Competition and Collusion in the Generic Drug Market*

Emily Cuddy†

December 2021

Abstract

This paper examines the role of competition in the US generic drug market. Long considered one of the rare bargains of the US healthcare system, the generic drug market experienced a series of unprecedented price spikes over the past decade, which have been linked to the collusive conduct of manufacturers in the so-called "largest corporate cartel in US history." Using unsealed records of the ring's activity, I examine how collusion—in combination with other contemporaneous features of the market—affected upstream drug prices. I design a model of retail drug procurement, wherein generic manufacturers submit bids to supply national pharmacies with their drugs, and estimate this model using an estimator which accepts highly aggregated data of firms' winning bids. In doing so, I recover novel estimates of manufacturers' costs. Counterfactuals indicate that the collusive ring generated over $12 billion in surplus for itself over 18 months and, moreover, that the unprecedented backlog of generic drug applications at the Food and Drug Administration in this period exacerbated the situation. I conclude by discussing the effects of recently proposed pro-competitive policies in view of my findings.

I thank Bruno Baranek, Stefano Baratuche, Nick Buccholz, Matt Cocci, Janet Currie, Kate Ho, Bo Honoré, Adam Kapor, Jakub Kastl, Parth Parihar, Rob Porter, and Pëllumb Reshidi for advice and suggestions that greatly benefited this research. I also thank many participants in seminars at ASU, BU, Chicago Fed, Columbia University, HBS, US Department of Justice, US Department of the Treasury, Duke, Northwestern, NYU Stern, Princeton, Stanford University, University of Georgia, University of Louisville, University of Maryland, UC Berkeley, UC Berkeley Haas, and UC Irvine. Additional thanks to Robin Feldman for sharing data, John Coburn for helpful discussions, and the many generic drug executives for their willingness to speak with me. All errors are my own.

†Kellogg School of Management, Northwestern University. Email: emily.cuddy@kellogg.northwestern.edu

1 Introduction

The role of competition in the generic drug market has attracted considerable attention in recent years. Long considered one of the rare bargains in the US healthcare system, generic drugs have been widely accepted by US consumers. Over 90% of pharmaceutical drug prescriptions filled in 2017 were for generic formulations, up from just 10% in the 1980s (Aitken and Kleinrock, 2019; Berndt and Aitken, 2011). With widespread adoption has come massive cost savings, to the tune of more than $293 billion annually, in large part due to robust competition among generic manufacturers in the market.[1]

Nevertheless, conventional wisdom regarding the competitiveness of the generic market is now being challenged. Over the past decade, the market has started to show signs of distress, beginning in 2009 with recurring drug shortages and continuing into 2013 with a series of unexpected price increases for hundreds of generic drugs (Berndt et al., 2018). For instance, the price of a 30-day prescription of clomipramine hydrochloride 50 mg capsules—a tricyclic antidepressant used for obsessive compulsive disorder—went from $0.34 per capsule to $8.43 in just the first quarter of 2013 (GAO, 2016). These phenomena have drawn considerable attention from the public and researchers alike in hopes of identifying the root causes of the market dysfunction.

In this paper, I study the role that competition plays in the generic drug market when it is functioning well and and when it is not. I develop and estimate a model of the retail pharmacy procurement process, wherein generic manufacturers submit bids to supply national pharmacies with their drugs, and estimate the model using both public and private data from the Centers for Medicare and Medicaid Services (CMS), the Food and Drug Administration (FDA), and a large private health insurance plan provider. I supplement these data with qualitative data from interviews I conducted with generic drug executives.[2]

In a world where generic manufacturers' costs are a closely guarded secret, my model allows me to estimate them directly. Indeed, one contribution of the paper is to show how manufacturers' costs vary across different drugs and retail pharmacies. Another is methodological: a modification of the estimator of Laffont et al. (1995) (henceforth, LOV) permits the use of highly aggregated bid data—the average winning bid in the market—in the auction estimation routine. This modification may be applicable in many other empirical auction settings, where the winning bid is trade secret and unavailable to researchers. A third contribution is in identifying and quantifying the mechanism by which competition lowers prices in equilibrium in this market. In doing so, I shows that entry is not a panacea for high upstream prices, as there is a marked decline in the marginal effect of competition on prices beyond the third entrant. That said, the usefulness of my structural approach is most apparent in subsequent counterfactual analyses, where I use the recovered distributions of manufacturers' costs to analyze how recent "anti-competitive" developments in the generic market

[1] These savings refer to the difference in price between a branded and generic version of a drug. Note that this statistic is from 2018, as it is the most recent year for which this estimate is available.

[2] I held a series of anonymous informational interviews with executives at five different generic manufacturers during the spring of 2018. These interviews helped me to capture the most important features of the procurement process and manufacturer costs in my model.

may have affected upstream prices.

There were two concurrent issues at hand in the US generic market in the early 2010s: insufficient entry into generic markets and an active collusive ring of generic manufacturers who formed the "largest corporate cartel in US history."[3] In the case of the former, an underfunded and understaffed FDA was unable to keep pace with the flood of Abbreviated New Drug Applications (ANDAs)[4] it received as several blockbuster branded drugs' marketing exclusivity expired.[5] Within months, a backlog of over 3,000 applications had accumulated, which meant that new entry into generic markets was effectively at a standstill. In the case of the latter, a ring of 16 generic firms, among the largest in terms of ANDA holdings, actively colluded in order to maintain supra competitive prices while retaining their market share. Price data from this period suggests they may have been successful in their efforts, as there were sustained price increases in over 200 drug markets for the 18-month period when the cartel was active. In fact, it is likely that the ring would have carried on had it not been discovered through the concerted efforts of the State of Connecticut to identify the source of the suspicious price increases.[6]

In my two main counterfactual exercises, I use the estimated model to simulate the equilibrium effects of (i) reducing the FDA's application backlog by six months and (ii) removing the collusive ring from the market. To do so, I construct counterfactual price series with which to evaluate the ensuing changes in retail pharmacy spending. My results suggest that both the backlog and the collusive scheme resulted in significantly higher upstream prices in the generic market; however, the collusive ring was, by far, more damaging. I find that a six-month delay in the approval (and subsequent entry) of a new manufacturer cost pharmacies an average of $60 thousand per drug whereas six months of collusive activity cost pharmacies an average of $2.5 million per drug. Indeed, for my sample alone, I estimate the damages of the collusive ring during its 18-month existence to be $2.2 billion. However, considering that my data include only a subset of total generic drug purchases in the US, a back-of-the-envelope estimate of the total market damage imposed by the ring is over $12 billion. Thus, while my findings provide support for the FDA's recent efforts to facilitate entry into the generic drug market, they also provide clarity of the stakes for identifying policies to manage extraordinary manufacturer conduct.

Contribution to the literature. This paper contributes to two distinct literatures. The first comprises analyses of the US prescription drug market. Many early studies are, in large part, motivated by understanding firms decision making in the time of initial generic entry, such as which markets to enter or what prices to charge (Morton, 1999; Acemoglu and Linn, 2004; Saha et al.,

[3]This is a quote by William Tong, the Attorney General of Connecticut, who is currently in charge of the pending lawsuits against the ring. The quote is from a November 2019 interview, which can be found here: https://medicine.yale.edu/news-article/21752.

[4]A generic manufacturer must receive FDA approval of an ANDA before it can begin marketing a drug in the US.

[5]This period is often referred to as the first "patent cliff" in the generic drug market; a second patent cliff began in late 2019.

[6]The non-public investigation commenced in July 2014; 48 states and US territories would eventually join the investigation.

2006; Grabowski and Vernon, 1992; Frank and Salkever, 1997). Newer studies look at how pharmaceutical firms react to major policy changes. These papers examine not only demand-side policies, such as how branded drug prices fell in response to the introduction of Medicare Part D (Duggan and Scott Morton, 2010), but also supply-side policies, such as how firms reacted to the FDA's new fee program as part of its Generic Drug User Fee Amendments (GDUFA) I and II initiatives (Berndt et al., 2018).

While my paper speaks to some of the issues raised in these earlier papers, it contributes more directly to a nascent strand of the literature on equilibrium pricing behavior in the generic market (Dave et al., 2017; Conrad and Lutter, 2019). In practice, these papers tend to be more descriptive, using claims data to document systematic differences between average invoice prices and competition, as proxied by the number of manufacturers who appear in the data; however, there are two notable exceptions. Reiffen and Ward (2002) estimate a system of simultaneous equations to estimate the relationship between entry and profitability, and Olson and Wendling (2013) use an instrumental variables strategy which exploits the 180-day marketing exclusivity period granted to generic firms in order to estimate the causal effect of entry on prices.[7]

The approach I take in my paper—modeling the upstream market as a series of simultaneous procurement auctions—breaks from the typical assumptions made about how generic firms compete in generic markets. This modeling choice was deliberate. In the interviews I conducted with generic manufacturers, they revealed the widespread use of procurement systems in the retail pharmacy industry. Incorporating this insight, I build a flexible, realistic framework through which I am able to study not only the equilibrium relationship between competition and prices but also two of the most significant developments in the generic market over the past decade. To my knowledge, I am the first to examine the economic impact of the collusive ring's activities in the market, which may be unsurprising as many of the lawsuits against the ring are still pending in court.

My paper also contributes to a vast literature on empirical auction estimation (see Athey and Haile (2002) for a survey). As noted earlier, I follow LOV in using information on the winning bids from auctions to back out a parametric distribution of firm costs; however, I depart from their framework in order to accommodate the limited data available in my setting. Here, my innovation is to show (via simulation) that an average of winning bids in the market is sufficient to recover parametric distributions of firms' costs.

In its treatment of the collusive ring, my paper also adds to the existing literature on collusive behavior in an auction setting (see Hendricks and Porter (1989) for an early survey). While many of these papers are focused on detection of collusion in auctions (Athey et al., 2011; Bajari and Ye, 2003; Porter and Zona, 1993, 1999), I follow the approach of Asker (2010), using proprietary data from law enforcement agencies on the inner workings of a previously identified cartel to quantify damages.

The paper proceeds as follows. Section 2 provides background on the industry and introduces

[7]These challenges are termed "Paragraph IV" ANDAs. This provision of exclusivity was provided under the the Hatch Waxman Act of 1984, which regulates FDA approval requirements for generic drugs, in order to protect against questionable patents filed by branded drug manufacturers.

the data used in the analysis; Section 3 introduces a model of generic drug procurement at retail pharmacies. Section 4 discusses how I bring the model to the data, discussing my estimation strategy in great detail. In Section 5 and Section 6, I present my main results and context and estimates for my counterfactuals, respectively. Section 7 discusses the policy implications of this work and concludes.

2 Empirical Setting

In this paper, I examine the upstream retail generic prescription drug market, wherein generic manufacturers sell their products to retail pharmacies.[8] A simplified representation of the market is presented in Appendix Figure A.1.

2.1 Background

Retail pharmacy chains, like CVS and Walgreens, acquire their generic drugs through the use of a nationwide procurement auction.[9] Internally, each pharmacy chain maintains a list of every available generic drug product (e.g., escitalopram 10 mg oral tablets), as well as its expected annual usage.[10]

Upon the entry or exit of a manufacturer, a pharmacy chain will distribute electronically, often via email, a formal request for proposals from every generic manufacturer actively marketing that particular drug. In their proposal, manufacturers must provide information on the best possible price (per pill) that they are willing to sell the drug, as well as the share of the reported annual usage amount that they can reasonably provide on a timely basis. After the pharmacy receives all the bids from manufacturers, it awards the contract to whichever pharmacy submits the lowest per-pill price.[11]

Although the timing and existence of an auctions, as well as the particular details of the contract, are public knowledge—manufacturers receive the same information in the request for proposals–manufacturers' bids remain private and reflect their private costs of goods sold. As I discuss in my empirical application in Section 3, this information structure is important for modeling the auction mechanism as a first-price sealed-bid reverse auction with private values.

2.2 Data

In the project, I use five distinct data sets. The first three yield inputs for the main auction estimation; the remaining two provide institutional details for the counterfactuals. In what follows,

[8] There are separate procurement streams for facility settings; we do not examine those here.

[9] Prior to the use of the procurement system, the system was more relationship-based, whereby one firm would contract specifically with a single pharmacy for all of its products.

[10] Proprietary data suggests that these annual usage amounts are highly correlated year-over-year.

[11] With the exception of the most popular generic drugs with extraordinary annual volumes, a pharmacy will single-source a product from one generic manufacturer.

I describe how each dataset is used in the analysis. Additional details on the data construction process are available in Appendix Section B.

2.2.1 Auction Data

Winning Bids. Information on the winning bids are from the National Average Drug Acquisition Cost (NADAC) survey administered by the Centers for Medicare and Medicaid Services (CMS). These data report the weekly national average acquisition cost for each outpatient drug covered by Medicaid beginning in October 2012.[12] Each month, CMS contacts a random sample of 500-600 retail pharmacies—both independent and chain pharmacies—and requests the invoice data associated with their previous month's drug purchases.[13] After processing and quality assurance checks, the price data are grouped by active ingredient, strength, dosage form, and route of administration and then an average price is calculated for each group. Importantly, there is only one reported price per generic drug, so price observations from different manufacturers' versions of a particular product are averaged together.

Given this sampling strategy, the reported price in the NADAC survey is effectively a weighted average of pharmacy chain prices, where the weights correspond to the sampling probabilities of each pharmacy chain. I convert these price series data into average winning bid data by identifying distinct periods corresponding to a cluster of auctions for a particular drug, where a cluster refers to a group of auctions held at different pharmacies in the same time period. Most often, these periods begin with the entry (or exit) of a new (old) firm, consistent with what was told to me in interviews with manufacturers. I define the average winning bid price for a particular cluster of auctions as the average NADAC over that period, excluding the first three months of the price series.[14] I show this process graphically for two representative auction clusters in Appendix Figure B.3.

Pharmacy Weights. Information on the sampling weights used in the NADAC calculation are from the National Plan and Provider Enumeration System (NPPES), the same registry used by CMS to draw its monthly sample. This database includes detailed information, including name, address, parent company, and taxonomy, for each retail pharmacy by its National Provider Identifier (NPI). I use historical extracts of the NPPES database on the Wayback Machine to identify (1) the entry and exit dates of each retail pharmacy location and (2) the chain to which the pharmacy belongs.[15] I define the monthly sampling weight of a pharmacy chain as its share of pharmacy locations in that month, where only pharmacies with active NPPES records are included in the calculation. As shown in Appendix Figure B.5, these shares are extremely stable over time.[16]

[12] CMS began publishing NADAC "draft" files in October 2012, in order for the public to review and to comment on the survey design. The first official file was published in December 2013.

[13] Although the survey is administered on a monthly basis, the price data are updated each week to reflect any intra-month price changes which are reported by pharmacies directly to CMS via its NADAC help desk.

[14] I experimented with other lead in times. The results were robust to this decision.

[15] Independent pharmacies are aggregated into a single purchasing group, approximating a large wholesaler.

[16] That said, I do observe a jump in CVS's market share in the wake of its acquisition of Target's pharmacy business.

Auction Characteristics. Information on the characteristics of each drug product are from the National Drug Code (NDC) database maintained by the FDA. This database includes detailed information on each approved drug product at the granularity of the NDC. Given that records are removed from the database as soon as a drug is withdrawn from the market, I use an internal FDA version of the NDC database, which also includes deactivated drug records.[17] From these data, I extract all relevant features to control for auction heterogeneity in the estimation. In particular, I include the form (e.g., capsule), delivery mechanism (e.g., orally dissolving), strength (e.g., 5 mg), and active ingredient (e.g., atorvastatin), which I use to identify the drug's Anatomical Therapeutic Chemical (ATC) classification. The data also include the drug's labeler and ANDA—I use these variables to exclude any drugs marketed by repackaging firms.[18]

Auction Volume and Identity of Winning Bidder. Information on the volume of drug product sold at each pharmacy is estimated directly from a national sample of claims data from a large private health insurance provider, which sponsors both small and large employer-sponsored health plans as well as public insurance plans on behalf of state exchanges, Medicaid, and Medicare Advantage. Enrollees in these plans are geographically diverse, as the insurer has significant market share in all 50 states. The breadth of these claims is critical, insomuch as they serve as a quasi-census of the drugs on each pharmacy chain's shelves. In particular, for each date, I identify which manufacturer's products are being dispensed at which particular pharmacy chain in order to determine which firm won that pharmacy's procurement auction.

While comprehensive, these data do have the limitation that they represent only a subset of all retail pharmacy purchases. To the extent that the insurer's consumers' pharmacy demand is not systematically different from that of other US consumers, my use of a subset of the population should not introduce bias into the estimation.[19] That said, using a subset of retail pharmacy purchases, however large, does make estimating market-level quantities (e.g., damages) more complicated, as it is unclear which benchmark to use for scaling up results. In Section 6.4, I describe and apply one such benchmark, which is based on the insurer's population share. In total, the claims data comprise the drug purchases of over 320 million individuals, where each claim includes the NDC, date of purchase, and NPI of the rendering provider (i.e., the pharmacy which dispensed

See here: https://corporate.target.com/article/2015/12/cvs-target-acquisition-complete.

[17] These data contain information on all drugs which appeared in the FDA NDC Files before November 2017, the month in which I was sent the extract.

[18] I exclude these products from the analysis because they are distributed through non-retail channels, i.e., at physician offices or facilities. Typically, a repackaging firm will purchase a drug directly from a manufacturer and then repackage the drug into unit doses, i.e., 500 pill bottle to 500 individually packaged bar-coded, non-reusable containers.

[19] One potential source of systematic differences in pharmacy demand are preferred pharmacy networks, where the insurer requires enrollees to pay more for their drugs if they choose to buy them at pharmacies outside of the preferred network. Starc and Swanson (2018) show that plans with these networks do have "steering ability," insomuch as they are successful in directing consumers to chains within the network. Nevertheless, despite widespread adoption in Medicare Part D plans, these types of networks were not adopted by this insurer until late into the sample period. I discuss a robustness check of this assumption in the results section.

the prescription).

Summary statistics. For the purposes of the analysis, I define a drug as the unique combination of molecule-dosage form-strength (e.g., atorvastatin 25 mg tablet). Although this product definition is more granular than that used in other contemporaneous studies of the generic drug market (see, for example, Berndt et al. (2017)), it is the one which is most applicable to the retail pharmacy procurement setting.[20]

Notably, the drugs in my sample do not represent the universe of generic drugs, as I apply several restrictions in the process of constructing the final data set. Most importantly, I include only oral solid drugs, i.e., pills. This choice was deliberate. Oral solid drugs share a common production process and so—as will be shown in Section 3—it is more natural to assume that manufacturers' costs can be captured in a single parametric expression (see Sacher and Khinast (2016) for a survey).[21] Moreover, as these drugs are shelf-stable, they are the ones most commonly dispensed in a retail pharmacy setting.[22] From the drugs that remain, I exclude all branded generic products, as they are, by definition, produced by a single firm and thus are procured by pharmacies in a different manner (see, for example, Shrank et al. (2009); Kesselheim et al. (2016)).

The main estimation sample comprises 593 generic drugs, which correspond to 4,515 unique auction clusters. Summary statistics for these drugs are presented in Table 1. In total, the average drug in the sample has an annual volume of 9.6 million units and annual sales $2.365 million, yielding an average price per pill—the price metric used throughout the rest of the paper—of just over 50 cents. None of these drugs are novel molecules; the branded drugs which preceded these generics were brought to market nearly two decades ago. Given that most branded exclusivity periods last around six to eight years, this implies that most of these generic drugs have been around for over a decade. Around three-quarters of the drugs are available in capsule form, and about one-fifth have some sort of specialized delivery mechanism. Often, a drug's active ingredient is available in multiple forms—4.6 on average–reflecting unique delivery mechanisms (e.g., extended release and immediate release) or, more commonly, strengths (e.g., 10 mg and 45 mg). Recall that these forms are treated as separate products.

[20] As part of the interviews with manufacturers, I reviewed the bid request forms from two of the largest pharmacy chains. These forms requested bids at the level of molecule-dosage form-strength. Consistent with this fact, manufacturers noted that they often did not win a pharmacy's business for all strengths of a drug, even though they submitted bids for all strengths.

[21] Inhalants, opthalmic, and topical drugs, in contrast, tend to be produced by firms who specialize in only those dosage forms.

[22] Injectable and infusible drugs are often purchased and used in outpatient and facility settings and require special storage procedures, such as refrigeration.

Table 1: Summary Statistics

Variable	**Mean**
Annual Volume (mil.)	9.632
Average Sales (mil.)	2.365
Years Since Initial Branded Entry (as of 2016)	19.099
Average Price per Pill	0.518
Capsule	0.261
Complex	0.125
Delivery Mechanism	0.207
Extended Release	0.157
Orally Dissolving Tablet	0.024
Delayed Release	0.020
Other	0.007
Unique Forms	4.460
Average N Bidders	4.737
Appearances	1.548
Entry: Post 2010	0.766
During Collusion	0.317
After Collusion	0.578
N	593

Notes: The above table provides summary statistics for the drug products which comprise my main estimation sample. Data on drug volume come from a private insurer's pharmacy claims data, and generic manufacturer entry from FDA Orange Book files. *Capsule-Other* and *Entry* are indicators.

Table 2 shows a breakdown by the one-digit World Health Organization's (WHO) Anatomic Therapeutic Classification (ATC). The largest classes are drugs treating ailments within the cardiovascular system (e.g., statins) and nervous system (e.g., antidepressants), both in terms of sales volume and revenue. Of note, there is one major ATC missing in the sample, which is dermatologicals. Their absence is due to the fact that they are almost exclusively offered in topical preparations, and so they do not meet the pill form requirement to be included in the sample.

Table 2: Sample breakdown by ATC-1

	Percentage	Q-Weight	$-Weight
A: Alimentary tract and metabolism	6.4	8.4	6.3
B: Blood and blood forming organs	0.8	0.0	0.2
C: Cardiovascular system	24.8	35.5	20.5
G: Genito-urinary system and sex hormones	2.4	0.2	0.4
H: Systemic hormonal preparations	0.8	0.5	1.1
J: Antiinfectives for systemic use	5.7	2.5	5.7
L: Antineoplastic and immunomodulating agents	0.8	0.7	0.9
M: Musculo-skeletal system	2.9	1.8	2.0
Multiple ATC	15.0	15.1	11.9
N: Nervous system	36.3	34.3	49.2
R: Respiratory system	2.4	0.8	1.0
V: Other	1.7	0.1	0.9
Total	100	100	100

Notes: The above table provides the breakdown by the World Health Organization's (WHO) Level-1 Anatomic Therapeutic Classification (ATC) for the drug products which comprise my main estimation sample. Each number represents the share of drugs which fall into that class, where *Percentage* refers to the product share, *Q-Weight* refers to the volume share, and *$-Weight* refers to the average annual revenue share (in $2013).

Market characteristics in the bottom panel of Table 1 suggest that, despite their age, there is still significant entry into these markets. In particular, over 75% of the markets have at least one firm enter since 2010, of which 41% experienced entry during the height of the collusive ring's activity. Although I will examine the effects of these entrants on market prices directly in Section 6, one critical role of entry is to trigger pharmacies to hold new sets of auctions.[23] This explains why, for each product, I observe, on average, 1.55 unique auction clusters, i.e., a set of auctions held at each pharmacy chain at approximately the same time.

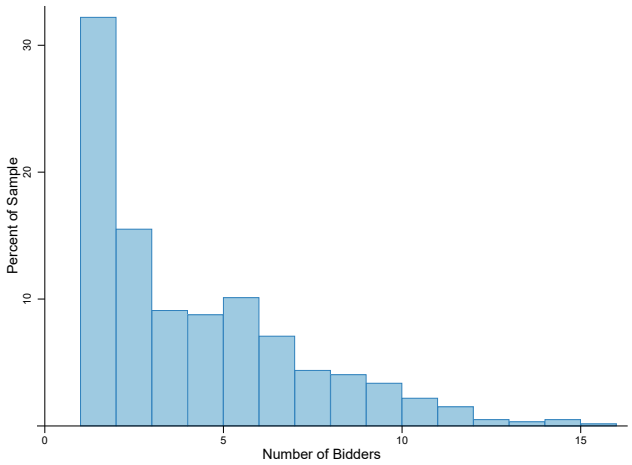
Figure 1 presents the distribution of the number of bidders in each auction. The modal number of bidders in any auction is two; although, there is significant variation across products. Indeed, the average market has 4.7 manufacturers. This is consistent with the findings of Berndt et al. (2017), who estimated the number of manufacturers in orally formulated molecules at around five using national sales data from IQVIA, the industry's gold standard resource for tracking pharmaceutical sales. Importantly, they also find that markets for orally formulated products tend to be more competitive than those of other drug formulations, insomuch as the majority have at least two active manufacturers.

Finally, recall that no auction is held if there is only a single manufacturer.[24] Thus, the market dynamics which arise in markets with monopolist suppliers, while important, will not feature in my analysis.

[23]This trigger was highlighted during the interviews with pharmaceutical executives, where they noted that entry reliably is followed by a pharmacy requesting new bids from both incumbent manufacturers and the entrant.

[24]There is a monopolist supplier in approximately one-quarter of markets with oral formulations.

Figure 1: Number of bidders present across auctions



Notes: The above figure shows the distribution of the number of manufacturers bidding in each auction comprising my main estimation sample. Data are constructed using a private insurer's pharmacy claims data and FDA Orange Book files.

An outstanding issue is how to identify competitive and collusive markets. As will be discussed in Section 3, my model captures equilibrium bidding behavior consistent with non-collusive conduct and the estimation adopts this same conduct assumption. In the following section, I discuss the data which allow me to differentiate between the two types of markets.

2.2.2 Counterfactual Data

Collusive Ring Information on the activities of the collusive ring are from the now unsealed indictments filed by States Attorneys in 2016, 2018, and 2019 against the collusive firms and the participating executives, as well as follow-up lawsuits by several national pharmacy chains and insurers in the subsequent years.[25] These reports are incredibly detailed, as they were compiled using not only information from thousands of documents produced by generic firms and an industry-wide phone call database with millions of phone calls among collusive ring members but also the testimony of unnamed participating members. In short, the evidence amounts to what the FTC has previously called "smoking gun" evidence implicating the firms. Most relevant for my purposes, the report names the firms accused of being participants in the collusive ring as well as the drugs where there was documented collusive activity. I use these lists of firms and drugs to flag markets as collusive.[26] Later on, in Section 6, I will use the reports' discussions of how the collusive ring was originally formed, as its origins are important in justifying the validity of the damages counterfactual.

[25] Humana, UnitedHealthcare, Kroger, Albertsons, and H.E.B. all have pending lawsuits against the ring.

[26] I experimented with various definitions of a collusive market, from the exact product named in the report (e.g., amitriptyline tablets) to all products falling under the umbrella of a named ingredient (e.g., amitriptyline). In practice, there were only a handful of drugs where the product was available in both tablet and capsule formulations, so it did not appreciably affect any results.

Application (Filing) Dates Information on the application dates of generic firms are from Feldman et al. (2017). As part of their study, the authors develop a methodology to identify the initial application date of a generic drug and use it on the universe of generic drugs which was approved by the FDA between 2006 and 2015.[27] Although the FDA publishes approval dates for every drug it approves on its website, it does not publish the corresponding application dates, i.e., the date in which a generic firm files its initial paperwork. These data are crucial for the counterfactual simulations, insomuch as they allow me to examine the extent to which each drug in my sample was affected by the FDA's application backlog.

3 Model

In this section, I introduce a model of the retail drug procurement process, wherein generic manufacturers submit bids to retail pharmacies in order to supply them with a particular drug. The model adapts the LOV model with an important distinction: I allow for correlation in firm costs within clusters of auctions. In the end, the main question to be answered by the model is how firms' costs of goods delivered vary according to observable characteristics of the auctioned drug contracts.

3.1 Set-up

A single and indivisible drug supply contract a is auctioned off by a pharmacy.[28,29] The pharmacy collects all bids b_{ia} from firms $i \in \{1, 2, \ldots, I_a\}$ simultaneously and ultimately awards the contract to the firm who submits the lowest bid b_a^w, where I_a represents the total number of bidders participating in auction a.

I assume that bidder i's cost in auction a is drawn from an auction-specific distribution $F_a(\cdot)$ and are private information.[30] I further assume that each firm knows the number of firms participating in the auction; however, they do not know the exact realizations of those firms' costs—only that they are being drawn from the same distribution $F_a(\cdot)$. This is consistent with my application, wherein the FDA publishes an official notice whenever a new firm receives approval to begin marketing a drug whilst pharmacies hold secretive auctions for their business.

Let $F_a(\cdot)$ represent the distribution of cost of goods delivered. This distribution is assumed to be absolutely continuous with respect to the Lebesgue measure and has a density $f(\cdot)$ and a support given by $C \subset [0, +\infty)$. Under the assumption that firms are risk neutral, the symmetric

[27] I refer the reader to the original paper for additional details.

[28] In reality, some large pharmacy chains will contract with multiple firms simultaneously for the same product. This multi-winner auction ensures that they have a backup supplier—a so-called "secondary supplier"—in cases of shortage or contamination. I exclude these auctions in my empirical analysis.

[29] The structure of the contract varies a lot in practice. At larger chains, the entire volume is not delivered at once, but rather divided over a pre-specified time frame. Smaller pharmacies, in contrast, often use "estimated quantity tenders" through which they place orders as needed. I will return to this point in Section 5.

[30] Note that I will formalize the exact timing of the cost draw in Section 3.2.

Bayesian Nash equilibrium of the preceding model yields a bidding function for firm i of:

$$b_{ia} = e(c_{ia}, I_a, F_a) = c_{ia} + \frac{1}{(1 - F_a(c_{ia}))^{I_a - 1}} \int_0^{c_{ia}} (1 - F_a(x))^{I_a - 1} dx \tag{1}$$

(see Maskin and Riley (1984)). Put into words, the equilibrium bid of firm i is an increasing function of its private cost c_{ia} and a decreasing function of the total number of bidders I_a participating in the auction. The latter is what drives the cost savings associated with additional entry. Of particular note in equation (1) is the strict monotonicity of the equilibrium bidding function with respect to c_{ia}. This strict monotonicity implies that the winner of the auction is, in fact, the firm with the lowest cost. In other words, I have:

$$b_a^w = e(c_{(1)a}, I_a, F_a) \tag{2}$$

where the winning bid, b_a^w, in auction a is itself a function of $c_{(1)a}$, i.e., the minimum cost draw across the I_a bidders present in the auction.

What is unique in my setting relative to a typical first price auction is the lack of a reservation bid. That is, pharmacies must accept whichever price arises in the auction, even if that implies that they make losses on the drug in the short term.[31] Incorporating this feature in the setup implies that the expected revenue from the auction is simply the expected winning bid.[32]

3.2 Application

Correlation in firm costs. Most empirical auction models assume independence in firms costs along two dimensions: within and across auctions. In the former, it is assumed that firms draw their costs independently of one another; in the latter, it is assumed that firms draw their costs independently of the costs they drew in previous auctions. In my setting, the latter assumption is too strong, as firms' costs are likely to be correlated across auctions which occur in the same time period, reflecting their current production capabilities.

In light of this, I allow for correlation in firms' costs within clusters of auctions, where a cluster refers to a set of auctions which are for the same product, in the same period, but at different pharmacies.[33] In practice, I assume that firms face two cost shocks per auction. The first, a so-called "idiosyncratic" shock ($\sigma_{id} u_{ia}$), is drawn by a bidder once per auction, and it serves as the typical i.i.d. shock that is present in most empirical auction settings. The second, a so-called "cluster" shock ($\sigma_{cl} u_{ik}$), is atypical. This shock is drawn once per *cluster* of auctions by each bidder. As will be discussed in Section 4.4, the cluster shock is what permits correlation in firms' costs across auctions, so that—all else equal—there is a higher likelihood of one firm winning multiple

[31] This is, in fact, an increasingly common occurence in the retail pharmacy setting. See, for example, https://www.npr.org/sections/health-shots/2015/10/22/450600567/how-generic-drugs-can-cost-small-pharmacies-big-bucks.

[32] See equation 6 in LOV.

[33] Imagine, for example, the set of auctions for tramodol 50 mg tablets in March 2012 at CVS, Walgreen's, and RiteAid.

auctions within a given cluster of auctions.

The fact that a firm's cluster shock applies across multiple auctions raises an important issue regarding how information on auction outcomes is used by bidders. In other words, with a persistent cost shock, it is possible that firms will use a more complicated dynamic equilibrium bidding strategy, which makes use of information from previous auctions (Jofre-Bonet and Pesendorfer, 2000, 2003). That said, in my setting, these concerns are not relevant, as there is no opportunity for information leakage across auctions within a cluster. During the interviews, generic executives emphasized that the auctions often happen in quick succession and, furthermore, that information about the winning firm and its particular winning bid is not released. This lack of information leakage, coupled with the distributional assumptions made on these two shocks in my application (see Section 4), imply that the optimal equilibrium bidding strategy remains unchanged from equation 1.

Cost distribution. There is considerable heterogeneity among drug contracts. As discussed in Section 2.2.1, the drugs vary significantly in their characteristics, which implies a certain degree of variation in the auctioned contracts. There is also inherent heterogeneity due to the fact that these contracts are offered by different pharmacies, with distinct customers and corporate policies. In grouping together auctions for the estimation, it is important to control for this heterogeneity. I adopt a parametric cost distribution, which controls for all characteristics I observe of the auctioned drug contracts. In particular, firm i's costs for auction a are drawn from $F_a(\cdot)$:

$$F_a(\cdot) = F(\cdot|z_a, \theta) \tag{3}$$

where θ is an unknown parameter vector in $\mathbb{R}^k$ and z_a is a vector of variables affecting firms' costs through the distribution of private values. I assume that the vector z_a is fully observed and, moreover, that the number of bidders I_a is known.

4 Estimation

In this section, I describe how I will estimate the structural model outlined in Section 3.

4.1 Distributional Assumptions on Costs

As noted in the previous discussion in Section 3.2, I adopt a parametric assumption on the distribution of manufacturer cost of goods delivered. Specifically, I assume that the cost c_{ia} for manufacturer i in auction a is log-normally distributed, where:

$$c_{ia} = e^{\mu_a + \sigma u_{ia}} \tag{4}$$

and u_{ia} is $\mathcal{N}(0,1)$, $\sigma^2 = \sigma_{cl}^2 + \sigma_{id}^2$, and μ_a is a linear function of a set of observable exogenous characteristics of auction a. In particular, I define:

$$E \log c_a = \mu_a = \theta_1 + \theta_2 units_a + \theta_3 units_a^2 + \theta_4 capsule_a + \theta_5 complexity_a + \theta_6 delivery_a + \gamma_a + \omega_a \tag{5}$$

where *units* (*units*^{2}) refer to the annual volume (volume-squared) of the contract, *capsule* is an indicator for a capsule formulation (i.e., versus tablet), *complexity* is an indicator for a complex formulation (e.g., involving complex active ingredients or formulations), and *delivery* is an indicator for a special delivery mechanism (e.g., extended release).[34] Note that γ_a are fixed effects which correspond to the World Health Organization's (WHO) Anatomical Therapeutic Chemical (ATC) classification system, and ω are pharmacy fixed effects (N=6).[35]

Most of these controls, excluding ω, are included in order to capture the inherent heterogeneity in the production costs associated with the contract of the drugs being auctioned. I allow for returns to scale in production by controlling non-linearly for the volume of the auctioned contract. In one of the only comprehensive studies on generic drug manufacturing costs, Hill et al. (2018) find that complex drugs and drugs with modified delivery mechanisms are more expensive to produce. That said, in my interviews with generic drug executives, they emphasized that active pharmaceutical ingredients (API) are their "first-order" cost driver. This motivates the inclusion of the ATC1 fixed effects in the baseline specification and ATC2 fixed effects in my preferred specification, where I re-estimate the model separately for each ATC1 category.[36]

While the other variables can be interpreted as shifting manufacturers' cost of goods delivered from a production standpoint, the pharmacy fixed effects control for systematic differences in the competitive environment. Note that these fixed effects comprise any pharmacy-specific features which influence firms' effective costs. For example, qualitative reports suggest that some pharmacy chains may provide more favorable contract terms, including prompt delivery bonuses, more flexible renegotiation terms, or volume discounts (Hill et al., 2018). These types of manufacturer-friendly policies may prompt submission of lower bids on the margin, as manufacturers expect to benefit from these policies for the duration of the contract.

As noted in Section 3.2, I impose the correlated cost structure by assuming that firms draw two cost shocks. Next, I add the relevant structure for these shocks. Whereas the idiosyncractic shock is drawn once per auction by each bidder from $\mathcal{N}(0, \sigma_{id}^2)$, the cluster shock is drawn once per cluster k of auctions by each bidder from $\mathcal{N}(0, \sigma_{cl}^2)$.[37] Given that μ_a is shared by all firms, the particular realizations of the cost shocks determine the winner of an auction. In other words, the winner of auction a is the bidder who has the lowest *total* cost shock $s_{(1)a}$, or:

$$s_{(1)a} = \min_{i \in I_a} \sigma_{cl} u_{ik} + \sigma_{id} u_{ia} \tag{6}$$

[34] I follow Gupta et al. (2018) in my definition of a "complex" drug.

[35] I mask the identities of the pharmacy chains in the estimation. The six categories are Chain 1-4 (national pharmacy chains), Independent (agglomeration of independent pharmacies), and Wholesaler.

[36] In Section 5.2, I re-estimate the baseline specification separately for each of the five largest ATC1 categories in my sample: A, C, G, J, and N. This allows me to include more granular fixed effects. For example, instead of ATC1 = C for all "Cardiovascular system" drugs, I can separately control for ATC2 = C01 ("Cardiac therapy"), C02 ("Antihypertensives"), ... C10 ("Lipid modifying agents").

[37] Note that the *id* and *cl* are used to differentiate between the "idiosyncratic" and "cluster" variances, respectively.

where both u_{ik} and u_{ia} are independent draws for bidder i in auction a, belonging to auction cluster k, from $\mathcal{N}(0,1)$.[38] Because each cost draw is an independent normally distributed random variable, their sum is also normally distributed, and their variance is given by the sum of their respective variances. That is, the total cost draw is distributed $\mathcal{N}(0,\sigma)$, where $\sigma = \sigma_{cl}^2 + \sigma_{id}^2$.

Before discussing how I use the data to estimate the structural model, notice that there are two sets of parameters to be estimated: the parameters of μ_a (i.e., θ) and σ^2 and the parameters representing the variances of the two cost shocks, σ_{cl}^2 and σ_{id}^2. Under my maintained assumptions that (i) costs are log-normally distributed, (ii) μ_a is given by equation 5, and (iii) the cluster and idiosyncratic cost shock distributions are distributed $\mathcal{N}(0,\sigma_{id}^2)$ and $\mathcal{N}(0,\sigma_{cl}^2)$, respectively, I have 24 parameters, in total, to estimate. Each parameter is listed in Table 3.

Table 3: Parameters to be estimated

Parameters	#	Description
		Main estimation
θ_1	1	Constant
θ_2	1	Annual volume (100k)
θ_3	1	Squared annual volume (100k)
θ_4	1	Capsule
θ_5	1	Complexity
θ_6	1	Delivery mechanism
γ	11	ATC-1 fixed effects
ω	5	Pharmacy fixed effects
σ_{id}^2	1	Variance of idiosyncratic cost distribution
σ_{cl}^2	1	Variance of cluster cost distribution
		Drug-class specific estimations (N=5)
ω_A	3	
ω_C	8	
ω_G	3	ATC-2 fixed effects
ω_J	2	
ω_N	6	

Notes: The above table displays the 24 parameters in the main estimation, as well as 22 additional fixed effects to be estimated in the drug-class specific estimations (see Section 5.2). *ATC-1* and *ATC-2* fixed effects comprise the World Health Organization's (WHO) Level-1 and 2 Anatomic Therapeutic Classification (ATC), where the latter include only a subset of ATC-1 categories: {A, C, G, J, and N}.

Recall that the result that the expected winning bid is equal to the expected second order statistic assumes competitive bidding behavior; therefore, I will estimate these parameters on the subset of auctions which were not flagged as being collusive in any of the indictments.

In practice, the estimation itself has two steps: first, I estimate the ratio of the two variances determining the cost shock distributions. Then, using this ratio, I estimate the parameters of the mean cost expression as well as the variance of the total cost shock distribution. I obtain

[38] I use this notation to be consistent with the estimation.

bootstrapped estimates of standard errors by resampling auction clusters, redrawing firms' standard normal cost shocks, and then reestimating the cost parameters. I now discuss the details of each estimation step, beginning with the estimation of (μ_a, σ^2). A complete description of the estimation algorithm appears in Appendix Section A.2.

4.2 Estimating μ_a, σ^2

The estimation strategy for (μ_a, σ^2) draws from LOV but introduces a novel feature: the ability to use highly aggregated data. I do this out of practical necessity. Whereas LOV had data on the winning bid for each auction in their sample, I observe only the average of the winning bids across multiple auctions in my sample. This difference, while subtle, is important, insomuch as there are many settings where even the winning bid from an auction is unavailable. In my setting, for instance, the winning bid is a trade secret of the pharmacy.

Relying on the revenue equivalence theorem, LOV make use of the first moment of the wining bid in forming their NLLS estimator given its mapping to the seller's expected revenue in the auction (see equation 7). Under these parametric assumptions and equation 2, they define the expectation of the winning bid in auction a as:

$$E[b_a^w] = \exp(\mu_a) \left(\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \exp(\sigma u_{(2)}) \phi(u_1) \dots \phi(u_{I_a}) du_1 \dots du_{I_a} \right) \tag{7}$$

where $\mu_a = z_a'\theta$, while $\{u_1, \dots, u_{I_a}\}$ represent the I_a bidders cost draws from a standard normal distribution with density $\phi(\cdot)$, and $u_{(2)}$ is the second-order statistic of these I_a draws.

I do the same here using the average winning bid. In particular, let $E\left[\sum_{j\in\gamma(a)} \alpha_j b_j^w\right] \equiv m\left((X_j)_{j\in\gamma(a)}, (I_j)_{j\in\gamma(a)}; \theta, \alpha(a)\right)$ denote the conditional expectation of $\sum_{i\in\gamma(a)} \alpha_i b_i^w$ which is a convex combination of the winning bids in auctions $j \in \gamma(a)$ with the exogenous and observable weights α_j. Here, $\gamma(a)$ refers to the cluster of auctions to which auction a belongs. For ease of notation, I will denote such a cluster by k going forward. Because I do not have a closed form solution for $m(\cdot)$, I will simulate it, denote this simulator by $\overline{\overline{X}}_k(\theta)$. Substituting the simulator into the usual NLLS objective function yields:

$$Q_{S,K}^* = \frac{1}{K} \sum_{k=1}^{K} \left(\bar{b}_k - \overline{\overline{X}}_k(\theta) \right)^2 \tag{8}$$

where $\bar{b}_k$ is the average winning bid from the data, that is $\bar{b}_k = \sum_{a\in k} \alpha_a b_a^w$.

The unbiased simulator of the average winning bid, $\overline{\overline{X}}_k(\theta)$, is a modified version of the simulator in LOV. In my setting, because I do not observe the winning bid, my simulator must account for the aggregation of the data. Thus, whereas their simulator $\overline{X}_a(\theta)$ simulates the expected winning bid in auction a, my simulator $\overline{\overline{X}}_k(\theta) = \sum_{j\in k} \alpha_j \overline{X}_j(\theta)$, simulates the expected average winning bid in auction cluster k, where the average is calculated using the observable αs. I refer the reader to Appendix Section A.2 for details of the construction of the simulator.

The remaining issue is how to account for the inconsistency of the estimator due to the use

of a simulator for $m(\cdot)$. Recall that minimizing the objective function given in equation 8 yields an inconsistent estimator for any fixed S as L goes to infinity, where S refers to the number of simulated observations used in constructing the simulator. Following the proof in Appendix A of LOV, I derive the appropriate correction term for my modified estimator, taking into account the correlation among costs within a cluster. Subtracting this correction term from equation 8 yields my modified simulated NLLS objective function:

$$\frac{1}{S(S-1)}\frac{1}{K}\sum_{k=1}^{K}\left(\sum_{a\in k}\alpha_a^2\sum_{S}\left(X_{sa}(\theta)-\bar{X}_a(\theta)\right)^2\right)+ \\ \frac{1}{S(S-1)}\frac{1}{K}\sum_{k=1}^{K}\left(2\sum_{m.n\in k,\, m\neq n}\sum\alpha_m\alpha_n\left(X_{sm}(\theta)-\bar{X}_m(\theta)\right)\left(X_{sn}(\theta)-\bar{X}_n(\theta)\right)\right) \quad (9)$$

where S is the number of simulations, α_a is the pharmacy a's weight in the weighted average bid calculation, k is the particular cluster of auctions used in the average calculation, and K is the total number of auction clusters.

To investigate whether I am able to recover the structural parameters using the proposed estimation method, I conduct a detailed Monte Carlo study. The results of this exercise are presented in Appendix Section A.2.3. Summarizing, I find that the estimation works well, insomuch as I am able to estimate well not only each parameter in the expression for μ_a but also σ^2. I also find that the variance of the estimated parameters in the Monte Carlo study is typically lower for my modified estimator than for the original LOV estimator for sample sizes approximating my own.[39] This may be due to the fact that each average winning bid observation comprises multiple winning bid observations but, as an average, is less noisy.[40]

4.3 Estimating $\sigma_{cl}^2, \sigma_{id}^2$

The previous estimation approach allows us to identify the variance of the total cost shock; however, I also want to be able to estimate its component variances, i.e., for the cluster and idiosyncratic cost shock distributions. To do so, I estimate the ratio of the cluster to idiosyncratic shock variances via a simulated method of moments (SMM) estimator based on the moments listed in Table 4.

The first set of moments match the expected maximum number of auctions won among the bidders present in the cluster; the second set match the expected number of auctions won among the bidders present in the cluster. Notice that the number of bidders present in the auction generates variation in both sets of moments. Bidders in auctions with more (fewer) bidders have lower (higher) expected and expected maximum number of auction wins within the cluster. Put differently, it is far more likely for a single firm to win multiple auctions in an auction cluster when

[39] As I discuss in Appendix Section A.2.3, the comparison I make here is between the LOV estimator which treats each observation in the Monte Carlo as a winning bid observation whereas the modified estimator treats each observation as an average winning bid observation.

[40] I leave this as a future proof.

it is competing against fewer firms. Given this predictable variation, I condition each moment on the number of bidders I_k in the cluster k, for $I_k = 2, \ldots, 12$. Importantly, in my setting, the same firms compete in every auction within a cluster, or $I_a = I_k \forall a \in k$. Therefore, in total, I use 22 moments to identify the ratio of variances.

Table 4: Moments Used in SMM estimation

Moments	#	Description
$\max_{i \in I_k} \left(\sum_{a \in k} \mathbf{1}_{i,a} \right) - \max_{i \in I_k} \left(\sum_{a \in k} \hat{\mathbf{1}}_{i,a} \right)$	11	Maximum Auctions Won
$\mathbf{1}(I_k = n) \frac{1}{I_k} \sum_{i=1}^{I_k} \sum_{a \in k} \mathbf{1}_{i,a} - \mathbf{1}(I_k = n) \frac{1}{I_k} \sum_{i=1}^{I_k} \sum_{a \in k} \hat{\mathbf{1}}_{i,a}$	11	Expected Auctions Won

Notes: The above table displays the 22 moments used to estimate the ratio of the variance parameters σ^2_{cl} and σ^2_{id} in the SMM estimation. The term $\mathbf{1}_{i,a}$ is shorthand for the indicator function $\mathbf{1}$(Bidder i won auction a), where $\mathbf{1}_{i,a}$ are data and $\hat{\mathbf{1}}_{i,a}$ denote model predictions. The data used in the estimation are compiled using claims data from a private insurer.

4.4 Identification

Given the two-part estimation routine, a separate identification argument applies for each half of the estimation. Identification of each of the parameters in the expression for μ_a (excluding θ_0) is possible due to variation in the characteristics of the auctions themselves.

Identification of θ_0 and σ^2 is more complicated, as it requires variation in the number of bidders across auctions (and clusters). Without this variation, one cannot separately identify the two, as was the case in LOV.

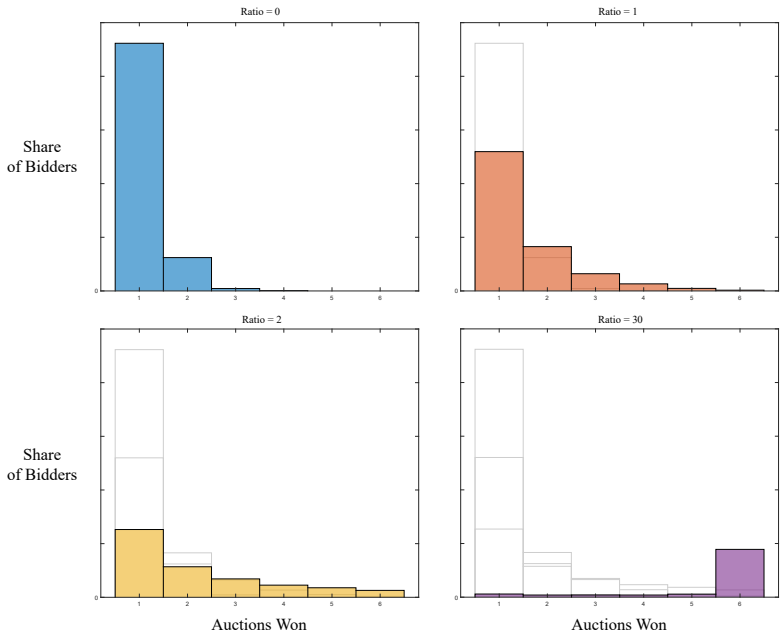
For example, take two auctions A and B which are observably identical, except for the number of bidders present ($I_A > I_B$). Equation (7) implies these two auctions will have different expected winning bids, where both the magnitude of the expected wining bids and their difference are functions of (θ_0, σ^2). In particular, note that the expressions for the expected winning bids are themselves increasing in θ_0 and decreasing in σ^2 whereas the gap ($b^w_A - b^w_B > 0$) is increasing in both θ_0 and σ^2. It follows that the pair (θ_0, σ^2) is uniquely identified by the level and difference in the winning bids.[41]

The ratio of the variances of the cluster and idiosyncratic shock distributions are identified through the empirical probability distribution of the number of auctions won by each bidder within an auction cluster. Intuitively, if there were no cluster shock, then each firm would be equally likely to win each auction within the cluster. Introducing a cluster shock introduces a wedge in the

[41] Suppose, for a contradiction, that this were not the case, i.e., there was another (θ_0', σ'^2) consistent with the same data. Without loss of generality, assume that $\theta_0' > \theta_0$. Because the expected winning bid is increasing in θ_0, to match the magnitude of the expected winning bid, σ'^2 must also be bigger; however, a larger σ'^2 implies that the difference in the expected winning bids would be too large. Analogous arguments can be made for θ_0' smaller and σ'^2 smaller or bigger.

distribution, insomuch as the firm with the smallest cluster shock (i.e., in absolute terms) is more likely, *a priori*, to win each auction within the cluster.

Figure 2: Simulated distribution of auction wins within an auction cluster



Notes: The above graph shows four empirical probability distributions of the number of auction wins within an auction cluster, conditional on winning at least one auction. Each distribution is generated by simulating my model 10,000 times for a fixed variance ratio (labeled *Ratio* above each tile). In each case, the auction cluster comprises six auctions, and each auction has six bidders.

Figure 2 depicts four empirical probability distributions from 10,000 simulations of my model using a fixed variance ratio ($\frac{\sigma_{cl}}{\sigma_{id}}$),[42] where I condition on a bidder having at least one win.[43] Each distribution pertains to auction clusters with six bidders, and—as in my application—clusters of six auctions.

Note that, in the upper-left panel of Figure 2, the ratio is zero, so there is only an idiosyncratic cost shock ($\sigma_{id}u_{ia}$), and each bidder is equally likely to win each auction. As expected, the modal number of wins is one. As the magnitude of the ratio of the cluster to idiosyncratic shock variance grows, so too does the probability of *multiple* wins within an auction cluster. Indeed, at a certain point (represented here by a ratio of 30 in the bottom-right panel of the same figure), the cluster shock ($\sigma_{cl}u_{ik}$) becomes so dominant that one firm is disproportionately likely to win every auction within the auction cluster. Finally, note also that the identification power from these moments is

[42]The simulation proceeds as follows: I normalize the variance of the idiosyncratic shock distribution to be equal to 1, then I scale the variance of the cluster shock distribution by the ratio. Next, I independently draw one cluster shock and six idiosyncratic shocks per bidder per auction cluster from their respective distributions. For each auction cluster, I determine the winner of each auction within the cluster, and then I count up the number of auctions won by each bidder within the cluster. This represents one iteration of the simulation. I repeat this process 10,000 times.

[43]This is why the total area under each histogram varies across the four panels. If I were to include the "no win" category, then the area would be the same under all.

from the two extremes of the distribution, where the relationship is strictly monotonic. That said, the moments also make use of intermediate wins for additional power.

5 Results

5.1 Baseline estimates

Baseline estimation results are presented in Table 5. Given the assumption of log-normal costs, each coefficient can be interpreted as the change in log costs associated with a one-unit increase in the characteristic. For ease of exposition, I exponentiate each coefficient in the discussion that follows.

Table 5: Main estimation results

Cost per Pill	
Constant	0.55***
	(0.12)
Annual volume (100,000s)	−0.28*
	(0.11)
Annual volume, squared	0.0005
	(0.0014)
Capsule	0.28***
	(0.08)
Complexity	0.13
	(0.10)
Delivery mechanism	0.53***
	(0.13)
Wholesaler	0.06
	(0.08)
Independent	0.21***
	(0.07)
Chain 1	0.04
	(0.09)
Chain 2	0.08
	(0.11)
Chain 3	0.05
	(0.10)
σ_{id}	0.14*
	(0.07)
σ_{cl}	0.13*
	(0.07)
ATC-1 FE	YES
N	4515

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

Notes: The above table summarizes estimates of the 24 parameters from Table 3 corresponding to the main estimation. Standard errors, shown in parentheses, are computed using 5,000 bootstrap samples of auction clusters. Fixed effects for ATC-1 drug classes (n=11) are suppressed.

The negative coefficient on units suggests that there are returns to scale in drug production. In particular, an increase of 100 thousand units in annual volume is associated with 24% lower costs. Consistent with the previously cited qualitative evidence on drug production, I find that drugs

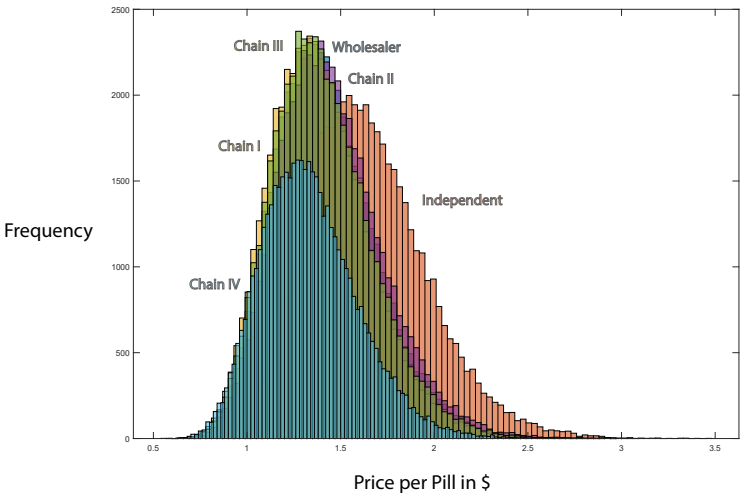
available as capsules and drugs with more sophisticated (or, non-standard) delivery mechanisms are also associated with higher manufacturer costs. The latter effect is particularly large, increasing costs by almost 70% relative to a standard delivery mechanism (i.e., immediate release).

Turning to the pharmacy fixed effects, I find largely insignificant effects, as the estimates are noisy. The estimated cost of supplying a drug contract at an independent pharmacy is highest, i.e., roughly 25% higher than the cost of supplying a contract at Chain 4, the omitted category. I show this graphically in Figure 3. Recall that the monotonicity of the equilibrium bidding function implies these higher costs translate into higher bids, so independent pharmacies end up paying more for their generic drugs. Such a finding is supported by recent reports of independent pharmacies failing to cover their ingredient costs and their rising levels of bankruptcy (Lieberman and Ginsburg, 2018; NPCA, 2019).

As noted in Section 4.1, there are several explanations for why manufacturers' costs of goods delivered are higher at independent pharmacies. Another explanation worth highlighting here relates to how independent pharmacies structure their procurement contracts, with respect to the timing of drug delivery. In short, they use "estimated quantity tenders," whereby they request deliveries from manufacturers as soon as their inventory of a particular product dwindles. Thus, it could be that the higher estimated costs could reflect manufacturers' distaste for the "lumpy" demand of independent pharmacies. If facing unpredictable demand makes it difficult for manufacturers to allocate their production capacities across products optimally, then they may require a higher price per pill, as compensation for the hassle.[44]

[44]Standard pharmacy operating procedure is to keep generic drug inventory as low as possible, in order to allow for a large inventory of branded drugs. Note these types of inventory concerns are likely much less pressing for a large national pharmacy chain like CVS or Walgreens, as they operate at such high volumes to require their own wholesaling operations.

Figure 3: Simulated manufacturer cost distributions across pharmacy chains



Notes: The above graph shows the simulated distributions of manufacturer costs resulting from simulating my model 10,000 times for a fixed drug contract at each pharmacy (i.e., all non-pharmacy characteristics of the contract (i.e., volume and drug characteristics) remain the same in each simulation. The estimates I use for the simulation are shown in Table 5, where I use the base categories of all indicator variables and an annual volume of 100,000 units as the fixed contract.

Relatedly, another important source of heterogeneity in estimated costs across pharmacies is due to systematic differences in the volume of drugs the chains purchase. Figure 4 shows this cost heterogeneity for the four national pharmacy chains and wholesaler.[45] Note that manufacturer costs for supplying Chain III are systematically lower than for any other pharmacy because it is the largest purchaser of generic drugs.

Turning to the estimated variances of the cluster and idiosyncratic cost distributions, I see that the two are very similar in magnitude. Per my earlier simulations (see Figure 2), this suggests that there is significant cost persistence across auctions within a cluster, especially as most auction clusters in my sample have fewer bidders than in my simulations.[46] I interpret this persistence as reflecting a short-term comparative production advantage, such as a temporary reduction in input costs from an upstream API supplier.

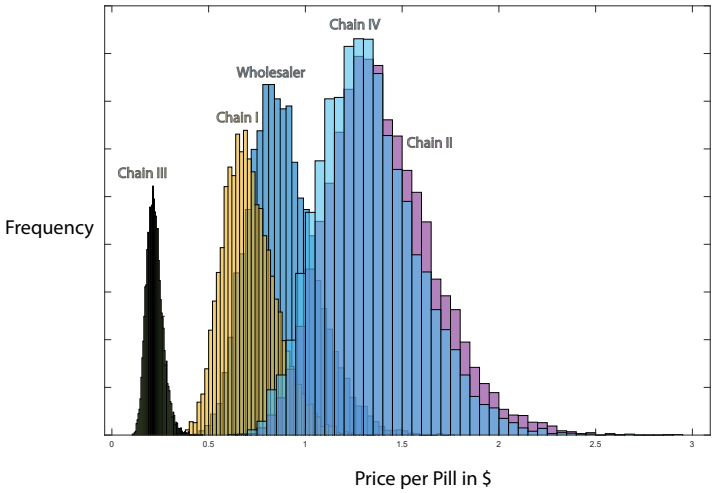
At this point, it is important to emphasize the implicit assumption being made in the baseline model: cost heterogeneity across drug molecules is captured by the level of mean log cost. More precisely, the ATC1 fixed effects included in the mean log cost expression given by equation 5 fully control for any cost differences across drugs with different active ingredients. That said, it is possible that the broadly defined ATC1 therapeutic category controls fail to account for the meaningful cost heterogeneity which exists across molecules and, moreover, do not capture any cost complementarities due to interactions between ingredients and other auction characteristics.

[45] I omit the independent pharmacies from this figure, as the independent pharmacy demand in my application represents the composite demand of many independent pharmacies.

[46] Recall that the effect of the ratio on the number of auction wins within an auction cluster is decreasing in the number of bidders in the auction cluster.

While fully interacting all characteristics with the ingredient fixed effects is impractical given the size of the sample, I can estimate the model separately for the five largest therapeutic categories. I present these results in the following subsection.

Figure 4: Simulated manufacturer cost distributions across pharmacy chains, with varying contract size



Notes: The above graph shows the simulated distributions of manufacturer costs resulting from simulating my model 10,000 times for a specific drug contract at each pharmacy. For each contract, I scale the volume such that it reflects that pharmacy's share of the total volume in the market; all drug characteristics of the contract remain unchanged. The estimates I use for the simulation are shown in Table 5, where I use the base categories of all indicator variables and an annual volume of 100,000 units as the base contract, corresponding to Chain 3.

5.2 Heterogeneity

I estimate my model on the five largest therapeutic categories in the data: (A) Alimentary tract and metabolism, (C) Cardiovascular system, (G) Genito-urinary system and sex hormones, (J) Anti-infectives for systemic use, and (N) Nervous system. Note that because I run the model separately by ATC1 classification I are able to include a limited set of ATC2 fixed effects for each sub sample.[47] For example, in the nervous system drug estimation, I control for anesthetics (N01), analgesics (N02), antiepileptics (N03), anti-Parkinsons (N04), psycholeptics (N05), psychoanaleptics (N06), and other nervous system drugs (N07) with fixed effects. I refer the reader to Appendix Table A.1 for the full set of estimates.

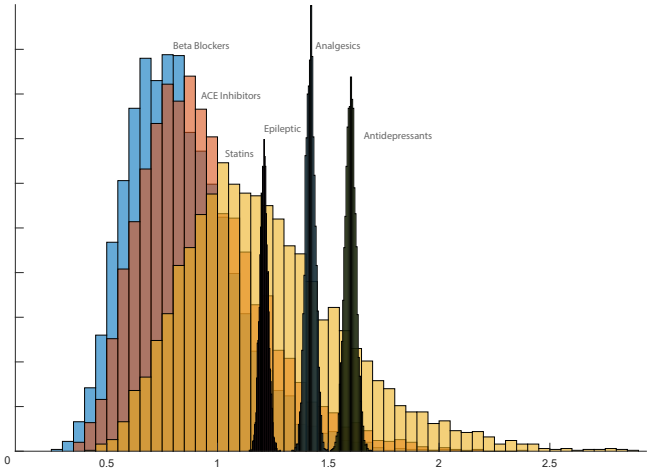
Many of the findings from the baseline specification remain true in the drug-class specific models, although the estimates are even noisier due to the use of a smaller sample. In particular, across the five classes of drugs, there is still evidence of returns to scale in production; although, all coefficients are smaller in magnitude. Increasing the annual sales volume is estimated to decrease costs by 15-20% (down from 25%). Moreover, complex molecules and capsules remain associated with higher

[47] I include any ATC-2 category with at least 20 member drugs.

estimated costs, as do contracts for independent pharmacies, particularly among nervous system drugs.

Perhaps most striking among these results, however, are the stark differences in mean costs, both across and within therapeutic class, as well as in the variance of costs, as captured by the estimated variances of the two cost shock distributions. To provide some intuition, Figure 5 shows the estimated cost distributions for the two largest classes of drugs—nervous and cardiovascular systems—as well as the three most common sub-classes of each.[48] Consistent with manufacturers' claims, the active ingredient is shown to generate significant cost heterogeneity. Cardiovascular drugs are, on average, cheaper than nervous system drugs; however, among cardiovascular drugs, beta blockers and ACE inhibitors are cheaper to produce than statins.[49]

Figure 5: Simulated manufacturer cost distributions across nervous system and cardiovascular drugs



Notes: The above graph shows the simulated distributions of manufacturer costs resulting from simulating my model 10,000 times for specific sub-classes of nervous system (ATC-1=N)—epileptic, analgesics, and antidepressants—and cardiovascular drugs (ATC-1=C)—beta blockers, ACE inhibitors, and statins. For each contract, I use the base categories of all indicator variables (except those corresponding to the drug class) and an annual volume of 100,000 units. Because I overlay the sets of distributions in a single graph, the y-axis is scaled for each set of distributions. The estimates I use for the simulation are shown in Appendix Table A.1.

Figure 5 also shows that there is significantly more variation in cardiovascular drug costs than there is among nervous system drugs. This is important, insofar as what it implies regarding the gains to entry in these markets. Recall that there are higher potential returns to entry where there is more variation in the underlying cost distribution. Given that the expected winning bid is equal

[48]For the purpose of generating these distributions, I assume an annual sales volume of 100,000 units and the most basic pill form, i.e., a non-complex tablet with an immediate release formulation. I further assume that the auction contract which is being auctioned is at Chain 4.

[49]The patterns illustrated here are largely consistent with the limited available research on generic drug production costs (see, for example, Hill et al. (2018)).

to the expected second lowest cost, the expected *difference* in the winning bid upon entry grows as the variation in costs grows. I explore this idea further in Section 6.1.

5.3 Model fit and robustness

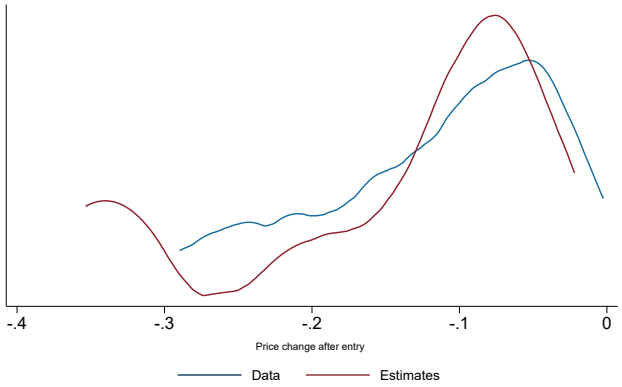
To evaluate the goodness-of-fit of my estimated model, I calculate an R^2 statistic based on my objective function. Let $R^2 = 1 - Q^*_{S,K}(\hat{\theta})/\hat{\text{Var}}(\overline{b_k})$, where $\overline{b}_k = \sum_{a \in k} \alpha_a b_a^w$. On average, in the baseline model, I explain approximately 34% of the variation in the average winning bid across auctions despite a parsimonious set of controls for the active pharmaceutical ingredient. I can also calculate this R^2 statistic for the set of drug-class specific models, which include more granular active ingredient controls. As expected, the model fit improves significantly. In particular, estimated R^2's are between 60-70%.

I can also examine how the model performs out of sample. So far, I have assumed that firms' cost draws are independent *across* auction clusters, and so different auction clusters for the same drug are treated as separate observations in the estimation. If my model is well specified, then I ought to be able to recover the expected change in the average weighted winning bid upon entry of an additional firm given the characteristics of the markets pre- and post-entry. Because I observe markets where there is entry in the data, a natural test of the model is to compare the predicted and realized change in price. Figure 6 presents the empirical probability distributions among cardiovascular system drugs.[50] Overall, the predicted pattern of prices is similar to the realized pattern; although, I note that the model does not capture the long tail of price changes in the data. To the extent that some of these extreme changes reflect idiosyncratic market features which are not captured in the model—like a temporary drug shortage, such a result is to be expected.[51]

[50] This is the drug class with the largest number of available entry observations (N=90).

[51] Cardiovascular drug shortages comprise a disproportionate share of total drug shortages. In fact, between 2001-2014, there have been over 180 shortages involving cardiovascular drugs (Reed et al., 2016; GAO, 2014).

Figure 6: Simulated density of percent change in price after entry



Notes: Each line presents the density corresponding to the percent change in price after the entry of a new firm into a market (N=90). The figure includes only drugs belonging to the class of Cardiovascular drugs (ATC-1=C). I plot the realized changes (*Data*) in blue and simulated changes (*Estimates*) in red. A value of -0.1 refers to a 10% reduction in price relative to the pre-entry price. The estimates I use for the simulation are shown in Appendix Table A.1.

Estimating the model outlined in Section 4 required making several assumptions. Here, I discuss the sensitivity of the results to these assumptions. As suggested by Figure 1, most auctions involve between two and six bidders; however, I use auctions with up to 12 bidders in the estimation. It is possible that bidders in the most competitive markets may temporarily withdraw from the market if they are unsuccessful in winning a supply contract such that the effective I_a in these markets is smaller than reported, implying that my model is misspecified in these markets. To test this, I re-estimate the model on the subset of auctions with six or fewer bidders. The estimates are qualitatively similar using this subset of data; although, estimates of the cost shock variances are somewhat larger.

Recall that in the analysis, the insurance claims data serve as a quasi-census of the pharmacy demand for different drugs. To the extent that the pharmacy demand of these insured patients is not systematically different from other consumers, my use of a sub-sample of the prescription drug population should not introduce bias in the estimation—the drug volume drug at each pharmacy is simply a fixed share of the total volume. In 2017, many of the sample plans removed CVS from their preferred pharmacy networks, leading to a differential drop in traffic at CVS, which violates my assumption. Thus, I re-estimate the model on the subset of auctions which occur before 2017. Although estimates are noisier due to the reduced sample size, the results are quite similar.

Finally, at the crux of the estimation is the assumption that the behavior of firms is consistent with my specified model, i.e., that bidders are using the equilibrium bidding strategy given by (1). This motivates my use of the markets which are *not* named explicitly in any of the recent indictments against firms in the collusive ring. That said, while the indictments provide a full list of the drug molecules affected by the collusive activity, they do not always specify the formulation

or delivery mechanism. Out of an abundance of caution, I re-estimate the model, excluding all drugs with active ingredients mentioned in the indictments. Again, while the estimates are noisier, they are qualitatively unchanged.

6 Counterfactuals

Is there heterogeneity in the benefits of entry across markets? How costly was the application backlog at the FDA? What was the magnitude of damages from the collusive ring's activity in the market? In this section, I use the estimates from the structural model presented in Section 5 to address each of these questions in turn.

6.1 Distributional effects of entry

Encouraging entry has become one of the major policy objectives of the FDA. In the wake of the shortages in 2009 and the price spikes in 2013-2015, FDA Commissioner Dr. Scott Gottlieb spearheaded efforts to spur competition in the generic drug market. In 2017, the FDA announced its Drug Competition Action Plan (DCAP) to "remove barriers to generic drug development and market entry in an effort to spur competition so that consumers can get access to the medicines they need at affordable prices." Indeed, one of the first initiatives under DCAP was the creation of the *List of Off-Patent, Off-Exclusivity Drugs without an Approved Generic*, a list maintained by the FDA in order to alert generic drug manufacturers to markets with insufficient competitors.[52]

That said, all entry may not be equal, in terms of its effect on market prices. Given the striking cost heterogeneity across markets, it follows that there could be equally heterogeneous benefits to entry. To evaluate this premise, I estimate the change in the upstream price associated with the entry of one additional firm in each of the markets in my sample. Note that this counterfactual assumes that this marginal firm is not any different *a priori* from the other firms present in the market. In particular, in the context of the model, its costs of goods delivered are being drawn from the same distribution and its equilibrium bidding function is no different from that of the other firms.

Simulations confirm that entry is associated with lower per pill prices; however, the magnitude of the effect is quite varied. While the addition of a firm to the market is associated with an 8.5% drop on average in upstream prices, the inter-quartile range of effects is 4.3-16.7%.

As shown in Table 6, the largest gains to entry are found in the markets with the fewest bidders. In particular, adding a third bidder in the average market with two bidders is associated with over $300 thousand dollars in cost savings annually. There is a steep decline in savings as the number of bidders in the market before entry increases. For example, adding an eleventh bidder in a market of 10 yields just $13 thousand dollars in cost savings annually.

[52] The list is available online here: https://www.fda.gov/drugs/abbreviated-new-drug-application-anda/list-patent-exclusivity-drugs-without-approved-generic.

Table 6: Annual cost savings from an additional entrant

N of Bidders	Mean	Median	25th	75th
2	304.8	149.0	48.4	427.2
3	131.0	82.8	31.2	202.2
4	79.0	59.6	17.4	119.0
5	68.0	52.8	14.9	90.7
6	43.7	39.5	11	69.7
7	25.2	16.6	2.5	39.2
8	25.7	26.3	7.8	39.2
9	18.5	13.0	1.2	29.3
10	13.4	8.2	0.3	15.8

Notes: The above table shows the estimated cost savings from an additional entrant into the market, where each row includes all markets of that pre-entry market size. Each estimate is in thousands of $2013. The column titles *25th* and *75th* correspond to the 25th and 75th percentiles of the distribution, respectively. The estimates I use for the simulation are shown in Appendix Table A.1.

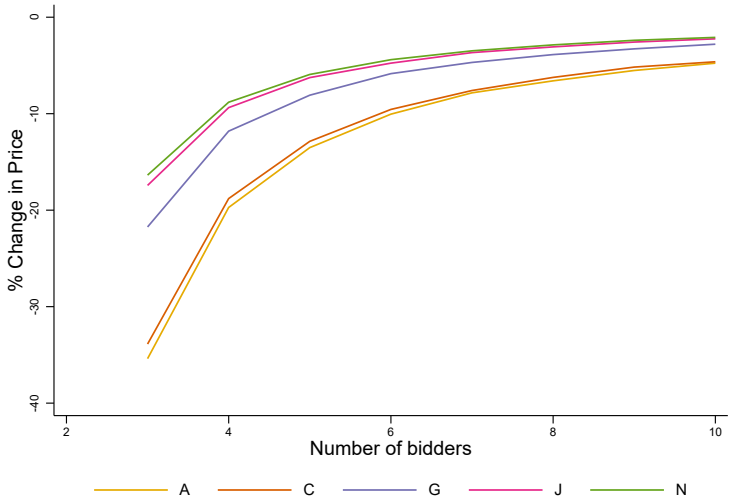
Note, however, that these aggregate statistics may mask important heterogeneity, insomuch as the composition of drugs within markets of different sizes may themselves be very different. Suppose, for instance, that the drugs with the highest returns to entry were also the drugs with the fewest manufacturers because they had very high fixed costs of entry. While in the data, I do not observe every possible market outcome with respect to entry; with the model, I can.

Figure 7 shows the percentage change in price from the marginal entrant given $I_a = \{2, \ldots, 10\}$ initial firms in the market, where I repeat the exercise for each ATC1 class.[53] As before, I see that the largest returns to entry are from the third bidder[54] and decline for each subsequent entrant; however, there are stark differences across ATC1 classes. In particular, alimentary tract and metabolism (A) and cardiovascular system (C) drugs have considerably higher returns to entry for all marginal entrants, but especially for early entrants. In the context of the model, this pattern is a direct result of the higher underlying variances in costs for firms manufacturing drugs in these two classes.

[53] Recall that the structure of my parametric costs implies that all drugs within the class share the same path when defined as a percentage change, i.e., the mean cost cancels out.

[54] My model does not generate predictions for monopolist markets; therefore, the first entrant I can model is the third.

Figure 7: Percent change in price after entry, by ATC-1



Notes: Each line depicts the average simulated percentage change in price as the market size changes (due to entry) for the subset of drugs within a particular WHO Level-1 ATC group. The groups include A = *Alimentary tract and metabolism*, C=*Cardiovascular system*, G=*Genito-urinary system and sex hormones*, *J=Antiinfectives for systemic use*, and N=*Nervous system*. The estimates I use for the simulation are shown in Appendix Table A.1.

The preceding counterfactual shows how the estimated model can be used to simulate the effect of entry on upstream prices, where there may not have been entry. Here, I extend this exercise in order to look at the effect of entry on upstream prices, where entry *was* realized, albeit with a significant delay.

6.2 Effect of the FDA application backlog

Beginning in 2011 and 2012, many blockbuster drug markets opened to generic firms, as branded firms lost their marketing exclusivity status as their patents expired. In the midst of this so-called "patent cliff" expiration, the FDA became overwhelmed with processing generic drug applications. Understaffed, the agency could not process as many applications as they were receiving, and so a backlog of applications grew.[55] To address this question, I turn to the issue of the FDA application backlog. This backlog was extremely troubling not only to the FDA but also to generic manufacturers, as it meant that they were indefinitely blocked from marketing any new products. Indeed, by October 2012, the FDA had over 3,000 applications that were pending review, resulting in approval wait times of over 30 months.[56]

In order to determine the effect of the backlog on upstream prices, I use my model to simulate

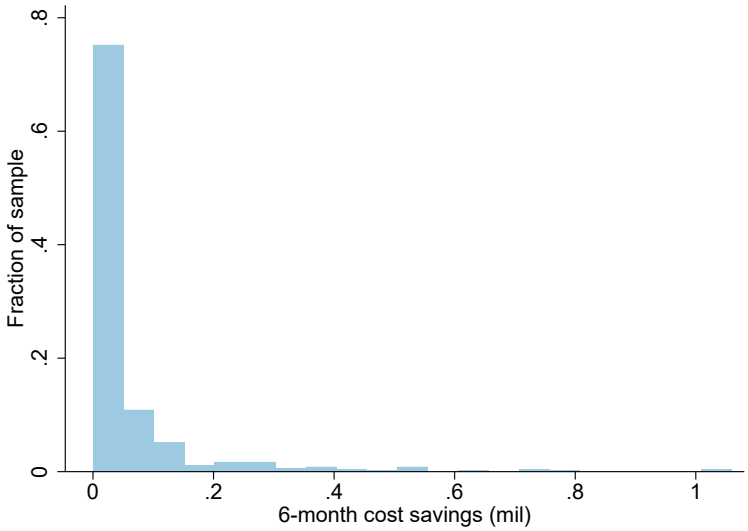
[55]There were other contributing factors to the backlog growing, including the proliferation of pay-for-delay schemes (Bokhari, 2013) and Citizen Petitions (Feldman et al., 2017).

[56]Previously, wait times were approximately 8-12 months, so the lag induced by the backlog was significant.

prices in a world where the backlog was less severe. To do so, I use the estimates presented in Section 5 and market characteristics to simulate a counterfactual price series, in which each firm's entry is sped up by a fixed amount of time. The difference in the simulated prices is weighted by the total volume (in units) in the market during the time interval, in order to approximate the total cost of delaying entry. In what follows, I present results corresponding to a scenario, in which the backlog was shortened by six months.[57]

Figure 8 shows the predicted cost savings from reducing the application review time by six months for the 341 unique products which experienced entry between December 2012 and January 2016.[58] The average cost savings across markets is $60 thousand dollars, which represents approximately 10-15% of the total revenue in these markets over the same six-month time horizon.

Figure 8: Distribution of cost savings from 6-month improvement in FDA backlog



Notes: The above figure shows the distribution of simulated cost savings in millions ($2013) from a six-month reduction in the FDA's application backlog. The estimates I use for the simulation are shown in Appendix Table A.1.

While these effects are economically significant, they represent modest losses in light of the results in Section 6.1. Evidence to why this is the case is shown in Figure 9. In particular, over 75% of the entrants stuck in the application backlog were would-be fifth, sixth, etc. entrants into their respective markets.[59] Once again, my simulations show that the markets with the highest counterfactual cost savings are those awaiting the third entrant, where the average expected cost savings per six-month reduction in the backlog reach nearly $200 thousand per entrant.

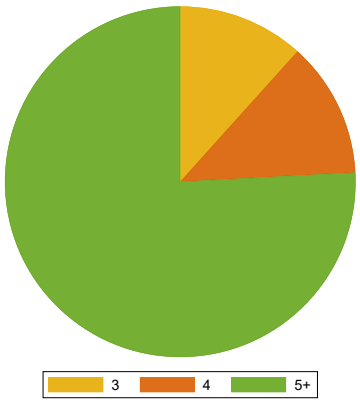
[57] It is difficult to approximate what "fully" eliminating the backlog would entail in this scenario, as there is a natural ebb and flow of approval wait times at the FDA corresponding to periods of low and high application rates, and indeed, this period was a time of historically high application rates. I choose six months in order to be conservative; however, preliminary estimates of longer wait times do not change the qualitative findings.

[58] I use this date range because the FDA officially cleared its backlog in January 2016.

[59] In its report on the backlog, Pew Charitable Trusts (2019) calls these late entrants "subsequent generics."

Before moving to the estimation of collusive damages, it is important to highlight the necessity of the model in the preceding counterfactuals. One could argue that similar conclusions could be reached by looking at the price change which was realized upon entry in the data. Obviously, as in the first counterfactual, this is not always possible because not all markets realize all levels of entry. And yet, in the second counterfactual, this argument does not hold. There, the value of the model relative to reduced-form methods—say, an event study—is that it allows me to consider scenarios which are increasingly out-of-sample. For example, it is possible that the equilibrium price which pertains when the backlogged entrant enters two years later is considerably different than the equilibrium price one and one half years later to the extent that the characteristics of the markets (beyond I_a) are different. Traditional reduced-form methods cannot account for such a change.

Figure 9: Breakdown of markets affected by FDA backlog, by entrant



Notes: The above pie chart shows the distribution of entrants whose entry was delayed by the backlog, according to the order in which they would eventually enter their respective markets. That is, a value of three implies that the firm was the third entrant into the market. All entrants beyond the fifth are grouped into the 5+ category. Moreover, entrants who would have been the second generic firm to enter the market are excluded from the figure. Data on entry are from the FDA's Orange Book files.

6.3 Collusion

6.3.1 Background

Throughout the paper so far, I have assumed the competitive conduct of manufacturers. As discussed in the introduction, however, this was often not the case in many generic markets in the early 2010s. As William Tong, the Attorney General in charge of the pending lawsuits, said in a November 2019 interview, the generic market was home to the "largest US corporate cartel in this period."[60]

The series of indictments filed by States Attorneys General shed light on the collusive activities of the ring. Central to their scheme was the mutual understanding of "fair share," whereby each

[60]The original interview can be found here: https://medicine.yale.edu/news-article/21752.

firm in the ring was entitled to a certain share of the retail pharmacy market for a particular drug.[61] The ring codified its fair share principle in an Excel spreadsheet: a firm's share of the market was a simple function of the number of participants in the market and the order in which that firm had entered—all else equal, earlier entrants were entitled to larger shares of the market; although, all firms were made to cede market share upon the entry of new entrants to the ring.[62]

This market allocation scheme provided the ring the structure it needed in order to accomplish its primary goal—raising prices without eroding market share. Nevertheless, raising prices required considerable coordination on behalf of the ring members. They were in near constant communication in order to avoid raising any flags among the pharmacy chains. Given the existing procurement system, avoiding suspicion required falsifying bids and, in several instances, selectively refusing to bid, with them falsely claiming supply chain issues. In the end, over 200 drugs—162 oral solid dosage drugs—were affected by the collusive behavior of these firms.[63] Notably, the collusive scheme spanned every therapeutic category.[64]

6.3.2 Damages from the Collusive Ring

While explicitly modeling the collusive rings' conduct is beyond the scope of this paper, it is possible to use the previously estimated competitive model to estimate the damages associated with the ring's activity. To do so, I use the recovered parametric cost estimates presented in Section 5 and the characteristics of each collusive market, including the *true* number of bidders present, to re-simulate each auction. This procedure allows me to estimate a counterfactual average winning bid for each market. Comparing the average winning bid in the data to this counterfactual average winning bid then yields a damage estimate due to the collusive efforts of manufacturers.[65]

The validity of this damage calculation relies on two strong assumptions: first, that the collusive markets are not different *a priori* from the non-collusive (i.e., competitive) markets in unobservable ways and, second, that the collusion itself did not induce changes *ex post* in the collusive markets beyond the equilibrium price.[66] I next provide evidence to demonstrate that both assumptions are reasonable given the data available.

A key feature of the indictments is their careful accounting of the formation of the ring. Early emails exchanged among ring members suggest that the markets which were involved in the scheme were selected specifically on the basis of the number of "quality competitors" present.[67] That

[61] By share, I refer to the share of total volume sold of a particular drug in the retail pharmacy market. Given the nature of the generic drug supply chain (shown in Appendix Figure A.1, a firm's market share is a function of which pharmacy contracts it holds and the volume demanded by those pharmacies.

[62] See Appendix Section B.1 for a reproduction of the table included in the May 2019 indictment.

[63] See Appendix Section B.2 for a full list of all drugs which were identified in at least one indictment as being affected by the collusive scheme.

[64] Although dermatological drugs are omitted from my analysis, they were also affected by the ring's operation. Indeed, there was an indictment filed in June 2020 which covered solely that therapeutic class.

[65] Note that although I label this difference as "damages," I make no *a priori* assumptions about the sign of the effect.

[66] Note that I can control for any observable differences between the collusive and competitive markets directly.

[67] See page 162 of the May 10, 2019 complaint.

said, the term "quality competitor" was simply an alias for a member of the ring. This strategy of targeting markets where the only firms present were ring members follows naturally from the nature of the procurement auctions themselves. For example, if the ring tried to raise prices collectively in the presence of a non-ring firm, then a non-ring firm potentially could underbid them and increase its own market share. I will discuss this idea more in Section 7.

In addition to this qualitative evidence, it is possible to look at other characteristics of these collusive markets, which I do not use in the estimation, but might indicate some inherent differences between the collusive and competitive markets. Table 7 presents the results of this exercise. There is only one statistically significant difference between markets with competitive conduct and without: the price of the products themselves. Of course, this increased revenue in collusive markets was a direct result of the ring's actions.

Table 7: Characteristics of competitive and collusive drug markets

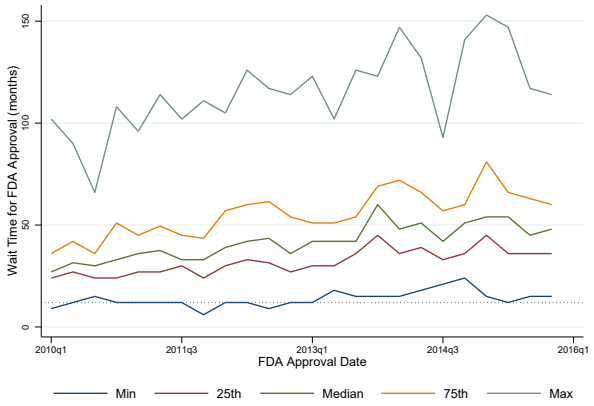
	Full Sample	Competitive	Collusive	P(T>t)
Annual Volume (mil.)	9.573	9.632	9.459	0.906
Average Annual Sales ($mil)	2.915	2.365	3.985	0.002
Age of Market	22.384	22.392	22.364	0.981
Entrants, Post-2010	0.774	0.766	0.790	0.405
During Collusive	0.304	0.317	0.279	0.237
After Collusive	0.592	0.578	0.620	0.234
# Bidders	4.599	4.737	4.331	0.043

Notes: The above table provides summary statistics for the drug products which comprise the full, competitive, and collusive estimation samples. Data on drug volume come from a private insurer's pharmacy claims data, drug characteristics from the FDA National Drug Code (NDC) files, and generic manufacturer entry from FDA Orange Book files. *Entrants* are indicators, where *During Collusive* refers to the 18-month period between July 2013-January 2019 when the collusive ring was at the height of its activity.

The second concern—whether or not the collusive conduct induced a structural change in those particular markets—is most relevant for entry and demand. The former relates back to the earlier literature which found that drug manufacturers are more likely to enter markets where revenues are currently high (Morton, 1999; Acemoglu and Linn, 2004). As a result of the astronomical price increases in the markets, revenues were historically high, and so one might expect that there was additional entry in these markets spurred by the rising revenues, i.e., *more* than would have occurred without the price increases.

Application data suggests this was not the case. Figure 10 shows the distribution of application dates for firms who entered in this period: all precede the creation of the collusive ring. In effect, due to the backlog, even firms who may have *wanted* to take advantage of the increased revenues in collusive markets were unable to do so. I discuss the implications of this finding in the next section.

Figure 10: Average wait times among entrants into collusive markets, 2012-2015



Notes: The above figure shows a selection of quantiles from the wait time distribution for drugs which are approved on a particular month. Wait time, here, is defined as the number of months between a manufacturer's filing of ANDA paperwork and the subsequent approval of the ANDA by the FDA, including any revisions. Data on wait times are from Feldman et al. (2017).

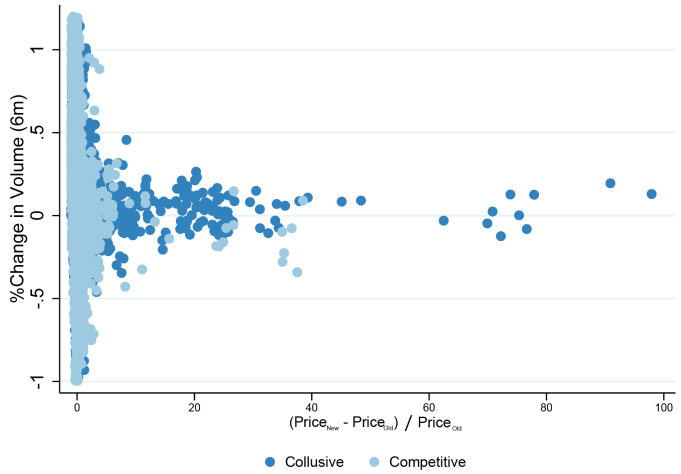
It is also possible that the price increases in collusive drug markets may have affected the underlying demand for those drugs. There is a long literature in economics estimating the price sensitivity of consumers for health care services, perhaps most famously in the RAND Health Insurance Experiment (Manning et al., 1987).[68] While most estimates suggest pharmaceutical demand is relatively inelastic, a recent paper by Yeung et al. (2018) finds large dispersion within elasticity estimates for pharmaceuticals using drug-specific price variation.

How and to what extent upstream price changes are passed through to consumers with drug insurance is unclear, especially in the short run. One potential mechanism is through coinsurance in the pharmacy plan benefit, whereby an individual is responsible for a fixed percentage of the underlying cost of the drug. While the use of coinsurance in lower tiers of a drug plan's formulary historically was rare, it has become increasingly more common. A recent Kaiser Family Foundation study found that 14-21% of employer-based plans used coinsurance on the lowest tier of their formularies (Claxton et al., 2019).

In practice, to test for such a demand channel, I examine the relationship between six-month price and demand changes for both collusive and competitive drugs. Figure 11 shows that the results of this exercise. Notably, I observe no systematic evidence of a reduction in demand following large increases in prices, even among the subset of collusive markets with price increases of over 4000%. In fact, for many of these outliers, demand increased.

[68] See Zweifel and Manning (2000) for an overview.

Figure 11: Percentage change in price versus percentage change in volume over 6-months, 2012-2015



Notes: The above figure shows the percentage change in volume versus the percentage chain in price over a six month period for both collusive (dark-blue) and competitive (light-blue) drug markets in 2012-2015. Data used in this graph are constructed from a private insurer's pharmacy claims data.

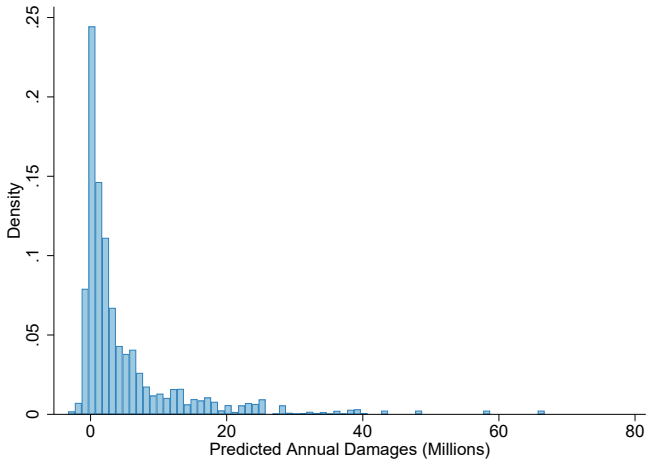
At this point, I have shown evidence to suggest that the collusive markets were not endogenously selected by the collusive ring and, further, that the collusive ring's behavior did not change the nature of those markets in a manner which would affect my model's predictions. With that, I present the damage counterfactuals.

Figure 12 presents the distribution of total annual damages for all collusive drug markets, and Table 8 provides a breakdown of total damages by ATC1. Notably, the predicted damages are nearly always positive (i.e., collusion rarely resulted in *lower* than the expected competitive price); although, there is significant variation conditional on being positive. Whereas some markets incurred tens of millions of dollars worth of damage, others were barely affected. In the average (median) market, though, the ring's activities resulted in damages of approximately $5 million ($2 million) dollars.

The total estimated cost of the collusive ring's activities is $1.4 billion dollars annually. Considering the ring was active for a period of 19 months, this suggests—for the patients in my sample alone—the total damage may have reached nearly $2.2 billion.

Importantly, this counterfactual assumes that the damages ceased as soon as the collusive ring was discovered. In reality, many of the markets where high prices were established as a result of coordination among the ring members still face elevated prices today. This suggests the estimates presented here are likely a lower bound, even for the subsample of drugs I consider in this analysis.

Figure 12: Estimated annual damages across collusive drug markets, 2012-2015



Notes: The above figure shows the distribution of simulated damages from the removal of the collusive ring from the market. The estimates I use for the simulation are shown in Appendix Table 5.

Table 8: Breakdown of estimated total annual damages by ATC-1

ATC1	Total Damages ($mil)	SE
A: Alimentary tract and metabolism	45.0	4.5
B: Blood and blood forming organs	5.8	3.8
C: Cardiovascular system	139.2	13.9
G: Genito-urinary system and sex hormones	88.9	5.2
H: Systemic hormonal preparations	6.9	2.2
J: Antiinfectives for systemic use	67.2	4.2
L: Antineoplastic and immunomodulating agents	46.1	1.0
M: Musculo-skeletal system	48.3	3.1
Multiple ATC	395.2	14.3
N: Nervous system	430.0	21.4
R: Respiratory system	26.5	0.6
V: Other	76.5	1.7
Total	1375.5	

Notes: The above figure shows the distribution of simulated damages from the removal of the collusive ring across markets according to the World Health Organization's (WHO) Level-1 Anatomic Therapeutic Classification (ATC). All dollar estimates are in million ($2013). Standard errors are estimated with via bootstrap (N=5000). The estimates I use for the simulation are shown in Appendix Table 5.

6.4 Benchmarking results

As discussed in Section 2.2, using one insurer's pharmacy claims data to estimate the model will eventually require the use of an external benchmark in order to scale estimates to the appropriate market level. The choice of benchmark is made difficult by the fact that many natural options

are not readily available. For instance, one promising benchmark would use the share of each pharmacy's prescriptions are dispensed to the insurer's enrollees; however, the total number of prescriptions filled at each pharmacy is unavailable. Alternatively, one could use the share of each drug's annual volume which are dispensed to the insurer's enrollees; however, again, the total volume of each drug is unavailable.

Given this dilemma, I use the following benchmark: the share of the insurer's patients in the US population. This measure has its own benefits and flaws. On the one hand, it is something that I can easily calculate and is tractable. On the other hand, it assumes away any heterogeneity in the prescription share across drugs. Given Medicare's near universal coverage of the elderly, it seems unlikely that the age distribution in my mainly privately insured sample is representative of the US population at large. That said, with the exception of any drugs which are used disproportionately by the elderly, I expect that the benchmark is otherwise acceptable.[69]

At the height of the collusive scheme—2013 and 2014, the insurer's population share was approximately 17%. Therefore, a back-of-the-envelope conversion from my data to the population at large involves multiplying each estimate by 5.8. This benchmark implies: (i) the market-level cost of delaying entry by six months for an average drug was $\$60,000 \times 5.8$, or \$352 thousand and (ii) the market-level damages imposed by the collusive ring were $\$2.2m \times 5.8$, or roughly \$12.9 billion.

7 Discussion and conclusion

Perhaps unsurprisingly, the volatility of prices in the generic market between 2012-2015 resulted in a robust response by policymakers. After the price spikes began, congressional hearings were held and investigations were launched in order to identify their underlying causes. These efforts were fruitful, insomuch as they eventually led to the discovery of the collusive ring. However, they also led to the development of new policies and legislation aimed at preventing what had just occurred from happening again. My findings provide a lens through which to evaluate their potential efficacy.

The FDA took a particularly active role in these efforts, working with Congress to implement the GDUFA programs in 2012 and 2017 (Berndt and Aitken, 2011). The GDUFA program collects over \$300 million annually from generic manufacturers to help, among other things, increase staffing at the FDA drug approval office such that it is more difficult for a new application backlog to form. My results in Section 6.2 provide context for whether such a level of spending is appropriate. In short, the answer depends on the composition of the FDA's application caseload: if, like during the backlog, the majority of pending applications are for late entrants, then this amount is surely too high. If the caseload comprises mainly second or third entrants, then it is surely too little.

Relatedly, in 2017, the FDA announced the FDA Reauthorization Act (FDARA), which established a priority review process for firms who were submitting generic drug applications for markets with fewer than three active manufacturers. This is likely to be an effective policy in light of my results from Section 6.1 because there are still large gains to be had from a third market participant.

[69] For this subset of drugs, my market-level damage estimate is likely too small.

That said, in many of the collusive markets, this policy would not have had any effect, as there were more than three firms present—these firms just happened to be colluding.

One could imagine an alternative policy for such a scenario, whereby the FDA grants similar priority review to firms who enter markets where the price has increased at an above-average rate without reasonable cause (e.g., shortage). In cases where there were no domestic firms available, the FDA could consider importing drugs from abroad under a mutual recognition program.[70] This type of policy relates back to the literature on collusion detection in auctions, where there are stark predictions for how prices ought to adjust upon entry of a non-collusive firm to a market with a "non-inclusive" collusive ring.[71] The indictments suggest that, this particular ring did not accept new members, and so encouraging entry into these markets may have helped offset some of the damages I estimated in Section 6.3, insomuch as these marginal competitive entrants may have exerted downward pressure on price. This is surely a fruitful area for future research.

Policymakers in Washington, D.C., reacted to the price spikes in a different manner, advocating for greater price transparency in drug markets, be it through "name and shame" policies or the development of a public price index.[72] In light of my results, it is unlikely that any name-and-shame policy will have its intended effect, both because there is so little brand awareness for generic firms and, beyond that, because there is so little responsiveness of demand to price to begin with. Similarly, if the ultimate goal is to prevent another collusive ring from forming in the generic (or, branded) market, price transparency efforts may have the unintended consequence of facilitating collusion: this relates back to the "observability problem" of Stigler (1964). Indeed, providing near-immediate drug price information at the pharmacy level might allow for easier detection of cheating within the ring.[73] That said, given the high concentration within the pharmacy sector, one could make the argument that even the average price reported in the NADAC survey by CMS is already too revealing in this respect.

To the extent that a system can be established which is "auto-correcting," it may be easier to avoid further disruptions to the market going forward. The reactionary entry policies mentioned above are one such solution; another involves increasing the elasticity of consumers to price changes. In some cases, the increase in generic prices were so significant that it actually would have been cheaper for patients to receive the branded version of the drug. Nevertheless, switching patients over was not always an option, as many generic firms no longer faced branded drug competition. In such instances, other formulary management methods might be applicable, including mandatory substitution. This is another important area for future research.

Taken together, my results support many of the FDA's current "pro-competitive" policies, particularly those aimed at reducing barriers to entry. Nevertheless, contrary to certain rhetoric in Washington, D.C., my results do not support encouraging entry indiscriminately—entry beyond a

[70] In these programs, the FDA would accept a firm's imported drugs from abroad so long as the firm had met the standards for efficacy and safety in their origin country. These "reciprocal" standards would be agreed upon by both agencies in advance. See Bollyky and Kesselheim (2017) for an overview.

[71] A non-inclusive ring is simply a collusive group which does not permit other members to join.

[72] See here for an example: https://www.nytimes.com/2018/05/17/health/drug-prices-generics-fda.html

[73] What has been suggested is akin to the price posting system studied in Sorensen (2000).

certain point does not generate significantly lower prices, and it may well be the case that the cost savings from these marginal entrants do not cover the estimated $1-5 million in fixed costs that these firms incur to enter the market (Berndt, 2002).

An alternative interpretation of my findings in this paper is that they support the notion that it is possible to improve the health care system *without* implementing new policies but rather by more effectively enforcing the policies we already have.[74] After all, the FDA's newest policies are being implemented so that they can more effectively fulfill their decades-old charter. Under that view, the question remains how to detect collusion and other misconduct by firms in the market—my results suggest that, with respect to these conduct issues, the stakes are incredibly high.

[74] See Scott-Morton (2019) for examples in other health care settings.

A Appendix

A.1 Supplemental tables and figures

Figure A.1: Simplified market structure of the US generic drug market

Notes: The above figure shows a simplified version of the US generic drug market. The focus of my analysis is on the circled upstream section, whereby manufacturers (represented here by Mylan, Aurobindo, and Teva) bid in drug procurement auctions hosted by pharmacies (represented here by Walmart, Walgreens, ..., AmerisourceBergen). The object that is being procured here is that pharmacy's business for a particular drug, say, atorvastatin 25 mg. tablets. Once the procurement auction is held, the lowest bidder begins providing the corresponding pharmacy with that drug (represented by the dotted arrows).

Table A.1: Drug-class specific estimations

	Cost per Pill				
	A	C	G	J	N
Constant	0.3149	0.1088***	0.0679	0.6513**	0.5829***
	(0.2082)	(0.0413)	(0.1539)	(0.2673)	(0.0999)
Annual Volume (100,000s)	-0.1920	−0.1976***	-0.2046	-0.1414	−0.1656*
	(0.1987)	(0.0689)	(0.3384)	(0.3166)	(0.0879)
Annual Volume Squared	0.0004	0.0004**	-0.1634	-0.0116	-0.0023
	(0.0007)	(0.0002)	(0.4412)	(0.2507)	(0.0158)
Capsule	0.1531	0.2326	0.3718	0.0725	0.3049***
	(0.1748)	(0.1609)	(0.3529)	(0.1257)	(0.1131)
Complexity	0.0336	0.2612***	0.0907	-0.0199	0.0390
	(0.0681)	(0.0907)	(0.1906)	(0.0160)	(0.1095)
Delivery Mechanism	0.2000	0.2612	1.1468***	0.2467	0.7044***
	(0.3467)	(0.1714)	(0.2975)	(0.2384)	(0.1304)
Wholesaler	0.0494	0.0132	-0.0174	0.1063	0.0680
	(0.0618)	(0.0257)	(0.1032)	(0.1108)	(0.0541)
Independent	0.1366	0.0545	-0.0724	0.2993	0.2431*
	(0.2152)	(0.1522)	(0.4220)	(0.3394)	(0.1394)
Chain 1	0.0233	0.0049	0.0251	0.0451	0.0487
	(0.0760)	(0.0300)	(0.1145)	(0.0902)	(0.0400)
Chain 2	0.0437	0.0117	0.0879	0.0724	0.0932*
	(0.0884)	(0.0344)	(0.1829)	(0.1165)	(0.0558)
Chain 3	0.0234	0.0077	0.0275	0.0503	0.0536
	(0.0649)	(0.0324)	(0.1010)	(0.0985)	(0.0416)
σ_{id}	0.5352*	0.4393***	0.2907**	0.2089	0.2228***
	(0.2733)	(0.1208)	(0.1286)	(0.1889)	(0.0778)
σ_{cl}	0.4060*	0.4674***	0.2751**	0.2382	0.1975***
	(0.2128)	(0.1156)	(0.1202)	(0.2067)	(0.0744)
ATC-2 FE	YES	YES	YES	YES	YES
N	302	733	191	257	1266

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

Notes: The above table summarizes estimates of the 24 parameters from Table 3 corresponding to the drug-class specific estimations. Each column represents a separate model which includes only the drugs listed in the column title above. Each WHO Level-1 ATC group is defined as follows: A = *Alimentary tract and metabolism*, C=*Cardiovascular system*, G=*Genito-urinary system and sex hormones*, *J=Antiinfectives for systemic use*, and N=*Nervous system.* Fixed effects for ATC-2 drug classes (n=2 to 8) are suppressed. Standard errors, shown in parentheses, are computed using 5,000 bootstrap samples of auction clusters.

A.2 Supplemental information on the estimation routine

A.2.1 Construction of the NLLS Simulator

Preliminaries

For each auction $a = 1, ..., A$:

- Draw S independent samples of size I_a from a standard normal distribution. Denote these sets of idiosyncratic cost shock draws $u_{s1a}, u_{s2a}, ...u_{sI_a a}$.

For each auction cluster $k = 1, ..., K$

- Draw S independent samples of size I_a from a standard normal distribution. Denote these sets of cluster cost shock draws $u_{s1k}, u_{s2k}, ...u_{sI_a k}$.

Note: These random cost draws are drawn before the estimation and are not functions of θ.

Construction of $\overline{\overline{X}}_k(\theta)$

1. For each auction $a = 1, ..., A$:

   - Form the simulator $\overline{X}_a(\theta)$ of $E(b_a^w)$:

   $$\bar{X}_a(\theta) = \frac{1}{S}\sum_{s=1}^{S} X_{sa}(\theta) \quad \text{where} \quad X_{sa}(\theta) = \min_{i \in I_a} \mu_a + \sigma_{id} u_{ik} + \sigma_{cl} u_{ik} \tag{10}$$

   given $\{Z_a, I_a\}$.

2. For each auction cluster $k = 1, ..., K$:

   - Form the simulator $\overline{\overline{X}}_k(\theta)$ of $E\left[\sum_{a \in k} \alpha_a b_a^w\right]$:

   $$\overline{\overline{X}}_k(\theta) = \sum_{a \in k} \alpha_a \overline{X}_a(\theta) \tag{11}$$

   where the exogenous weights α_a are known.[75]

[75] That is, each α_a corresponds to the market share of the pharmacy holding the auction and can be thought of as data for the purposes of the estimation.

A.2.2 Estimation Algorithm

Given an initial guess of $(\sigma_{id}^2, \sigma_{cl}^2)$ and a set of standard normal draws u_{sia} and u_{sik} drawn prior to the estimation as described in *Preliminaries* in Appendix Section A.2:

- Algorithm 1

  1. Select the parameter $\frac{\sigma_{cl}}{\sigma_{id}}$ which minimizes $\sum_{k=1}^{K}(g_1, g_2)'(g_1, g_2)$, where:

$$g_{1,n}(X_k; \frac{\sigma_{cl}}{\sigma_{id}}) = \max_{i \in I_k} \sum_{a \in k} \mathbf{1}(\text{Bidder i won auction } a) - \max_{i \in I_k} \sum_{a \in k} \hat{\mathbf{1}}(\text{Bidder i won auction } a)$$

$$g_{2,n}(X_k; \frac{\sigma_{cl}}{\sigma_{id}}) = \mathbf{1}(I_k = n)\frac{1}{I_k}\sum_{i=1}^{I_k}\sum_{a \in k} \mathbf{1}(\text{Bidder i won auction } a)) - \mathbf{1}(I_k = n)\frac{1}{I_k}\sum_{i=1}^{I_k}\sum_{a \in k} \hat{\mathbf{1}}(\text{Bidder i won auction } a))$$

  and $g_1 = [g_{1,2} \;\; g_{1,3} \;\; \dots \;\; g_{1,12}]$ and $g_2 = [g_{2,2} \;\; g_{2,3} \;\; \dots \;\; g_{2,12}]$

Given the estimated ratio $\frac{\hat{\sigma_{cl}}}{\sigma_{id}}$ and initial guess of (θ, σ^2):

- Algorithm 2

  1. Form the NLLS simulator using the algorithm in Appendix Section A.2
  2. Select the set of parameters (θ, σ^2) which minimize the SNLLS objective function, imposing that $\sigma^2 = \sigma_{id}^2 + \sigma_{id}^2 \times \frac{\sigma_{cl}^2}{\sigma_{id}^2}$ where the objective function is given by:

$$Q_{S,K}^{*} = \frac{1}{K}\sum_{k=1}^{K}\left(\bar{b}_k - \sum_{j \in \gamma(k)} \alpha_j \bar{X}_j(\theta)\right)^2 - \text{SimulationError} \tag{12}$$

and SimulationError is equal to:

$$\frac{1}{S(S-1)}\frac{1}{K}\sum_{k=1}^{K}\left(\sum_{a \in k}\alpha_a^2 \sum_{S}\left(X_{sa}(\theta) - \bar{X}_a(\theta)\right)^2\right) + \frac{1}{S(S-1)}\frac{1}{K}\sum_{k=1}^{K}\left(2\sum_{m.n \in k,\, m \neq n}\sum \alpha_m \alpha_n \left(X_{sm}(\theta) - \bar{X}_m(\theta)\right)\left(X_{sn}(\theta) - \bar{X}_n(\theta)\right)\right) \tag{13}$$

where

$$\bar{X}_a(\theta) = \frac{1}{S}\sum_{s=1}^{S} X_{sa}(\theta) \quad \text{and} \quad X_{sa}(\theta) = \min_{i \in I_a} \mu_a + \tilde{\sigma} u_{i\gamma(a)} + \sigma u_{ia} \tag{14}$$

A.2.3 Monte Carlo Simulation of Estimated Parameters

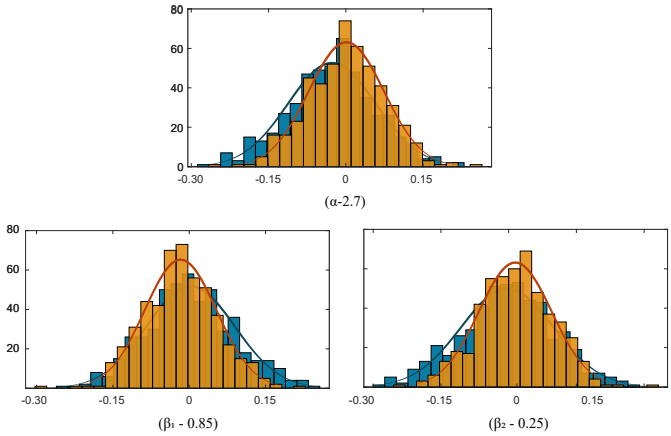
In this section, I present the results of the Monte Carlo study. In particular, I re-simulate the model $10,000$ times, with a sample of $N = 3,000$ average winning bids in each simulation. As in the data, each auction cluster comprises six auctions. I experimented with a wide range of parameter values; however, I present the results from one such Monte Carlo exercise below for ease of exposition. The qualitative findings are robust to the choice of parameter values. I generate log-normal costs, and follow the procedure described in Section 4.

$$c_i = e^{\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\varepsilon_i} \qquad \varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$$

Concretely in the example shown below the parameter values are:

$$\alpha = 2.7 \qquad \beta_1 = 0.85 \qquad \beta_2 = 0.25 \qquad \ldots$$

Figure A.2: Monte Carlo Simulation of Estimated Parameters



The graphs above represent the histogram of the mean subtracted estimated parameter values, when winning bids are observable, in blue, and when only average winning bids are observable, in orange.

As can be seen in Figure A.2, the estimated parameters are centered around the true values and approximately normally distributed.

$$\begin{array}{cccc} \hat{\alpha} = 2.7003 & \hat{\beta}_1 = 0.8483 & \hat{\beta}_2 = 0.2510 & \ldots \\ (0.0243) & (0.0235) & (0.0054) & \end{array}$$

What I persistently observe is that estimating the parameter values by observing N average winning bids, generates more efficient estimators of the parameter values compared to the case in which I observe N winning bids. Below I present the estimates from a Monte Carlo study once more with $10,000$ simulations, however, instead of $N = 3,000$ average winning bids, I now simply generate

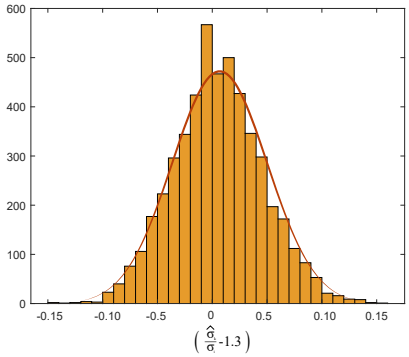
$N = 3,000$ actual winning bids. That is, the estimates below would result from a case in which the winning bid would be observable.

$$\tilde{\alpha} = 2.6910 \qquad \tilde{\beta}_1 = 0.8509 \qquad \tilde{\beta}_2 = 0.2498 \qquad ...$$
$$(0.0284) \qquad (0.0289) \qquad (0.0068)$$

As can be seen, these estimated parameters have higher variances compared to the initial simulated estimates. I suspect that what makes my estimation procedure more efficient is the following. Although information is lost by observing averaged winning bids, instead of each winning bid separately, these averaged winning bids still incorporate information from multiple winning bids. Thus, a dataset of $3,000$ averaged winning bids, with each average calculated from 6 winning bids, while not as informative as a dataset comprising $18,000$ winning bids, is still more informative than a dataset comprising $3,000$ winning bids.

Next, I investigate the properties of the $\frac{\sigma_{cl}}{\sigma_{id}}$ estimator described in Section 4. To do so, I once more re-simulate the model $5,000$ times, with a sample of $N = 3,000$ average winning bids in each simulation. As in the data, each auction cluster comprises six auctions. For exposition let $\sigma_{cl} = 1.3$ and $\sigma_{id} = 1$ leading to $\frac{\sigma_{cl}}{\sigma_{id}} = 1.3$.

Figure A.3: Monte Carlo Simulation of Estimated Parameters



The graphs above represent the histogram of the mean subtracted estimated $\left(\hat{\frac{\sigma_{cl}}{\sigma_{id}}}\right)$ value.

As can be seen in Figure A.3, the estimated ratio is centered around the true parameter value, with an approximate normal distribution. The utilized estimation approach once more does a rather good job in recovering the true parameter value. Concretely, the average estimated value of the ration, as well as its standard deviation are

$$\left(\hat{\frac{\sigma_{cl}}{\sigma_{id}}}\right) = 1.3057$$
$$(0.0412)$$

Once again, experimenting with different parameter values produced comparable qualitative results.

B Data Appendix

B.1 Additional information on the collusive ring

Figure B.1: "Fair-share" system used by collusive ring

Market Share - Fair Unit Share assumptions

Order of Entry Grid

Number of Competitors

Order of Entry \ Number of Competitors	1	2	3	4	5	6	7
1	100%	60%	45%	35%	30%	30%	30%
2		40%	35%	30%	25%	25%	25%
3			20%	20%	20%	20%	20%
4				15%	15%	15%	15%
5					10%	10%	10%
6						10%	10%
7							10%
Total	100%	100%	100%	100%	100%	100%	100%

Note: This is a chart used by one of the collusive firms in order to keep track of the agreed upon "fair share" market allocation scheme. Each cell within the table represents the market share that a firm is entitled to assuming they are the nth entrant (row) in a market with y competitors (column).

Source: Page 39 of the non-public version of the complaint filed in the US District Court for the District of Connecticut on May 10, 2019. See here: Link

B.2 List of collusive drug products in collusive sample

Acetazolamide
Acyclovir
Allopurinol
Amantadine HCL
Amiloride HCL/HCTZ
Amitriptyline
Amoxicillin/Clavulanate
Amphetamine/Dextroamphetamine
Atenolol Chlorthalidone
Atropine Sulfate
Azithromycin
Baclofen
Balsalazide Disodium
Benazepril HCTZ
Bethanechol Chloride
Bromocriptine Mesylate
Budesonide
Bumetanide
Buprenorphine
Buprenorphine Naloxone
Buspirone Hydrochloride
Cabergoline
Capecitabine
Captopril
Carbamazepine
Carisoprodol
Cefdinir
Cefprozil
Cefuroxime Axetil
Celecoxib
Chlorpromazine HCL
Cimetidine
Ciprofloxacin HCL Tablet
Clarithromycin
Clemastine Fumarate
Clomipramine HCL
Cyproheptadine HCL
Desmopressin Acetate
Desogestrel and Ethinyl Estradiol [Kariva]
Dexmethylphenidate HCL [Focalin]
Dextroamphetamine Sulfate ("Dex Sulfate")
Diclofenac Potassium
Dicloxacillin Sodium
Diflunisal
Digoxin
Diltiazem HCL
Diphenoxylate Atropine HCL
Disopyramide Phosphate
Disulfiram
Divalproex
Doxazosin Mesylate
Doxycycline
Doxycycline Hyclate
Doxycycline Monohydrate
Drospirenone and Ethinyl Estradiol
Enalapril Maleate
Entecavir
Eplerenone
Epitol
Estazolam
Estradiol
Estradiol and Norethindrone Acetate
Ethinyl Estradiol and Levonorgestrel
Ethosuximide
Etodolac
Exemestane
Fenofibrate
Fluconazole
Fluoxetine HCL
Flurbiprofen
Flutamide
Fluvastatin Sodium
Fosinopril HCTZ
Gabapentin
Glimepiride
Glipizide-Metformin
Glyburide
Glyburide-Metformin
Griseofulvin
Haloperidol
Hydralazine
Hydrocodone Acetaminophen
Hydroxyurea
Hydroxyzine Pamoate
Irbesartan
Isoniazid
Isosorbide Dinitrate
Isotretinoin
Ketoprofen
Ketorolac Tromethamine
Labetalol HCL
Lamivudine/Zidovudine
Lamotrigine
Leflunomide
Levothyroxine
Loperamide HCL
Medroxyprogesterone
Meprobamate
Metformin (F) ER
Methadone HCL
Methimazole
Methotrexate Sodium
Methylphenidate
Methylprednisolone
Modafinil
Moexipril HCL
Moexipril HCL HCTZ
Montelukast
Nabumetone
Nadolol
Naproxen Sodium
Niacin
Nimodipine
Nitrofurantoin
Norethindrone Acetate
Omega-3-Acid Ethyl Esters
Omeprazole-Sodium bicarbonate
Ondansetron
Oxaprozin
Oxybutynin Chloride
Oxycodone Acetaminophen
Oxycodone HCL
Paricalcitol
Paromomycin
Penicillin VK
Pentoxifylline
Perphenazine
Phenytoin Sodium
Pilocarpine HCL
Pioglitazone-Metformin
Piroxicam
Potassium Chloride ER
Pravastatin
Prazosin HCL
Prednisone
Prochlorperazine
Progesterone
Propranolol
Raloxifene HCL
Ranitidine HCL
Silver Sulfadiazine
Spironolactone HCTZ
Sumatriptan
Tamoxifen Citrate
Temozolomide
Timolol Maleate
Tizanidine HCL
Theophylline
Tolmetin Sodium
Tolterodine Tartate
Topiramate
Trazodone
Triamterene HCTZ
Trifluoperazine HCL
Ursodiol
Valganciclovir
Valsartan HCTZ
Verapamil
Warfarin Sodium

Table B.1 continued from previous page

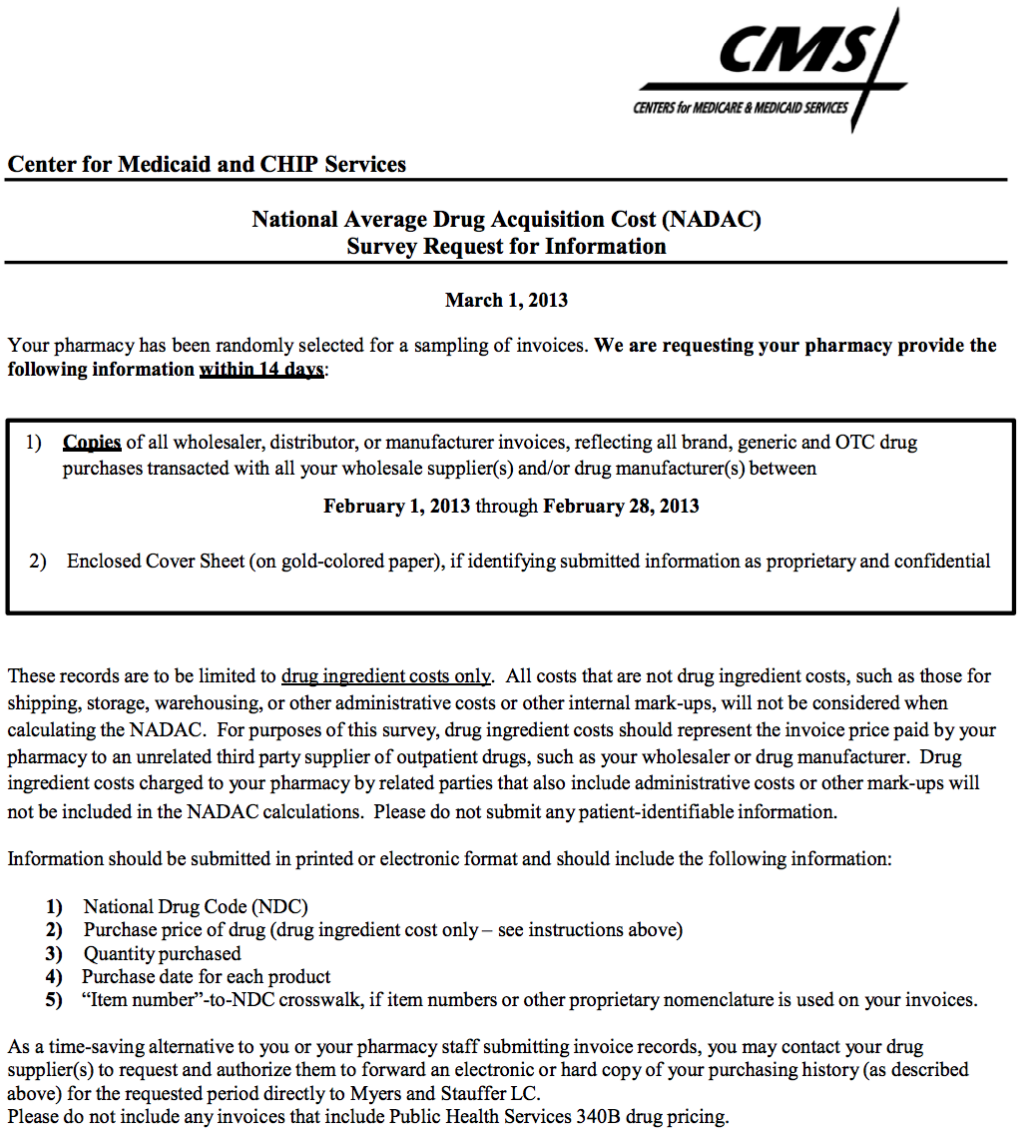
Acetazolamide	Drospirenone and Ethinyl Estradiol	Nabumetone

Notes: These drugs include all active ingredients which were mentioned in any of the federal indictments, as possibly being involved in the collusive ring. Drugs which were exclusively available in non-solid dose formulations were excluded. This yielded a total of 161 collusive drugs.

B.3 NADAC Data

B.3.1 Example NADAC survey questionnaire

Figure B.2: NADAC survey questionnaire distributed to pharmacies



CMS
CENTERS for MEDICARE & MEDICAID SERVICES

Center for Medicaid and CHIP Services

National Average Drug Acquisition Cost (NADAC)
Survey Request for Information

March 1, 2013

Your pharmacy has been randomly selected for a sampling of invoices. **We are requesting your pharmacy provide the following information within 14 days**:

1) **Copies** of all wholesaler, distributor, or manufacturer invoices, reflecting all brand, generic and OTC drug purchases transacted with all your wholesale supplier(s) and/or drug manufacturer(s) between

February 1, 2013 through **February 28, 2013**

2) Enclosed Cover Sheet (on gold-colored paper), if identifying submitted information as proprietary and confidential

These records are to be limited to drug ingredient costs only. All costs that are not drug ingredient costs, such as those for shipping, storage, warehousing, or other administrative costs or other internal mark-ups, will not be considered when calculating the NADAC. For purposes of this survey, drug ingredient costs should represent the invoice price paid by your pharmacy to an unrelated third party supplier of outpatient drugs, such as your wholesaler or drug manufacturer. Drug ingredient costs charged to your pharmacy by related parties that also include administrative costs or other mark-ups will not be included in the NADAC calculations. Please do not submit any patient-identifiable information.

Information should be submitted in printed or electronic format and should include the following information:

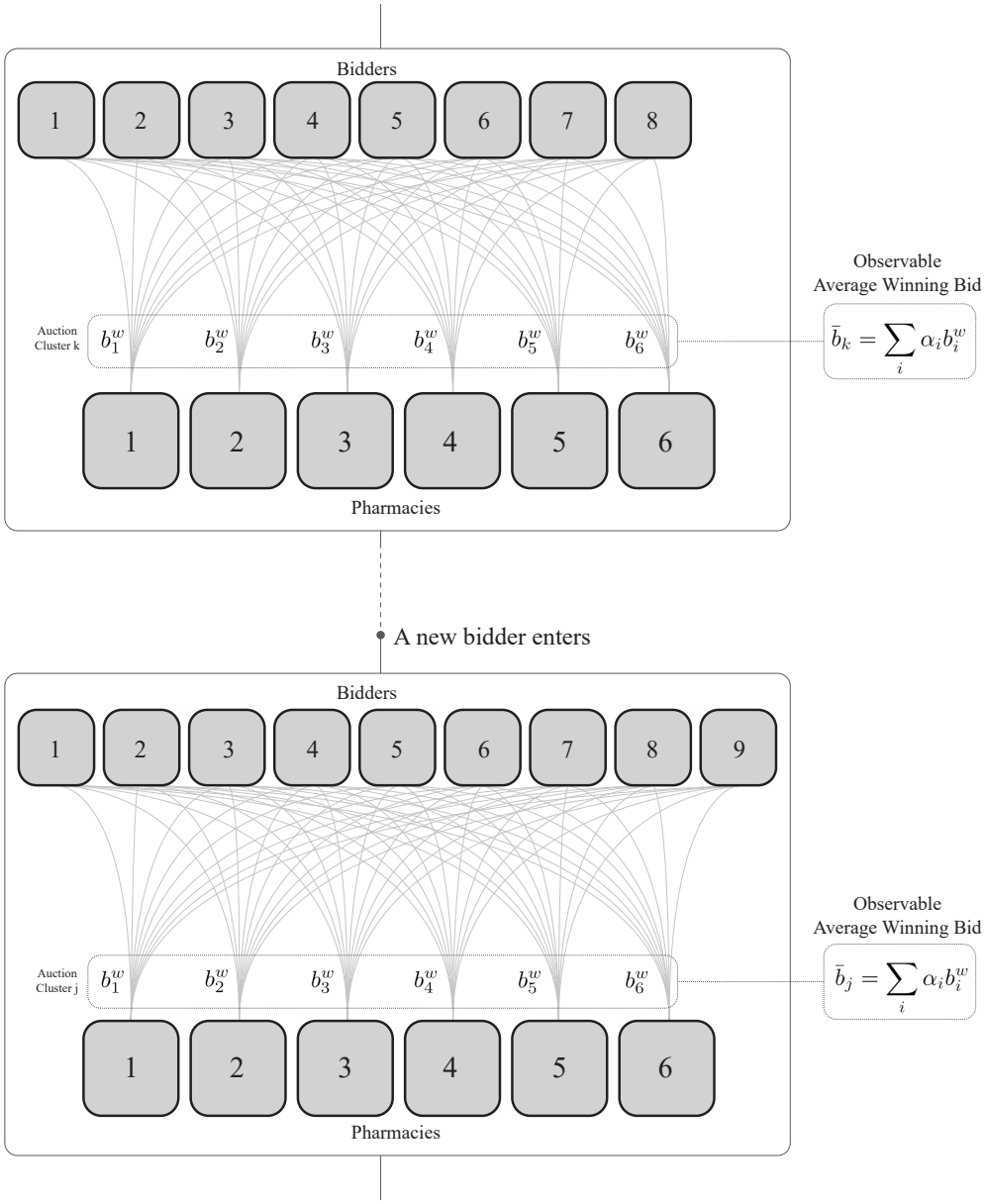
1) National Drug Code (NDC)
2) Purchase price of drug (drug ingredient cost only – see instructions above)
3) Quantity purchased
4) Purchase date for each product
5) "Item number"-to-NDC crosswalk, if item numbers or other proprietary nomenclature is used on your invoices.

As a time-saving alternative to you or your pharmacy staff submitting invoice records, you may contact your drug supplier(s) to request and authorize them to forward an electronic or hard copy of your purchasing history (as described above) for the requested period directly to Myers and Stauffer LC.
Please do not include any invoices that include Public Health Services 340B drug pricing.

Note: This is the survey letter which is mailed to pharmacies who have been selected to participate in the monthly NADAC survey. Often, the form is filled out at the corporate headquarters of the pharmacy. Source: CMS.

B.3.2 Data construction

Figure B.3: Data generating process from auction to NADAC survey



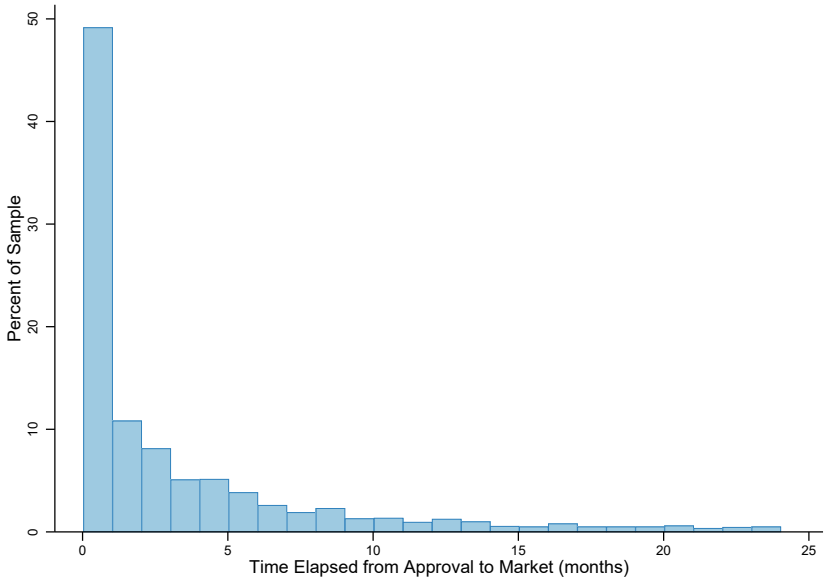
Notes: This figure shows how the average winning bid data are generated from the winning bid data at each pharmacy chain. There are two auction clusters shown: the top (*Auction Cluster k*) represents a series of auctions with eight bidders and the bottom (*Auction Cluster j*) represents a series of auctions with nine bidders. Note that both auction clusters are for the same drug product; however, the auctions held at each pharmacy across the two clusters are *not* the same. In particular, the details of the auctioned contract (e.g., volume) and the manufacturers' costs for fulfilling each auctioned contract are not the same across the two clusters. The winning bids from each auction within each cluster, denoted b_i^w for $i \in \{1, \dots 6\}$, are averaged using weights α_i to yield estimates of the average winning bid in each cluster: $\bar{b}_k$ and $\bar{b}_j$. These are the objects I observe in the NADAC survey data.

Other notes on the data construction process:

1. To convert the price series into a single winning bid, I divide the monthly NADAC price series into distinct periods which are characterized by having the same number of active firms, i.e., where the auction participants are fixed, and then calculate an average market price during that period.

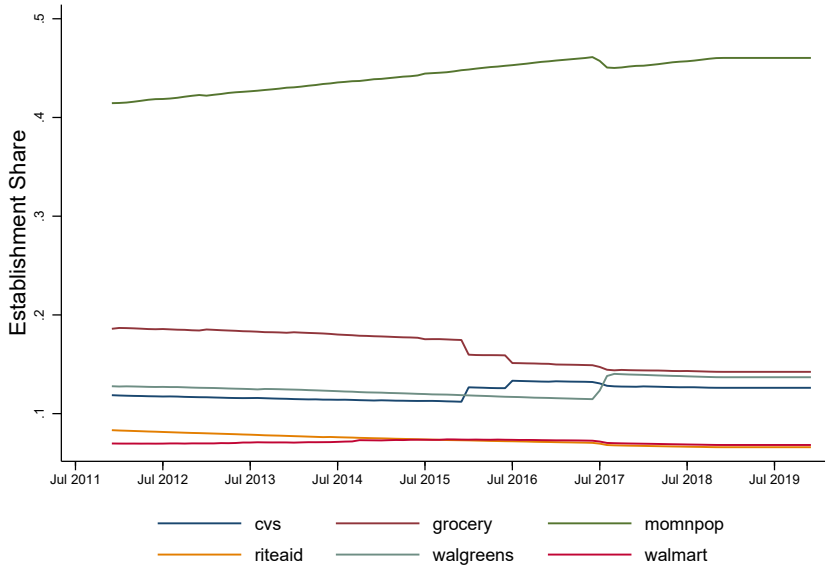
2. In cases where there is entry of a new firm, I limit the price observations in the average calculation to those occurring at least three months after the entering firm begins marketing its products, in order to allow for the market price to stabilize. As shown in Appendix Figure B.4, 90% of firms begin marketing their products within five months of FDA approval. In the claims data, I find another lag between the marketing date and the date upon which an entering firm's products end up on the shelves at retail pharmacies. This is likely due to a pharmacy needing to use up its existing inventory. This process varies by pharmacy chain but typically occurs within three months.

Figure B.4: Distribution of the time from ANDA approval to active marketing status



Notes: The above figure shows the distribution of time elapsed between the official FDA approval date of a generic firm's drug application and the first date that a unit of its product appears in the private insurer's pharmacy claims data. Data on ANDA approval dates are from the FDA's Orange Book files.

Figure B.5: NADAC weights between 2012-2018, by pharmacy chain



Notes: The above figure shows the pharmacy weights which are derived from NPPES data, in order to replicate the survey weights in CMS'S monthly NADAC survey. Each share represents the "establishment share" of each pharmacy outlet, or the number of physical locations of the pharmacy relative to all physical pharmacy locations. The shares serve as the α weights in Section 4, although we mask the identities of the chains in the estimation.

References

Acemoglu, D. and Linn, J. (2004). Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry*. *The Quarterly Journal of Economics*, 119(3):1049–1090.

Aitken, M. L. and Kleinrock, M. (2019). Medicines use and spending in the u.s.: A review of 2018 and outlook to 2023. *IQVIA Institute for Human Data Science*.

Asker, J. (2010). A study of the internal organization of a bidding cartel. *American Economic Review*, 100(3):724–62.

Athey, S. and Haile, P. A. (2002). Identification of standard auction models. *Econometrica*, 70(6):2107–2140.

Athey, S., Levin, J., and Seira, E. (2011). Comparing open and sealed bid auctions: Evidence from timber auctions. *The Quarterly Journal of Economics*, 126(1):207–257.

Bajari, P. and Ye, L. (2003). Deciding between competition and collusion. *Review of Economics and statistics*, 85(4):971–989.

Berndt, E. R. (2002). Pharmaceuticals in us health care: determinants of quantity and price. *Journal of Economic perspectives*, 16(4):45–66.

Berndt, E. R. and Aitken, M. L. (2011). Brand loyalty, generic entry and price competition in pharmaceuticals in the quarter century after the 1984 waxman-hatch legislation. *International Journal of the Economics of Business*, 18(2):177–201.

Berndt, E. R., Conti, R. M., and Murphy, S. J. (2017). The landscape of us generic prescription drug markets, 2004-2016. Working Paper 23640, National Bureau of Economic Research.

Berndt, E. R., Conti, R. M., and Murphy, S. J. (2018). The generic drug user fee amendments: an economic perspective. *Journal of Law and the Biosciences*, 5(1):103–141.

Bokhari, F. A. (2013). What is the price of pay-to-delay deals? *Journal of Competition Law and Economics*, 9(3):739–753.

Bollyky, T. and Kesselheim, A. (2017). Can drug importation address high generic drug prices? *Hutchins Center Working Papers*.

Claxton, G., Rae, M., Damico, A., Young, G., McDermott, D., and Whitmore, H. (2019). Health benefits in 2019: premiums inch higher, employers respond to federal policy. *Health Affairs*, 38(10):1752–1761.

Conrad, R. and Lutter, R. (2019). Generic competition and drug prices: New evidence linking greater generic competition and lower generic drug prices. Report, Food and Drug Administration.

Dave, C. V., Kesselheim, A. S., Fox, E. R., Qiu, P., and Hartzema, A. (2017). High generic drug prices and market competition: A retrospective cohort study. *Annals of internal medicine*, 167(3):145—151.

Duggan, M. and Scott Morton, F. (2010). The effect of medicare part d on pharmaceutical prices and utilization. *American Economic Review*, 100(1):590–607.

Feldman, R., Frondorf, E., Cordova, A. K., and Wang, C. (2017). Empirical evidence of drug pricing games-a citizen's pathway gone astray. *Stan. Tech. L. Rev.*, 20:39.

Frank, R. G. and Salkever, D. S. (1997). Generic Entry and the Pricing of Pharmaceuticals. *Journal of Economics & Management Strategy*, 6(1):75–90.

GAO (2014). Public health threat continues despite efforts to help ensure product availability. *Washington, DC: United States Government Accountability Office February*, pages 14–194.

GAO (2016). Part d generic drug prices declined overall, but some had extraordinary price increases. *Report to Congressional Requesters*.

Grabowski, H. G. and Vernon, J. M. (1992). Brand loyalty, entry, and price competition in pharmaceuticals after the 1984 drug act. *The Journal of Law & Economics*, 35(2):331–350.

Gupta, R., Bollyky, T. J., Cohen, M., Ross, J. S., and Kesselheim, A. S. (2018). Affordability and availability of off-patent drugs in the united states—the case for importing from abroad: observational study. *bmj*, 360.

Hendricks, K. and Porter, R. H. (1989). Collusion in auctions. *Annales d'Économie et de Statistique*, 15/16:217–230.

Hill, A. M., Barber, M. J., and Gotham, D. (2018). Estimated costs of production and potential prices for the who essential medicines list. *BMJ global health*, 3(1).

Jofre-Bonet, M. and Pesendorfer, M. (2000). Bidding behavior in a repeated procurement auction: A summary. *European Economic Review*, 44(4-6):1006–1020.

Jofre-Bonet, M. and Pesendorfer, M. (2003). Estimation of a dynamic auction game. *Econometrica*, 71(5):1443–1489.

Kesselheim, A. S., Gagne, J. J., Franklin, J. M., Eddings, W., Fulchino, L. A., Avorn, J., and Campbell, E. G. (2016). Variations in patients' perceptions and use of generic drugs: results of a national survey. *Journal of general internal medicine*, 31(6):609–614.

Laffont, J.-J., Ossard, H., and Vuong, Q. (1995). Econometrics of first-price auctions. *Econometrica: Journal of the Econometric Society*, pages 953–980.

Lieberman, S. M. and Ginsburg, P. B. (2018). Would price transparency for generic drugs lower costs for payers and patients. *Brookings Institute*, 14.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., and Leibowitz, A. (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. *The American economic review*, pages 251–277.

Maskin, E. and Riley, J. (1984). Optimal auctions with risk averse buyers. *Econometrica: Journal of the Econometric Society*, pages 1473–1518.

Morton, F. M. S. (1999). Entry decisions in the generic pharmaceutical industry. *The RAND Journal of Economics*, 30(3):421–440.

NPCA (2019). Health of independent pharmacy. Technical report, National Community Pharmacists Association.

Olson, L. M. and Wendling, B. W. (2013). *The effect of generic drug competition on generic drug prices during the Hatch-Waxman 180-day exclusivity period*. Bureau of Economics, Federal Trade Commission.

Pew Charitable Trusts (2019). Fda approves more generic drugs, but competition still lags. Technical report, Pew Charitable Trusts.

Porter, R. and Zona, D. (1999). Ohio school milk markets: An analysis of bidding. *RAND Journal of Economics*, 30(2).

Porter, R. H. and Zona, J. D. (1993). Detection of bid rigging in procurement auctions. *Journal of political economy*, 101(3):518–538.

Reed, B. N., Fox, E. R., Konig, M., Jackevicius, C. A., Masoudi, F. A., Rabinstein, A. A., and Page II, R. L. (2016). The impact of drug shortages on patients with cardiovascular disease: causes, consequences, and a call to action. *American Heart Journal*, 175:130–141.

Reiffen, D. and Ward, M. R. (2002). Recent empirical evidence on discrimination by regulated firms. *Review of network Economics*, 1(1).

Sacher, S. and Khinast, J. G. (2016). An overview of pharmaceutical manufacturing for solid dosage forms. In *Process Simulation and Data Modeling in Solid Oral Drug Development and Manufacture*, pages 311–383. Springer.

Saha, A., Grabowski, H., Birnbaum, H., Greenberg, P., and Bizan, O. (2006). Generic competition in the us pharmaceutical industry. *International Journal of the Economics of Business*, 13(1):15–38.

Scott-Morton, F. (2019). Diagnosing the problem: Exploring the effects of consolidation and anticompetitive conduct in health care markets. *House Committee on the Judiciary*.

Shrank, W. H., Cox, E. R., Fischer, M. A., Mehta, J., and Choudhry, N. K. (2009). Patients' perceptions of generic medications. *Health Affairs*, 28(2):546–556. PMID: 19276015.

Sorensen, A. T. (2000). Equilibrium price dispersion in retail markets for prescription drugs. *Journal of Political Economy*, 108(4):833–850.

Starc, A. and Swanson, A. (2018). Preferred pharmacy networks and drug costs. Working Paper 24862, National Bureau of Economic Research.

Stigler, G. J. (1964). A theory of oligopoly. *Journal of political Economy*, 72(1):44–61.

Yeung, K., Basu, A., Hansen, R. N., and Sullivan, S. D. (2018). Price elasticities of pharmaceuticals in a value based-formulary setting. *Health Economics*, 27(11):1788–1804.

Zweifel, P. and Manning, W. G. (2000). Chapter 8 - moral hazard and consumer incentives in health care. In Culyer, A. J. and Newhouse, J. P., editors, *Handbook of Health Economics*, volume 1 of *Handbook of Health Economics*, pages 409 – 459. Elsevier.