_t \setminus \mathcal{J}_{ht}^{search}. \tag{14}$$

Again, this assumption is weak in that it provides a smaller value than adding k to any other set, and also in that it is potentially consistent with both simultaneous and sequential models of search (and in the case of sequential, any order of search). To account for the possibility that $\mathcal{J}_{ht}^{search}$ may be much larger than $\mathcal{J}_{ht}$ (and maybe approaching $\mathcal{J}_t$), we further take the minimum of the bounds above over all potential $k \in \mathcal{J}_t \setminus \mathcal{J}_{ht}$.

For both bounds, estimating $AV_{jht}(\theta,\beta,\gamma;\emptyset\cup j)$ involves computing counterfactual equilibrium prices and quantities for the respective counterfactual choice sets (for each hospital-year). We do this using our estimated demand (θ) and supply (β,γ) parameters. We then construct distributions of the demand unobservables for brands both in and out of the observed consideration sets. In our current implementation, we use estimated ξ_{jht} and η_{jht} for any $j \in \mathcal{J}_{ht}$ and random draws from the conditional ξ distribution computed above and unconditional η distribution from the parameter estimates for $k \in \mathcal{J}_t \setminus \mathcal{J}_{ht}$.[44]

We use these bounds and parameterize search costs by:

$$sc_{jht} = \psi^0 + \psi^{t-1}1_{\{j\in\mathcal{J}_{ht-1}\}} + \psi^{far}Z_{jht}^{far} + \nu_{jht} \tag{15}$$

where ψ^0 estimates the mean search/contracting cost (in units of dollars of expected surplus

[44]Note our current implementation computes search costs (and later counterfactuals) as if the entire $\xi = \xi^o + \xi^u$ is known prior to contracting. We are currently working on parametric assumptions that would allow us to deconvolute these two sources of error, as well as less parametric robustness to different assumptions on the distribution of ξ^o (similar to Eizenberg (2014) in the context of firm entry). For example, for search costs estimation, the bounds can be weakened further by assuming the hospital has extreme beliefs in the portion of the demand unobservable it knows prior to search, e.g. $\xi_{jht}^o = \max_{ht}\xi_{jht}$ for $j \in \mathcal{J}_{ht}$ and $\xi_{kht}^o = \min_{ht}\xi_{kht}$ for $k \notin \mathcal{J}_{ht}$.

per patient) for a product/vendor that is new to the hospital, ψ^{t-1} and ψ^{far} allow for lower search costs for brands sourced by the hospital in the previous year or in other product categories in the current year, respectively. We implement the constrained optimization approach to computing the identified sets for this moment inequalities estimator described in Dong et al. (2017).

4 Quantifying Sources of Price Dispersion

4.1 Estimated parameters: demand, pricing, and search

Table 2 shows the estimated parameters from our demand and pricing models. The price means and coefficients of variation are shown for each product category for context. The first parameters of interest are the demand parameters. We summarize the estimated selection correction as the implied change in the expected value of the residual for a one standard deviation increase in $1 - \widehat{Pr[j \in \mathcal{J}_h]}$. The positive estimates indicate positive selection on unobservable demand – the less likely a product is to be in the choice set based on our measure of exposure in far away product categories, the more likely the hospital has a high unobservable taste draw ξ^o_{jht} for that brand.

There is a wide range over the estimated nesting parameter λ, which captures substitution to the outside good (recall λ is an approximate measure of within-category correlation in substitution, with 0 characterizing the plain logit model, and 1 implying no substitution to the outside good). The class average λ is similar for PPIs (0.28) and non-PPIs (0.33). However, class averages conceal a great deal of heterogeneity, reflecting variation across categories in the availability and closeness of the substitute treatment opportunities that comprise the outside option.

The price coefficient θ^p (that scales dollars into logit utils, and recall the extreme value type 1 normalization fixes the standard deviation of the error to $1 - \lambda * (\pi/\sqrt{6} \sim 1.28)$) is rather small across most categories, indicating little price sensitivity in product usage patterns in general. There are, however, large differences in price sensitivity across product categories. PPI usage is more than an order of magnitude less price-sensitive than non-PPIs (on both a level and percent change basis). This is consistent with what we would expect, given the relative amounts of brand-manufacturer-specific branding, and the relative importance of PPIs and non-PPIs in determining procedural outcomes. We find it reassuring that our fairly parsimonious demand model is able to empirically identify this anticipated feature of physician preference items via substitution patterns revealed by the data.

The other output of the demand estimation reported here is the consumer surplus com-

Table 2: Demand and Pricing Parameter Estimates

	p		$\frac{\partial E[\xi]}{\partial \sigma}$	λ	$\theta^p * 1,000$	AV^{CS}		B		$\frac{p-mc}{p}$	
	μ	$\frac{\sigma}{\mu}$				μ	$\frac{\sigma}{\mu}$	μ	$\frac{\sigma}{\mu}$	μ	$\frac{\sigma}{\mu}$
Other Medical/Surgical Supplies (Non-PPIs)											
Linen Underpads	$0.30	0.08	1.11	0.48	-16.243	$51	0.15	0.00	0.47	0.22	0.31
Isolation Gowns	$0.45	0.08	0.44	0.18	-16.461	$69	0.06	0.00	0.43	0.18	0.30
Surgical Gloves	$0.86	0.07	0.94	0.29	-14.532	$179	0.02	0.00	0.35	0.25	0.26
Pulse Oximeter Probes	$10	0.18	-0.33	0.55	-1.997	$427	0.14	0.15	0.58	0.33	0.34
Liquid Adhesives	$17	0.12	0.81	0.55	-5.721	$117	0.16	0.08	0.50	0.68	0.32
Pneumatic Compression Cuffs	$19	0.11	0.49	0.64	-7.353	$83	0.21	0.17	0.44	0.47	0.29
Trocars	$35	0.21	1.02	0.28	-6.267	$107	0.10	0.13	0.62	0.37	0.39
GI Staples	$133	0.18	0.31	0.23	-0.299	$3,036	0.03	0.05	0.19	1.00	0.00
Linear Staplers	$142	0.14	0.35	0.33	-1.029	$593	0.07	0.24	0.15	1.00	0.00
Orthopedic Fixation Systems	$396	0.20	1.31	0.09	-0.801	$886	0.09	0.18	0.45	0.59	0.27
Hemostatic Media	$447	0.06	0.34	0.46	-0.898	$670	0.17	0.10	0.36	0.22	0.26
Electrosurgical Forceps	$523	0.20	1.69	0.35	-0.237	$3,036	0.07	0.13	0.45	0.50	0.25
Ablation/Mapping Cath.	$880	0.14	0.30	0.06	-0.297	$2,792	0.05	0.11	0.41	0.45	0.27
Allografts	$956	0.14	0.68	0.00	-0.230	$4,200	0.05	0.10	0.45	0.31	0.32
Bone Grafts	$2,691	0.13	0.54	0.43	-0.108	$4,569	0.07	0.11	0.47	0.41	0.34
Average(15)	$417	0.14	0.67	0.33	-4.831	$1,388	0.10	0.10	0.42	0.47	0.26
Physician Preference Items (PPIs)											
Patellar Knee Prosth.	$414	0.24	0.61	0.65	-0.420	$3,402	0.28	0.22	0.62	0.38	0.36
Acetabular Hip Prosth.	$1,152	0.23	2.04	0.00	-0.154	$5,581	0.07	0.21	0.24	1.00	0.00
Drug Eluting Stents	$1,471	0.06	0.47	0.44	-0.168	$2,853	0.15	0.39	0.11	0.98	0.00
Humeral Shoulder Prosth.	$2,173	0.21	1.21	0.04	-0.202	$3,663	0.16	0.24	0.46	0.51	0.28
Average(4)	$1,303	0.18	1.08	0.28	-0.236	$3,875	0.16	0.27	0.36	0.72	0.16

ponent of added value – the additional surplus accruing to the physician-patient-hospital for having access to brand j – implied by the estimated utility model. In part due to the low price sensitivity of demand, these are relatively large in dollar value across all categories, but increasingly so in the preference item categories. Coefficients of variation of the added value range from 0.16 in PPIs to 0.10 in non-PPIs – slightly smaller than price coefficients of variation. These AV^{CS} estimates are the key input from the demand estimation that, combined with the granular price data, allows estimation of the bargaining model.

The results for the bargaining model are shown for the last four columns. The bargaining model provides estimates of marginal costs, which in turn define markups; and bargaining parameters, which rationalize the split of the total added value $AV^{TS} := AV^{CS} + p - mc$ between device vendors and hospitals, conditional on the consideration set. The bargaining parameters indicate that manufacturers of non-PPIs capture 10 percent of the total value added up for negotiation on average, while PPIs capture more, 27 percent on average. These parameters implicitly capture some of the unique agency relationships in the hospital supply setting. Recall our assumption that the preferences estimated in the demand model are the relevant preferences for measuring added value in the bargaining model. The bargaining residual captures the relative weight put on vendor and hospital surplus to explain the price variation as a function of added value variation. Thus, a smaller relative share to vendors can be driven by the purchasing agents involved in negotiation perceiving brands as closer substitutes than provider substitution patterns would indicate. In that light, one explanation consistent with prior work on PPIs would be that the larger bargaining split is driven by the greater ability/desire of physicians to transmit their preferences to purchasing (or conversely, the inability of purchasing to move physician market share) for PPIs than for non-PPIs.

As expected given the large price dispersion documented, estimated markups are large in most product categories, ranging from 18 percent of price in gowns to 100 in a few categories with estimated marginal costs of zero. The combination of lower price sensitivity and higher bargaining parameters leads to larger average markups of 72 percent in PPIs, vs. 47 percent in non-PPIs.

4.2 Search costs and breadth of buyer-supplier relationships

In Section 4.3, we will further examine the contributions of search, bargaining, and demand to the level and variation of markups. First, though, we consider the parameter estimates of the search cost functions. Of particular interest are the mean search cost parameters, but we also discuss the coefficients on our proxies for past contracting relationship ψ^{t-1} and current contracting relationship with the brand's vendor across dissimilar product categories ψ^{far}. The latter coefficient speaks to one channel via which supplier breadth is argued to potentially enhance welfare. Table 3 provides evidence on this matter.

Table 3: Determinants of Search/Contracting Costs

	ψ^0	ψ^{far}	ψ^{t-1}	$Z^{s,far}$		$1\{t-1\}$	
				μ	σ	μ	σ
Other Medical/Surgical Supplies (Non-PPIs)							
Linen Underpads	[0.1, 0.0]	[0.2, -0.1]	[0.1, -0.1]	0.09	0.08	0.60	0.26
Isolation Gowns	0.9	-0.0	-0.2	0.08	0.07	0.60	0.29
Surgical Gloves	0.2	-0.0	-0.0	0.08	0.08	0.62	0.29
Pulse Oximeter Probes	0.6	-0.0	-0.0	0.04	0.03	0.48	0.29
Liquid Adhesives	[0.4, 0.3]	[0.2, -0.1]	[0.0, -0.1]	0.05	0.06	0.53	0.32
Pneumatic Compression Cuffs	0.5	[0.2, -0.0]	[0.0, -0.1]	0.07	0.06	0.59	0.29
Trocars	1.2	-0.0	-0.0	0.06	0.06	0.60	0.33
GI Staples	32.4	-1.0	-0.4	0.06	0.06	0.54	0.37
Linear Staplers	12.6	-0.3	-0.2	0.06	0.06	0.60	0.36
Orthopedic Fixation Systems	10.5	-0.0	-0.1	0.14	0.08	0.53	0.38
Hemostatic Media	63.7	-19.1	0.2	0.05	0.06	0.57	0.35
Electrosurgical Forceps	31.1	-0.0	-1.2	0.03	0.03	0.44	0.36
Ablation/Mapping Cath.	56.8	0.0	-0.0	0.03	0.03	0.58	0.37
Allografts	37.7	-0.0	-0.7	0.02	0.02	0.45	0.33
Bone Grafts	64.4	-0.1	-0.6	0.10	0.07	0.54	0.35
Average(15)	[20.9, 20.8]	[-1.3, -1.4]	-0.2	0.06	0.06	0.55	0.33
Physician Preference Items (PPIs)							
Patellar Knee Prosth.	3.3	-0.0	-0.0	0.07	0.06	0.62	0.35
Acetabular Hip Prosth.	29.0	-0.0	-0.2	0.09	0.07	0.54	0.39
Drug Eluting Stents	91.3	1.9	-2.2	0.04	0.03	0.54	0.35
Humeral Shoulder Prosth.	83.8	-0.0	-0.8	0.08	0.05	0.49	0.39
Average(4)	51.9	0.5	-0.8	0.07	0.05	0.55	0.37

For each product category, the first three columns of Table 3 report the search cost parameter estimates, and the last four columns provide summary statistics (mean and standard deviation) for the extent of purchasing persistence ($1(q_{jh,t-1} > 0)$) and breadth ($Z^{s,far}$) to facilitate understanding of the quantities involved. Parameters are point identified in most cases.[45] Looking at the category averages, search/contracting costs are meaningful, but not overwhelming, averaging roughly 5 percent of price.[46]

It is difficult to tell directly from these parameter estimates the extent to which search/contracting frictions vs. product differentiation are the driving force behind the large estimated markups in these medical devices. Both seem to be nontrivial, but without an equilibrium model, it is not clear how to assess their relative role. Relatedly and similarly, it is difficult to characterize the relative contribution of demand versus bargaining heterogeneity in the observed price dispersion across hospitals based on the parameter estimates alone. The next Section computes several counterfactuals in order to shed further light on these issues.

4.3 Decomposing price variation

As a final exercise, we use our parameter estimates to examine various decompositions of the prices observed in the data. If the price variation we document, and the market power underlying estimated markups, are driven by true brand differentiation in quality (where quality could either be vertical quality for the average use, or horizontal use-specific match quality), then that has quite different welfare implications than physician-specific brand preferences or search frictions that limit the choice set. To disentangle these factors, we explore equilibrium prices and consumer surplus in several counterfactual scenarios.

First, to better understand the drivers of price variation across hospitals, we condition on the observed choice sets in the data, but we counterfactually shut down heterogeneity across hospitals in bargaining ($B_{jht} = \frac{\beta_{jt}}{\beta_{jt} + \frac{1}{|\mathcal{H}|}\sum_{h \in \mathcal{H}} \beta_h / \eta_{jht}}$) and demand ($\delta_{jht} = \theta_{jt} + \frac{1}{|\mathcal{H}|}\sum_{h \in \mathcal{H}}(\theta_h + \theta^f f(Z^s_{jht}) + \xi^u_{jht}) - \theta^p p_{jht}$), each in turn. We then compute equilibrium prices (and quantities) at these parameters, and recompute the coefficient of variation in price across hospitals (within brand-year, focusing on the year 2013). This provides a measure of the extent to which the price dispersion in the data is being driven by bargaining or demand, conditional on the observed choice sets.

[45] The estimates here follow the approach of using estimated ξ_{jht} and drawing counterfactual ξ_{kht} from the estimated distribution $f(\xi | k \notin \mathcal{J}_{ht})$. Parameter estimates for looser bounds using $\xi = \xi^{\max}_{\min}$ and also different bounds more akin to the "stability" inequalities in Ghili (2018) available by request. For the loosest bounds, most parameters are no longer point identified. Approaches to deconvoluting the distributions of ξ^o and ξ^u are currently in progress.

[46] In our current implementation, the estimated effects of vendor exposure in other product categories and presence in prior year's choice set are small; however, we offer this with the caveat that these parameters are especially sensitive to which moment inequalities we use. A primary area of ongoing work for this project is to better understand how different inequalities, and greater flexibility, affect search cost estimates.

In Table 4, the left panel repeats summary statistics from the data on choice set sizes and prices for reference. The middle panel (columns 5-6) shows the proportion of price dispersion ($\% \frac{\sigma}{\mu}$) remaining after the counterfactuals just discussed – taking out price variation driven by bargaining ($p^{-\sigma_B}$) and preferences ($p^{-\sigma_D}$), respectively. We find that in every product category except one, variation in prices across hospitals in the observed data is driven (typically much) more by variation in bargaining than by demand. This relative dominance of bargaining variation is slightly stronger in non-PPIs, but varies more within our product classes than across them; e.g., bargaining is far more important to price dispersion for surgical gloves than for linen underpads among the non-PPIs, and more important for shoulder prostheses than for drug-eluting stents among the PPIs. All of the categories show some covariance between bargaining and demand estimates, with sum of the price variation explained by each totaling more than the observed price variation in the data.

Table 4: Decomposing Variation in Prices and Markups

	$\overline{\lvert\mathcal{J}_h\rvert}$	p		Decompose $\sigma(p)$		Decrease Search Costs by Half			
				$p^{-\sigma_B}$	$p^{-\sigma_D}$	$\overline{\lvert\mathcal{J}_h^c\rvert}$	p		$\overline{\Delta CS(\mathcal{J}_h^c, \mathcal{J}_h)}$
		μ	$\frac{\sigma}{\mu}$	$\% \frac{\sigma}{\mu}$	$\% \frac{\sigma}{\mu}$		$\mu, q^w_{\mathcal{J}_h}$	$\mu, q^w_{\mathcal{J}_h^c}$	
Other Medical/Surgical Supplies (Non-PPIs)									
Linen Underpads	2	\$0.31	0.07	53.7	103.4	28	\$0.29	\$0.35	\$48
Isolation Gowns	2	\$0.47	0.10	10.5	99.3	9	\$0.45	\$0.91	\$48
Surgical Gloves	10	\$0.90	0.07	9.7	101.6	22	\$0.76	\$0.85	\$47
Pulse Oximeter Probes	6	\$9.92	0.22	16.9	102.2	30	\$8.72	\$64	\$402
Liquid Adhesives	4	\$19	0.10	49.6	113.7	20	\$14	\$22	\$107
Pneumatic Compression Cuffs	3	\$18	0.16	47.5	99.3	11	\$15	\$20	\$71
Trocars	9	\$33	0.21	5.0	100.2	35	\$32	\$37	\$130
GI Staples	6	\$145	0.16	16.0	104.8	23	\$134	\$137	\$2,787
Linear Staplers	7	\$154	0.11	33.2	109.7	27	\$141	\$151	\$791
Orthopedic Fixation Systems	25	\$410	0.22	0.9	100.0	27	\$409	\$392	\$436
Hemostatic Media	4	\$395	0.08	62.4	100.2	4	\$390	\$403	\$164
Electrosurgical Forceps	10	\$546	0.17	10.0	102.9	23	\$519	\$561	\$2,366
Ablation/Mapping Cath.	22	\$893	0.13	2.1	100.1	30	\$891	\$935	\$1,977
Allografts	21	\$1,003	0.15	2.0	100.0	113	\$996	\$1,318	\$6,604
Bone Grafts	10	\$2,683	0.16	8.2	102.1	16	\$2,595	\$2,244	\$3,633
Average(15)	9	\$421	0.14	21.8	102.6	28	\$410	\$419	\$1,307
Physician Preference Items (PPIs)									
Patellar Knee Prosth.	6	\$438	0.23	35.5	100.2	8	\$420	\$455	\$307
Acetabular Hip Prosth.	24	\$1,128	0.24	3.6	100.9	43	\$1,116	\$1,259	\$3,726
Drug Eluting Stents	4	\$1,361	0.06	119.8	176.8	4	\$1,351	\$1,352	\$127
Humeral Shoulder Prosth.	17	\$2,118	0.23	1.9	99.9	22	\$2,113	\$2,249	\$1,810
Average(4)	13	\$1,261	0.19	40.2	119.5	19	\$1,250	\$1,329	\$1,492

We also explore the comparative static of how choice sets and prices change as search costs decrease by half. Specifically, we compute counterfactual choice sets $\mathcal{J}_{ht}^c$, by drawing counterfactual ξ_{kht} for $k \notin \mathcal{J}_{ht}$, ordering these brands by the expected value of adding them to the choice set $AV_k(\mathcal{J}_{ht} \cup k) - sc_{kht}/2$, and adding them to the choice set until it is no longer beneficial to do so.[47] We then recompute prices and quantities in equilibrium in each

[47] $k^{(n+1)}$ such that $AV_{k^{(n+1)}}(\mathcal{J}_{ht}^{(n)}) - sc_{k^{(n+1)}ht}/2 < 0$, starting with $n = 0$ and $\mathcal{J}_{ht}^{(0)} = \mathcal{J}_{ht}$.

hospital-year for this new choice set, $\mathcal{J}^c_{ht}$.[48]

The results are summarized in the right panel (columns 7-10) of Table 4. When search costs are cut in half, choice sets increase by about 50 percent for PPIs and more than triple for non-PPIs, on average. Relating this back to the full set of potential suppliers, hospitals in this counterfactual now contract with about 25 percent of all potential suppliers, vs. about 10 percent in the data.

Columns 7 displays mean counterfactual equilibrium prices, but weighted by the quantities in the data (in particular, $q^w_k = 0$ for any counterfactual k added to the choice set from the data), thus focusing purely on the competitive effect of a large choice set on prices. This effect on prices is modest, decreasing average prices by 0-7 percent, depending on the product category. This is driven by the low price sensitivity estimates (added values do not drop dramatically as brands are added to the choice set) and low percentages of the bargaining split going to manufacturers (muting the effect of any change in added values on prices).

Columns 8 weights by the counterfactual quantities, showing the additional effect of new purchasing patterns. In all but a few categories, this increases average prices per unit relative to the weighting using the quantities from the data. This indicates that hospitals tend to purchase more expensive brands as their choice sets grow. Of course, by assumption of utility maximization, hospital consumer surplus is growing too, so that these higher prices are more than offset by higher quality of these more expensive brands. The final column of Table 4 measures the change in consumer surplus in moving from the data to this counterfactual, in dollars per device used. Related to the low price sensitivity and large added values estimated, the additional surplus from expanding the choice set is often large, making it clear that these consumer surplus benefits, rather than savings in prices paid, are the primary driver behind the growing choice sets in this counterfactual.

4.3.1 Heterogeneity in contracting cost effects across hospitals and brands

Given the large heterogeneity across hospitals and brands in the data, the average effects of decreasing search costs are likely to mask large heterogeneity in effects at these individual business establishments. We examine this issue by summarizing the heterogeneity in results of this counterfactual at the brand-hospital level across hospitals and brands. Table 5 below summarizes the distribution of these changes across hospitals and brands for each product category, in percent terms relative to the original spend in the data to facilitate comparisons across hospitals, brands, and categories.

The top panel of Table 5 shows that outcomes across hospitals in the counterfactual with

[48] Bargaining unobservables for new additions to the choice set are drawn from the empirical distribution of η estimates.

Table 5: Lowering Search Costs: Heterogeneity Across Hospitals and Brands

All numbers represent percent changes moving to low search cost counterfactuals from status quo:	μ	σ	p10	p50	p90
Hospital outcomes					
Non-PPIs					
$\lvert J_h \rvert$	4.17	5.37	0.61	2.63	9.41
$\bar{p}_h(q^w_{\mathcal{J}_h})$	-0.07	0.04	-0.13	-0.06	-0.02
$\bar{p}_h(q^w_{\mathcal{J}^c_h})$	0.61	0.73	-0.08	0.43	1.55
CS_h	12.16	43.61	0.59	2.00	13.63
PPIs					
$\lvert J_h \rvert$	0.74	1.58	0.04	0.25	1.85
$\bar{p}_h(q^w_{\mathcal{J}_h})$	-0.02	0.03	-0.04	-0.00	-0.00
$\bar{p}_h(q^w_{\mathcal{J}^c_h})$	0.07	0.13	-0.04	0.03	0.23
CS_h	0.62	0.86	0.06	0.34	1.50
Supplier outcomes					
Non-PPIs					
$\lvert \mathcal{H}_j \rvert / \lvert \mathcal{H} \rvert$	14.88	19.10	0.97	7.37	39.91
π_j	1343.30	4781.29	0.17	18.81	3052.33
PPIs					
$\lvert \mathcal{H}_j \rvert / \lvert \mathcal{H} \rvert$	5.20	10.61	0.08	1.35	13.89
π_j	30.47	108.17	-0.13	0.33	61.51

lower search costs are quite heterogeneous, and also quite skewed. Some hospitals in the bottom part of the distribution see modest changes, and some in the upper tail see enormous changes, especially in the number of suppliers contracted with and consumer surplus. These estimates come with a number of caveats already mentioned regarding robustness to assumptions in search cost estimation and counterfactual computation, but it is clear that there are a set of hospitals with ex-ante "bad" supplier sets who benefit greatly from adding more, better value suppliers.

Similarly, the lower panel of Table 5 shows results across individual brands, aggregated over product categories. The average effects are to greatly increase the percent of hospitals in which the average product is present (over double at the median for PPIs) and increase profits (33 percent for PPIs). However, these results are again quite heterogeneous with a long tail of enormous winners – those brands of relatively high value that were in few hospitals' choice sets ex-ante. Different from hospitals, though, the supplier market does see some losers – those brands of relatively low value that were in many hospitals' choice sets ex-ante (e.g. the tenth percentile of brands in PPIs see a 13 percent decrease in profits).

5 Conclusion

Price dispersion across buyers for the same exact product must come from dispersion in marginal costs of distribution or dispersion in markups. Thus, absent the former, price dispersion is an indicator of market power, and understanding the economic forces underlying the price dispersion is critical for understanding impediments to market efficiency. In business-to-business markets, price dispersion across buyers due to variation in markups is of antitrust interest per se, due to its potential impact on downstream competition.

In this paper, we explore price dispersion in a large and policy-relevant market: hospital supply contracting. In detailed data from hospital purchase orders across a variety of product markets, we document substantial price dispersion across hospitals for the same brands purchased from the same vendors. In spite of the fact that the demand side in this application solely includes hospitals, there is large variation across product categories in the preferences of end users, the concentration and bargaining power of suppliers, and the potential importance of information and search/contracting frictions. We document reduced form evidence suggesting that all of these features may play some role in the observed data.

We then develop a structural model allowing for heterogeneity across hospital buyers in demand for differentiated products, price negotiations, and frictions in the search/contracting process that determines who contracts with whom. We address the problem of potential selection of suppliers based on unobserved preferences by using a control function based on hospital exposure to a vendor in product categories that are likely unrelated in terms of user preferences, but potentially related at an administrative level through their impact on contracting costs. We also leverage exogenous variation in prices due to the introduction of benchmarking information to the hospitals in the sample. The "generic" nature of these identification strategies allows us to obtain credible estimates of demand and supply across a large variety of product categories. Finally, we estimate search/contracting frictions using a moment inequalities approach that is computationally feasible and accommodates various forms of the search/contracting process, which we do not observe directly.

Our estimates suggest that large markups are primarily driven by large perceived product differentiation and lack of price sensitivity among health care providers in their product usage decisions. This problem is especially severe in "physician preference items", where price sensitivity is more than an order of magnitude lower than in other non-PPI medical/surgical supplies. Hospital purchasing administrators are able to counteract this somewhat by exercising a large degree of monopsony power in their price negotiations, but this ability varies widely across hospitals, driving most of the observed price dispersion across hospitals.

We also compute counterfactual equilibrium choice sets, prices, and quantities under a one half reduction in estimated search costs. We find small to moderate average effects on choice sets and prices, but these effects are extremely heterogenous across hospitals and brands. Hospitals with previously small and low quality choice sets can see large benefits of lower search costs, as can brands that are high quality but also have high search costs.

Taken together, our results suggest that current markups are driven primarily by lack of price sensitivity, supporting initiatives to align physician incentives with hospital costs when possible. Some high quality but high cost brands seem disproportionately left out of choice sets, suggesting there may also be important work to be done aligning hospital administrator and patient/physician incentives. However, without better understanding the determinants of physician's revealed preferences, it is difficult to draw strong welfare conclusions on this tension. The potentially large health quality and costs at stake suggest unpacking these issues as an important area for further research.

References

Allen, J., Clark, R., and Houde, J.-F. (2013). The effect of mergers in search markets: Evidence from the canadian mortgage industry. *American Economic Review*, 104(10):3365–3396.

Allen, J., Clark, R., and Houde, J.-F. (2018). Search frictions and market power in negotiated price markets. forthcoming, Journal of Political Economy.

Attanasio, O. P., Goldberg, P. K., and Kyriazidou, E. (2008). Credit constraints in the market for consumer durables: Evidence from micro data on car loans. *International Economic Review*, 49(2).

Baker, L., Bundorf, M. K., and Kessler, D. (2014). Vertical integration: hospital ownership of physician practices is associated with higher prices and spending. *Health Affairs*, 33(5):756–63.

Berry, S. and Waldfogel, J. (1999). Free entry and social inefficiency in radio broadcasting. *RAND Journal of Economics*, 30(3):397–420.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2).

Berry, S. T., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Bloom, N., Sadun, R., and VanReenen, J. (2014). Does management matter in healthcare? Stanford Mimeo.

Chade, H. and Smith, L. (2006). You have full text access to this content simultaneous search. *Econometrica*, 74(5):1293–1307.

Chipty, T. and Snyder, C. M. (1999). The role of firm size in bilateral bargaining: A study of the cable television industry. *The Review of Economics and Statistics*, 81(2):326–340.

Collard-Wexler, A., Gowrisankaran, G., and Lee, R. S. (2017). Nash-In-Nash bargaining: A microfoundation for applied work. *forthcoming in Journal of Political Economy*.

Cooper, Z., Craig, S., Gaynor, M., and Van Reenen, J. (2018). The price ain't right? Hospital prices and health spending on the privately insured. *Quarterly Journal of Economics*. Accepted.

Craig, S. V., Grennan, M., and Swanson, A. (2018). Mergers and marginal costs: New evidence on hospital buyer power. NBER Working Paper No. 24926.

Crawford, G. and Yurukoglu, A. (2012). The welfare effects of bundling in multichannel television. *American Economic Review*, 102(2).

Dana, J. (2012). Buyer groups as strategic commitments. *Games and Economic Behavior*, 74(2):470–485.

Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. (2018). Consumer price search and platform design in internet commerce. *American Economic Review*, 108(7):1820–1859.

Dong, B., Hsieh, Y.-W., and Shum, M. (2017). Computing moment inequality models using constrained optimization. USC-INET Research Paper No. 17-21. Available at SSRN: https://ssrn.com/abstract=2990826.

Duggan, M. and Scott Morton, F. (2010). The impact of Medicare Part D on pharmaceutical prices and utilization. *American Economic Review*, 100(1):590–607.

Eizenberg, A. (2014). Upstream innovation and product variety in the us home pc market. *The Review of Economic Studies*, 81(3):1003–1045.

Evans, M. and Stevens, L. (2018). Amazon targets hospital supplies. *The Wall Street Journal*.

Galenianos, M. and Gavazza, A. (2018). Regulatory intervention in consumer search markets: The case of credit cards. Working paper.

Gans, J. S. and King, S. P. (2002). Exclusionary contracts and competition for large buyers. *International Journal of Industrial Organization*, 20(9):1363–1381.

Ghili, S. (2018). Network formation and bargaining in vertical markets: The case of narrow networks in health insurance. Working paper.

Gowrisankaran, G., Nevo, A., and Town, R. (2015). Mergers when prices are negotiated: Evidence from the hospital industry. *The American Economic Review*, 105(1):172–203.

Grennan, M. (2013). Price discrimination and bargaining: Empirical evidence from medical devices. *American Economic Review*, 103(1):145–177.

Grennan, M. (2014). Bargaining ability and competitive advantage: Empirical evidence from medical devices. *Management Science*, 60(12):3011–3025.

Grennan, M. and Swanson, A. (2019). Transparency and negotiated prices: The value of information in hospital-supplier bargaining. *Journal of Political Economy*. forthcoming.

Gruber, J. and McKnight, R. (2016). Controlling health care costs through limited network insurance plans: Evidence from Massachusetts state employees. *American Economic Journal: Economic Policy*, 8(2):219–50.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054.

Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In Bresnahan, T. and Gordon, R., editors, *The Economics of New Goods*, volume 58. National Bureau of Economic Research.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–162.

Hemphill, C. S. and Rose, N. L. (2018). Mergers that harm sellers. *The Yale Law Journal*, 127(1).

Ho, K. (2009). Insurer-provider networks in the medical care market. *American Economic Review*, 99(1):393–430.

Ho, K., Ho, J., and Mortimer, J. (2012). The use of full-line forcing contracts in the video rental industry. *American Economic Review*, 102(2):686–719.

Ho, K. and Lee, R. S. (2017). Insurer competition in health care markets. *Econometrica*, 85(2):379–417.

Ho, K. and Lee, R. S. (2018). Equilibrium provider networks: Bargaining and exclusion in health care markets. Manuscript.

Hong, H. and Shum, M. (2006). Using price distributions to estimate search costs. *RAND Journal of Economics*, 37(2).

Honka, E. (2014). Quantifying search and switching costs in the us auto insurance industry. *RAND Journal of Economics*, 45(4):847–884.

Horn, H. and Wolinsky, A. (1988). Bilateral monopolies and incentives for merger. *RAND Journal of Economics*, 19:408–419.

Hortacsu, A. and Syverson, C. (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of s&p 500 index funds. *Quarterly Journal of Economics*, 119(2):403–456.

Inderst, R. and Wey, C. (2007). Buyer power and supplier incentives. *European Economic Review*, 51:647–667.

Kaplan, G. and Menzio, G. (2015). The morphology of price dispersion. *International Economic Review*, 56(4).

Ketcham, J., Leger, P. T., and Lucarelli, C. (2009). Group incentives and standardization: An application to hospital physician gainsharing. Working paper.

Lewis, M. and Pflum, K. (2015). Diagnosing hospital system bargaining power in managed care networks. *American Economic Journal: Economic Policy*, 7(1):243–74.

Liebman, E. (2018). Bargaining in markets with exclusion: An analysis of health insurance networks. Working paper.

Marvel, H. P. and Yang, H. (2008). Group purchasing, nonlinear tariffs, and oligopoly. *International Journal of Industrial Organization*, 26(5):1090–1105.

McConnell, K. J., Lindrooth, R. C., Wholey, D. R., Maddox, T. M., and Bloom, N. (2013). Management practices and the quality of care in cardiac units. *JAMA Internal Medicine*, 173(8):684–692.

Montgomery, K. and Schneller, E. (2007). Hospitals' strategies for orchestrating selection of physician preference items. *Milbank Quarterly*, 85(2):307–335.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(307-42).

Noether, M. and May, S. (2017). Hospital merger benefits: Views from hospital leaders and econometric analysis. Technical report, Charles River Associates.

Nugent, W. C., Kilo, C. M., Ross, C. S., Marrin, A. S., Berwick, D. M., and Roessner, J. (1999). *Improving Outcomes and Reducing Costs in Adult Cardiac Surgery.* Institute for Healthcare Improvement, Boston, MA.

Petrin, A. and Train, K. (2009). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, xlvi.

Robinson, J. C. (2008). Value-based purchasing for medical devices. *Health Affairs*, 27(6).

Rubinstein, A. (1985). A bargaining model with incomplete information about time preferences. *Econometrica*, 53(5):1151–1172.

Salz, T. (2017). Intermediation and competition in search markets: An empirical case study. Available at SSRN: https://ssrn.com/abstract=2961795.

Schneller, E. S. (2009). The value of group purchasing - 2009: Meeting the need for strategic savings. *Health Care Sector Advances, Inc.*

Scholten, P. A. and Smith, S. (2002). Price dispersion then and now: Evidence from retail and e-tail markets. *The Economics of the Internet and E-commerce*, 11:63–88.

Sela, R. J. and Simonoff, J. S. (2011). Reemtree: Regression trees with random effects. R package version 0.90.3.

Snyder, C. M. (1996). A dynamic theory of countervailing power. *Rand Journal of Economics*, 27:747–769.

Snyder, C. M. (1998). Why do larger buyers pay lower prices? Intense supplier competition. *Economics Letters*, 58:205–209.

Sorensen, A. (2003). Insurer-hospital bargaining: Negotiated discounts in post-deregulation connecticut. *Journal of Industrial Economics.*

Starc, A. and Swanson, A. (2018). Preferred pharmacy networks and drug costs. Working paper.

Stigler, G. (1961). The economics of information. *Journal of Political Economy*, 69(3):213–225.

Stole, L. and Zwiebel, J. (1996). Intra-firm bargaining under non-binding contracts. *Review of Economic Studies*, 63:375–410.

Swanson, A. (2019). Physician investment in hospitals: Specialization, incentives, and the quality of cardiac care. *working paper*.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica*, 47(3).

ELECTRONIC APPENDICES – NOT FOR PRINT PUBLICATION

A Data Appendix

The data used in this project are from the PriceGuide™ benchmarking service offered by the ECRI Institute, a non-profit health care research organization. The PriceGuide data are expected to be of high quality for several reasons. First, the hospital supply transactions are typically transmitted as a direct extract from a hospital's materials management database to the PriceGuide benchmarking portal. Second, hospitals have strong incentives to report accurately because the analytics the benchmarking service's web portal provides are based on comparing the hospital's submitted data to that of others in the database. Finally, as discussed in Grennan and Swanson (2019), at least for coronary stents, the distribution of prices observed in hospitals' pre-join benchmarking data is similar to that observed in an external, representative market research dataset. Nevertheless, supply and demand analysis requires an analytic sample that is complete at the hospital-product category level, containing all relevant substitutes and *only* relevant substitutes, and with comparable prices and units across all observations. Thus, our analyses required several cleaning steps beyond the processing undertaken by ECRI for the purposes of their within-SKU PriceGuide analytics.[49] In this Appendix, we describe the process we used to construct the final analytic sample. This process is intended to focus on "important," high-spending categories with well-defined substitutes, reported in consistent units, and with manageable file sizes.

The raw transactions data contain 116 million observations for 2,876 members across 3,394 product categories and 2.7 million stock keeping units (SKUs). Our analyses include 19 important product categories, defined by their UMDNS codes. To arrive at this set of product categories, we looked within the top 100 categories by overall spending. From these, we excluded categories that were too broad (involving products that were not clearly substitutes) or where data consistency seemed to be an issue. Regarding the first cut, we selected eight categories by hand that seemed excessively broad based on their UMDNS names: Blood Collection Tubes, Clinical Reagents, Computer Supplies, In-Vitro Diagnostic (IVD) Kits, Industrial Supplies, Office Supplies, Patient Education Materials, and Pharmaceuticals.[50] Regarding the second cut, demand estimation requires that we analyze quantities across hospital-years, within each product category, for a well-defined unit. Although many medical and surgical product categories are sold by the unit (e.g., a single coronary stent), others

[49] PriceGuide reports accurate price comparisons across all observations within the same SKU. However, there is evidence that the data are missing some hospital-product category pairs. For example, we find it unrealistic that some broadly used categories (e.g., examination gloves) do not include data from all hospitals. This can occur when transactions are not submitted with informative free-text item descriptions and are accordingly not assigned UMDNS codes by ECRI.

[50] For example, "IVD Kits" include microbial detection kits costing $2.14 on average, as well as tests for antibiotic-resistant bacteria colonization costing $4,400 on average.

are sold in pairs, boxes, cases, etc. The transactions data indicates this distinction in the "unit of measure" field, and further notes how many subunits are in each unit of measure using a "conversion factor" field. Using these fields, we transformed the price and quantity variables into a price per single unit and quantity of single units purchased, respectively. We dropped transactions that were missing either unit of measure or conversion factor data, and, for each product category, we restricted the analysis to transactions that were reported in the modal unit of measure (i.e., if a product category is usually sold in boxes, we include only transactions reported in boxes in our analytic sample). Finally, we dropped product categories from the analysis if more than thirty percent of transactions were lost when we limited the sample to the modal unit of measure. This filter ensures that the included data are meaningfully representative of the category.[51]

Next, for reasons of practicality, we exclude product categories with prohibitively large datasets. We drop products requiring greater than 20GB of RAM for the random forest procedure we use to estimate market size[52] or greater than 50GB of RAM for supply and demand estimation.[53] Our supply and demand estimation also includes large matrices of fixed effects, many of which are inherently sparse. This requires two additional filters: to speed up the procedure that removes linearly dependent columns, we drop categories with $N_{jht} > 50,000$[54]; and we only keep categories such that all hospital fixed effects survive the above procedure.[55]

Finally, we dropped several product categories from our analysis due to incompatibilities with the identification strategy detailed in Section 3.1.1. In order to address the identification problem introduced by selection of brands into each hospital's consideration set, we use instruments based on exposure of hospitals to vendors across dissimilar product categories. This approach requires a measure of dissimilarity – a starting point for this classification approach is based on UMDNS codes' relative positions in a UMDNS hierarchy, which is missing for eleven of the remaining product categories.[56]

Our "one-size-fits-all" identification strategy is quite powerful for many heterogeneous

[51]At this stage, we dropped five additional categories: Examination Gloves, IV Solutions, MRI Contrast, Procedure Kits, and Surgical Packs.

[52]Cochlear Stimulators, Incontinence Neurostimulators, Mammary Prostheses, Skin Expanders, Vagus Nerve Stimulators, and Ventricular Assist Devices.

[53]Batteries, Dressings, Femoral Knee Prostheses, IV Administration Kits, IV Tubing Extensions, Surgical Staplers, Sutures, and Tibial Knee Prostheses.

[54]Balloon Catheters, Bone Screws, Drill Bits, Femoral Hip Prostheses, Guide Wires, Guiding Catheters, and Polyglactin Sutures.

[55]Atherectomy Catheters, Bone Plates, Reconstructive Tissue Material, Tissue Fusion Devices, and Total Knee Prostheses are excluded at this stage

[56]Antibiotic Orthopedic Cement, Cranial Bone Screws, IV Saline, Long Term IV Catheters, Spinal Bone Plates, Spinal Bone Screws, Spinal Rod Implants, Spinal Spacers, Trauma Bone Plates, Trauma Bone Screws, and Vascular Closure Devices.

product categories in our data, but (not unexpectedly) fails to be powerful for some product categories. For example, this may happen in cases with a very limited set of vendors and generally complete consideration sets; for such product categories, the search and contracting costs that are a primary focus of this paper are less meaningful. We only include in our analysis those product categories for which a joint F-test of significance of the excluded exposure instruments in our first stage has a p-value of less than 0.1. This results in our exclusion of 22 product categories that would not otherwise be excluded: twelve non-PPIs such as ECG Recorders, Laparoscopic Clip Appliers, and Oxygen Sensors; and ten PPIs such as Bare Metal Stents, Brain Stimulators, and Resynchronization Defibrillators. Similarly, given that our analysis relies crucially on estimates of hospitals' price sensitivity, we exclude several product categories for which $\hat{\theta}^p \geq -1^{-4}$.[57]

After applying the above filters, we are left with 19 important medical supply product categories: fifteen non-PPIs (e.g., Electrosurgical Forceps, Isolation Gowns, Surgical Gloves, and Trocars), and four PPIs (Acetabular Hip Prostheses, Drug Eluting Stents, Humeral Shoulder Prostheses, and Patellar Knee Prostheses).

A.1 Data Cleaning and Final Sample

Above, we describe the selection of the product categories we include in our analysis. Within the included set of product categories, we performed several additional refinements to the sample to address variable availability, suspected errors, and management of outliers.

- First, we limit to usable transaction data (with non-missing memberid, SKU, and manufacturer; and with positive quantity purchased). We also remove transactions with suspected price errors (i.e., brands with mean price an order of magnitude below the median price across brands, and transactions that are integer multiples of other common prices for the same SKU) and transactions with prices in the tails of the observed distribution. The main goal of the latter filter is to remove products that are erroneously included in the product category and/or which are not substitutable with the majority of products in the category. For example, the "surgical staplers" category includes many stapler refill cartridges. In practice, we limit to transactions with prices between the 5^{th} and 95^{th} percentiles of the price distribution.

- In addition to dropping transactions not in the modal unit of measure in each UMDNS code or with a missing conversion factor (e.g., 10 units per box), we also drop trans-

[57] Bone Implant Putty, Bone Nails, Cardiac Valve Prostheses, Circular Staplers, IV Infusion Pumps, Kyphoplasty Kits, Pacemakers, Polymeric Mesh, and Suture Anchors.

actions with a suspicious unit of measure: with a unit of measure of either "BX" or "CS" and with a conversion factor of 1.

- Keep only those manufacturers with sufficiently many hospital-brand-month-year observations ($N_{hjmy} > 200$) to allow us to assign brand IDs in the machine learning procedure detailed below.

- Limit to hospitals and health systems (as opposed to laboratories); such that we observe the member hospital's timing of database join in the login data and can use the price shock observed after database join to identify price elasticities; and that merge onto the AHA survey data with non-missing location (HRR), total admissions, outpatient admissions, overall bed size, bed size of each department (obstetric beds, cardiac ICU beds, etc.), and full-time-equivalent staff. The AHA variables are primarily used to construct the total market size in each category.

- Remove small hospital-years, for which demand estimation is less well-behaved and the assumption that the consideration set is equivalent to the set of brands purchased in that hospital-year is less palatable. In practice, we remove hospital-years with $q_{ht} < 30$ and hospital-years below the 5^{th} percentile in terms of total quantity.

A.2 Identifying Brands in the Transaction Data

We utilize machine learning methods to categorize SKUs into brand IDs, in order to appropriately control for brand-specific price trends. The absence of a brand identifier in the database creates a problem of sparsity, in which many SKUs are purchased by only a small number of hospitals, or in only a small number of months. The most thorough method we employed to identify brands, for a subset of products, involved examining manufacturer catalogs, finding likely brand names, searching for similar strings within the item description field, and validating SKUs for those brands against the catalog numbers. This was infeasible for all product categories due to the large number of manufacturers and SKUs. Additionally, many manufacturers' websites were found to be difficult to navigate, particularly once we extended the analysis beyond high-dollar physician preference items. Finally, the item description field was often uninformative as to brand. Hence, we used an algorithmic approach to assign brand identifiers for the other product categories.

Our preferred algorithm implements the Random Effect-Expectation Maximization (RE-EM) estimation method from Sela and Simonoff (2011), which is an adaptation of a recursive partitioning tree algorithm to allow for group effects. With no particular assumption made about the significance of each letter within a SKU, a recursive partitioning tree allows us to obtain overfitting-proof groupings that minimize a 10-fold cross validation error. Furthermore, the group effects in the RE-EM estimation method allow us to control for systematic heterogeneity in price across hospital-time.

Given a transaction $i = 1, \ldots, N$ where N is the size of the dataset within a UMDNS code, price p_i of the transaction, dummy matrix Z_i indicating each transaction's hospital-time group, group effect b_i, and attribute vector $D_i = \{d_{i1}, \ldots, d_{iL}\}$ where d_{il} is the lth digit of the SKU associated with transaction i, the RE-EM proceeds as follows:

1. Initialize estimated group effect $\hat{b}_i$ to zero.
2. Iterate through the following steps until the estimated hospital-time effect $\hat{b}_i$ converges.
   (a) Estimate a regression tree with recursive partitioning on price adjusted by hospital-time group effect, $p_i - Z_i\hat{b}_i$ with attributes D_i. Take the terminal nodes, $j \in J$, of the tree and create an indicator variable, $I(D_i \in j)$.
   (b) Fit a linear model, $p_i = Z_i b_i + I(D_i \in j)\mu_p + \epsilon_i$ and extract $\hat{b}_i$ from the model.
3. Once $\hat{b}_i$ converges, take the final grouping $j \in J$ and use it as the new product identifier for each i.

At each iteration of step (2a), the tree is pruned using 10-fold cross validation at each split; the model retains the simplest tree with cross validation error no more than one standard error away from the tree with the minimum cross validation error.

With this method, we categorized 12,760 SKUs across 19 UMDNS codes into 1,682 RE-EM brands. For surgical staplers and drug-eluting coronary stents, which we validated by hand, we identified 3.8 RE-EM brands per "true" stapler brand, and 0.8 RE-EM brands per "true" drug-eluting stent brand. In Appendix C, we also show sensitivity of our results to an alternative product categorization approach.

B Control Function Selection Correction

The logic behind our selection correction is easiest to follow in a simple model with a single product and an outside good. Consider such a model, where we observe data across hospitals on whether they purchase the product $1\{q_h > 0\}$, and if so, how much (q_h). The hospital's mean (across patients) utility for that product, relative to the outside good, is δ_h. In practice, δ_h can be inferred from quantity purchased and the market size of potential patients in a discrete choice demand estimation exercise; e.g., $\delta_h = \ln(\frac{q_h}{Q_h - q_h})$ for logit demand.

Thus, we observe the distribution of mean utility, conditional on purchase: $F(\delta_h|1\{q_h > 0\} = 1)$. For any counterfactuals involving what would happen if hospitals that do not purchase the product were to purchase it, we also need the complementary conditional distribution: $F(\delta_h|1\{q_h > 0\} = 0)$. This is a counterfactual object about which we have no information without parametric assumptions and/or a source of random assignment.

Our approach to this problem relies on a source of random assignment. Consider a variable z_h which is (conditionally) mean independent of utility $E[\delta_h|z_h] = 0$, but which forces a hospital to purchase the product with probability $Pr[q_h > 0|z_h] = g(z_h)$. Our approach generates a control function that uses the randomness in $1\{q_h > 0\}$ induced by z_h to correct for the selection bias and recover the unconditional distribution $F(\delta_h)$ from the data. We then combine this estimated unconditional distribution and the observed distribution conditional on purchase to obtain the distribution conditional on non-purchase that we need to consider counterfactual changes to hospital choice sets.

B.1 Monte Carlo Evidence on Performance of Control Function Selection Correction

As discussed in Section 3.1.1, we use a control function approach to correct for potential bias introduced by our estimation of preference parameters within an endogenously-formed consideration set. In this Appendix, we use a simple Monte Carlo simulation to illustrate the identification problem and to demonstrate how the control function performs in addressing it.

First, suppose that the consideration set formation process can be well-approximated by the following reduced-form index model:

$$\mathbb{1}(j \in \mathcal{J}_h) := f(\phi_h + \phi_j + Z_{jh}\phi^z + \epsilon_{jh}) \tag{16}$$

where we will specify f using a Probit link or, alternatively, a linear probability model (LPM). Second, in a slightly simplified representation of our demand model, suppose that end user h chooses brand j from the consideration set $\mathcal{J}_h$ upon each use opportunity i to maximize utility represented by:

$$u_{ijh} = \theta_h + \theta_j + \xi_{jh} + \varepsilon_{ijh} \tag{17}$$

The use-specific i.i.d. unobservable ε_{ijh} is the standard type I extreme value error term (with scale normalized to one), and ξ_{jh} is the unobserved average "match" value between hospital h and brand j.

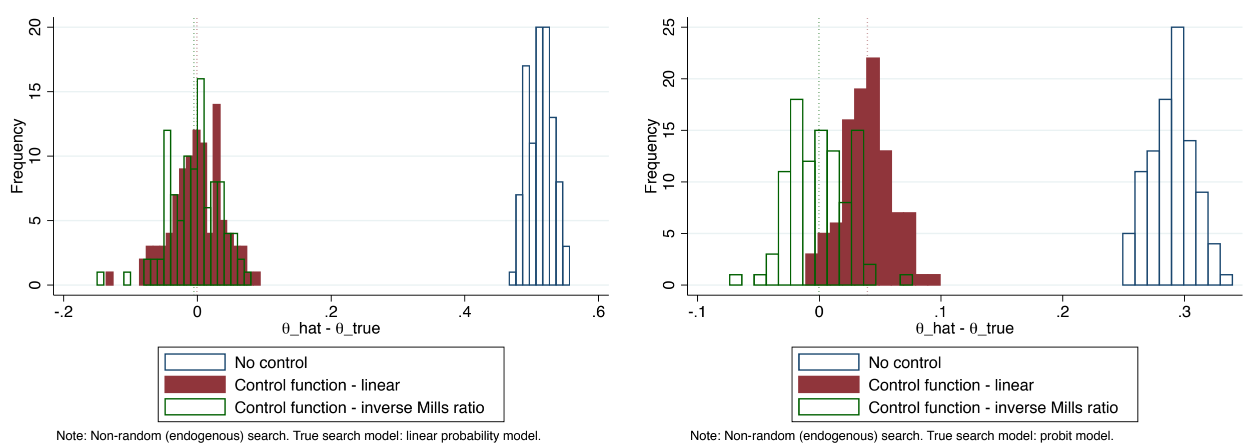
We allow for endogeneity between demand and "search" via correlation between the search cost shock ϵ_{jh} and ξ_{jh}:

$$\begin{pmatrix} \epsilon \\ \xi \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}\right]$$

In our Monte Carlo simulation, each iteration has $N_i = 500$ use cases for each of $N_h = 50$ hospitals forming consideration sets and then choosing among $N_j = 2$ brands. We draw each independent variable from the following distributions: $\theta_h \sim U[.1, .7]$; $\theta_j \sim U[.3, .7]$; $\phi_h \sim U[0, .2]$; $\phi_j \sim U[.2, .4]$; $Z \sim \mathcal{N}(.1, .1)$. Finally, we let ξ_{jh} and ϵ_{jh} be bivariate normal with means zero, variances of 0.5, and covariance $\rho = 0.5$. In Figure A1, we display the distribution of the bias in estimated brand fixed effects: $\hat{\theta}_j - \theta_j$. The left panel imposes that the true reduced-form search model is a linear probability model; the right panel imposes that the true search model is a Probit.

Figure A1: Monte Carlo Results: Distribution of $\hat{\theta}_j - \theta_j$

(a) *LPM* **(b)** *Probit*



The blue bars show the distribution of the bias in the estimated brand fixed effects when demand is estimated assuming no endogeneity. In each panel, the bias is positive, suggesting that a brand's average quality is overestimated when it is inferred only from those hospitals that liked it well enough to put it in their consideration sets. The red bars show the distribution of the bias when we implement a linear control function correction – we regress a dummy for consideration set inclusion linearly on hospital and brand dummies and our excluded instrument Z, then include the residual linearly as a control in the demand model. The green bars show the distribution of the bias when we implement a Probit-based control function correction – we perform a Probit regression of consideration set inclusion on hospital and brand dummies and our excluded instrument Z, then include the inverse Mills ratio of the predicted index as a control in the demand model.

As shown in the Figure, the procedure does a good job of correcting the bias due to consideration set endogeneity when the control function is implemented correctly based on the "true" search model. That is, the estimates based on the linear control function are unbiased when the search model was truly linear, and the estimates based on the inverse Mills ratio control function are unbiased when the search model was truly a Probit. The inverse Mills ratio control function also works well when the search model was truly linear, but the estimates based on the linear control function are still slightly biased upward when the search model was truly a Probit. Thus, the inverse Mills ratio procedure is somewhat more robust to model misspecification.

C Main Results With Different Levels of Product Categorization

In this Section, we present summary statistics, and supply and demand estimates under an alternative model of brand identification. In the "Vendor" results described in the following tables, we instead simply assume that supply and demand are each determined at the vendor level.

Several patterns stand out. In Table A1, we observe smaller choice $\mathcal{J}$ and consideration $\mathcal{J}_h$ sets, as expected. We observe much larger dispersion for the vendor level of aggregation then we did for our RE-EM brands; this is not unexpected, as the vendor level of aggregation involves coarser controls for time trends.

Table A1: Vendor – Summary of Purchasing Categories

	N_h	Annual Spend $1000s	p		$\|\mathcal{J}\|$	$\|\mathcal{J}_h\|$		$Pr[j^* \in \mathcal{J}_h]$	$Pr[j^* = j_h^*]$
			μ	$\frac{\sigma}{\mu}$		μ	$\frac{\sigma}{\mu}$		
Non-PPIs									
Surgical Gloves	758	$85	$0.86	0.26	29	2	0.46	0.38	0.22
Pulse Oximeter Probes	366	$158	$10	0.32	43	3	0.49	0.22	0.15
Liquid Adhesives	696	$60	$17	0.21	46	2	0.52	0.37	0.28
Pneumatic Compression Cuffs	351	$103	$19	0.18	17	1	0.41	0.34	0.30
Trocars	669	$64	$35	0.24	47	4	0.45	0.49	0.19
GI Staples	609	$126	$133	0.23	24	2	0.42	0.61	0.21
Linear Staplers	583	$86	$142	0.22	25	2	0.48	0.30	0.24
Electrosurgical Forceps	453	$168	$523	0.42	18	5	0.43	0.79	0.48
Ablation/Mapping Cath.	324	$432	$880	0.31	12	5	0.35	0.96	0.49
Allografts	369	$230	$956	0.33	34	6	0.42	0.82	0.36
Bone Grafts	393	$446	$2,691	0.32	16	4	0.45	0.97	0.75
Average(11)	506	$178	$492	0.28	28	3	0.44	0.57	0.33
Physician Preference Items									
Patellar Knee Prosth.	470	$100	$414	0.24	14	3	0.47	0.66	0.28
Acetabular Hip Prosth.	516	$276	$1,152	0.23	14	4	0.41	0.80	0.35
Average(2)	493	$188	$783	0.24	14	3	0.44	0.73	0.31

In Table A2, we observe similar supply and demand parameters as in our baseline results in Table 2, with the primary exceptions that we observe higher nesting parameters for both non-PPIs and PPIs, and we estimate higher bargaining splits for non-PPIs. This suggests that there is more substitution among vendors than to the outside good than we observe at the brand level, and we likely underestimate product substitutability when we aggregate to the vendor level by insufficiently controlling for within-vendor variation in quality.

Table A2: Demand and Pricing Parameter Estimates

	p		$\frac{\partial E[\xi]}{\partial \sigma}$	λ	$\theta^p * 1,000$	AV^{CS}		B		$\frac{p-mc}{p}$	
	μ	$\frac{\sigma}{\mu}$				μ	$\frac{\sigma}{\mu}$	μ	$\frac{\sigma}{\mu}$	μ	$\frac{\sigma}{\mu}$
Other Medical/Surgical Supplies (Non-PPIs)											
Surgical Gloves	$0.86	0.26	0.13	0.85	-47.956	$22	0.40	0.17	0.50	1.00	0.00
Pulse Oximeter Probes	$10	0.32	-0.13	0.70	-1.215	$605	0.25	0.20	0.46	1.00	0.00
Liquid Adhesives	$17	0.21	0.35	0.54	-3.257	$241	0.18	0.08	0.36	0.89	0.04
Pneumatic Compression Cuffs	$19	0.18	0.46	0.95	-0.928	$747	0.58	0.07	0.96	0.84	0.04
Trocars	$35	0.24	0.39	0.41	-5.082	$123	0.14	0.14	0.56	0.65	0.27
GI Staples	$133	0.23	0.11	0.69	-0.957	$513	0.36	0.19	0.50	0.73	0.11
Linear Staplers	$142	0.22	0.00	0.71	-0.938	$638	0.40	0.15	0.74	0.47	0.29
Electrosurgical Forceps	$523	0.42	0.40	0.44	-0.204	$3,196	0.19	0.13	0.69	0.69	0.25
Ablation/Mapping Cath.	$880	0.31	0.72	0.00	-0.307	$2,569	0.12	0.18	0.46	0.81	0.18
Allografts	$956	0.33	0.87	0.14	-0.175	$4,699	0.14	0.23	0.38	0.92	0.03
Bone Grafts	$2,691	0.32	0.51	0.15	-0.103	$6,878	0.08	0.16	0.32	1.00	0.00
Average(11)	$492	0.28	0.35	0.51	-5.557	$1,839	0.26	0.16	0.54	0.82	0.11
Physician Preference Items (PPIs)											
Patellar Knee Prosth.	$414	0.24	0.14	0.79	-0.247	$3,810	0.37	0.15	0.73	0.40	0.39
Acetabular Hip Prosth.	$1,152	0.23	0.30	0.56	-0.176	$2,987	0.23	0.19	0.55	0.57	0.24
Average(2)	$783	0.24	0.22	0.68	-0.211	$3,398	0.30	0.17	0.64	0.49	0.31

D Role of Strategic Exclusion

One of our key observations in the raw data is that each hospital sources its medical and surgical supplies from only a small subset of the vendors available in the market. One explanation for this observation is that search and contracting frictions prevent hospitals from making purchases from many different vendors. Another potentially important explanation is that hospitals strategically exclude some vendors in order to strengthen their bargaining leverage. As noted, we do not believe that it is computationally feasible for us to embed both of these mechanisms in our model. In this paper, we focus on the role of contracting costs. This Appendix explores the potential role of strategic exclusion, given the patterns in our data.

We begin by testing whether the observed consideration sets, demand realizations, and prices are consistent with a stability condition on the buyer-supplier networks in the data. In the Nash-in-Nash with Threat of Replacement model of strategic exclusion in Ho and Lee (2018) (hereafter, NiNTR), a given buyer-supplier network $\mathcal{J}_h$ is stable under NiNTR prices if, for every supplier j in h's network, higher bilateral surplus is generated for pair hj by $j \in \mathcal{J}_h$ than by replacement of j with any $k \in (\mathcal{J} \setminus \mathcal{J}_h)$ (holding all other agreements $-hj$ fixed). That is, buyer h's network cannot exclude any supplier k that generates greater bilateral surplus than any included supplier j. We examine an analogous condition, by testing whether hospital h would like to unilaterally deviate by terminating its contract with one of the included brands $j \in \mathcal{J}_h$ and replacing it with one of the excluded brands $k \in \mathcal{J} \setminus \mathcal{J}_h$, paying the brand's reservation price (the minimum price k would be willing to accept to be included in h's consideration set). That is, we ask whether:

$$\pi_h(\mathcal{J}_h, p_{hj}) - \pi_h(\mathcal{J}_h \setminus j, p_{-hj}) \geq \max_{k \in \mathcal{J} \setminus \mathcal{J}_h} \{\pi_h((\mathcal{J}_h \setminus j) \cup k, \{p_{-hj}, p_{hk}^{res}\}) - \pi_h(\mathcal{J}_h \setminus j, p_{-hj})\}$$

where the price of the excluded brand p_{hk}^{res} is one that makes the brand's vendor indifferent between selling and not selling. Note that this condition deviates somewhat from the condition in Ho and Lee (2018) in that we examine network stability at *observed* prices.[58]

To understand how well this stability condition holds in our data, we calculate gains-from-trade for each hospital and brand pair using the demand and cost estimates presented in Section 4. As with our estimation of search costs in Section 3.3 and our decompositions in Section 4.3, this exercise requires that we calculate gains-from-trade for counterfactual hospital-brand pairs. Naturally, the implied stability violations hinge crucially on whether ξ_{hk} for $k \in \mathcal{J} \setminus \mathcal{J}_h$ is particularly low. Accordingly, we explore a range of alternative

[58] Also unlike in Ho and Lee (2018), in our model, negotiated payments do not enter linearly into buyers' profits, making bilateral surplus a more complicated function of price.

assumptions regarding hospital h's unobservable match value ξ_{kh} for brand k not in their observed choice sets. We then calculate the fraction of markets (hospital-years) that violate the stability condition for each product category under each set of assumptions.

Each of these exercises relies on the procedure outlined in Section 3.3 that identifies the *conditional distribution* of ξ_{hk} for $k \in \mathcal{J} \setminus \mathcal{J}_h$, correcting for selection using plausibly exogenous variation in consideration sets introduced by our exposure instruments. When we perform this test with each ξ_{kht} generated by a random draw from the estimated conditional distribution, almost every market violates stability. Though this seems to us like the correct intuitive exercise, most hospital choice sets are at least an order of magnitude smaller than the full possible set of suppliers, implying that it is almost guaranteed that at least one of these draws will make it worth adding the product with the high draw in place of the worst product in a choice set. In a way, this speaks how difficult it would be to rationalize the data with strategic exclusion alone and no search/contracting costs.

To shed more light on this issue, we next explore just how extremely negative the match values ξ for unobserved hospital-brand combinations would have to be to rationalize the data with strategic exclusion alone. We perform the same test, fixing all of these counterfactual ξ to relatively low values, specifically the 25th and 1st percentiles of the estimated conditional distribution of ξ_k for $k \in \mathcal{J} \setminus \mathcal{J}_h$. Figure A2 Panel (a) shows the fraction of markets violating stability for these two tests. Each bar around the markers represents one standard error estimated from a nonparametric bootstrap, resampling at the market level.

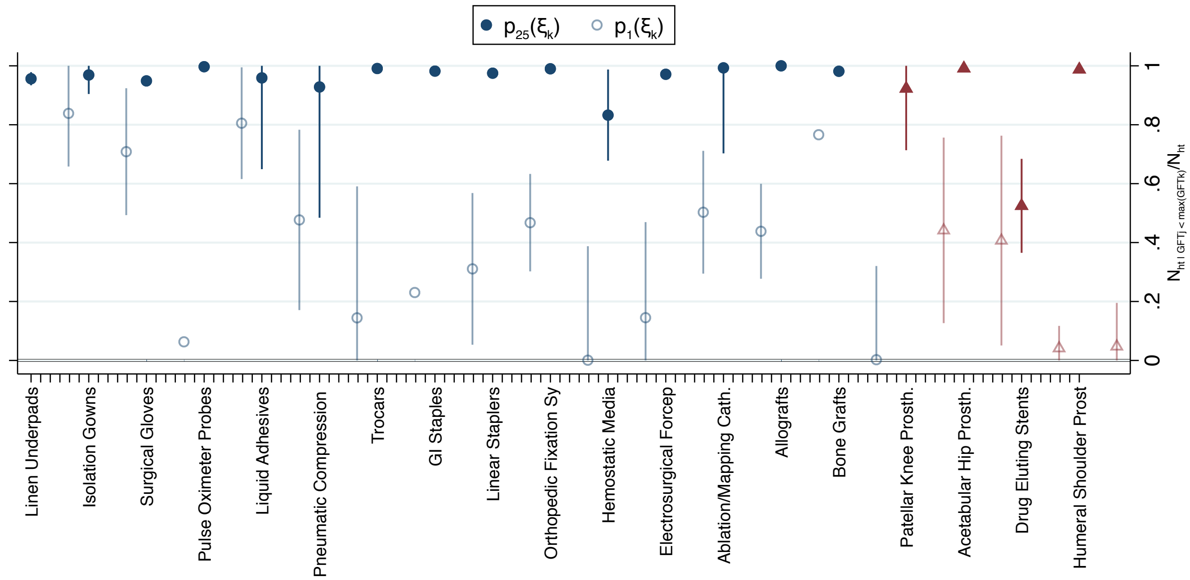
The solid markers summarize the results assuming the matches are all at the 25th percentile, and all product categories but one still violate stability in over 80 percent of markets. The single exception is drug-eluting stents, for which the total market is small $|\mathcal{J}| = 10$ and consideration sets are relatively large $|\mathcal{J}_h| = 4$ on average. That is, even given these pessimistic match values, it is still true that there is almost always an unpurchased product of higher value than some purchased product. The hollow markers summarize the results using the even more extreme 1st percentile, and here we see 8 of 19 categories start to see fewer than 25 percent violations, while 4 are still at 75 percent or above.

Figure A2 Panel (b) summarizes how extreme unobserved match values ξ_{kht} must be for fewer than 25 percent of hospital-year markets to be estimated to be unstable in each product category. For PPIs, these draws must be fixed at least 1.43 standard deviations below the mean; the analogous number for non-PPIs is 2.03 standard deviations.

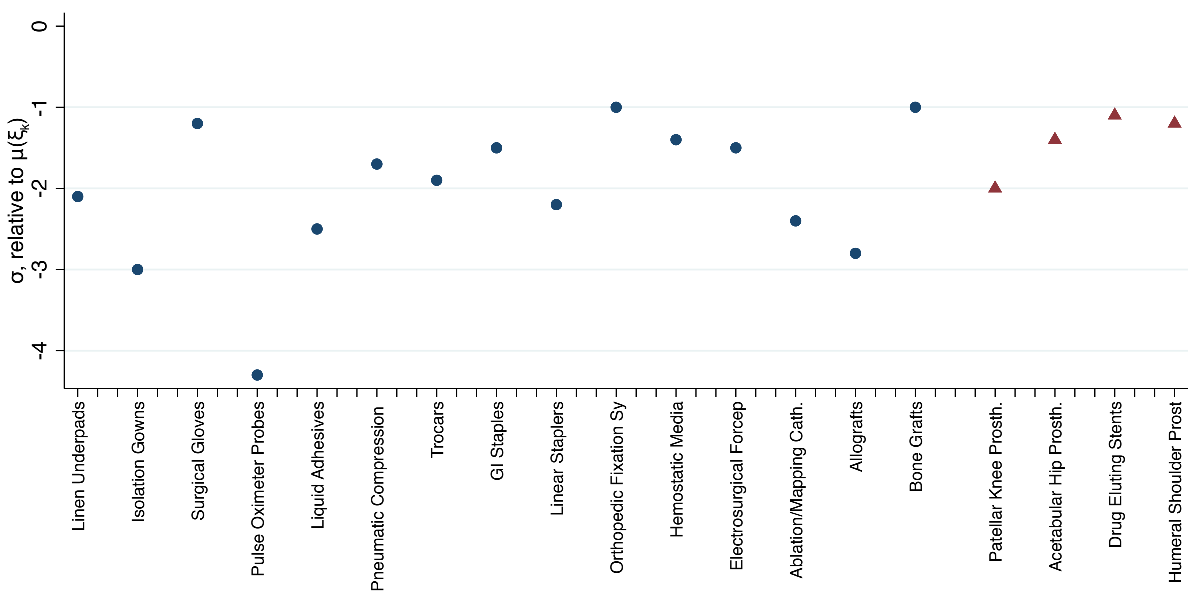
These results provide suggestive evidence that it is difficult to reconcile the consideration sets and prices we observe with a model of strategic exclusion *alone*. That said, the tests have caveats: in particular, they rely on supply and demand parameters inferred from an empirical Nash-in-Nash procedure, and it may be that small consideration sets are driven

Figure A2: Stability violations

(a) ξ_k fixed at $p_{25}(\xi_k)$, $p_1(\xi_k)$



(b) SD below $\mu(\xi_k)$ s.t. <25% markets unstable



by a combination of strategic exclusion and search/contracting frictions. We consider this an important topic for future research.

E Other Tables and Figures

Table A3: Correlation between p_{jht} and X_{ht} – Detail

	Hosp. Beds	Sys. Beds	Hosp. q	Sys. q	Full Std.	Almost Std.	Any GPO	Big GPO
Bone Grafts	-.0079*	-.0025	.0038	-.0135***	-.222	-.0099	-.0123	.0075
	(0.003)	(0.004)	(0.004)	(0.004)	(0.172)	(0.013)	(0.010)	(0.007)
Ablation/Mapping Cath.	-.0018	-.0083	-.0112*	-.0014	.1206**	.0107	-.0267	.0373***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.040)	(0.035)	(0.015)	(0.009)
Allografts	-.0109	-.0116*	.0042	-.0084	.1751*	.0082	-.016	.0149
	(0.006)	(0.005)	(0.007)	(0.007)	(0.084)	(0.015)	(0.015)	(0.013)
Electrosurgical Forceps	-.005	-.0033	-.0085	.0014	-.0384	-.0078	.0134	-.0133
	(0.005)	(0.004)	(0.006)	(0.004)	(0.022)	(0.009)	(0.016)	(0.010)
Pulse Oximeter Probes	-.0009	-.0058	-.0151	-.0135	-.0328	.0009	-.0327	.0341
	(0.008)	(0.008)	(0.010)	(0.010)	(0.053)	(0.015)	(0.033)	(0.022)
Orthopedic Fixation Systems	-.0187*	-.0075	-.0247*	-.0111	0	.0126	.0272	-.0005
	(0.008)	(0.009)	(0.013)	(0.009)	(0.000)	(0.040)	(0.027)	(0.015)
GI Staples	.011	-.0216***	.0041	-.0066	-.0189	.0164	-.0068	.0125
	(0.008)	(0.005)	(0.006)	(0.005)	(0.032)	(0.010)	(0.026)	(0.014)
Hemostatic Media	-.006	-.0056	.0039	-.022**	.0058	0	-.0005	.0148
	(0.004)	(0.004)	(0.005)	(0.007)	(0.014)	(0.006)	(0.010)	(0.008)
Pneumatic Compression Cuffs	-.0097	-.0341*	.0154	-.0124	-.0218	-.0084	-.0009	.0397
	(0.010)	(0.016)	(0.012)	(0.015)	(0.028)	(0.022)	(0.050)	(0.027)
Surgical Gloves	.0035	-.0074**	-.0006	-.001	.0462	.0117	.0134	-.0206**
	(0.003)	(0.003)	(0.001)	(0.001)	(0.025)	(0.009)	(0.012)	(0.008)
Linear Staplers	.0004	-.0093	.0012	-.0091	-.0605	-.0193	-.0109	.0011
	(0.005)	(0.006)	(0.005)	(0.005)	(0.038)	(0.018)	(0.017)	(0.013)
Trocars	.0017	-.0176**	-.0036	-.0062	-.1388***	.0419*	-.0104	.0079
	(0.008)	(0.006)	(0.007)	(0.007)	(0.041)	(0.021)	(0.020)	(0.014)
Liquid Adhesives	.0107	-.0104	-.0236	-.0168*	.0134	-.0147	-.0783***	.0679***
	(0.007)	(0.007)	(0.014)	(0.007)	(0.016)	(0.010)	(0.023)	(0.018)
Linen Underpads	-.0074	-.0033	-.0103*	.0046	-.0081	.0064	.0337	-.0589***
	(0.004)	(0.006)	(0.005)	(0.007)	(0.012)	(0.009)	(0.020)	(0.017)
Isolation Gowns	-.0121	-.0147**	-.0064	-.0207**	-.0071	-.0041	-.0215	.0372*
	(0.006)	(0.005)	(0.007)	(0.007)	(0.016)	(0.013)	(0.025)	(0.015)
Non-PPI	-.0038*	-.0074***	-.0067**	-.0077***	-.0113	.0038	-.0085	.0131***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.008)	(0.004)	(0.006)	(0.003)
Drug Eluting Stents	.0007	-.0054	-.001	-.0086**	-.0026	.0016	-.0105	.004
	(0.002)	(0.004)	(0.003)	(0.003)	(0.018)	(0.005)	(0.008)	(0.007)
Acetabular Hip Prosth.	.017*	.0052	-.0588***	-.0484***	.0206	-.0516	.0248	.0309
	(0.007)	(0.007)	(0.012)	(0.008)	(0.133)	(0.040)	(0.023)	(0.017)
Humeral Shoulder Prosth.	-.017	-.0001	-.0137	-.0321***	0	.0185	.0531	.0052
	(0.014)	(0.009)	(0.011)	(0.007)	(0.000)	(0.052)	(0.035)	(0.019)
Patellar Knee Prosth.	.0056	.0114	.006	-.0658***	-.1795***	.0381	.039	.1***
	(0.013)	(0.012)	(0.021)	(0.012)	(0.043)	(0.021)	(0.032)	(0.026)
PPI	.0107*	.0045	-.0315***	-.0374***	-.0574*	-.0079	.0289	.022*
	(0.005)	(0.005)	(0.007)	(0.005)	(0.023)	(0.009)	(0.016)	(0.011)

Table A4: Hospital Characteristics Statistics

	N_{jht}	N_h	Hosp. Beds		Sys. Beds		Hosp. q		Sys. q		Full	Almost	Any	Big
			μ	σ	μ	σ	μ	σ	μ	σ	Std.	Std.	GPO	GPO
Bone Grafts	10,199	393	434	269	1,412	1,439	173	184	326	320	0.00	0.07	0.93	0.76
Ablation/Mapping Cath.	18,071	324	484	274	1,485	1,426	466	482	906	1,009	0.00	0.01	0.94	0.77
Allografts	15,145	369	424	266	1,498	1,447	224	238	497	560	0.00	0.05	0.93	0.73
Electrosurgical Forceps	10,734	453	417	274	1,521	1,495	257	285	609	634	0.01	0.23	0.94	0.76
Pulse Oximeter Probes	4,596	366	402	297	1,306	1,249	14,198	28,502	29,652	54,933	0.10	0.65	0.94	0.76
Orthopedic Fixation Systems	22,509	441	438	268	1,531	1,480	314	463	663	719	0.00	0.01	0.94	0.75
GI Staples	12,257	609	359	256	1,454	1,532	907	1,479	2,257	2,846	0.02	0.12	0.93	0.73
Hemostatic Media	2,340	290	469	283	1,420	1,356	299	412	540	763	0.12	0.32	0.91	0.75
Pneumatic Compression Cuff	2,565	351	315	256	1,194	1,451	5,474	8,380	9,998	12,290	0.22	0.74	0.94	0.76
Surgical Gloves	17,300	758	307	255	1,444	1,525	101,077	1,450,752	370,888	3,195,868	0.04	0.18	0.93	0.71
Linear Staplers	10,996	583	340	249	1,377	1,477	580	847	1,480	1,790	0.06	0.13	0.94	0.73
Trocars	15,404	669	330	253	1,440	1,501	1,811	2,214	5,485	6,747	0.01	0.07	0.93	0.70
Liquid Adhesives	4,972	696	319	256	1,319	1,417	3,210	5,991	9,130	12,682	0.27	0.67	0.93	0.71
Linen Underpads	3,720	602	323	258	1,336	1,470	96,804	187,758	291,943	639,621	0.39	0.71	0.93	0.75
Isolation Gowns	2,261	501	320	263	1,249	1,373	41,521	121,072	102,492	209,221	0.52	0.83	0.94	0.72
Non-PPI	10,205	494	386	271	1,431	1,463	15,210	491,668	52,450	1,089,831	0.05	0.19	0.93	0.74
Drug Eluting Stents	4,430	351	441	277	1,508	1,437	677	635	1,556	1,758	0.02	0.20	0.94	0.74
Acetabular Hip Prosth.	28,305	516	360	257	1,418	1,473	239	275	652	765	0.00	0.03	0.94	0.73
Humeral Shoulder Prosth.	13,501	321	424	260	1,340	1,268	102	75	206	207	0.00	0.00	0.94	0.73
Patellar Knee Prosth.	6,310	470	335	249	1,400	1,462	226	514	613	806	0.07	0.30	0.94	0.71
PPI	13,136	414	407	271	1,459	1,438	463	562	1,102	1,466	0.01	0.14	0.94	0.73