NBER WORKING PAPER SERIES

DOES ENTRY REMEDY COLLUSION? EVIDENCE FROM THE GENERIC PRESCRIPTION DRUG CARTEL

Amanda Starc
Thomas G. Wollmann

Working Paper 29886
http://www.nber.org/papers/w29886

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2022

We thank John Asker, Emily Cuddy, Kathleen Hui, Michi Igami, Rob Porter, Fiona Scott Morton, Mike Sinkinson, Bob Topel, Brett Wendling, as well as seminar participants at the Wisconsin, Stanford, Chicago Booth, 2020 Midwest IO Fest, and 2022 ASSA/AEA Annual Meeting for their comments. We benefited immensely from conversations with Doni Bloomfield, with whom we have related work. Paulo Ramos provided not only superb research assistance but also substantive contributions to the model and its estimation. Paloma Avendano also provided excellent research assistance. Wollmann thanks the Becker Friedman Institute's Industrial Organization Initiative for support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Does Entry Remedy Collusion? Evidence from the Generic Prescription Drug Cartel
Amanda Starc and Thomas G. Wollmann
NBER Working Paper No. 29886
March 2022
JEL No. L11,L41,L65

ABSTRACT

Entry represents a fundamental threat to cartels engaged in price fixing. We study the extent and effect of this behavior in the largest price fixing case in US history, which involves generic drugmakers. To do so, we link information on the cartel's internal operations to regulatory filings and market data. We find that collusion induces significant entry, which in turn reduces prices. However, regulatory approvals delay most entrants by 2-4 years. We then estimate a structural model to assess counterfactual policies. We find that reducing regulatory delays by just 1-2 years equates to consumer compensating variation of $597 million-$1.52 billion.

Amanda Starc
Kellogg School of Management
Northwestern University
2211 Campus Drive
Evanston, IL 60208
and NBER
amanda.starc@kellogg.northwestern.edu

Thomas G. Wollmann
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
Thomas.Wollmann@chicagobooth.edu

1 Introduction

Cartels thwart competition, even in modern economies. The US Department of Justice has prosecuted price fixing by all three major canned tuna brands, 70 auto parts suppliers, 15 global financial institutions trading foreign currency, over 100 real estate investors bidding in foreclosure auctions, and many others in the last few years alone. However, as cartels raise prices and profits, they may also attract uncooperative entrants, whose efforts to gain market share will undercut the incumbents' agreements. Hence, entry can serve as a fundamental safeguard against sustained collusion. In many markets, though, entry is a slow and expensive process, so the likelihood that entry restores competitive prices depends critically on the size of costs and the length of delays. Despite this, and despite the policy relevance, little to no work explores the extent and effect of entry in cartelized markets. We study it in the context of the largest price fixing case in US history, which involves generic prescription drugmakers.

Historically, most economists and policymakers thought of generic drugs as a competition success story. When branded drugs lose patent protection, an influx of generics typically follows, capturing more than half of the market at less than half the branded equivalent's price (Scott-Morton (1999); Wendling et al. (2011)). In recent years, however, prices of many generics rose substantially. According to court documents described below, many of these increases can be traced to a single precipitating event: in 2013, Teva Pharmaceuticals, the largest generic drugmaker, hired NP, a marketing executive with especially strong industry relationships, and tasked her with "price increase implementation."[1] Over an 18-month period, industry participants exchanged thousands of calls and texts—alongside countless LinkedIn, Facebook, WhatsApp, and face-to-face conversations—with contacts at rival firms to coordinate the increases (Complaint, page 322).[2] Following this period, prescription drug expenditures by governments, private insurers, and individuals rose sharply by billions of dollars.

In this paper, we measure the effect of price fixing on market entry, estimate a structural model of generic drug competition, and use our estimates to assess counterfactual policies that would reduce regulatory costs and delays. We exploit detailed information on the cartel's internal operations, which were revealed when a complaint against the drugmakers was filed in May 2019.[3] The complaint presents witness testimonies, private communications within and between rival firms, and internal documents collected in the course of an extensive government investigation. It includes a list of the drugs for which NP fixes prices, the dates on which those prices are fixed, and the criteria NP uses to select which prices to fix. Conveniently, it also reports NP's own numerical measure of the strength of executive relationships with competing drugmakers, which is critical for predicting collusion.

Information from the complaint is crucial to our research design, which compares cartelized and uncartelized drugs before and after collusion. Such comparisons typically raise concerns about unobservable differences across markets and time that relate to both the likelihood of collusion and outcomes of interest. However, due to the government's investigation, we know how NP selects markets and observe the variables on which the selection depends, which are both stable over time and independent of the relevant outcomes. Also, the vast majority of the price increases were implemented within about a year of NP's first anniversary

[1] We identify individuals using their initials. Their names are discoverable in court documents but irrelevant to our analysis.

[2] Throughout the paper, we use the word "cartel" or similar terms to characterize the behavior and economic arrangement, not reflect the findings of the court. The case against Teva, its former employees, and several co-conspirators is ongoing at the time of writing, so from a legal perspective, this is an "alleged cartel" with respect to their involvement. However, the facts presented in the complaint, which we take at face value, satisfy any substantive definition of collusion. See Section 2 for details.

[3] See https://portal.ct.gov/AG/Press-Releases/2019-Press-Releases/DRUG-PRICE-FIXING-COMPLAINT-UNSEALED.

of joining Teva, mitigating concerns that they were timed to coincide with unobservable environmental changes. Moreover, observable outcomes in cartelized and uncartelized markets track extremely closely with each other leading up to collusion (but clearly diverge sharply afterwards).

Our sample consists of drugs manufactured by Teva in the quarter just prior to NP joining the firm (where "drugs" refers to substance-delivery-release-strength combinations).[4] We measure quantity using prescriptions, which we obtain from IQVIA and Medicaid, and we measure point-of-sale price in dollars per prescription, which we obtain from Medicaid. Before manufacturing a drug, firms must file an Abbreviated New Drug Application (ANDA) and have it approved by the Food and Drug Administration (FDA). ANDA filings closely correspond to entry. We obtain the filings and their respective dates from the FDA.[5] The dataset we build covers 2008 to 2019.

We begin by documenting four key reduced form findings. To do so, we combine information from the complaint with regulatory filings and market outcomes. First, collusion leads to abrupt price increases, which average 40-50%. Second, the cartel induces significant entry. Following collusion, firms file 30-40% more ANDAs. Third, regulation delays competition from newly filed ANDAs for years. On average it takes several years to receive an approval so, for example, a firm filing an ANDA in 2013 might not enter before 2017. Fourth, entry precipitates price declines.

These facts inform our structural model, which emphasizes two main decisions that generic drugmakers face. In the first stage, firms *without* regulatory approval file ANDAs if the sum of expected discounted profits associated with entering exceeds the sunk costs.[6] Entry is not immediate, since ANDA approvals involve significant, stochastic delays. In the second stage, firms *with* regulatory approval set prices. Pricing decisions depend on the history of play as well as incentive compatibility constraints: cartel members choose prices that maximize joint profits so long as no member has deviated in the past and every member finds it more valuable to cooperate. The model highlights how cartel formation raises prices, which in turn increases incentives to enter. Moreover, it emphasizes that entry exerts downward pressure on price, given that not all entrants are cartel members. However, it also underscores the point that observed entry will not "unravel" the collusive agreement (if it did, prices would return to competitive levels, so any "market clearing" entry would have already occurred prior to cartel formation).

Next, we take the model to the data. We estimate demand, use these estimates to recover marginal costs, and combine these results to forecast the profits that firms would earn for any hypothetical market structure under either collusion or competition. We use these forecasts in conjunction with the empirical distribution of entry delays to compute the value of entry, which equals the sum of discounted expected future profits. Firms enter if and only if the value of entry exceeds the sunk cost of doing so, so observed entry decisions map to the parameters that determine those costs. Since two types of firms (i.e., cartel members and nonmembers) make entry decisions, multiplicity is the rule rather than exception. Thus, we rely only on the necessary conditions implied by Nash equilibrium, bounding the parameters of interest using moment *inequalities* (Tamer, 2003; Pakes et al., 2015).

[4]Since Teva is by far the largest generic drugmaker, more than one-third of all drugs remain in the sample after the second restriction. In any case, merely to preserve comparability across drugs, we make other minor restrictions. For example, we exclude drugs that go "off patent" in the middle of the panel, since their price paths reflect very different dynamics. See Section 3 for details.

[5]All approval dates are reported by the agency, but only a subset of filing dates are. However, ANDA numbers are issued approximately sequentially, so filing dates can be inferred without meaningful error. See the Online Appendix for a detailed description of the process.

[6]This condition implies that cartel members do not coordinate entry decisions, even though they may cooperatively set prices. In this sense, the cartel is *semicollusive* (Fershtman and Muller, 1986). See Section 2 for more detail.

Our estimates reflect an assumption that collusive prices persist despite the government's investigation (and ongoing prosecution). While this restriction partly reflects the stability of drug prices around the time the investigation is publicized, it is mainly due to the results of a formal conduct test. The test provides a pairwise comparison between collusive and competitive pricing, following a general methodology laid out by Backus, Conlon, and Sinkinson (2021), which we adapt to our setting. To summarize the result, we find that the data strongly rejects competitive pricing post-2015 in favor of continued collusion. In other words, enforcement did not undermine the existing cartel agreements. The finding is consistent with the idea that some price fixing agreements are hard to initiate but easy to perpetuate, which is consistent with with related statements made in the complaint, and further emphasizes the importance of entry.

With estimates of the parameters governing demand, prices, and entry in hand, we simulate counterfactual outcomes under alternative policies. In particular, we consider reducing the sunk cost of entry and reducing the delays associated with entry. The counterfactuals are logical and important alternatives to consider for several reasons. First, there are slow and expensive drug-specific government approvals required for generic production. Firms must prepare and file an ANDA with the FDA, and the approval process can take years and cost millions of dollars. Second, these factors lie within the government's control: the FDA has varied fees considerably over the past decade and experimented with various expedited approval programs. Former commissioner Scott Gottlieb has lamented both the cost and time associated with entry (Gottlieb (2016)). Third and perhaps most important, regulatory costs and delays are basic features of many large industries. Examples range from air travel and power generation to defense contracting and retail banking. Hence, our broad conclusions are likely to extend beyond prescription drugs.

We estimate that sunk entry costs average $3.2 million per ANDA for the drugs in our sample. Reassuringly, the figure is consistent with with previous statements made by FDA officials. We also estimate that entry costs are significantly higher for ANDAs that cover multiple strengths and non-standard delivery methods (e.g., syrups and chewables rather than "pills"). Our first set of counterfactual experiments study the effect of reducing sunk costs by $400,000-$800,000 per ANDA—approximately one-eighth to one-quarter of the average expense required to enter a substance-delivery-release combination. We find that there is much more entry into cartelized markets witness much more entry and that consumer compensating variation equals $142-374 million. Our second set of counterfactual experiments study the effect of reducing delays by 1-2 years. We find that this has a similar effect on entry but a much greater effect on consumer compensating variation, which equals $596 million to $1.5 billion.

Our findings show that entry can play a key role in disciplining cartels, with a first order impact on consumer welfare. Nonetheless, the discipline it exerts may be incomplete and slow, resulting from high costs and long delays, many of which can be attributed to regulation. To be clear, we do not in any way mean to imply that the time and expense associated with generic drug evaluations are wasteful, only that limiting fees and hastening approvals generates significant surplus for buyers when incumbent firms are playing cooperatively. Relatedly, our estimates alone cannot advocate for one policy over another. Lower fees may draw resources away from other oversight activities, while quicker approvals may require additional staff (or, again, result in lax enforcement). Forecasting the costs associated with these changes is far beyond the scope of the present paper, which focuses on the relationships between cartel formation, entry, pricing, and purchases in equilibrium.

We contribute to the growing body of empirical work on cartels by incorporating equilibrium entry. While foundational work by Stigler (1964) names both cheating and entry as threats to collusive agreements,

most early and influential formalizations of the cartel problem (e.g., Green and Porter (1984)) study the former threat but rule the latter one out. Fershtman and Pakes (2000) relax this restriction by way of simulations, numerically solving Markov perfect equilibria that permit for entry, exit, and quality differences across firms. The authors report a strong industry relationship between collusion and subsequent entry—a relationship they state "has largely been ignored in the literature and has important implications for the welfare analysis of collusive behavior."[7] In empirical work, authors have carefully chosen markets where it is reasonable to assume that entry is either unlikely to occur in the near-term (Miller and Weinberg (2017)) or that entry is determined by factors that are "outside" the model (Porter (1983), Byrne and de Roos (2019), Igami and Sugaya (2020)).[8] By contrast, we not only allow for entry but endogenize the behavior, drawing on work by Borenstein (1989), Bresnahan and Reiss (1991), Mazzeo (2002), Seim (2006), and the large, growing literature that followed.

This paper also improves our understanding of prescription drug pricing (see, for example, Berndt et al., 2018). Sco (2000) shows that larger revenue markets, markets with more hospital sales, and products that treat chronic conditions attract more entry. At the same time, Ganapati and McKibbin (2021) show that the entry—the traditional mechanism of reducing market power—has been limited in the United States in recent years. We expand this literature by exploring the strategic incentives faced by generic manufacturers and their impact on consumers. Cuddy (2020) models this market as a series of simultaneous procurement auctions in order to study the equilibrium relationship between competition and prices. She estimates large damages from cartel behavior using a counterfactual auction model of competitive behavior. Similarly, Clark et al. (2021) estimate large damages using a reduced form approach. We focus on the market forces that may serve to alleviate these harms.

The paper is organized as follows. In the next section we describe the institutional setting and in section 3, the data. In section 4 we present our reduced form findings and in section 5 we outline the model. We report estimates of the structural parameters in section 6 and present our predictions under counterfactual policies in section 7. Section 8 is the conclusion.

2 Market setting and cartel operations

2.1 The US generic prescription drug industry

Generic drugs are a competition success story. When a branded drug loses exclusivity, generic entry drives prices down towards marginal cost. For this reason, the market is the "most dynamic and cost-effective in the world" (Scott-Morton and Boller (2017)). Historically, barriers to entry were relatively low. Under

[7] In particular, Fershtman and Pakes (2000) report, "The contrast between the entry states in the model that allows for collusion and the entry states in the model that does not is quite striking. The [relationships between entry and the state variables] simply *disappear* when collusion is not allowed" (emphasis in original). The authors emphasize the universally positive effect collusion has on entry; however their statement contains a slight but especially interesting caveat due to work by by Asker (2010), who studies bidding rings among stamp deals and shows how their arrangement induces *overbidding*, which deters entry.

[8] For example, Igami and Sugaya (2020) study vitamin cartels in the 1990s. In their setting, exogenous entry is an especially natural approach, given that competition unexpectedly arose from technology change, which benefited fringe suppliers. Alternatively, Miller and Weinberg (2017) study the effect of the Miller-Coors joint venture on the US beer market. In their setting, the emergence of another major brewer is very unlikely, at least in the near future, given that the leading brands have accumulated brand equity over decades of advertising, which insulates them from startup competition. Moreover, while the "craft" segment has grown over the past three decades, their Table I shows the top two firms have nearly two-thirds of nationwide share. Entry barriers are not necessary, however, to restrict entry. For instance, Harrington (1989) shows that sufficiently patient incumbents can credibly threaten entrants, while Scott-Morton (1997) finds evidence of predation by turn-of-last-century British shipping cartels.

the Hatch-Waxman Act, firms can enter by filing an ANDA that shows that the active ingredient, delivery mechanism, strength, and dose of the generic drug are the same as the branded drug. The generic drug must be "bioequivalent" to the branded drug. Generally, the first generic entrant will price its product slightly lower than the branded drug, and the second generic entrant will reduce the price to approximately 50% of the branded drug price. Conditional on having a large number of entrants, prices fall to around 20% of the price of the branded counterpart.

Generic drug manufacturers compete with each other to sell the generic drugs they produce to wholesalers, distributors, and in some cases, directly to retail pharmacy chains, mail-order and specialty pharmacies, hospital chains, and some health plans. Due to complex "cost-plus" reimbursement rules, higher wholesale prices may weakly benefit these market participants.[9] In recent years, wholesale prices for some generic molecules have increased substantially. The increases have been attributed to both supply shortages and anti-competitive behavior (Cuddy, 2020).

2.2 The cartel

The cartel we study traces back to a single change in industry leadership: on April 22, 2013, NP joined Teva Pharmaceuticals.[10] This event was special for several reasons. First, in the years leading up to the cartel's formation, NP forged uniquely strong relationships. She worked as the Director of Global Generic Sourcing for Amerisource Bergen (ABC), one of the three major US drug distributors. The role led to "routine interaction with representatives from every major generic drug manufacturer" (Complaint, page 158). Second, Teva was—and still is—the world's largest generic drugmaker. By early 2013, for example, it produced about one out of every three generic tablets and capsules.[11] Third, NP's role as Director of Strategic Customer Marketing involved, in her own words, "price increase implementation" (Complaint, page 158). The significance of this move was not lost on other industry leaders. At Taro Pharmaceuticals, another leading generic drugmaker, the Vice President of Sales and Marketing emailed the COO just days after NP left ABC to say, "[NP] Going to Teva—Hush Hush for now" (Complaint, page 159).

The complaint states that just 8 days after joining Teva, NP began identifying target markets. The process was highly structured. She started by assigning each generic drugmaker an individual "quality" rating, which ranged from -3 to 3 and reflected the strength of her relationships with their sales and marketing executives. Next, she combined these ratings into a drug-specific score, which was the most important element in the selection process. Finally, she factored in the number of firms in the market and other minor considerations. In an internally distributed 2014 document, she summarizes "Candidate Identification" as "Target 2-4 total players, where quality of competitor is high" and "[where] Teva has majority share and quality of competitors is high" (Complaint, page 217).[12]

[9] McKesson 2014 10-K, Cardinal 2014 10-K, ABC 2014 Annual Summary all explicitly state that their profits are positively affected by manufacturer price increases (due to cost- plus arrangements).

[10] Note that while the complaint focuses on drugs affected by NP, it alleges other segments of the industry were not entirely immune to coordinated behavior. These allegations do not affect our analysis. Our sample includes only drugs that Teva historically produced. As the complaint states, Teva was uncooperative before NP joined, so these markets are presumably competitive prior to her actions. Besides that, the complaint describes antitrust violations in the other segments that are qualitatively and comparatively unimportant.

[11] Most generic drugs are manufactured in a pill/capsule form, which is the delivery mechanism we study. Injectible drug markets differ in a host of ways—manufacturing processes are very different and customers are typically hospitals rather than retail pharmacies.

[12] The complaint reports NP's quality ratings and we observe market structure, so we can confirm NP used precisely these criteria. She also mentions targeting "Exclusive items" (i.e., those where Teva is the only supplier). Our focus is collusion, so we exclude these cases from our sample. They are rare, so this restriction is without any meaningful loss of generality.

The complaint states that the outcome of the process was a spreadsheet titled "Immediate PI File" (where "PI" stands for "Price Increase"). It contained a list of drugs for which NP expected to cooperatively raise prices. NP forwarded the list to supervisors on May 24, and Teva changed prices on July 3, preceded or followed closely by other drugmakers. The magnitudes of the increases are especially noteworthy. Many commonly prescribed medications—drugs treating cancer, bacterial infections, arthritis pain, and high blood pressure, to name a few—doubled or more in price. Figure I plots the increases in calendar time. Prices in cartelized and uncartelized drugs track closely until the date NP is hired, at which point they diverge sharply.[13] By 2016, the former rise 50% relative to the latter.

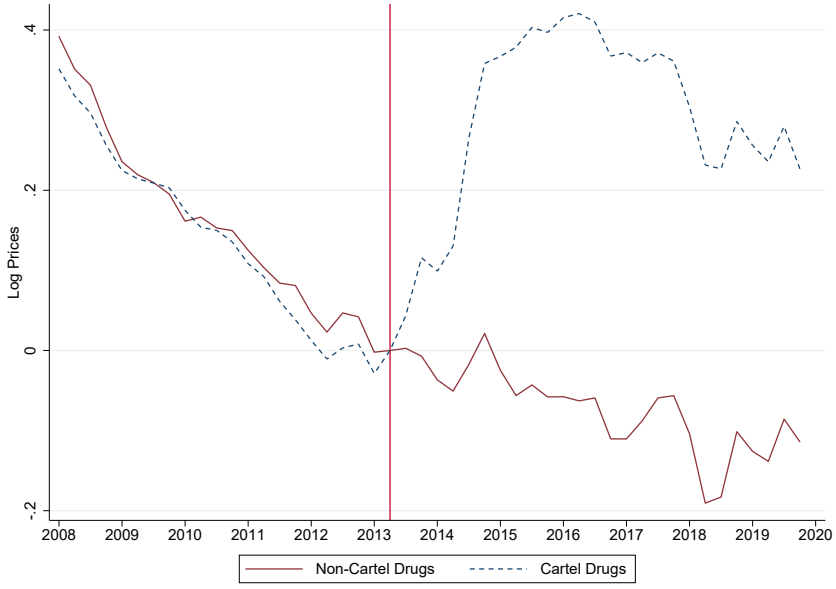

Figure I: *Prices in calendar time*

This figure plots log prices on the y-axis against calendar time on the x-axis. The dashed and solid lines correspond to cartelized and uncartelized drug markets, respectively. The vertical red line corresponds to the first quarter of 2013—the period in which NP joined Teva. Prices are normalized to zero in that quarter.

Implementing these increases required considerable coordination, which in turn required frequent communication. Leading up to the increases, phone records reveal thousands of calls and messages between NP and her counterparts at firms producing the drugs for which she sought to raise price. For instance, the day before an increase, NP exchanged 15 calls with 3 individuals.[14] Interestingly, since the firms could not

[13]Notice that prices trend down throughout the panel. They fall 20 percentage points in the first year and then deflate at about 3 percentage points per year thereafter. The first pattern reflects two patterns in the data. The panel includes drugs that lose exclusivity as late as 2007, so a subset experiences declines that are typical around patent expiry (and "Paragraph IV" challenges). Also, data quality improves in the first year, and when quantities are undermeasured, prices are mechanically higher. Since our estimates rely on relative comparisons between cartelized and uncartelized drugs, which exhibit similar patterns, we opted to include the first year's data in our analyses, letting it further demonstrate the comparability between the treatment and control groups. Modest deflation thereafter may reflect many underlying factors (e.g., outsourcing production to low wage countries, which reduces marginal cost).

[14]Though it does not affect our analysis, it's nonetheless interesting to note that most conspirators knew their behavior was illegal. For instance, according to the complaint, when NP described communications with rival firms during a 2013 internal meeting, MP, another Teva executive, "smiled, put her hands over her ears, and pretended that she could not hear what was being said" (Complaint, page 337). In other instances, executives deliberately avoided written communications. When a senior executive at Taro asked about the arrangement, a fellow executive replied, "No emails please. Phone call. [Redacted] let's discuss" (Complaint, pages 49-50). In yet

perfectly synchronize the price changes, sharp increases by one drugmaker often prompted its customers to approach other producers, who were forced to decline the business. In some cases, drugmakers outright declined, citing fictitious supply problems (Complaint, pages 239 and 265), while in others they quoted inflated prices, frustrating customers. Teva's "fluff pricing" to Schnucks, a Midwest grocery chain, was so egregious that the customer stated they felt "so insulted" (Complaint, page 146-7).[15]

The complaint states that NP coordinated five main rounds of price increases between July 2013 and January 2015. After the first round, opportunities for subsequent increases emerged for a host of reasons. One source was leadership changes. If a sales and marketing executive with a close relationship to NP moved from Firm A to Firm B, then the likelihood of Firm B fixing their prices increased. Zydus provides one such example. NP originally assigned the firm a quality rating of -3, indicating she lacked relationships with Zydus personnel. However, NP's colleague at Teva, KG, moved to Zydus in 2014, prompting her to change the firm's rating to 2. Subsequent phone and text records indicate that the two communicated extensively (Complaint, page 272-273). A second source of opportunities was supply disruptions. If a firm whose sales and marketing executives lacked relationships with NP was forced out of a particular drug market due to, for instance, an interruption in access to the active pharmaceutical ingredient, then the likelihood of fixing the price of that drug increased. Notably, the price increases would have been even more compressed were it not for NP's maternity leave, which resulted in a "hiatus" of over six months (Complaint, page 212).

According to the Connecticut Attorney General (AG), "suspicious price increases of certain generic drugs" prompted the state in July 2014 to begin an investigation, which is still underway. [16] On May 10, 2019, 43 US states and territories filed a complaint, which was unsealed the following month. It alleges a "horizontal conspiracy" to fix prices for multiple generic drugs in violation of Section 1 of the Sherman Act. Alongside the states' civil suits, the US Department of Justice brought criminal charges against several firms and individuals.[17]

The complaint states that NP's supervisor was "aware of the government investigations that had been commenced" by early 2015. She "told [NP] that the government was showing up on people's doorsteps," and "warned [NP] to be careful about communicating with competitors" (Complaint, page 341). Further according to the complaint, news of the investigation ended the price increases (Complaint, page 338). However, the news does not appear to have undermined existing agreements, which the firms may have maintained tacitly. This is consistent with NP's assertions that generic drug price fixing is hard to initiate but easy to sustain (Complaint, page 160) as well as internal documents referencing cooperative arrangements years after the government launched its investigation (Complaint, page 50). We address this issue directly in Section 6.

It is important to note that when we reference "agreements," "coordination," and especially "collusion," we more precisely mean semicollusive behavior—competitive play with with respect to long-run choices but

other instances, executives used personal email accounts to transfer illegally obtained information and deleted illegal communications when they learned of the government investigation (Complaint, page 341).

[15]"Fluff pricing" is NP's term for this practice. See Complaint page 146.

[16]To be precise, the investigation traces back to a *New York Times* article titled "Rapid Price Increases for Some Generic Drugs Catch Users by Surprise" (Rosenthal, 2014), which the supervisor of the Connecticut AG office's unit of antitrust and fraud read and forwarded to a staff attorney, who subsequently sought subpoenas (Pazniokas, 2019).

[17]These are serious offenses. For example, two executives at Heritage Pharmaceuticals who plead guilty to their role in the conspiracy currently await sentencing and face up to 10 years in prison. At the time of writing, federal authorities have charged 7 firms and 5 individuals. Teva was charged in August 2020. Several parties have settled. For example, Taro admitted its role in the conspiracy and paid a $205.7 million penalty in July 2020.

cooperative play with respect to short-run ones (Fershtman and Muller, 1986). Specifically, we assume that colluding firms set prices to maximize joint profits. Supernormal profits are an outcome of this behavior and are necessary but not sufficient to induce entry.[18] The complaint suggests that members understood the value of sticking to the agreement. It also demonstrates that potential entrants were attracted to supernormal *equilibrium* profits, not simply high prices. Finally, it is worth noting what type of conduct is absent from the complaint and related filings. Despite thousands of pages of court documents recounting almost countless conversations about prices, no statements suggest that cartel members coordinated entry. In other words, it seems very unlikely that members reached agreements not to enter each others' markets. Consistent with that view, when entry by a cartel member into a cartelized market did occur, incumbent producers responded by accommodating rather than retaliating.[19]

3 Data

3.1 Sources

The data come from several sources. The National Drug Code (NDC) Directory, which is published by the FDA, provides a current list of all prescription pharmaceutical products manufactured for sale in the US. Each product is identified by a unique NDC code, which identifies the substance-delivery-release-strength combination associated with the product as well as the firm that produces it and the ANDA that authorizes that production. The FDA updates the list daily. Using the Internet Archive, we take annual snapshots of the directory, which are the starting point for our panel dataset.

Medicaid State Drug Utilization Data, which is published by Centers for Medicare & Medicaid Services, is the primary source of price and quantity information. Each quarter, all states and the District of Columbia report the number of prescriptions filled by Medicaid enrollees and the amount of corresponding expenditure. We download national aggregate statistics, and we merge to the NDC Directory data, described above, at the NDC-quarter level. Since NDCs are specific to drug packaging (e.g., 1000-count bottles, 14-count blister packages, etc.), firms are associated with multiple NDCs per drug. Thus, we sum over NDCs to form drug-firm-quarter observations.

We also acquired data from from IQVIA, a private provider whose National Prescription Audit covers 92% of US pharamcies. We first use IQVIA data to scale up the quantities reported in State Drug Utilization Data, which covers only Medicaid, to US totals. To do so, we compute the ratio of total IQVIA prescriptions to total Medicaid prescriptions, and we scale up by that factor. Second, we use IQVIA to carefully check whether patterns in Medicaid are representative of industry-wide purchasing behavior. Total purchases track extremely closely with one another over time, and their responses to cartel-induced price changes are almost identical (see Section 10 of the Online Appendix).

The complaint, which is described in the previous section, provides information about cartel formation. As it directly relates to dataset construction, the complaint lists the drugs for which NP fixed prices, the

[18] Further evidence comes from Civica Rx, a startup generic drugmaker owned by a consortium of hospitals. In a New York Times article titled "Fed Up With Drug Companies, Hospitals Decide to Start Their Own" (Abelson and Thomas, 2018), the firm stated that when it enters the market it expects incumbents to respond by "quickly dropping the price of the drugs in question." Further, the firm stated that were it to exit subsequently, it expects incumbents to respond by "raising them again later."

[19] ANDA launches by Aurobindo, Lupin, and Actavis provide particularly clear examples. See Complaint, pages 74, 81-82, and 103-104, respectively.

dates on which those prices were fixed, and NP's relationships. From this information, we construct an indicator variable for whether a cartel was ever formed in a particular drug market. Within each drug market where a cartel is formed, we also construct an indicator for whether a particular period precedes or follows cartel formation. Finally, we construct an indicator variable for cartel membership.

We then merge in ANDA filing and approval dates. To obtain filing dates, we "scrape" the FDA website for all available ANDA approval letters, most of which state the date the application was filed. Although not all ANDA approvals are published, we can infer the missing filing dates, since ANDA numbers are issued sequentially, with a few exceptions (see Section 10 of the Online Appendix, which shows that we measure ANDA filing dates without meaningful error). We obtain approval dates from the Orange Book, which is published by the FDA.

Our sample covers 2008 to 2019 and consists of all drugs manufactured by Teva in the quarter prior to NP joining the firm, subject to some exceptions. We omit drugs that lose exclusivity during the sample period.[20] We also omit a small number of drugs for which NP raised prices twice. This decision drops only 11 substance-delivery-release combinations, and avoids issues stemming from ill-defined cartel formation dates. Last, we drop 38 injectable drugs; injectables occupy a very different segment of the industry, require radically different production methods, have short shelf lives, are mostly manufactured by different producers than the ones we study, and most importantly are never mentioned in the complaint.

3.2 Summary Statistics

Table I summarizes the data at the drug-year level. There are 4,992 observations spread over 416 drugs and 12 years. The average number of prescriptions filled is 1.9 million, and the average price is $31.66. Average expenditures are $31.1 million, or about $255 billion over the full panel. The mean number of manufacturers in the market is 4.15, while the mean number of ANDA filings is 0.21, highlighting that entry into mature generic drug markets is typically rare, even though these markets support several competitors in equilibrium. Most ANDA filings can be attributed to cartel "nonmembers" (i.e., drugmakers that are not named in the Complaint). The group accounts for about 60% of entry over the full sample and closer to 69% in the periods following NP's joining Teva, regardless of whether we consider all drug markets or just those where cartels are eventually formed (not shown). Cartels are formed in the markets for 113 drugs, or 27.2% of the sample.

4 Reduced form analysis

In this section, we describe the patterns in the data that motivate the structural analysis. The analysis reveals that collusion dramatically increases drug prices; that collusion induces entry, evidenced by a sharp increase in ANDA filings and eventual approvals; that regulatory approval involves significant delays, and that entry reduces price in the cartelized markets.

[20] They exhibit drastic price declines that ruin comparability across units in our analysis and reflect idiosyncrasies of patent challenges and expiration, which are unrelated to cartel formation. Paragraph IV challenges are one of many idiosyncrasies. See also Footnote 13.

Table I: *Summary statistics*

	Count	Mean	Std. dev.	Minimum	Maximum
Price	4,992	31.66	40.42	2.32	628.53
Quantity (in thousands)	4,992	1,904.66	3,618.83	0.15	29,571.61
Expenditure (in millions)	4,992	31.10	110.51	0.01	4,188.54
Number of firms	4,992	4.15	1.94	1.00	12.00
Number of ANDA filings	4,992	0.21	0.54	0.00	5.00
Cartelized drug	4,992	0.27	0.44	0.00	1.00
Cartelized drug $\times$ Post cartel formation	4,992	0.14	0.34	0.00	1.00

The unit of observation is a drug-quarter. Price refers to dollars per prescription. Quantity is measured in thousands of prescriptions. Expenditure is measured in millions of dollars. "Cartelized drug" is an indicator for whether a cartel formed in the drug market at any point in the sample. "Cartel drug×Post cartel formation" is the interaction of "Cartelized drug" and an indicator for periods following cartel formation (i.e., the indicator is "on" if the cartel for a particular drug formed in April 2013 and the observation summarizes Q3 2013 or later).

4.1 Collusion raises price sharply

In Section 2, above, Figure I provides initial evidence in calendar time that collusion affects drug prices. In this section, we begin by measuring the effect in event time. To do so, we split the sample into cartelized and uncartelized drugs, calculate weighted average prices at the drug-year level (weighting each firm's product by quantity sold), and plot how prices evolve around cartel formation. Formally, we estimate

$$Y_{kt} = \sum_{\tau=-21}^{26} \beta^{\tau} x_{kt}^{\tau} + a_k + b_t + e_{kt}. \tag{1}$$

The outcome variable, Y, represents log point-of-sale price from the Medicaid data and k indexes products in quarter t. a_k and b_t represent product- and time-specific fixed effects. x_{kt}^{τ} is an indicator variable that equals one if and only if k is a cartelized drug and t is τ periods from the start of collusion. β^{τ} represents the difference in the prices of cartelized ("treated") and uncartelized ("control") drugs relative to NP being hired. Given the timing of the price increases and the time period covered by our data, τ can take values between -21 and 26. We normalize $\beta^{-1} = 0$, so levels of the coefficients represent price differences relative to the period immediately preceding NP's hire.

Figure II plots estimates of β^{τ}. Two features of the graph stand out. First, prices in cartelized and uncartelized drug markets evolve similarly prior to collusion, evidenced by near-zero point estimates on the left-hand side of the graph. In other words, there are no pre-event trends. Second, prices of cartelized and noncartelized drugs diverge sharply when collusion begins. Two years after cartel formation, for example, prices are nearly 50 percent higher in the treated group relative to the control.

Cartel-induced price increases occurred within a narrow window. Most occurred within months of NP joining Teva, and all were completed within 18 months of that event. Moreover, as explained in Section 1, opportunities to form cartels arose for plausibly exogenous reasons—the movement of personnel or supply disruptions. Further, since our structural estimates rely on annual observations, one can conceptualize almost all the price increases as occurring in a single period. In other words, the changes were abrupt.

Nonetheless, it is important to show these results are not driven by variation across groups receiving

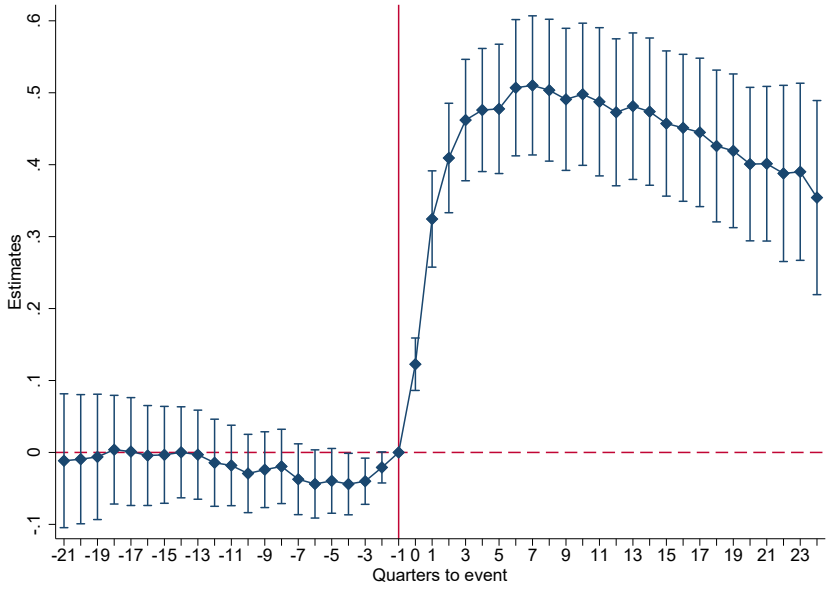

Figure II: *The cartel increases price sharply and abruptly.*

This figure plots β^{τ}, which is obtained by estimating equation 1, on the y-axis against event time on the x-axis. The unit of observation is a drug-year-firm. Log prices are the outcome of interest, and the unit of observation is a product (i.e., drug-year-firm). The vertical red line at event time -1 corresponds to the year immediately prior to cartel formation. Vertical bars around the point estimates show 95 percent confidence intervals for those coefficients, based upon standard errors that are clustered by drug.

treatment at different times. To do so, we follow an approach recently proposed by Sun and Abraham (2020), which accounts for potential contamination of leading and lagging coefficients, and then we plot the resulting estimates. Figure XIV in the Online Appendix reports the result, which very closely resembles Figure II.

We also confirm these results are not sensitive to our definition of treatment. To do so, we define treated drugs not as those where cartels actually formed but as those where NP had especially strong industry relationships. We then replicate the steps used to generate Figure II. Figure XII in the Online Appendix reports the result, which, again, very closely matches the patterns described in the prior graphs.

4.2 Collusion induces entry

Next, we turn attention to entry. Each firm must prepare and file an ANDA and obtain FDA approval of its application before it launches a generic drug product. In this process, the first action we observe a firm take is filing. Thus, for an initial marker of how entry responds to collusion, we begin by plotting ANDA filings in calendar time. The filing dates for some ANDAs are not available on the FDA website, so we need to infer them (in the Online Appendix, we describe this procedure and show in Figure XI that the missing dates can be inferred from the available ones without any meaningful error). We then replicate Figure I, replacing average log price per drug with the average number of ANDA filings per drug and collapsing the data down to the year rather than quarter level, to improve legibility (since ANDA filings are relatively

infrequent, especially in the absence of cartels).

Figure III illustrates the result. Entry exhibits similar patterns to price. Prior to NP joining Teva, average ANDA filings for cartelized and uncartelized drugs evolved similarly, and uncartelized drugs trend smoothly through her hiring date, but the two groups diverge sharply in 2013. In other words, cartel formation induces substantial entry.

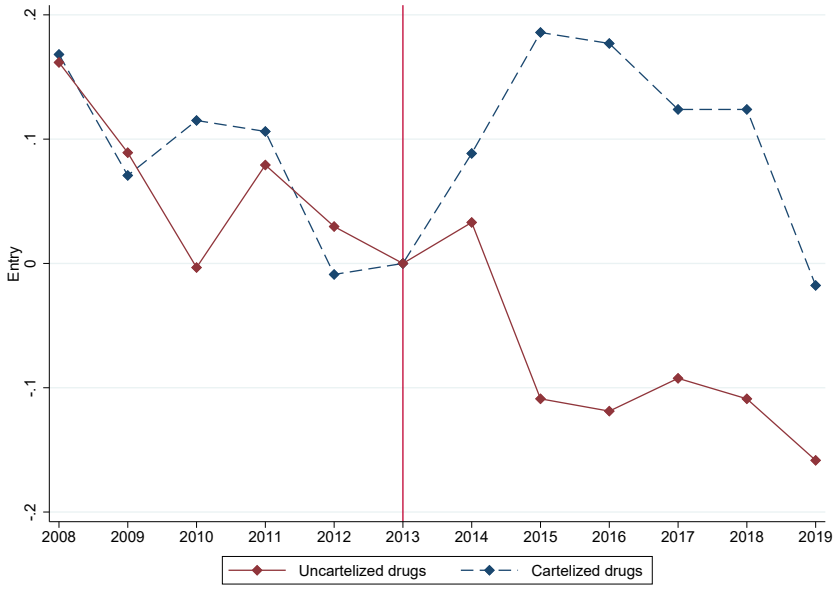

Figure III: *Entry in calendar time*

This figure plots the number of ANDA filings per drug-year on the y-axis against calendar time on the x-axis. The solid and dashed lines correspond to uncartelized and cartelized drugs, respectively. The vertical line at 2013 corresponds to the year that NP joined Teva. ANDA filings are normalized to zero in that year.

To quantify the divergence, we measure the effect of collusion on entry in an event study framework. We re-estimate equation 1 and plot the coefficients in event time, as in Figure II, but we replace the outcome variable, Y, with number of ANDA filings per drug-year (not log price at the product-quarter level). The result is shown in Panel A of Figure IV. The average number of ANDA filings per market evolves similarly over time for cartelized and uncartelized drugs leading up to collusion (i.e., no pre-event trends are apparent). Then the figures diverge sharply. One year after the start of collusion, the number of ANDAs filed in cartelized markets is already economically and statistically distinct from the number filed in uncartelized markets. The difference grows over time, presumably reflecting the lead times for the the application, which includes laboratory tests to demonstrate bioequivalence. Within two years of a drug being cartelized, there in an average of 0.4 additional ANDA filings for that drug.

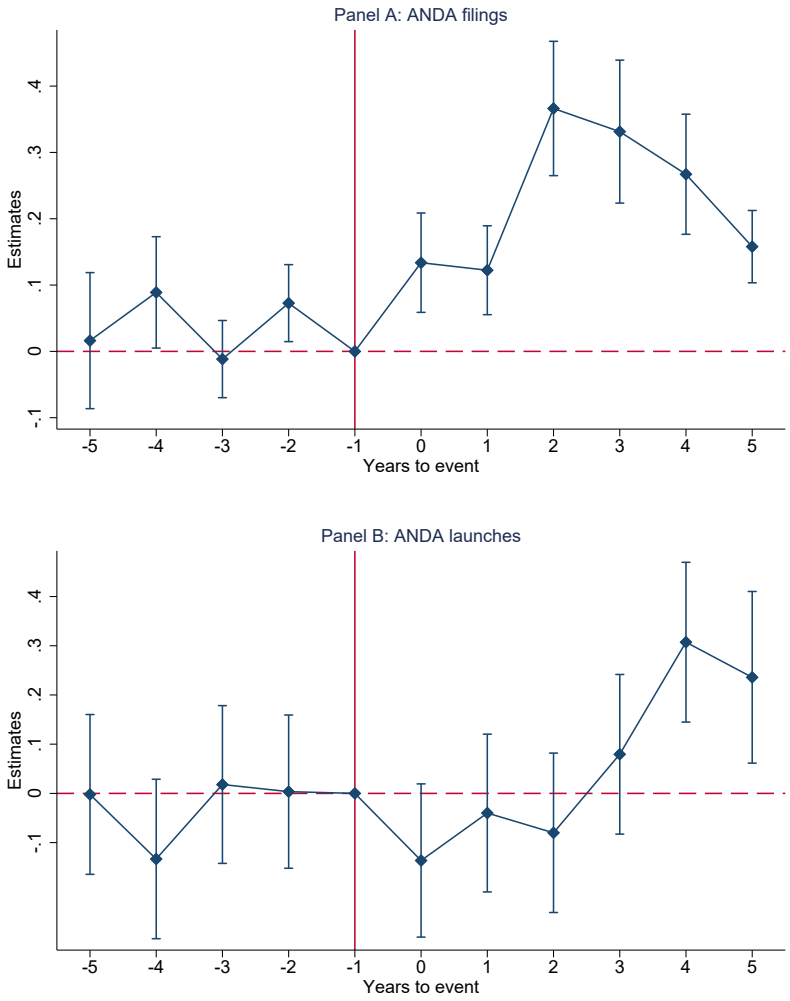

Figure IV: *The cartel induces significant entry.*

This figure plots coefficients obtained by estimating equation 1 on the y-axis against event time on the x-axis. The unit of observation is a drug-year. In Panel A, ANDA filings are the outcome of interest. In Panel B, ANDA filings are replaced with ANDA launches. The vertical red line at event time zero corresponds to the year in which Teva hired NP. Vertical bars around the point estimates show 95 percent confidence intervals for those coefficients, based upon standard errors that are clustered by drug.

The filing of an ANDA does not permit immediate entry. The FDA must review and approve the ANDA, which involves delays. For example, a generic drugmaker who files an ANDA in 2014 may only receive approval to produce the drug in 2017. As a result, we additionally consider ANDA launches, where a "launch" is defined as the first year in which a firm sells the drug authorized by the ANDA.[21] That is, we re-estimate equation 1, replacing Y with the number of ANDA launches per drug-year. Panel B of Figure IV reports the result. Once again, the graph does not exhibit pre-trends. Consistent with regulatory delays, the graph also does not exhibit differences between cartelized and uncartelized markets in the years

[21] ANDA launches are a subset of ANDA filings because applications filed towards the very end of our sample may not receive approvals by the time the sample ends.

immediately following collusion. However, ANDA launches rise sharply four years after collusion begins.

Infrequently, producers re-enter markets using dormant ANDAs (i.e., approved filings that are no longer associated with production but once were). This might be especially common in markets where cartels have recently formed, since they drive price rises that could encourage not only *de novo* entry in the form of ANDA filings but also induce inactive firms to become active once again. To evaluate this possibility we re-estimate equation 1, replacing Y with the number of "re-entries" per drug-year. Due to the relatively small number of these occurrences, the result is in the Online Appendix. As with the previous findings, there are no pre-trends. Re-entry spikes in the year immediately following collusion. This pattern is consistent with the cartel turning some unprofitable markets profitable once again, and it squares with the fact that already-approved applications can enter without delay.

4.3 Regulatory approvals involve delays

Figure IV illustrates lags between ANDA filings and launches. These lags have important consequences if entrants exert downward pressure on prices, driving the cartelized markets closer to pre-cartel equilibrium conditions. To examine the delays more carefully, we plot the density of the time between ANDA filing and regulatory approval. Figure V reports the result. Mean and median delays are about 2.5 and 2 years, respectively.

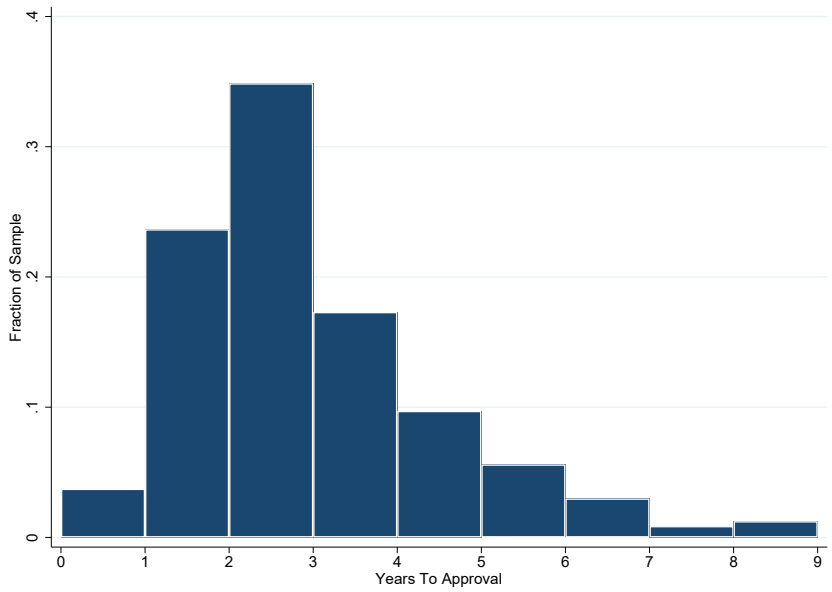

Figure V: *Distribution of time between ANDA filing and approval.*

This figure plots the density of delays, defined as the number of years between filing and approval. The unit of observation is an ANDA.

4.4 Entry disciplines cartel prices

So long as entrants are not brought into the collusive agreement, they have incentives to undercut cartel members on price and steal market share. To evaluate their impact and avoid obvious endogeneity that arises from comparing markets that do and do not experience post-collusion entry, we exploit variation in a measure of market size. In particular, we use each drug's total revenue in 2012—the year just prior to Teva hiring NP. Intuitively, collusion in large markets should attract more entrants than in small ones, so the likelihood of post-collusion entry is a function of pre-collusion market size.[22]

To validate our approach, we examine the relationship between market size and entry at the drug level. That is, we narrow the sample to cartelized drugs and plot post-collusion ANDA filings (on the y-axis) against the log of total 2012 revenue in millions of dollars (on the x-axis). Figure VI plots the result. The top panel measures entry using the probability that a market experiences an ANDA filing, while the bottom one measures it using the number of ANDA filings per market. Regardless of the measure, market size and entry are closely related. Only about 8% of drugs with around $1 million in revenue attract any entry at all. However, drugs with over $1 billion in revenue almost always attract entry, with an average of three firms filing ANDAs following cartel formation.

[22] If collusion in small generic drug markets is easier to sustain than in large ones, then market size could directly affect post-collusion price paths. However, the collusive agreements are not in jeopardy. See Section 6, which shows, e.g., that incentive compatibility constraints are easily met and that observed prices strongly suggest cartel members cooperate through the end of our sample. We thank Kathleen Hui for pointing this out.

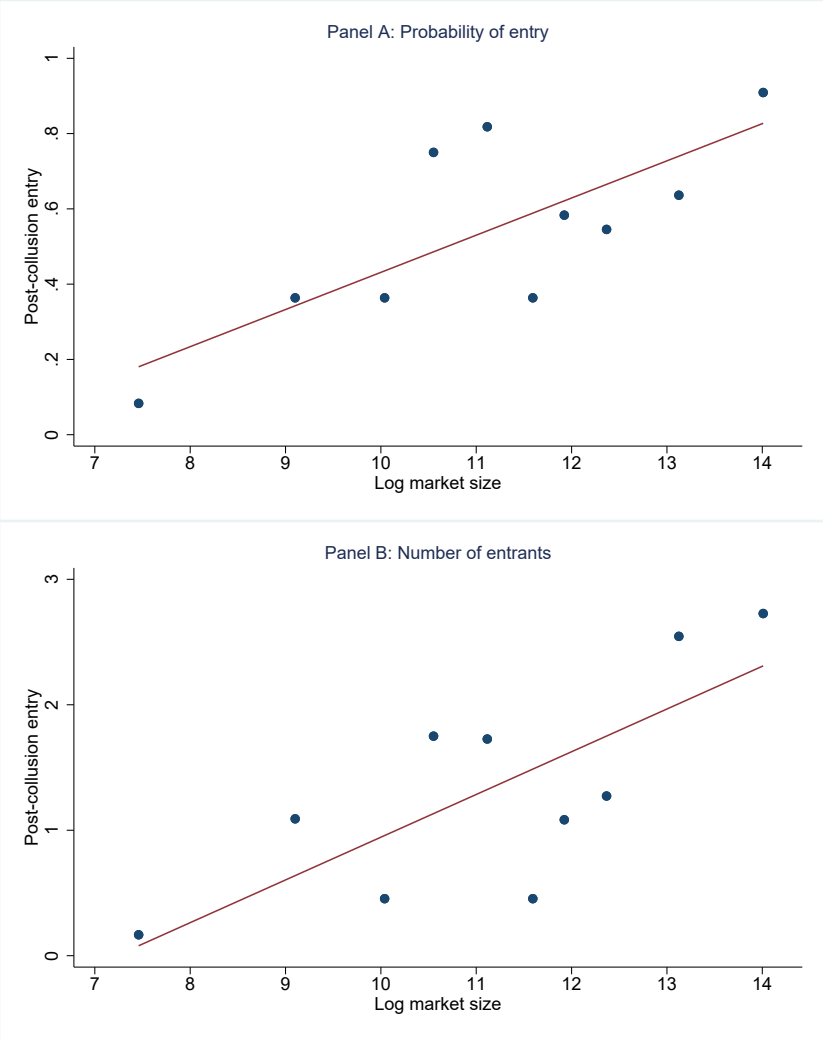

Figure VI: *Larger markets attract more entry.*

This figure plots log market size on the x-axis against the number of entrants in the post-collusion period. The unit of observation is a drug. Market size is measured in total revenue in 2012, the year immediately prior to NP joining Teva. Data are binned according to x-axis values, so averages within the bin are plotted (i.e., the graph represents a "binscatter").

We then use the same approach to compare the path of log prices for drugs with above- and below-median pre-collusion market size. The motivation for this comparison is straightforward. Assuming entry costs do not scale exactly with market size, cartelization of large markets will be more likely to induce entry than the cartelization of small markets. If entry reduces price, then prices will fall in large markets but remain relatively stable in small markets.

Specifically, we separate cartelized drugs into two groups based upon whether their 2012 revenue was above or below the median drug's 2012 revenue, and we plot average log prices in calendar time separately for the two groups. For the sake of comparison, we also plot average log prices of uncartelized drugs. Figure VII reports the result. Several features of the graph are noteworthy. First, prices of "small market"

and "large market" drugs trend similarly into 2013, when Teva hires NP. Second, prices of drugs in both groups rise sharply upon hiring, especially relative to uncartelized drugs. In fact, by mid-2014, the log changes are nearly identical. However, after that point, high prices of "small market" drugs persist, whereas the prices of "large market" drugs fall consistently. By the end of the panel, below-median cartelized drug prices are 50% higher than those in the uncartelized group, while above-median cartelized drug prices have converged to within 10% of it. These findings strongly support the idea that entrants drive down prices.

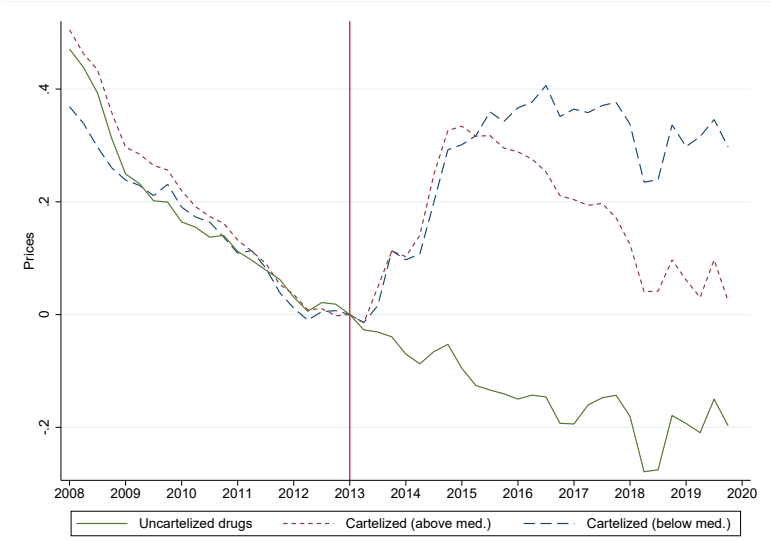

Figure VII: *Entry disciplines cartel-induced price increases*

This figure plots log prices on the y-axis against calendar time on the x-axis. The long-dashed and short-dashed line correspond to cartelized drugs whose revenues in 2012 were above and below the median drug in the sample. The solid line corresponds to uncartelized drugs. The vertical red line corresponds to the first quarter of 2013—the period in which NP joined Teva. Prices are normalized to zero in that quarter.

This result does not offer any guidance as to how low entry costs would need to be to, for example, induce similar price changes among "small market" drugs or induce even greater price declines among "large market" drugs. Accurately predicting how welfare-relevant outcomes would evolve under reduced fees or shorter delays requires forecasting the equilibrium response of drugmakers and buyers. We turn to a structural model of entry, pricing, and demand to provide those forecasts below.

5 Model

In this section, we model the market's response to cartel formation in mature segments of the US generic prescription drug industry. Collusion has two key effects. First, firms ordinarily choose prices competitively, maximizing current period profits, but under collusion, cartel members might choose prices cooperatively. Second, while entry into mature drug markets is typically very rare, cartel-induced price increases may raise profits so much that firms that are not currently active in the market enter, even though doing so involves potentially large costs and long delays. Entry may mitigate the harms associated with collusion.

Each period (year) has two "stages" of play. In the first stage, drugmakers file ANDAs for markets (drugs) they would like to enter. These filings are approved by the FDA with a delay of uncertain duration. Applications take at least a year to be approved so no drugmaker enters a market in the year it files an ANDA. In the second stage, conditional on the market structure, drugmakers set prices, while consumers choose which product to consume. Firms that hold active ANDAs manufacture the drug and earn profits, which equal the product of the per-unit profit margin, market share, and market size. That is, $\pi = (p - mc)s(p)A$, where p denotes the price the firm charges, mc denotes the firm's marginal cost, s denotes market share, which is a function of price, and A denotes the market size. Marginal costs depend on production technology, which we assume is outside the firm's immediate control.

By calculating per-period profits under collusion and competition, drugmakers can assess their incentives to cooperate, and by calculating the sum of discounted expected future profits, they can evaluate whether the value of entering exceeds the sunk cost of doing so. In summary, because entry depends on pricing, which in turn depends on demand, we work backwards from the consumer's decision when we describe the model below.

5.1 Demand

Although patients are the final consumers of prescription medications, they are typically not the ones to decide which generic drugmaker's product to purchase. Those decisions are made by intermediaries, which include wholesalers (e.g., Cardinal Health), group purchasing organizations (e.g., MMCAP, a cooperative formed to source drugs for state-owned facilities), and large retail chains (e.g., CVS).[23] These buyers hold procurement auctions for each drug, scoring the offers made by the drugmakers along several dimensions besides price. For example, Cardinal Health states that it values high fill rates, short lead times, frequent product availability reports, infrequent recalls, accurate invoicing, on-time deliveries, carefully marked pallets, and backhaul utilization (Cardinal Health, 2020).[24]

We index buyers by i, drugs by d, and drugmakers by f. Following Miller (2014) and Miller and Sheu (2020), if i contracts with f, then it receives a payoff equal to the gross value of the product, denoted v_{idf}, less the disutility of the price paid, which equals $\alpha_d p_{ifd}$. Alternatively, if i chooses not to buy from any drugmaker, then it forgoes downstream sales of the drug and receives a payoff normalized to zero.[25] Buyers' decisions maximize their payoffs.

Gross value is given by

$$v_{idf} = \lambda_d + \xi_{df} + \zeta_{df} + (1 - \sigma)\epsilon_{idf}. \tag{2}$$

Here, λ_d is a drug-specific dummy variable. Firm specific shocks are denoted by ξ_{df}, while ϵ_{idf} denotes an i.i.d. product-buyer shock. We assume ϵ is distributed Type 1 extreme value. We nest the "inside" goods separately from the no-purchase option, allow the gross value of the goods within the nests to be correlated with one another, and assume ζ_{df} is distributed such that $[\zeta_{df} + (1 - \sigma)\epsilon_{idf}]$ is also drawn from a Type 1 extreme value distribution. Both errors are known to the buyers but not drugmakers.

[23] Some buyers have combined their sourcing operations, typically by forming joint ventures. For instance, McKesson and Walmart formed ClarusONE. See also Footnote 26.

[24] Backhaul utilization refers to using Cardinal's dedicated fleet of tractor-trailers to move product. It requires that the product can easily be transported to certain shipping lanes, so it is specific to the buyer, seller, and drug.

[25] In our setting, the outside good represents, in part, buyer-drug specific stockouts, which are very common throughout this period. See, e.g., the Food and Drug Administration's 2020 report for a summary, available at https://www.fda.gov/media/150409/download.

To make the problem tractable (and because we do not observe buyer-specific purchases and contract-specific prices), we assume that each drugmaker charges one price to all buyers and that buyers are sufficiently small that s_{dft} is continuous and twice continuously differentiable in its arguments. The probability that i purchases f is then given by

$$s_{idf} = \frac{e^{\lambda_d + \alpha_d p_{df} + \xi_{df}}}{\Lambda^\sigma (1+\Lambda)^{1-\sigma}}, \tag{3}$$

where $\Lambda = \sum_{f' \in \mathcal{F}_d} e^{\lambda_d + \alpha_d p_{df'} + \xi_{df'}}$. Analogously, the probability that i chooses the outside option is given by

$$s_{idf} = \frac{e^{\lambda_d + \alpha_d p_{df} + \xi_{df}}}{1 + \Lambda^{1-\sigma}}. \tag{4}$$

Summing over buyers and adding time subscripts, t, yields market shares, denoted s_{dft}.[26][27]

5.2 Pricing

Firms' profits are given by

$$\pi_{dft} = (p_{dt} - mc_{dft})\, s_{dft} A_{dt}. \tag{5}$$

In the absence of collusion, firms choose prices competitively. That is, they maximize π_{dft} in the current period, setting the first-order condition of the right-hand side of equation 5 equal to zero. Assuming consumers dislike price, α_d is negative, and markups are strictly positive. Let p^B_{dft} denote the resulting competitive price, and π^B_{dft} denotes the profits (where "B" stands for "Bertrand-Nash").

In the presence of collusion, cartel members—firms named as defendants in the complaint—coordinate decisions to internalize business-stealing effects, and prices are "fixed" in the sense that cartel members set them cooperatively. Nonmembers can undercut the members' prices but never unravel the collusive agreement, as described below, and the cartel responds in equilibrium by reducing their price.[28] We define $p^{C,M}_{dft}$ as the price charged by cartel members under collusion, and we define $p^{C,N}_{dft}$ as the nonmembers' best response to that price. ("C" stands for "collusive," and "M" and "N" stand for member and nonmember, respectively.) Analogously, when these prices are chosen, we define the profits earned by cartel members and nonmembers as $\pi^{C,M}_{dft}$ and $\pi^{C,N}_{dft}$, respectively.

We assume that that until the cartel is formed, firms play p^B_{dft}, but upon cartel formation, prices depend on the dynamic incentives of the members. To be specific, we assume cartel members play a trigger strategy

[26] Some buyers are large. For instance, in its 2013 annual report, Teva reports that its largest buyer accounts for 17% of its sales. In practice, Teva might offer it a lower price. Our model cannot capture these nuances. Further, buyers have consolidated over the sample. However, large GPOs procure a secondary supply of most generic drugs via another auction, which limits their power. To summarize, we recognize that drug demand is complicated and our assumptions ignore potentially interesting features of the industry, but we think the model provides a close enough approximation to produce useful predictions about prices, quantities, profits, and entry, which are described in Sections 6 and 7.

[27] These assumptions let us recast the seller's problem as price setting in the face of "nested logit"-type demand rather than bidding in a very large number of scoring auctions with independent private values. See Einav (2003) for a more thorough description of how these models relate.

[28] So long as ICCs are satisfied, the members' and nonmembers' pricing problems are isomorphic to ones found in competitive markets with multi-product firms. That is, when all firms are cartel members, the cartel behaves like a multi-product monopolist. When only some firms are cartel members, the cartel behaves like a multi-product firm in oligopoly, where all members' products are in the firm's portfolio, while nonmembers each behave like single product firms.

and that prices set at t are public information by $t+1$: if any member deviates from $p^{C,M}_{dft}$, then all members revert to p^{B}_{dft} forever after. Consistent with this timing and strategy, private communications published in the complaint suggest the cartel did not tolerate any non-cooperation. For example, in 2013, one member, Glenmark, temporarily undercut another member, Teva, apparently due to a misunderstanding. The response was immediate. A confused Teva employee sent an email whose only contents were "???" to NP. Exactly five minutes later, NP replied, "I know...made the call already." The next day NP spoke to her counterpart at Glenmark, and later that same day Glenmark withdrew the lower price, which required rescinding an offer to a major distributor (complaint, page 134).

The cartel's agreement is unenforceable in court and physical coercion is impractical, so cooperation requires $p^{C,M}_{dft}$ maximize the sum of discounted expected future cash flows for all members at every point in the future. In other words, retaliation must be costly enough that collusive prices are compatible with each firm's individual incentives. Given the extreme slack in the constraints at or near the parameters we will estimate, and in the interests of simplifying the notation considerably, the remainder of this section implicitly assumes that collusion continues once it is initiated. Further discussion of the Incentive Compatibility Constraints (ICCs) is postponed until the next section.[29]

5.3 Entry

When cartels form, prices rise, which induces entry if sunk costs are low enough. Indeed, as Section 4 shows, collusion in generic drug markets leads to a sharp rise in ANDA filings followed by a much later rise in the number of manufacturers producing the drug. We model this behavior explicitly. Since ANDAs apply to all strengths within a substance-delivery-release, our analysis of entry occurs at that level. When a cartel forms, firms without FDA approval face a one-time decision to start the application process, which involves a delay. For tractability, we assume that other market structure changes are determined outside the model. Since entry into generic drug markets is otherwise rare and exit is usually precipitated by supply disruptions outside the firm's control, this assumption greatly reduces the computational burden but is unlikely to have a material impact on our conclusions.

Since each drug (i.e., substance-delivery-release-strength) maps into one substance-delivery-release combination, we require additional notation. We index each substance-delivery-release by j and denote the set of d in j by $\mathscr{J}$. We also define t_d as the year the cartel forms in d. Since all cartels in j form in the same year, $t_d \equiv t_j$ for all d in j.

Any firm that has not filed an ANDA in t_j has the opportunity to do so. If f opts to file an ANDA, then it enters all d in $\mathscr{J}$ with the same delay. Since cartel members are bound by the collusive agreement, they set different prices than nonmembers when they enter, so their value of entering differs from that of nonmembers. The value of a cartel member filing an ANDA in j is given by $VE^{M}_{j}(\chi^{M}, \chi^{N}) = \sum_{d \in \mathscr{J}} VE^{M}_{d}(\chi^{M}, \chi^{N})$, where VE^{M}_{d} represents the value it receives from entering all d in j, where the subscript d is understood to incorporate information about M_d and N_d, and where χ^{k} represents the

[29] As we show later in the paper, even if entry is much more abundant than what we observe in the data, ICCs are always satisfied by a wide margin for drugs named in the complaint. This depends on several factors, including the speed at which prices become public information and the frequency with which drugmakers can "rebid" for contracts. However, we caution readers that ICCs should always be checked, especially under counterfactual policies that reduce their slack, such as ones that promote entry. For an especially clear illustration, see Igami and Sugaya (2020), who show how ICCs that were once satisfied can be violated due to changes in market conditions.

number of entrants of type k. $VE^N(\chi^M, \chi^N)$ is defined analogously. If f opts not to file an ANDA, it receives payoffs normalized to zero.

The value of entry equals the sum of discounted expected future profits. We assume that when firms are deciding whether to file an ANDA, they know the distributions of entry delays as well as demand and marginal cost shocks, but they do not know the future realized values of those variables. Under this information structure,

$$VE_d^M(\chi^M, \chi^N) = \sum_{t=1}^{\infty} \delta^t F_D(t) \sum_{e_M=0}^{\chi_M - 1} \sum_{e_N=0}^{\chi_N} \Big[\rho(\chi_M - 1, e_M, t; F_D) \rho(\chi_N, e_N, t; F_D) \int_{\xi} \int_{\omega} \pi_{d,f,t_d+t}^{C,M}(M_{d,t_d+t} + e_m + 1, N_{d,t_d+t} + e_n, \xi_{d,t_d+t}, \omega_{d,t_d+t}) dF_{\xi} dF_{\omega} \Big]. \quad (6)$$

Here, δ^t is the discount factor, which is set to 0.9 throughout our analysis. D, the delay, equals the number of periods between t_d and the time the manufacturer begins producing the drug. F_D is its cumulative mass function—it equals the probability that the potential entrant is active t periods after it decides to enter. For example, if t_d is the year 2013 and $F_D(3) = 0.5$, then there is a 50% chance that the potential entrant is earning profits in d in 2016. M_d is a vector whose scalar elements represent the number of incumbent cartel members that are active in future periods, and N_d is an analogous vector counting the numbers of incumbent nonmembers.[30] F_{ξ} and F_{ω} represent the cumulative density functions governing demand and marginal cost shocks respectively.

$\rho(\chi_M - 1, e_M, t; F_D)$ is the probability that e_M other cartel member entrants are active t periods after the cartel is formed, while $\rho(\chi_N, e_N, t; F_D)$ is an analogous probability for nonmember entrants. Both are a binomial expansions given by

$$\rho(a, b, t; F_D) = \frac{a!}{(a-b)!\, b!} F_D(t)^b [1 - F_D(t)]^{a-b}. \quad (7)$$

The final term represents the expected profits earned by the potential entrant, assuming it is active at $t_d + t$. $VE_d^N(\chi^M, \chi^N)$ is given by an expression analogous to the one that appears on the right-hand side of equation 6.

If f is a member of the cartel, then it files an ANDA in j if and only if $VE_j^M - \mathscr{E}[\theta_{jM} | \mathscr{I}_{jM}] \geq 0$, where θ_{jM} denotes the true sunk cost of entry, $\mathscr{E}$ denotes the firm's expectation operator, and $\mathscr{I}_{jM}$ denotes its information set. Similarly, if f is a nonmember, then it enters if and only if $VE_j^N - \mathscr{E}[\theta_{jN} | \mathscr{I}_{jN}] \geq 0$. Sunk costs include regulatory and non-regulatory costs, and we assume they are common knowledge. Entry decisions form a simultaneous move Nash equilibrium.

[30] For instance, the t^{th} element of M_d—call it M_{dt}—represents the number of incumbent cartel members that hold active ANDAs at time t_d and are active at time $t_d + t$.

6 Estimation and estimates

6.1 Demand

To estimate the demand system, we rely on the market share inversion proposed by Berry (1994) and estimate

$$ln(s_{dft}) - ln(s_{d0t}) = \lambda_d + \lambda_t + \bar{\alpha} p_{dft} + \sigma ln(s_{dft|dgt}) + \xi_{dft}, \tag{8}$$

where $s_{dft|dgt}$ denotes f's share of the inside good, which equals $s_{dft}/(1 - s_{d0t})$.The left-hand side of equation 8 is a straightforward transformation of the data, and the right-hand side is linear in parameters to be estimated—λ_d, λ_t, σ, and $\bar{\alpha}$—and an error term. The initial specification restricts $\alpha_d = \bar{\alpha}$, although we partially relax this restriction below.

Since firms likely know ξ when they set p, prices are endogenously determined in equation 8, so naively regressing left-hand side values on the right-hand side variables produces biased results. We identify $\bar{\alpha}$ using price variation induced by the cartel, which we documented in Section 4. Our price instrument equals the product of a dummy for cartelized drug markets and a dummy for years greater than the year each drug market was cartelized. The intuition behind the identification strategy is transparent. Around cartel formation, the consumption of cartelized drugs declines compared to uncartelized drugs, while the prices of cartelized drugs rise compared to uncartelized drugs. The decline in drug consumption relative to the rise in drug prices reflects sensitivity to price, which maps to the parameter $\bar{\alpha}$. Given the origin of the cartel, the selection of drugs, and the timing, it is reasonable to assume that values of the instrument are uncorrelated with drug-firm-year specific demand shocks.[31] Note that in nested logit models, an additional right-hand side term—$s_{dft|dgt}$—is also endogenously determined. Our instrument counts the number of products in the market, so σ is identified from share changes following entry and exit.[32]

Table II reports the demand estimates. Column 1 corresponds exactly to equation 8. We find that preferences are correlated within the "nest" of inside goods, with $\sigma = 0.57$. As expected, we also find buyers dislike price, with $\bar{\alpha} = -0.11$. Standard errors are clustered at the drug level, and both coefficients are significant at the 1% level.

In Column 2, we assess the importance of accounting for "authorized generic" products. Authorized generics are generic versions of drugs that are manufactured by the patent owner, which are often introduced around the point at which it loses exclusivity. Throughout our analysis, we do not distinguish these products from "ordinary" generics. To ensure this distinction is empirically unimportant, we append equation 8 to include dummy for authorized generics. Reassuringly, estimates of σ and α are entirely unchanged, and the coefficient on the authorized generic dummy is small and not significant.

We observe that two drug classes experience larger-than-expected price changes at the onset of collusion (as shown in Figure XIII in the Online Appendix). We also observe that they witness more subsequent entry than one would predict on the basis of market size alone. Both patterns are potentially consistent

[31] To avoid taking a stand on whether the years in which the cartels are formed are "treated" or "control" periods, we omit observations for which $t = t_d$. Also, to focus on periods proximate to cartel formation, we restrict attention to t such that $|t - t_d| \leq 5$. These restrictions have a negligible effect on our conclusions.

[32] To be precise, the instrument equals $M_{dt} + N_{dt}$, which is equivalent to $M_{dt} + N_{dt} - \bar{M}_d - \bar{N}_d$ in the presence of drug specific dummy variables, which all demand specifications include. Though the parameters are jointly determined, one can build intuition around identification by considering entry, which increases the instrument's value and shifts shares. If the entrant steals a relatively large share from other drugs instead of the outside option, this implies σ is relatively large. Exit provides analogous underlying variation.

Table II: *Demand estimates*

VARIABLES	(1) IV	(2) IV	(3) IV	(4) IV
Log of inside share	0.56***	0.56***	0.73**	0.73**
	(0.16)	(0.16)	(0.31)	(0.30)
Price	-0.11***	-0.11***	-0.21**	-0.21**
	(0.031)	(0.031)	(0.10)	(0.10)
Price X β-blocker			0.16*	0.16*
			(0.089)	(0.088)
Price X Anticonvulsant			0.20**	0.20**
			(0.094)	(0.094)
Indicator for authorized generic		-0.043		-0.067
		(0.17)		(0.31)
Observations	19,299	19,299	19,299	19,299
Drug FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Number of drugs	415	415	415	415

, **, and * denote significance at the 10%, 5%, and 1% levels, respectively.*
The unit of observation is a drug-year-firm. Standard errors are clustered at the drug level.

with less elastic buyers. Thus, to accurately represent demand and correctly forecast entry incentives, we allow α_d to vary based on whether the drug is an β-blocker, anticonvulsant, or other generic. That is, we replace $\bar{\alpha}$ on the right-hand side of equation 8 with $\alpha_0 + \alpha_1 \mathbb{1}\{\beta\text{-blocker}\} + \alpha_2 \mathbb{1}\{\text{antiepileptic}\}$, where $\mathbb{1}\{\cdot\}$ denotes an indicator variable.[33]

Column 3 reports the result. We find that $\alpha_1 = 0.16$ and $\alpha_2 = 0.19$, consistent with less elastic buyers in the separately named drug classes. As expected, we find a higher baseline price coefficient (in absolute value terms), with $\alpha_0 = -0.21$. We also estimate a stronger correlation of products within the nest of inside goods. All estimates are significant at the 5% level except α_1, which is significant at the 10% level. Column 4—included mostly for completeness—shows that column 3's results are equally as insensitive as column 1's to distinguishing between authorized and ordinary generics. Since accounting for heterogeneity in price sensitivity is important, while separating out authorized generics is not, we rely on estimates from column 3 for the remainder of the paper.

These estimates imply markups that very closely align with figures reported by Teva in their financial statements, as shown below. However, these comparisons depend in part on costs and conduct, so we discuss these issues first.

6.2 Marginal costs

As in the prior subsection, we parameterize marginal cost such that its log value is linear in drug and time dummy variables and a disturbance, which is i.i.d. over d, f, and t. Under this parameterization, we can recover the relevant parameters using only data from competitive markets. Estimation has three steps. First,

[33] Allowing α to vary more flexibly may be desirable, but generates noisy estimates.

we restrict attention to drug-year observations such that either (a) NP never cartelizes d or (b) NP cartelizes d at t_d but we consider only $t < t_d$. Second, we compute

$$\widehat{mc}_{dft} = p_{dft} - (\frac{\partial s_{dft}}{\partial p_{dft}})^{-1} s_{dft},$$

which is a function of estimated demand parameters and data. Third, we estimate

$$ln(\widehat{mc}_{dft}) = \gamma_d + \gamma_t + \omega_{dft}.$$

We find that the average marginal cost of a drug in our sample is \$10.45. We also observe heterogeneity across drugs and over time. For a sense of dispersion as well as how the "treatment" and "control" groups of drugs compare to one another, we separate cartelized and uncartelized drugs, compute average predicted log marginal cost (i.e., $\hat{\gamma}_d + \hat{\gamma}_t$), and plot the resulting densities in Figure VIII.

From Figure VIII we can observe that most of the support lies between 1 and 4.5. Accounting for the fact that ω enters the marginal cost function nonlinearly, marginal cost range between \$3 and \$90. In spite of this degree of dispersion, the mean, variance, and shape of the distributions are similar for cartelized and uncartelized drugs, further supporting the idea that the two groups are comparable to one another.

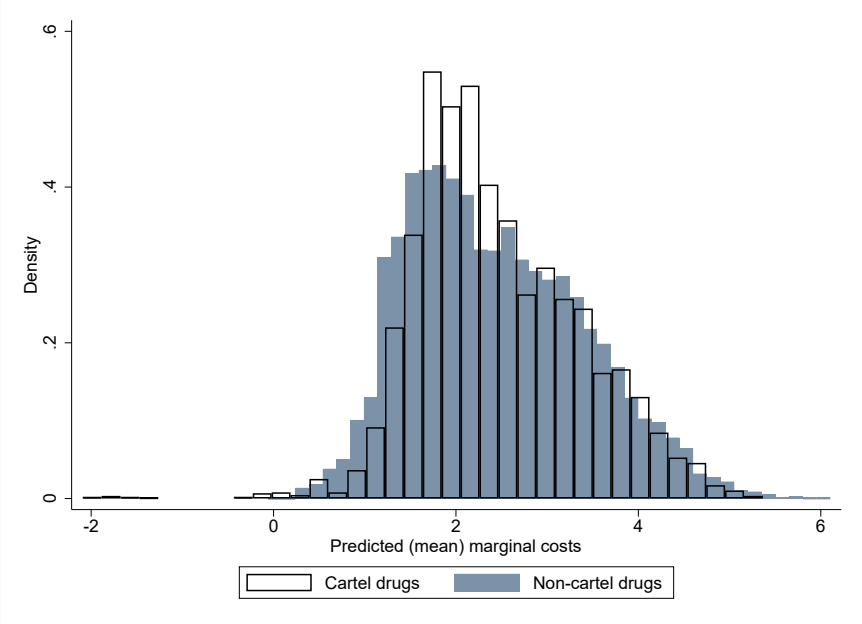

Figure VIII: *Distribution of average predicted log marginal cost.*

This figure plots the density of predicted log marginal cost, $\hat{\gamma}_d + \hat{\gamma}_t$. All drug-year observations are included; however, to avoid taking a stand on conduct before formal testing, we estimate the parameters on the subset of observation where (a) NP never cartelizes d or (b) NP cartelizes d at t_d but we consider only $t < t_d$.

If we already knew exactly which drug-year observations reflect collusive prices, then we could also draw

inference about the marginal cost parameters from markets where cartel members set prices cooperative.[34] However, even though we know precisely which drug markets NP cartelizes and when she does so, we cannot assume consistent conduct through our entire sample period. We investigate how and whether the cartel sustained cooperation immediately below.

6.3 Incentives to collude, conduct testing, markups, and damages

Incentive compatibility constraints. As we noted in Section 5, the collusive agreements are unenforceable in court or by means of coercion, so all members of the cartel must find it in their self-interest to set prices cooperatively. Cartel members will always find cheating a myopically profitable option, so sustained collusion exists only through the grim threat of returning to competition and the fact that collusive profits exceed competitive ones. To be precise, ICCs are satisfied if and only if the present value of expected deviation profits—which are accrued until cheating is discovered—plus the present value of expected Bertrand-Nash profits do not exceed the present value of expected perpetual collusive profits.[35]

With period length set to one year (to mirror the aggregation of our data), ICCs typically hold, often by a wide margin. Except for very low probability events (e.g., extreme demand shocks),[36] the value of deviation never exceeds the value of cooperation by more than 20%. Slack in the constraints depends on many factors, but in this setting, the relative willingness to substitute across drugmakers' products, the stable evolution of demand and cost determinants, and the small number of firms make the largest contributions.

In generic drug markets, though, deviation profits would never accrue for a full year. Cartel members would discover cheating quickly and could retaliate nearly as fast. Recall, for instance, the example from Section 5 where Glenmark inadvertently undercut Teva: Teva learned of Glenmark's deviation almost immediately, and Glenmark revised its pricing the very next day.[37] This is important, since the value of deviating is sensitive to the time during which deviation profits accrue.[38] In our setting, the timing assumption is innocuous: a period length as long as a month ensures the value of cooperation exceeds the value of deviating *by around an order of magnitude*.

Conduct testing. We next assess whether the government investigation affected firm conduct. Based on the data, we reject the notion that the investigation affected behavior and instead conclude that firms continue to collude, perhaps tacitly, after 2015. Our conclusion, though perhaps initially surprising, is sensible: if ICCs are satisfied, demand and costs are stable, and accurate information about cartel members' behavior

[34] To do so, we would compute $\widehat{mc}_{dt} = p_{dt} - \Delta_{dt}^{-1} s_{dt}$, where mc_{dt}, p_{dt}, and s_{dt} denote vectors of marginal costs, prices, and market shares, and where Δ_{dt} is a matrix determined by ownership and substitution patterns. Rows in mc, p, and s correspond to firms, arbitrarily ordered, while rows and columns in Δ correspond to different firms, ordered in the same way. The (i, j) element of Δ_{dt} equals $\partial s_{dit}/\partial p_{djt}$ if $i = j$ or if both i and j are cartel members, and it equals zero otherwise.

[35] If we put aside the prospect of entry to ease exposition, slightly abuse notation, and assume cheating is discovered the next period, we can quickly formalize the ICC. Assuming $f \in \mathscr{M}$ cooperates and all other cartel members cooperate, if f cooperates, then it earns $V_{dft}^{Coop.}$, which equals $\sum_{t=0}^{\infty} \delta^t \int_{\xi'\omega'} \pi_{dft}^{C,M}(\xi', \omega') dF_{\xi'} dF_{\omega'}$. However, if f deviates, then it earns $V_{dft}^{Dev.}$, which equals the sum of $\int_{\xi'\omega'} \pi_{dft}^{Dev.}(\xi', \omega')$ and $\sum_{t=1}^{\infty} \delta^t \int_{\xi'\omega'} \pi_{dft}^{N}(\xi', \omega') dF_{\xi'} dF_{\omega'}$ where $\pi_{dft}^{Dev.}(\cdot)$ is the "deviation" profit, which results from f best responding to cooperation among the other members. Collusion requires $V_{dft}^{Dev.} < V_{dft}^{Coop.}$.

[36] Very large draws of ξ can create enough asymmetry in current period profits to "break" the ICC of the member receiving the draw.

[37] See also the case of Par, Teva, and the market for labetalol hydrochloride tablets (Complaint, 227-278).

[38] To illustrate, suppose that "deviation" profits are twice as large as Bertrand-Nash profits and that collusive profits are 1.1 times as large. If deviation profits accrue for one year, then ICCs are exactly met, but if the deviation profits accrue for a month, the *the value of cooperating is an order of magnitude larger than the value of deviating*.

is rapidly available, then collusion should be easy to maintain even if it is hard to initiate. As one Teva executive states, "Price increases tend to stick and markets settle quickly when suppliers increase within a short time frame," (Complaint, page 160).

Our assessment relies on an intuitive test of conduct proposed by Backus et al. (2021) which adapts the non-nesting framework proposed by Rivers and Vuong (2002) to exclusion restrictions suggested by Berry and Haile (2014). The test is based on the idea that if conduct is correctly specified, then variables that determine markups but do not influence marginal costs will be uncorrelated with predicted marginal cost disturbances implied by that model. The test provides pairwise comparisons, and it does not require that either model of conduct is exactly correct. Informally, the test requires the econometrician to (a) calculate markups implied by each model of conduct, (b) subtract implied markups from observed prices to arrive at implied marginal costs, (c) compute marginal cost residuals, (d) calculate the correlation between the resulting residuals and variables that enter markups but not marginal costs, and (e) compare the correlations. Weaker correlations correspond to greater consistency of the conduct assumption and the data.

To illustrate how their test operates, consider a slightly simpler setting. For the sake of an example, suppose there are only two periods, and suppose drug markets fall into one of two groups. In group X, firms compete in both periods, while in group Y, firms compete in period 1 but fix prices in period 2. Next, construct an indicator variable that takes a value of one in group Y in period 2 and zero otherwise. The indicator variable is a valid instrument in the sense that it is strongly correlated with markups but does not directly affect marginal cost shocks, since price fixing does not impact production costs. Finally, compare the relative performance of two models of conduct. In the first model (A1), we correctly specify conduct. In an alternative model (A2), we incorrectly specify that firms compete in all periods and markets. By estimating the first model, one recovers true marginal cost shocks (up to estimation error). Since these disturbances are i.i.d. across markets and time, they are *uncorrelated* with the indicator variable. That is, the correlation between marginal cost shocks and the instrument is very low in A1. In estimating the second model, one recovers markups that are too low in group Y in period 2, which translate into marginal costs that are too high, which in turn translate into marginal costs shocks that are on average positive. In other words, when the indicator variable is turned on, marginal cost shocks are especially large. The instrument and marginal cost shocks are positively correlated in A2. Therefore, by comparison, the test favors A1.

We conduct two tests, which are summarized by Table III. Both tests assume that firms set Bertrand-Nash prices prior to the date that the cartel is alleged to have formed. In the first test (Test 1), we assess whether the prices of drugs named in the Complaint are consistent with competition or collusion between t_d and 2015. In the second test (Test 2), we assess whether those same prices are consistent with competition or collusion between 2016 and the end of the sample. That is, the first test evaluates conduct from the start of collusion until the government's investigation became public, whereas the second examines conduct after that point. Each amounts to a pairwise comparison of two models, which we have arbitrarily named A1 and A2.

One concern with our modeling approach is that cartel members may fold non-members into the cartel post-entry. If unmodelled, this behavior would affect our profit calculations and, therefore, the measure of sunk costs. We note that a single entrant should never expect to earn more from joining the cartel than they could earn bidding against the cartel. (Otherwise, the cartel should kick them out.) However, this need not be the case for larger groups of entrants. Using our framework, we can also evaluate the claim

Table III: *Behavioral assumptions for conduct tests*

Test 1.				
Markets:	Drugs named in the Complaint		Other drugs in the sample	
Model:	A1	A2	A1	A2
2008—(t_d-1)	Own-profit max.	Own-profit max.	Own-profit max.	Own-profit max.
t_d—2015	Collusion	Own-profit max.	Own-profit max.	Own-profit max.
2016—2018	–	–	–	–
Test 2.				
Markets:	Drugs named in the Complaint		Other drugs in the sample	
Model:	A1	A2	A1	A2
2008—(t_d-1)	Own-profit max.	Own-profit max.	Own-profit max.	Own-profit max.
t_d—2015	–	–	–	–
2016—2018	Collusion	Own-profit max.	Own-profit max.	Own-profit max.

"Own-profit maximization" refers to Bertrand-Nash pricing, meaning that $p = p^B$. "Collusion" corresponds to cartel members setting price equal to p^B and nonmembers setting it equal to $p^{C,N}$.

that cartel nonmembers best respond to the members prices (rather than going along with the collusive agreements upon their formation).[39] To do so, we introduce a third model, A3, in which all firms set prices cooperatively, not just cartel members. We then replicate Tests 1 and 2, replacing A2 with A3. In both sets of pairwise comparisons, the data strongly favor A1 over A3, consistent with the hypothesis that nonmembers are not "folded into" the cartels. The corresponding test statistics are -3.78 and -4.34, respectively, and we can reject the hypothesis. The appendix provides details.

Adapted to our setting, the steps proposed by Backus et al. (2021) to make pairwise comparisons between A1 and A2 are as follows:

1. Compute markups under models $A1$ and $A2$. Denote the markups mk^{A1}_{dft} and mk^{A2}_{dft}, respectively. Then compute $mc^{A1}_{dft} = p_{dft} - mk^{A1}_{dft}$ and $mc^{A2}_{dft} = p_{dft} - mk^{A2}_{dft}$.
2. Estimate $ln(mc^{\Omega}_{dft}) = \gamma^{\Omega}_d + \gamma^{\Omega}_t + \omega^{\Omega}_{dft}$ for $\Omega \in \{A1, A2\}$ to recover residuals $\hat{\omega}^{A1}_{dft}$ and $\hat{\omega}^{A2}_{dft}$.
3. Estimate $mk^{A1}_{dft} - mk^{A2}_{dft} = g(z_{dt}) + \zeta_{dft}$. $g(.)$ and z are context specific and described below.
4. Compute $\tilde{Q}(mk^{\Omega}) = \left(R^{-1}1\sum_{f,t}\hat{\omega}^{\Omega}_{dft}\hat{g}(z_{ft})\right)^2$ for $\Omega \in \{A1, A2\}$, where R is the number of products.
5. Denote the standard error of the difference between $\tilde{Q}(mk^{A1})$ and $\tilde{Q}(mk^{A2})$ as $\hat{\sigma}/R$; estimate $\hat{\sigma}/R$ by repeating the previous steps on bootstrapped samples.
6. Compute the test statistic, which equals $\hat{\sigma}^{-1}\sqrt{R}(\tilde{Q}(mk^{A1}) - \tilde{Q}(mk^{A2}))$ and is distributed $\mathcal{N}(0,1)$.

In Step 2, we specify that the log of marginal costs is linear in drug- and year-specific dummy variables, and we maintain that specification throughout the rest of the paper. Following BCS, z includes these dummies

[39] We thank Rob Porter for suggesting this exercise.

as well as an instrument for markups. The instrument is an indicator variable that takes a value of one if and only if the observation is associated with a cartelized drug and with a year that is greater than or equal to the year in which the Complaint alleges that drug was cartelized. The predicted difference in marginal costs, $g(\cdot)$, is linear in drug-specific dummy variables, year-specific dummy variables, and the instrument.

Table IV reports the results. In both tests, $\tilde{Q}(mk^{A2})$ far exceeds $\tilde{Q}(mk^{A1})$, yielding test statistics that are less than -10. To interpret these findings, recall that smaller values of $\tilde{Q}$ indicate better fit between the conduct assumption and observed behavior and that the test statistic is distributed standard normal. Thus, for drugs listed in the complaint, the data strongly reject competition in favor of collusion from the formation of the cartel through the end of the panel.

Table IV: *Results of conduct tests*

	$\tilde{Q}(mk^{A1})$ $\times$ 100	$\tilde{Q}(mk^{A2})$ $\times$ 100	Test statistic	p-value
Test 1. (Collusion pre-investigation?)	.68064	1.9315	-9.65	< .0001
Test 2. (Collusion post-investigation?)	.49900	1.7637	-10.84	< .0001

This table reports the results of the testing procedure proposed by Backus et al. (2021). The test statistic is distributed standard normal. The standard error of the difference between $\tilde{Q}(mk^{A1})$ and $\tilde{Q}(mk^{A2})$ is obtained via bootstrapping. See Table III for the definitions of A1 and A2 under Tests 1 and 2.

Evaluating markups. The preceding estimates imply sensible price-cost margins that closely match figures reported by Teva in their financial statements. To obtain values implied by our model, we set the first order condition of the profit function with respect to price equal to zero, solve for $p_{dft} - mc_{dft}$, divide the resulting markups by prices, and average over drugs manufactured by Teva, weighting by revenue. To obtain analogous figures from Teva's annual reports, we extract segment-specific income statements and compute the ratio of operating profits to total revenue for their generic division.[40] Our model assumes competitive pricing and implies that profit margins average 19.7%, while Teva's financial statements imply 20.0% in the two years prior to NP joining Teva. In the two years after NP joins Teva, our model assumes NP has cartelized many drug markets and implies that profit margins average 39.6%, while Teva's financial statements imply 38.6%. In other words, forecasts from the model not only match profit rates in levels but also changes around cartel formation.[41]

Assessing damages. Using our model and demand estimates, we can compute damages to consumers. For each product in a cartelized drug market, we compute equilibrium prices under competition and collusion, and we multiply the difference by the number of observed prescriptions. The median price differences are \$9.21, \$8.70, and \$9.45 per prescription for 2013, 2014, and 2015, respectively. Mean price differences are

[40] Operating margin is the right choice, given how Teva reports its income. Operating profit reduces total revenue by cost of goods sold and selling/marketing expenses, which are mostly variable, but not general/administrative expenses, (e.g., executive compensation, headquarters operations, etc.), which are mostly fixed/sunk.

[41] Although careful demand estimation contributed to this result, we believe that *such a close correspondence* between the model's predictions and the financial statement analysis is, at least in part, coincidental. The goal of this exercise was to see if the model was in the neighborhood of the annual reports—not whether it was a close match.

slightly higher at \$11.72, \$13.92, and \$12.68, respectively. Damages total \$978 million, \$1.9 billion, and \$2.5 billion, respectively. That is, damages total \$5.4 billion over the three year period, which averages out to \$13.7 million per drug per year.

Our figures are very similar to those reported in two other recent studies. Cuddy (2020) finds that collusion induced nationwide damages of \$7.9779 billion per year, or \$49.5 million for each substance-delivery-release combination she studies.[42] Even though our data and models differ, we arrive at a nearly identical figure—at their peak, the cartels cause damages of \$45.9 million per substance-delivery-release combination. Clark et al. (2021) study six substance-delivery-release combinations that were affected by price fixing, estimating damages using a carefully constructed difference-in-difference research design. Again, we reach similar estimates.[43] Since the source of our quantity data is the same as theirs, we predict nearly identical damages for the substance-delivery-release combinations for which we overlap.

6.4 Sunk costs

Setup. The necessary conditions of a simultaneous move Nash equilibrium imply that firms enter if and only if the value of doing so, conditional on the decisions of rival firms, exceeds the sunk costs. These restrictions imply the following bounds on sunk costs. For each cartel member that enters and does not enter j, respectively, we have

$$\sum_{d \in \mathscr{J}} VE_d^M(\chi^M, \chi^N) \geq \mathscr{E}[\theta_{jM} | \mathscr{I}_{jM}] \tag{9}$$

and

$$\sum_{d \in \mathscr{J}} VE_d^M(\chi^M + 1, \chi^N) < \mathscr{E}[\theta_{jM} | \mathscr{I}_{jM}]. \tag{10}$$

For each nonmember that enters and does not enter j, respectively, we have

$$\sum_{d \in \mathscr{J}} VE_d^N(\chi^M, \chi^N) \geq \mathscr{E}[\theta_{jN} | \mathscr{I}_{jN}] \tag{11}$$

and

$$\sum_{d \in \mathscr{J}} VE_d^N(\chi^M, \chi^N + 1) < \mathscr{E}[\theta_{jN} | \mathscr{I}_{jN}]. \tag{12}$$

Here, $\mathscr{M}$ and $\mathscr{N}$ denote the set of cartel members and nonmembers, respectively, and f, f', f'', and f''' are understood to be firms that did not have approved ANDAs covering j (or else they would not be contemplating entry in the first place). Inequalities 9-12 are the basis for estimation.

We allow sunk costs to vary with two important observable determinants. First, we allow them to depend on the number of strengths associated with the substance-delivery-release combination.[44] When

[42] To arrive at the \$7.9779 billion figure, we multiply \$1.3755 billion (from her Table 8) by 5.8 (from her Section 6.4), which the author states will scale her sample to nationwide consumption.

[43] Whereas they estimate 44.2% and 13.5% price increases for nystatin and theophylline, respectively, our structural model predicts 41.4% and 20.8% changes. To arrive at these figures, we divide the estimated damages per defined daily dose by pre-collusion prices, both of which are reported by the authors in their Table 7. Specifically, we define \$0.21 by \$1.561 and \$0.155 by \$0.350.

[44] Production requires know-how that some firms may lack. The counterfactual policies we consider attract additional entrants, so we require that at least that many potential entrants possess this know-how. The requirement is easily met in our setting, which studies relatively "simple" drugs like orally administered solids (i.e., tablets and capsules), but it could bind other segments of the market.

preparing an ANDA for filing, firms must demonstrate "bioequivalence" to the innovator drug at each strength level. The process is costly, as it involves measuring the time it takes for a given amount of the substance to reach the bloodstream in healthy volunteers. Separately, once approval is granted, most delivery methods require firms to design distinct packaging and install (or repurpose) separate equipment for each strength. Second, we allow sunk costs to vary with delivery method. Compared to tablets and capsules, which are by far the most common dosage forms among orally administered medications, irregular delivery methods such as syrups, solutions, and chewables involve specialized equipment and more complicated packaging. Moreover, these methods are much more susceptible to bacteria growth, which necessitates sterile packing conditions. We also allow sunk costs to depend on a symmetric i.i.d. disturbance that is unobserved by the econometrician but known by the firms when they decide whether to enter. We assume it is independent of the substance-delivery-release combination characteristics (i.e., the number of strengths and the delivery method).

Formally, we parameterize sunk costs such that

$$\theta_{jk} = \theta_0 + \theta_1 r_j + \theta_2 \ell_j + \eta_j, \tag{13}$$

where $k \in \{M, N\}$, r denotes the number of additional strengths associated with the substance-delivery-release combination, ℓ is an indicator variable that equals one for drugs with an irregular delivery method. η_j represents the disturbance term known to the firms when they contemplate entry. This term permits the sunk costs of entering the one market (e.g., warfarin tablets) to differ from the sunk costs of entering another (e.g., cefdinir capsules) in ways that influence behavior. However, it also introduces a selection problem. Entry is especially common in markets with low η_j but uncommon in markets with high η_j. To illustrate the problem that arises in estimation, suppose we naively construct exactly one instance of inequality 9 or 11 for each substance-delivery-release combination and entrant, substitute measured objects for true values, pool the instances together, and calculate their mean. The resulting inequality includes a weighted average of η_j, where the weights are determined by the number of entrants, which are negatively correlated with η_j. Since the weighted average is negative but unobserved to the econometrician, the observed portion of the inequality produces a biased bound.

If the predictions of the model and the actions of the firms differ beyond the flexibility provided for in equation 13, the differences are rationalized by mean-zero expectation errors. To be specific, their presence reconciles a small number of cases where entry by a nonmember implies that a member should enter but does not (and vice versa). Since these disturbances rationalize differences in the realized payoffs of cartel members relative to nonmembers, we can ascribe them to either group without loss of generality. We pin them to cartel nonmembers. Formally, we define $\theta_{jk} = \mathscr{E}[\theta_{jk}|\mathscr{I}_{jk}] + \nu_{jN}$, where $k \in \{M, N\}$.[45]

Finally, we assume that expected profits and the distribution of delays are measured without error. To be precise, we assume that

$$\int_{\xi'} \int_{\omega'} \hat{\pi}^k_{d,t}(M, N, \xi, \omega) d\hat{F}_\xi d\hat{F}_\omega = \int_{\xi'} \int_{\omega'} \pi^k_{d,t}(M, N, \xi, \omega) dF_\xi dF_\omega, \tag{14}$$

[45]For readers familiar with notation introduced by Pakes et al. (2015), our ν correspond to their "$\nu_{1,\cdot}$-type errors," while our η correspond to their "$\nu_{2,\cdot}$-type errors."

with $k \in \{M, N\}$, and $\hat{F}_D = F_D$. Here, $\hat{F}_D$, $\hat{F}_\xi$, and $\hat{F}_\omega$ denote the empirical distributions of delays, demand shocks, and marginal cost shocks, and $\hat{\pi}^k_{d,t}(M, N, \xi, \omega)$ denotes $\pi^k_{d,t}(M, N, \xi, \omega)$ evaluated at our estimates of λ_d, λ_t, α_d, σ, γ_d, and γ_t rather than the true parameters.[46] The first restriction is commonly employed in the literature, since it greatly reduces the computational burden of calculating the bounds, and a reasonable approximation to reality in our setting as well. It is supported by, for example, the fact that our quasi-experimental variation produces demand parameters that imply sensible profit margins, which are consistent with the audited financial statements (in both levels and in changes around the variation we exploit). The second restriction is equally reasonable. We observe the filing, approval, and entry dates of ANDAs submitted to the FDA over more than a decade, so our empirical distribution reflects several thousand applications, and our information closely resembles that which is available to the managers of the drugmakers.

Moments. The solution to the selection problem introduced by the structural errors lies in the fact that although η's conditional expectation varies with observed entry, its unconditional expectation is nonetheless mean zero (Ishii, 2005; Ho, 2009; Pakes et al., 2015). To see this conceptually, suppose for the sake of example that at least one cartel member enters each substance-delivery-release combination. Further, suppose that we construct precisely one instance of inequality 9 for each substance-delivery-release combination, substitute measured objects for true values, ignore the error terms, pool the instances together, and calculate their mean. This process collects one η_j from each j, yielding an unselected set of disturbances whose expected value is zero.[47]

First, we estimate lower bounds on the parameters of interest, which reflect firms that opt not to enter upon cartel formation. The first set of moments is given by

$$\frac{1}{J}\sum_j \frac{1}{\mu_j}\sum_{d\in\mathscr{J}} \frac{1}{2}\sum_{k\in\{M,N\}} h^i_j\Big[\widehat{VE}^k_d(\chi^M + \mathbb{1}\{k = M\}, \chi^N + \mathbb{1}\{k = N\}) - \theta_0 - \theta_1 r_j - \theta_2 \ell_j\Big] < 0. \tag{15}$$

Moments are indexed by i and formed by weighting observations using h^i_j, described immediately below. J denotes the number of unique substance-delivery-release combinations, and μ_j denotes the size of set $\mathscr{J}$ (i.e., the number of strengths associated with substance-delivery-release combination j). $\widehat{VE}^M_d(\cdot)$ equals $VE^M_d(\cdot)$ evaluated at our estimates of $\pi(\cdot)$, F_ξ, F_ω, and F_D rather than the true values, and $\widehat{VE}^N_d(\cdot)$ is defined analogously for nonmembers. In the appendix, we prove inequality 15 produces consistent bounds.

The weights applied in inequality 15 depend on substance-delivery-release characteristics. One moment i is formed by setting elements of h^i_j equal to one (i.e., a constant). Two moments are formed by setting h^i_j equal to $\mathbb{1}\{\ell_j = 0\}$ and $\mathbb{1}\{\ell_j = 1\}$. Three more moments are formed by setting h^i_j equal to $\mathbb{1}\{r_j = 0\}$,

[46] Notice that we require $\hat{\pi}^k_{dt}$ for t in the future. In practice, we assume that the values of M_{dt}, N_{dt}, λ_t, and γ_t in 2020 and beyond are fixed at their 2019 values. This is a reasonable approximation. M_{dt} is unlikely to change beyond that point, since re-entry is very infrequent four years or more after cartel formation. The same is true of N_{dt}, since exit is rare throughout the panel. Also, λ_t and γ_t do not exhibit meaningful trends in the final three years of the sample.

[47] This approach exploits the "ordered choice" nature of the problem. That is, since profits are declining in the number of competitors, if one observes X number of firms enter a market, then one can infer that the profits for X exceed the sunk costs of entry but the profits for $X + 1$ do not. This logic yields an upper and lower bound, demonstrated by Ishii (2005). We differ from her approach by having two types of entrants—cartel members and nonmembers—and permitting expectational error to reconcile differences in the sunk costs implied by their decisions. See Section III.C of Wollmann (2018) for a general discussion of ways to relax this assumption. To name one, the econometrician could specify the shape of the structural error and take the "probability inequality" approach, proposed by Tamer (2003), though this approach is computationally infeasible in our setting.

$\mathbb{1}\{r_j = 1\}$, and $\mathbb{1}\{r_j > 1\}$. Importantly, $h_j^i \geq 0$ for all j, $h_j^i > 0$ for at least one j in each i, and the variables used to construct h^i are all independent of η.

Second, we estimate upper bounds on the parameters of interest. This estimation strategy here is complicated by the fact that not every substance-delivery-release combination experiences entry. This feature of the data precludes us from constructing moments with exactly one η_j per j. By extension, if we form evenly weighted averages of η_j for the j in which we do observe entry, then we reintroduce the selection problem, since missing disturbances are on average unfavorable to entry. To solve this problem, we rely on an idea proposed by Pakes et al. (2015) and based on Powell (1986), which exploits the symmetry of η_j.

To illustrate using round numbers, suppose there are 100 markets of which 70 experience entry. Further, suppose that we employ an estimation strategy that is analogous to the one used to recover lower bounds on the sunk cost parameters (i.e., each substance-delivery-release combination contributes exactly one instance of the sunk cost disturbance term whenever possible). Then, the resulting moment includes a term that subtracts the average of 70 realizations of η for which we observe entry. Though η is unconditionally mean zero, our moment collects 70 realizations of η that are favorable to entry, so their values are negatively selected. This is the essence of the problem.

From the standpoint of obtaining consistent bounds, the worst case corresponds to our moment including the 70 smallest η (i.e., the negative selection is as bad as possible). Since we cannot rule this case out, we must work under the assumption that it is true. The solution involves subtracting out 30 values that are at least as small as the 30 smallest realizations of η, leaving us with the 40 most central draws of η plus some weakly positive differences.[48] Since the mean of a set of draws symmetrically distributed about zero converges in probability to zero, the term in our moment that includes the structural error converges in probability to a weakly positive value.

To obtain 30 values that are at least as small as the 30 smallest realizations of η, we draw in information from firms that did not enter. Ignoring expectational errors, which will average out in a way that does not affect the consistency of our bounds, the *observable* components of inequalities associated with decisions not to enter provide upper bounds on the draws of η. For instance, the most negative observable component is at least as small as the smallest η (since one more firm would have chosen to enter if it were greater greater than or equal to η). By extension of that logic, the 30 most negative observable components are at least as small as the 30 smallest realizations of η.

To formalize the approach, we require additional notation. Let L be the set of j with at least one entrant, let J_L be the size of that set, and let w^i be a positive valued function of r_j and ℓ_j. Also, define

$$VE_j^+ = \frac{1}{2} \sum_{d \in \mathscr{J}} \sum_{k \in \{M,N\}} \left[\widehat{VE}_d^k(\chi^M, \chi^N + \mathbb{1}\{k = N\}) \right], \tag{16}$$

which represents the average of the cartel members' and nonmembers' entry values in j. Finally, for each moment i, order j by their values of $w_j^i VE_j^+$, and let Ψ_{wVE} denote the set of j that correspond to the $J - J_L$

[48] To be precise, due to the presence of ν errors, the 30 draws are *on average* the 30 smallest, and the remaining ones are *on average* the 40 most central.

smallest values. The second set of moments is then given by

$$\frac{1}{J}\sum_{j\in L}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\mathbb{1}\{\chi^k\geq 1\}\widehat{VE}_d^k(\chi^M+\mathbb{1},\chi^N)}{\mathbb{1}\{\chi^M\geq 1\}+\mathbb{1}\{\chi^N\geq 1\}+\mathbb{1}\{\chi^M=0,\chi^N=0\}}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]$$
$$-\frac{1}{J}\sum_{j\in\Psi_{wVE}}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\widehat{VE}_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})}{2}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]<0. \quad (17)$$

In the Online Appendix, we prove inequality 17 produces consistent bounds.

Our moments provide one-sided restrictions on the parameters of interest, the most informative of which are the greatest lower bound and the least upper bound. Since our inference procedure is based on maxima and minima, respectively, rather than averages, we cannot rely on the central limit theorem. To obtain 95% confidence intervals around the true parameters, we follow Andrews and Soares (2010).[49] They propose inverting an Anderson-Rubin type test at each candidate vector $[\phi_0\ \phi_1\ \phi_2]$ (Chernozhukov et al., 2007). That is, for each candidate, we calculate each moment and its standard deviation. We then divide the first by the second and sum the resulting values to arrive at a test statistic. Then, we draw 2,000 bootstrap samples and compute analogous statistics to obtain a critical value (for our desired confidence level, i.e., 95%). Finally, we comprehensively search over a 3-dimensional grid and include in our confidence interval each candidate whose test statistic is less than or equal to the relevant critical value.

Results. Table V reports sunk cost estimates. The final column provides 95% confidence intervals, all of which exclude zero. We find that the coefficient on the constant term, θ_0, is bounded between \$260,000 and \$1.88 million. Reflecting the fact that sunk costs scale with the number of strengths per substance-delivery-release combinations, we find that ϕ_1 is bounded between \$0.92 million and \$2.02 million. Consistent with irregular delivery methods involving substantially higher sunk costs, we find that ϕ_2 is between \$950,000 and \$6.925 million. Calculated at the midpoint of each identified set, entry costs range between \$1.07 million and \$12.83 million, depending on the characteristics of the substance-delivery-release combination. The average across drugs, weighting each equally, is \$3.175 million.

Our entry cost estimates align with statements made by agency officials and medical researchers. For instance, FDA Commissioner Gottlieb states, "Filing a generic application requires an average of about \$5 million and can cost as much as \$15 million"—very close to the mean estimate reported in row 5 of Table V.[50] Similarly, our estimates fall squarely within the bounds reported by Scott-Morton (1999), who surveyed FDA officials around 1999 and found that sunk costs range from \$382,000 to \$31 million (in 2019 dollars, i.e., \$250,000 to \$20 million in the original text). Even more to the point, Baker-Smith et al. (2008) ran six bioequivalence studies between 1997 and 2004 and documented their per-study expenditures.[51] Baker-Smith et al. (2008) spent between \$807,000 to \$1.25 million per study (in 2019 constant dollars), which

[49] Tests based on inequalities depend on the degree to which the moments are binding. One consequence is that including uninformative moments (i.e., ones that are satisfied for a very wide range of parameters) typically widens confidence intervals. The main innovation of Andrews and Soares (2010) is a procedure for, loosely speaking, deciding which moments are sufficiently uninformative to be discarded. The authors have found moment selection incredibly important in other settings. We deemphasize it here only because it does not have a big effect on our results.

[50] See, e.g., Gottlieb (2016) and November 20, 2014 testimony before the Senate Subcommittee on Primary Health and Aging.

[51] Putting aside one-time production setup costs, pharmacological studies are the principal reason that entry costs depend on the number of strengths per substance-delivery-release combination, so they are an especially good benchmark for our estimates of the coefficient on r.

Table V: *Sunk cost estimates*

Variable	Parameter	95% confid. interval
Constant	ϕ_0	[0.26,1.88]
Number of additional strengths [r]	ϕ_1	[0.92,2.02]
Indicator for irregular delivery [ℓ]	ϕ_2	[0.95,6.925]
Observations	220	
Moments	10	
Minimum sunk cost of entry	1.07	
Average sunk cost of entry	3.175	
Maximum sunk cost of entry	12.83	

Sunk cost estimates are reported in millions of dollars. The bounds reported in the rightmost column are intervals in which the true parameters lie 95% of the time. To calculate the minimum, mean, and maximum sunk costs across the substance-delivery-release combinations, we set ϕ_0, ϕ_1, and ϕ_2 equal to their respective midpoints.

aligns with our estimates, which bound per-strength entry costs between about $1-2 million.

7 Policy analysis

In this section, we use the model estimates to perform policy counterfactual experiments. The preceding results indicate that entrants discipline collusive prices, but entry is limited and slow. Thus, we ask two questions. First, how much would American consumers pay to reduce the sunk costs of entry, and second, how much would they pay to accelerate entry?

These are policy relevant questions, since much of the time and expense associated with entry is within the federal government's control. In terms of expenses, the FDA collects fees from drugmakers directly under the Generic Drug User Fee Act. In 2022, fees totaled $225,712 per ANDA.[52] On top of direct transfers to the agency, firms must hire skilled staff to ensure compliance or outsource compliance to highly compensated third-party consultants. Some delays result from a lengthy approval process coupled with backlogged reviews, while others simply reflect the time-consuming process of preparing an application, which includes the clinical work that establishes bioequivalence to a reference drug. Notably, the FDA has varied fees considerably over the past decade and repeatedly experimented with expedited review programs, proving that fast approval is feasible (presumably without compromising the integrity of the process).[53]

We first simulate the effect of reducing entry costs by $400,000 and $800,000 per ANDA. This saves

[52]Drugmakers pay the FDA a base rate of $1.5 million per year and $200,000 per factory per year. If they wish to keep certain aspects of their operations confidential, especially as they relate to the production of active pharmaceutical ingredients (i.e., the "substance" of the drug), then drugmakers pay an additional $75,000 for the creation of a Drug Master File.

[53]For example, the agency established a "Priority" review program, which is designed to be completed within 6 months of submission.

potential entrants approximately one-eighth to one-quarter of the sunk cost of entering an average drug. Since the decision to enter depends on a comparison between discounted expected future profits and the sunk costs of entry, reducing the latter unambiguously makes entry more attractive. How many more entrants it attracts and the impact of those entrants on consumer welfare, however, are empirical questions. To answer them, we recompute equilibrium at the estimated parameters, using the midpoints of the identified sets to determine sunk costs.[54]

Columns 1-2 of Table VI report the result of this policy change. When entry costs are reduced by $400,000, an average of 42 out of 113 drug markets experience additional entry, and consumer compensating variation totals $143 million. When costs are reduced by twice that amount, an average of 59 out of 113 drug markets experience additional entry, and consumer compensating variation totals $374 million.

Table VI: *Counterfactual policy comparisons*

	Sunk costs reduced by:		Delays shortened by:	
	$400,000	$800,000	1 year	2 years
Mean sunk cost (millions of dollars)	3.91	3.51	4.31	4.31
Mean delay (years)	5.63	5.63	4.63	3.63
Drugs with additional entry (count)	42	59	41	64
Additional entrants (count)	50	115	45	95
Forgone agency revenue (millions of dollars)	60.8	121.6		
Consumer comp. variation (millions of dollars)	142.73	374.21	596.48	1520.02

Columns correspond to distinct counterfactual policy experiments. Columns 1-2 correspond to reducing the sunk entry costs; columns 3-4 report the result of shortening entry delays.

We then simulate the effect of shortening delays by one to two years. This policy change can benefit consumers two ways. First, quicker entry means fewer periods of cartel members charging monopoly prices. Nonmembers undercut the cartel in an effort to steal share, and members best-respond to the nonmembers prices by lowering their prices as well. Second, shorter delays means entrants can recoup their investment faster, making entry more attractive.

Columns 3-4 of Table VI report the result. When entry delays are reduced from an average of 5.63 to 4.63 years, there is additional entry in an average of 41 markets, and consumer compensating variation reaches $596 million. When entry delays are reduced by two years instead of one, there is entry in an average of 95 additional markets, and consumer compensating variation totals $1.52 billion.[55] Again, the

[54]Multiple equilibria are possible, so we require an equilibrium selection rule. In the data, entrants are more frequent cartel nonmembers than members. Thus, in two-thirds of our simulations, we let nonmembers decide if they would like to enter, and then we let members decide. In the remaining one-third of our simulations, we reverse the order. The resulting average ratio of members to nonmembers roughly corresponds to the data.

[55]The consumer compensating variation associated with shorter regulatory delays may seem especially large. Yet it is easy to arrive at similar figures using back-of-the-envelope calculations. Recall that an average of 1.9 million prescriptions are sold per drug per year at an average price of about $30 (see Table I). Also, recall that the cartel raises price by about 50% (see Figures I and II). Entry reduces the price increase by about half in half of the cartelized drug markets (see VII). In other words, observed entry generates a "flow" of surplus to consumers of approximately $7.125 million per year per cartelized drug (15*1.9*.5*.5), or $805 million in total. To a first approximation, shortening delays by a year provides one more year of this surplus. This occurs, on average, about 5 years out, so it's worth about $475 million in present value terms. Yet hastening entry increases the value of entry. In doing so, shorter delays attract

consumer welfare gains exceed reduced government fees by an order of magnitude. Put differently, the taxpayer would gladly finance the additional resources (e.g., additional staff) needed to expedite reviews and reduce industry fees.[56]

8 Conclusion

Stigler (1964) cites entry as one of two fundamental threats to collusive agreements. Yet, little to no prior work studies the extent and effect of entry in cartelized markets. In this paper, we investigate the issue in the context of the largest price fixing case in US history. An entrant who "breaks" the cartel—leading to Bertrand competition—will earn no more than they would entering the uncartelized industry. Therefore, to profitably enter, the entrant must successfully bid or price against the cartel. The latter will likely lead to a slow decline in prices.

Our reduced form results show that the cartel raises prices dramatically, but also attracts new potential entrants. The new entrants are concentrated in more profitable markets, implying that entry barriers limit the effect of entry. This is especially problematic because tacit collusion continues even after an antitrust investigation. Moreover, regulatory delays last several years, so even in markets where entrants do eventually undercut the collusive agreement, monopoly-like prices persist for long periods.

Turning to our structural estimates, we show that entry is not only slow but expensive. Our estimates suggest that the sunk cost of entry ranges between $1.1 million and $12.8 million per substance-delivery-release combination. While some of these costs cannot be avoided, policymakers can impact some of the entry costs and regulatory delays. We show that a reduction in either has a large impact on consumer surplus because we see entry in many more cartelized markets where prices remain high post-investigation. A reduction in time-to-market of a year and a 35% reduction in sunk costs have approximately the same impact. This provides guidance to policymakers considering optimal regulation of this important industry.

References

Barriers to entry, brand advertising, and generic entry in the us pharmaceutical industry. International Journal of Industrial Organization, 18(7):1085 – 1104, 2000.

R. Abelson and K. Thomas. Fed up with drug companies, hospitals decide to start their own. The New York Times, 18, 2018.

D. W. Andrews and G. Soares. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica, 78(1):119–157, 2010.

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

more entrants in about 43 markets. If their aggregate effect is like the impact of observed entry, reported in column 1, we can add $143 million. Then the total benefit to consumers is about $618 million—very close to the $596 million reported in the third column. Moreover, if we repeat this exercise but reduce delays by two years, then we arrive at roughly $1.32 billion, which is also near the $1.52 billion reported in the fourth column.

[56] Obviously, there are important caveats to this conclusion, including ensuring product quality. (Grennan and Town, 2020) study product testing requirements explicitly.

M. Backus, C. Conlon, and M. Sinkinson. Common ownership and competition in the ready-to-eat cereal industry. Technical report, National Bureau of Economic Research, 2021.

C. M. Baker-Smith, D. K. Benjamin Jr, H. G. Grabowski, E. D. Reid, B. Mangum, J. V. Goldsmith, M. D. Murphy, R. Edwards, E. L. Eisenstein, J. Sun, et al. The economic returns of pediatric clinical trials of antihypertensive drugs. American Heart Journal, 156(4):682–688, 2008.

S. T. Berry. Estimating discrete-choice models of product differentiation. The RAND Journal of Economics, pages 242–262, 1994.

S. T. Berry and P. A. Haile. Identification in differentiated products markets using market level data. Econometrica, 82(5):1749–1797, 2014.

S. Borenstein. Hubs and high fares: dominance and market power in the us airline industry. The RAND Journal of Economics, pages 344–365, 1989.

T. F. Bresnahan and P. C. Reiss. Entry and competition in concentrated markets. Journal of political economy, 99(5):977–1009, 1991.

D. P. Byrne and N. de Roos. Learning to coordinate: A study in retail gasoline. American Economic Review, 109(2):591–619, February 2019.

Cardinal Health. Manufacturer reference manual: For manufacturers to the pharmaceutical segment distribution supply chain. Technical report, 2020 2020. URL https://www.cardinalhealth.com/content/dam/corp/web/documents/Manual/cardinalhealth-pharma-supplier-guidebook.pdf.

V. Chernozhukov, H. Hong, and E. Tamer. Estimation and confidence regions for parameter sets in econometric models 1. Econometrica, 75(5):1243–1284, 2007.

R. Clark, C. A. Fabiilli, L. Lasio, et al. Collusion in the US generic drug industry. Technical report, 2021.

E. Cuddy. Competition and collusion in the generic drug market. Technical report, 2020.

L. Einav. A note on the analogies between empirical models of auctions and of differentiated product markets. Technical report, August 2003.

C. Fershtman and E. Muller. Capital investments and price agreements in semicollusive markets. The Rand Journal of Economics, pages 214–226, 1986.

C. Fershtman and A. Pakes. A dynamic oligopoly with collusion and price wars. The RAND Journal of Economics, 31(2):207–236, 2000.

S. Ganapati and R. McKibbin. Markups and fixed costs in generic and off-patent pharmaceutical markets. Technical report, National Bureau of Economic Research, 2021.

S. Gottlieb. How Obama's FDA Keeps Generic Drugs Off the Market. Missouri Medicine, 113(6):444–445, 2016.

E. J. Green and R. H. Porter. Noncooperative collusion under imperfect price information. Econometrica, 52(1):87–100, 1984.

M. Grennan and R. J. Town. Regulating innovation with uncertain quality: information, risk, and access in medical devices. American Economic Review, 110(1):120–61, 2020.

J. E. Harrington. Collusion among asymmetric firms: The case of different discount factors. International Journal of Industrial Organization, 7(2):289 – 307, 1989.

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

M. Igami and T. Sugaya. Measuring the incentive to collude: The vitamin cartels, 1990–1999. Working Paper, 2020.

J. Ishii. Compatibility, competition, and investment in network industries: ATM networks in the banking industry. Unpublished working paper, 2005.

M. J. Mazzeo. Product choice and oligopoly market structure. RAND Journal of Economics, pages 221–242, 2002.

N. Miller and G. Sheu. Quantitative methods for evaluating the unilateral effects of mergers. 2020.

N. H. Miller. Modeling the effects of mergers in procurement. International Journal of Industrial Organization, 37:201–208, 2014.

N. H. Miller and M. C. Weinberg. Understanding the Price Effects of the MillerCoors Joint Venture. Econometrica, 85(6):1763–1791, 2017.

A. Pakes, J. Porter, K. Ho, and J. Ishii. Moment inequalities and their application. Econometrica, 83(1): 315–334, 2015.

M. Pazniokas. Drug price-fixing lawsuit pushes CT probe into national spotlight. CT Post, May 2019.

R. H. Porter. A study of cartel stability: The joint executive committee, 1880-1886. The Bell Journal of Economics, 14(2):301–314, 1983.

J. L. Powell. Symmetrically trimmed least squares estimation for tobit models. Econometrica, 54(6): 1435–1460, 1986.

D. Rivers and Q. Vuong. Model selection tests for nonlinear dynamic models. The Econometrics Journal, 5 (1):1–39, 2002.

E. Rosenthal. Rapid price increases for some generic drugs catch users by surprise. New York Times, July 2014.

F. Scott-Morton. Entry and predation: British shipping cartels 1879–1929. Journal of Economics & Management Strategy, 6(4):679–724, 1997.

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

F. Scott-Morton and L. Boller. Enabling competition in pharmaceutical markets. Hutchins Center Working Paper, (30), 2017.

K. Seim. An empirical model of firm entry with endogenous product-type choices. The RAND Journal of Economics, 37(3):619–640, 2006.

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

L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 2020.

E. Tamer. Incomplete simultaneous discrete response model with multiple equilibria. The Review of Economic Studies, 70(1):147–165, 2003.

B. W. Wendling, D. R. Schmidt, M. Hertzendorf, L. M. Olson, H. Hadeishi, K. Cagatay, G. Kuczura, E. Connelly, J. Towey, K. A. Goldman, W. E. Cohen, K. A. Grimm, C. N. Bryan, and R. A. Nelson Jr. Authorized generic drugs: Short-term and long-term impact. Technical report, Federal Trade Commission, 2011.

T. G. Wollmann. Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. American Economic Review, 108(6):1364–1406, 2018.

9 [For Online Publication] Appendix: Proofs

THEOREM I.

Moments indexed by i and given by

$$\frac{1}{J}\sum_j \frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\frac{1}{2}\sum_{k\in\{M,N\}} h_j^i\Big[\widehat{VE}_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]<0 \tag{18}$$

produce consistent upper bounds.

PROOF.

For each moment indexed by i, we have

$$\begin{aligned}
&\frac{1}{J}\sum_j \frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\frac{1}{2}\sum_{k\in\{M,N\}} h_j^i\Big[\widehat{VE}_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]\\
&\quad=\frac{1}{J}\sum_j \frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\frac{1}{2}\sum_{k\in\{M,N\}} h_j^i\Big[VE_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]\\
&\quad=\frac{1}{J}\sum_j \frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\frac{1}{2}\sum_{k\in\{M,N\}} h_j^i\Big[VE_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})-\theta_{jk}+\eta_j\Big]\\
&\quad=\frac{1}{J}\sum_j \frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\frac{1}{2}\sum_{k\in\{M,N\}} h_j^i\Big[VE_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})-\mathscr{E}(\theta_{jk}|\mathscr{I}_{jk})-\nu_{jN}+\eta_j\Big]\\
&\quad<\frac{1}{J}\sum_j\Big[h_j^i\eta_j\Big]-\frac{1}{J}\sum_j\Big[h_j^i\nu_{jN}\Big]\xrightarrow{p}\mathbb{E}[h^i\eta]-\mathbb{E}[h^i\nu_{jN}]=\mathbb{E}[h^i]\mathbb{E}[\eta]-\mathbb{E}[h^i]\mathbb{E}[\nu_{jN}]=0.
\end{aligned} \tag{19}$$

The first equality results from replacing $\widehat{VE}_d^k(\cdot)$ with $VE_d^k(\cdot)$. $\widehat{VE}_d^k(\cdot)$ is a function of $\hat{\pi}_{d,t}^k(\cdot)$, $\hat{F}_\xi$, $\hat{F}_\omega$, and $\hat{F}_D$, and $VE_d^k(\cdot)$ is a function of $\pi_{d,t}^k(\cdot)$, F_ξ, F_ω, and F_D. Since $\pi_{d,t}^k(M,N,\xi_{dt},\omega_{dt})$, F_ξ, F_ω, and F_D are measured without error, and since the value entry depends on only those objects, entry values are measured without error. The second equality results from replacing $\theta_0+\theta_1 r_j+\theta_2\ell_j$ with $\theta_{jk}-\eta_j$, which follows directly from equation 13. The third equality results from replacing θ_{jk} with $\mathscr{E}[\theta_{jk}|\mathscr{I}_{jk}]+\nu_{jN}$, which follows directly from the definition of an expectational error. The inequality follows from the necessary conditions of a simultaneous move Nash equilibrium, which require $VE_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})-\mathscr{E}(\theta_{jk}|\mathscr{I}_{jk})<0$. (If this were condition were false, then another firm would have expected to profitably enter.) The next step follows from the law of large numbers. Since h^i depends on r and ℓ, η is independent of r and ℓ, ν is independent of r and ℓ, the fourth inequality holds. To arrive at the final inequality, notice that η and ν are both unconditionally mean zero (i.e., $\mathbb{E}[\eta]=0$ and $\mathbb{E}[\nu]=0$). ■

THEOREM II.

Moments indexed by i and given by

$$\frac{1}{J}\sum_{j\in L}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\mathbb{1}\{\chi^k\geq 1\}\widehat{VE}_d^k(\chi^M+\mathbb{1},\chi^N)}{\mathbb{1}\{\chi^M\geq 1\}+\mathbb{1}\{\chi^N\geq 1\}+\mathbb{1}\{\chi^M=0,\chi^N=0\}}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]$$
$$-\frac{1}{J}\sum_{j\in\Psi_{wVE}}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\widehat{VE}_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})}{2}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]\geq 0 \quad (20)$$

produce consistent lower bounds.

PROOF.

For each moment i, order j by their value of $w_j^i\eta_j$, let $L_{w\eta}$ denote the set of j that correspond to the smallest J values, and let $\Psi_{w\eta}$ correspond to the smallest $J-J_L$ values. For each moment indexed by i, we have

$$\begin{aligned}
&\frac{1}{J}\sum_{j\in L}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\mathbb{1}\{\chi^k\geq 1\}\widehat{VE}_d^k(\chi^M+\mathbb{1},\chi^N)}{\mathbb{1}\{\chi^M\geq 1\}+\mathbb{1}\{\chi^N\geq 1\}+\mathbb{1}\{\chi^M=0,\chi^N=0\}}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]\\
&\quad-\frac{1}{J}\sum_{j\in\Psi_{w\eta}}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\widehat{VE}_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})}{2}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]\\
&\geq\frac{1}{J}\sum_{j\in L}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\mathbb{1}\{\chi^k\geq 1\}\widehat{VE}_d^k(\chi^M+\mathbb{1},\chi^N)}{\mathbb{1}\{\chi^M\geq 1\}+\mathbb{1}\{\chi^N\geq 1\}+\mathbb{1}\{\chi^M=0,\chi^N=0\}}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]\\
&\quad-\frac{1}{J}\sum_{j\in\Psi_{w\eta}}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\widehat{VE}_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})}{2}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]\\
&=\frac{1}{J}\sum_{j\in L}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\mathbb{1}\{\chi^k\geq 1\}VE_d^k(\chi^M+\mathbb{1},\chi^N)}{\mathbb{1}\{\chi^M\geq 1\}+\mathbb{1}\{\chi^N\geq 1\}+\mathbb{1}\{\chi^M=0,\chi^N=0\}}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]\\
&\quad-\frac{1}{J}\sum_{j\in\Psi_{w\eta}}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{VE_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})}{2}-\theta_0-\theta_1 r_j-\theta_2\ell_j\Big]\\
&\quad=\frac{1}{J}\sum_{j\in L}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\mathbb{1}\{\chi^k\geq 1\}VE_d^k(\chi^M+\mathbb{1},\chi^N)}{\mathbb{1}\{\chi^M\geq 1\}+\mathbb{1}\{\chi^N\geq 1\}+\mathbb{1}\{\chi^M=0,\chi^N=0\}}-\theta_{jk}+\eta_j\Big]\\
&\quad-\frac{1}{J}\sum_{j\in\Psi_{w\eta}}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{VE_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})}{2}-\mathbb{E}(\theta_{jk}|\mathscr{I}_{jk})+\eta_j\Big]\\
&\quad=\frac{1}{J}\sum_{j\in L}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{\mathbb{1}\{\chi^k\geq 1\}VE_d^k(\chi^M+\mathbb{1},\chi^N)}{\mathbb{1}\{\chi^M\geq 1\}+\mathbb{1}\{\chi^N\geq 1\}+\mathbb{1}\{\chi^M=0,\chi^N=0\}}-\theta_{jk}+\eta_j-\nu_{jN}\Big]\\
&\quad-\frac{1}{J}\sum_{j\in\Psi_{w\eta}}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\frac{VE_d^k(\chi^M+\mathbb{1}\{k=M\},\chi^N+\mathbb{1}\{k=N\})}{2}-\mathbb{E}(\theta_{jk}|\mathscr{I}_{jk})+\eta_j-\nu_{jN}\Big]\\
&\quad\geq\frac{1}{J}\sum_{j\in L}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\eta_j-\nu_{jN}\Big]-\frac{1}{J}\sum_{j\in\Psi_{w\eta}}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w_j^i\Big[\eta_j-\nu_{jN}\Big]. \quad (21)
\end{aligned}$$

The first inequality follows from the construction of set Ψ_{wVE}. The first equality results from replacing $\widehat{VE}_d^k(\cdot)$ with $VE_d^k(\cdot)$. $\widehat{VE}_d^k(\cdot)$ is a function of $\hat{\pi}_{d,t}^k(\cdot)$, $\hat{F}_\xi$, $\hat{F}_\omega$, and $\hat{F}_D$, and $VE_d^k(\cdot)$ is a function of $\pi_{d,t}^k(\cdot)$, F_ξ, F_ω, and F_D. Since

$\pi^k_{d,t}(M, N, \xi_{dt}, \omega_{dt})$, F_ξ, F_ω, and F_D are measured without error, and since the value entry depends on only those objects, entry values are measured without error. The second equality results from replacing $\theta_0 + \theta_1 r_j + \theta_2 \ell_j$ with $\theta_{jk} - \eta_j$, which follows directly from equation 13. The third equality results from replacing θ_{jk} with $\mathscr{E}[\theta_{jk}|\mathscr{I}_{jk}] + \nu_{jN}$, which follows directly from the definition of an expectational error. The second inequality follows from the necessary conditions of a simultaneous move Nash equilibrium. That is, for $k \in \{M, N\}$, these conditions require $VE^k_d(\chi^M, \chi^N) - \mathscr{E}(\theta_{jk}|\mathscr{I}_{jk}) \geq 0$ as well as $VE^k_d(\chi^M + \mathbb{1}\{k = M\}, \chi^N + \mathbb{1}\{k = N\}) - \mathscr{E}(\theta_{jk}|\mathscr{I}_{jk}) < 0$.

Assume that $J_L/J \xrightarrow{p} q$ by the law of large numbers. We then have

$$
\begin{aligned}
&\frac{1}{J}\sum_{j\in L}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w^i_j\Big[\eta_j - \nu_{jN}\Big] - \frac{1}{J}\sum_{j\in\Psi_{w\eta}}\frac{1}{\mu_j}\sum_{d\in\mathscr{J}}\sum_{k\in\{M,N\}} w^i_j\Big[\eta_j - \nu_{jN}\Big] \\
&\qquad = \frac{1}{J}\sum_{j\in L}\Big[w^i_j\eta_j\Big] - \frac{1}{J}\sum_{j\in\Psi_{w\eta}}\Big[w^i_j\eta_j\Big] - \left(\frac{1}{J}\sum_{j\in L}\Big[w^i_j\nu_{jN}\Big] - \frac{1}{J}\sum_{j\in\Psi_{w\eta}}\Big[w^i_j\nu_{jN}\Big]\right) \\
&\qquad \geq \frac{1}{J}\sum_{j\in L_{w\eta}}\Big[w^i_j\eta_j\Big] - \frac{1}{J}\sum_{j\in\Psi_{w\eta}}\Big[w^i_j\eta_j\Big] - \left(\frac{1}{J}\sum_{j\in L}\Big[w^i_j\nu_{jN}\Big] - \frac{1}{J}\sum_{j\in\Psi_{w\eta}}\Big[w^i_j\nu_{jN}\Big]\right) \\
&\xrightarrow{p} \mathbb{E}[\eta|\eta<F^{-1}(q), w^i{=}1] - \mathbb{E}[\eta|\eta<F^{-1}(1-q), w^i{=}1] - (\mathbb{E}[\nu|\eta<F^{-1}(q), w^i{=}1] - \mathbb{E}[\nu|\eta<F^{-1}(1-q), w^i{=}1]) \\
&\qquad = \mathbb{E}[\eta|\eta<F^{-1}(q)] - \mathbb{E}[\eta|\eta<F^{-1}(1-q)] - (\mathbb{E}[\nu] - \mathbb{E}[\nu]) = 0. \qquad (22)
\end{aligned}
$$

The first equality follows from the fact that w^i and η_j, and ν_{jN} do not depend on d or k. The first inequality follows from the construction of $L_{w\eta}$. The third step follows from the law of large numbers. The second equality follows from the fact that η and ν are independent of r and ℓ, on which w^i depends. Thus, for example, $\mathbb{E}[\eta|\eta<F^{-1}(q), w^i{=}1] = \mathbb{E}[\eta|\eta<F^{-1}(q)]$. To arrive at the final step, notice that $\mathbb{E}[\eta|\eta<F^{-1}(q)]$ and $\mathbb{E}[\eta|\eta<F^{-1}(1-q)]$ are values that are equidistant from zero, so their difference is zero. Also, notice that ν is independent of η and is unconditionally mean zero. ■

10 [For Online Publication] Supplementary tables and figures

This section presents the following figures and tables.

- Figure IX assesses the suitability of Medicaid utilization data for demand estimation by way of comparisons to data we purchased from IQVIA, the "gold standard" for measuring prescription drug quantities.
  - Our sample reports the number of dispensed prescriptions nationally at the drug-month-year level, subject to two limitations.
  - First, it covers the first quarter of 2011 through the fourth quarter of 2017, inclusive. In other words, we do not span the period studied in the body of the main text (i.e., 2008-2019 inclusive). We were limited by cost as well as historical availability, so we focused on acquiring data around cartel formation.
  - Second, the data aggregates tablet and capsule purchases. In a small number of cases, substances are delivered in both forms, so we are forced to drop those drugs. The data does not distinguish between immediate and extended release versions of theophylline and etodolac. Again, we drop those drugs. Observing these distinctions would require more granular data, which was *much* more expensive, and since the omissions reflect random features of the sample, they will not affect the comparisons we derive them.
  - In terms of model predictions, the most influential feature of the quantity data is the mean change around cartel formation in cartelized markets relative to uncartelized ones. As a result, we plot log quantity in event time separately for two sources.
  - Panel A reflects IQVIA's National Prescription Audit, while Panel B reflects Medicaid utilization. Despite how differently the underlying observations are collected, the sources present very similar graphs. We observe (a) a very slightly positive pre-event trend one to three years prior to cartel formation, (b) no appreciable pre-event trend just prior to cartel formation, and (c) a clear decline in quantity thereafter, (d) culminating in a statistically significant decrease of 12-15%.
- Figure X addresses re-entry. In the body of the main text, we find that cartel formation attracts new entrants (i.e., ANDA filings). However, re-entry is also possible. Firms with dormant ANDAs—ones that were once associated with positive production but no longer are—might re-enter the market when cartels form. To study this possibility, we plot re-entry in event time for cartelized and uncartelized drugs. We find that cartel formation induces economically and statistically significant entry, and that re-entry occurs very soon after collusion begins. The stark contrast of this figure and the one that reports ANDA launches highlights the effect of approval delays.
- Figure XI shows that while there is no comprehensive correspondence between ANDA numbers and filing dates, the latter can be inferred without meaningful error.
  - To obtain filing dates, we downloaded all available approval letters from the FDA website and parsed out filing dates from the PDFs. Since 2000, the agency has issued three "waves" of ANDA numbers. Within each wave, numbers are assigned in chronological order. Specifically, the agency issued numbers in the 70,000s from 2000 to 2008, in the 90,000s from 2008 to 2010, and in the 200,000s thereafter.
  - In Panel A, we plot filing dates on the x-axis against ANDA numbers on the y-axis. The graph reflects 8,185 ANDAs for which we were able to obtain filing dates from parsed PDFs. The three waves are clearly visible. Putting aside a small number of very early ANDAs, which are filed years before our sample starts, there are only two obvious parsing errors.
  - Panel A shows that the sample period corresponds to the second and third wave of ANDA numbers. Thus, in Panels B and C, we isolate the waves separately. Panel B depicts ANDA numbers in the 90,000s, while Panel C depicts them in the 200,000s. In each panel, we plot ANDA numbers on the x-axis and filing dates on the y-axis. Both graphs illustrate the linearity of this relationship.

  - Given the aforementioned linearity, we regress ANDA numbers on filing dates within each wave and then predict filing dates for the remaining ANDA numbers. To assess overall fit, we compute the difference between actual and predicted dates measured in months and plot the density in Panel D. Substantially all of the parsed dates fall within 3 months of the predicted dates. Especially given that we aggregate ANDAs to the annual level, we conclude that ANDAs filing dates are measured accurately.

- Figure XIV evaluates the robustness of our result to an alternative event study specification. Sun and Abraham (2020) point out that when multiple cohorts are combined with heterogeneous treatment effects, leading and lagging coefficients are contaminated by effects from different periods. They propose estimating treatment effects that are cohort specific and then averaging these estimates using the cohort shares of the sample as weights. We take their approach and plot the resulting estimates in event time, which results in a figure that is nearly identical to Figure II.

- Table XIII reports price sensitivity by drug class. We plot $1/\hat{\alpha}_d$ on the x-axis against the drug class name on the y-axis. The graph shows that buyers in two classes in particular—other antiepileptics and β-blockers—are especially inelastic.

- Tables IX and IX assesses whether nonmembers are "folded into" the cartel.

  - In the body of the main text, we follow BCS to test conduct. In particular, we make pairwise comparisons between Model A1, which follows the assumptions we lay out in Section 5, and Model A2, where firms set Bertrand-Nash prices for all drugs in all years. We then introduce a third model, A3, which assumes that nonmembers also abide by collusive agreements, and make pairwise comparisons between A1 and A3. The details of the A1/A3 tests are provided here.
  - Models A1 and A3 are identical except for one difference. In A1, cartel members cooperate with one another, and nonmembers best respond. In A3, nonmembers join the cartel upon its formation. (This contradicts statements in the Complaint, but is nonetheless interesting to empirically evaluate.) Tests 3 and 4, summarized by Table VIII, replicate Tests 1 and 2 in the body of the main text, but replace A2 with A3. That is, in Test 3, we assess whether the data favors A1 over A3 or vice versa between t_d and 2015, and in Test 4, we make the same assessment between 2015 and the end of the sample.
  - Table IX reports the relevant results, which show that *the data strongly reject the model where nonmembers abide by the collusive agreement in favor of the model where nonmembers best respond.* Tests 3 and 4 yield test statistics of -3.78 and -4.34, which equate to very small p-values (<0.0001).

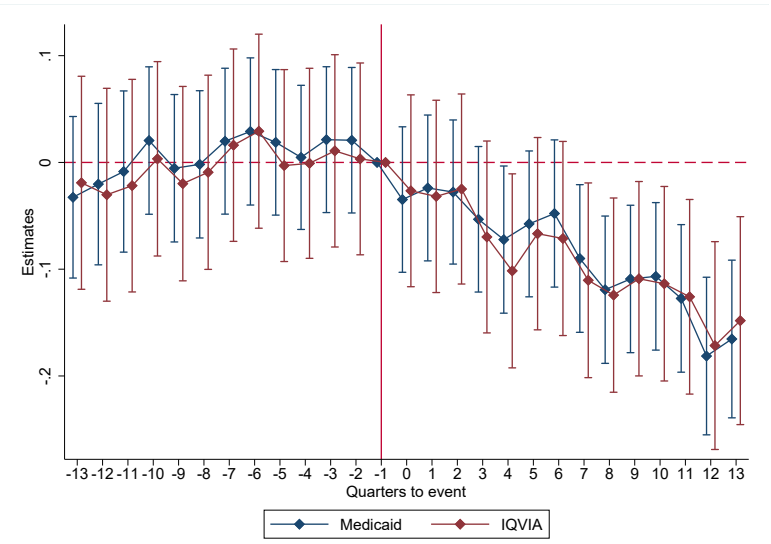

Figure IX: *The quantity responses reported in IQVIA and Medicare are very similar.*

This figure plots coefficients obtained by estimating equation 1 on the y-axis against event time on the x-axis. Log prescriptions are the outcome of interest, and the unit of observation is a drug-year. The vertical red line at event time -1 corresponds to the year just prior to cartel formation. Vertical bars around the point estimates show 95 percent confidence intervals, based upon standard errors that are clustered by drug. Notice that while the quantity decline in the Medicare data following cartel formation lags the one evidenced by IQVIA data, the delayed response does not have a meaningful effect on our results; when we estimate demand, we omit observations from the year in which each cartel is formed.

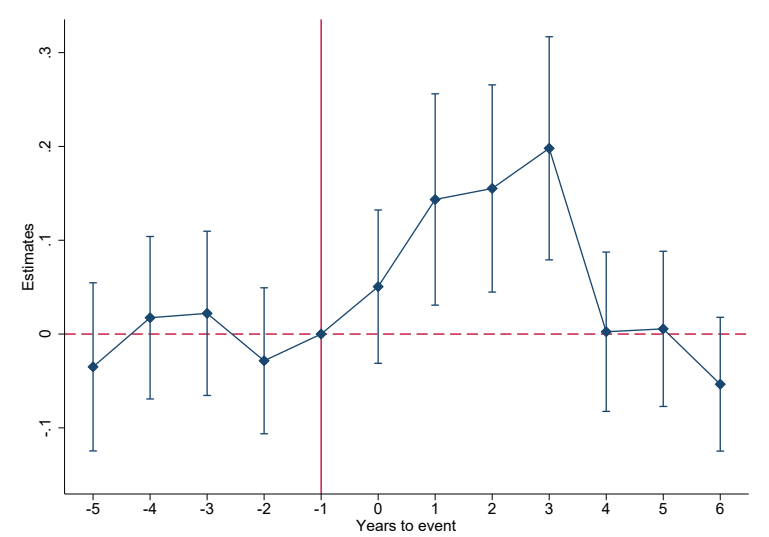

Figure X: *The cartel induces significant re-entry.*

This figure plots coefficients obtained by estimating equation 1 on the y-axis against event time on the x-axis. The unit of observation is a drug-year, and the outcome of interest is a ANDA re-entry. Re-entering ANDAs are those where the ANDA was associated with some output, then went at least one year without being associated with output, and then re-entered (i.e., was once again associated with output). The vertical red line at event time zero corresponds to the year in which Teva hired NP. Vertical bars around the point estimates show 95 percent confidence intervals for those coefficients, based upon standard errors that are clustered by drug.

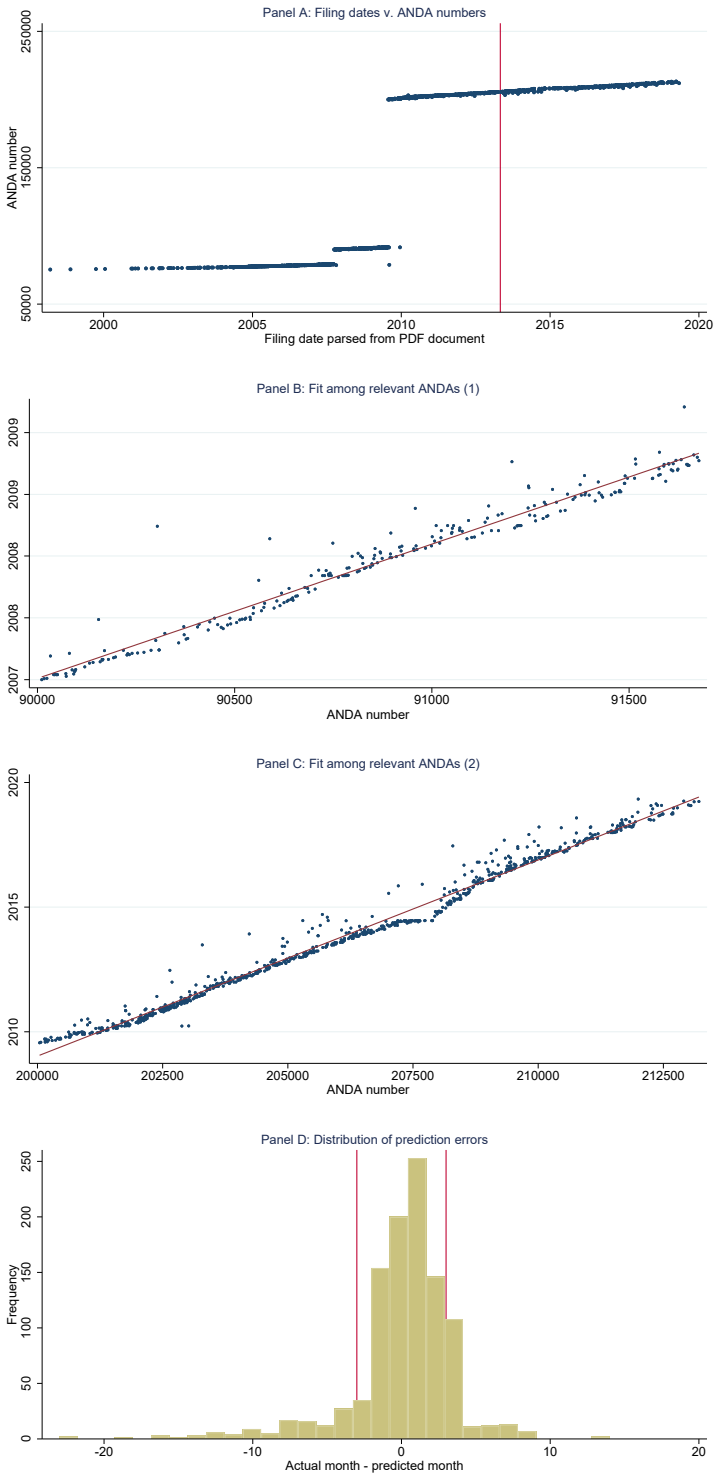

Figure XI: *Filing dates are inferred without meaningful error.*

Panel A plots parsed ANDA filing dates on the x-axis and ANDA numbers on the y-axis. Panels B and C plot ANDA numbers on the x-axis and parsed filing dates on the y-axis for the relevant "waves" of ANDAs. Panel D plots the difference between the actual and predicted filing months.

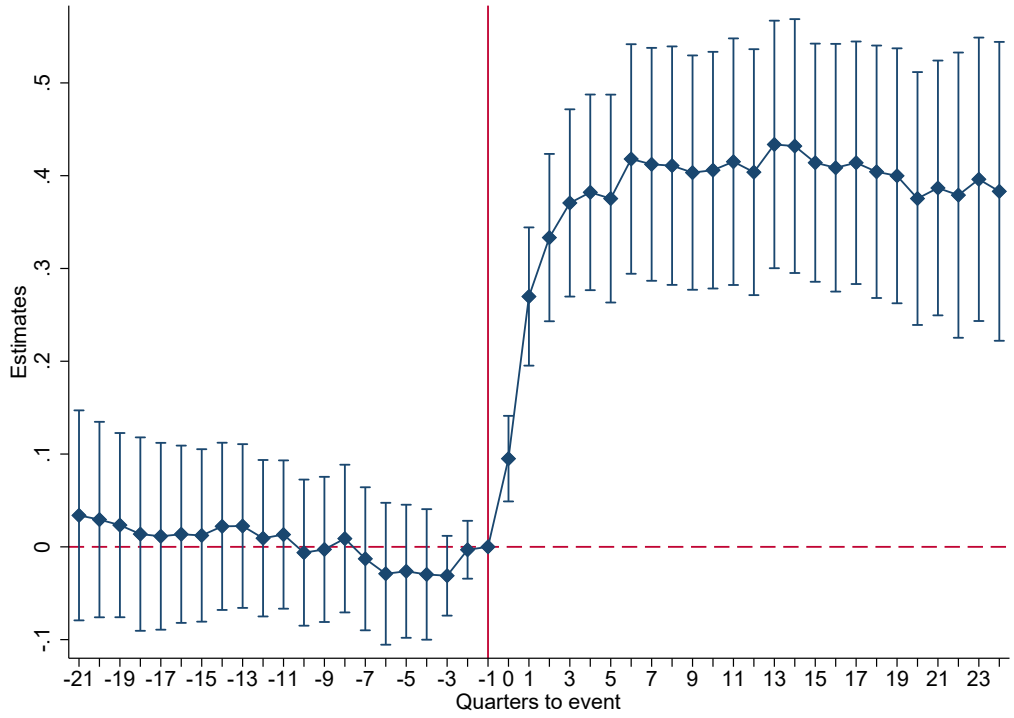

Figure XII: *Cartels are formed where NP has strong relationships.*

This figure plots β^{τ}, which is obtained by estimating equation 1, on the y-axis against event time on the x-axis. Unlike prior graphs, x_{kt}^{τ} is an indicator variable that equals one if and only if k is a drug for which NP's average quality score meets or exceeds 2 (i.e., the sales and marketing executives of the other firms selling the drug have strong relationships with NP). The unit of observation is a drug-year-firm. Log prices are the outcome of interest, and the unit of observation is a product (i.e., drug-year-firm). The vertical red line at event time -1 corresponds to the year immediately prior to cartel formation. Vertical bars around the point estimates show 95 percent confidence intervals for those coefficients, based upon standard errors that are clustered by drug.

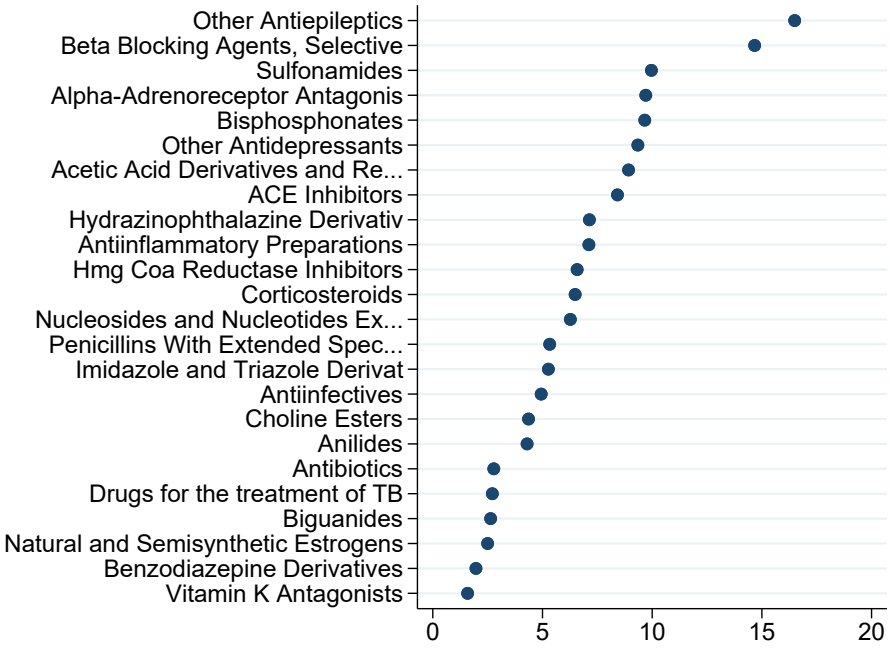

Figure XIII: *Patients consuming antiepileptics and β-blockers are inelastic.*

This figure plots inverse price coefficients (i.e., $1/\alpha$) on the x-axis against corresponding drug classes on the y-axis. Buyers of two classes of drugs—other antiepileptics and β-blockers—are especially inelastic, so we incorporate this heterogeneity in the demand system. See Section 6 for more details.

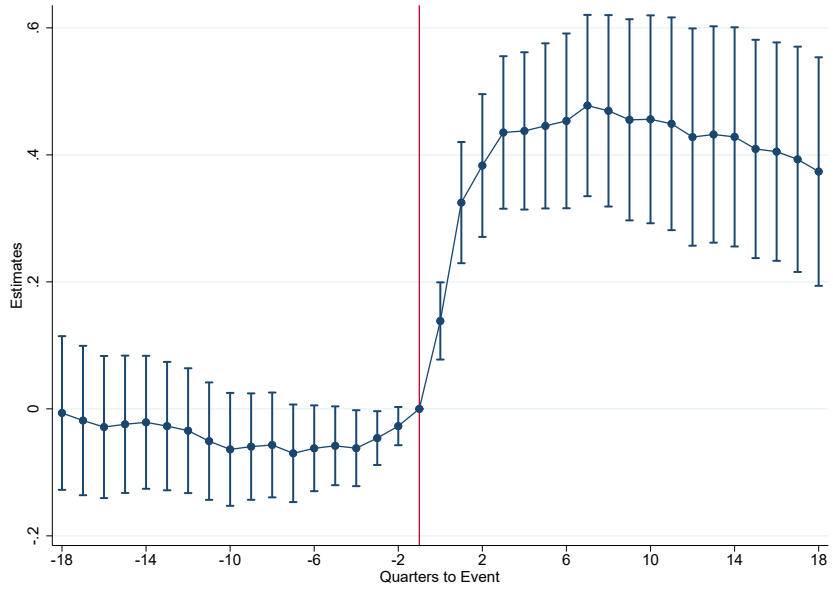

Figure XIV: *Prices in event time using the approach of Sun and Abraham (2020)*

This figure replicates Figure II but follows Sun and Abraham (2020), who propose weighting cohort-specific estimates in event study research designs. That is, observations are aggregated by relative quarter, using sample shares of each cohort as weights.

Table VII: *Summary of cartelized and uncartelized markets from 2008 to 2012*

	Cartelized	Uncartelized	Total
Price	27.36	33.86	32.10
	(27.31)	(35.51)	(33.60)
Quantity (in thousands)	867.9	1572.7	1381.3
	(1382.3)	(2851.5)	(2557.0)
Expenditure (in millions)	16.06	26.96	24.00
	(29.47)	(52.85)	(47.88)
Number of firms	3.732	4.362	4.191
	(1.625)	(2.216)	(2.091)
Number of ANDA filings	0.179	0.288	0.258
	(0.467)	(0.639)	(0.599)

The unit of observation is a drug-quarter. Price refers to dollars per prescription. Quantity equals Medicare market share multiplied by market size measured in prescriptions. Expenditures equal the product of price and quantity. Variable means and standard deviations are reported, with the latter in parentheses.

Table VIII: *Behavioral assumptions for supplementary conduct tests*

Test 3.				
Markets:	Drugs named in the Complaint		Other drugs in the sample	
Model:	A1	A3	A1	A3
2008—(t_d-1)	Own-profit max.	Own-profit max.	Own-profit max.	Own-profit max.
t_d—2015	Collusion	All cooperate	Own-profit max.	Own-profit max.
2016—2018	–	–	–	–
Test 4.				
Markets:	Drugs named in the Complaint		Other drugs in the sample	
Model:	A1	A3	A1	A3
2008—(t_d-1)	Own-profit max.	Own-profit max.	Own-profit max.	Own-profit max.
t_d—2015	–	–	–	–
2016—2018	Collusion	All cooperate	Own-profit max.	Own-profit max.

"Collusion" corresponds to cartel members setting price equal to p^B and nonmembers setting it equal to $p^{C,N}$. "All join cartel" corresponds to all firms cooperating, meaning nonmembers abide by collusive agreement, which results in monopoly prices.

Table IX: *Result of pairwise comparisons of A1 and A3*

	$\tilde{Q}(mk^{A1})$ $\times$ 100	$\tilde{Q}(mk^{A3})$ $\times$ 100	Test statistic	p-value
Test 3. (Collusion pre-investigation?)	.01384	.01707	-3.78	< .0001
Test 4. (Collusion post-investigation?)	.01180	.01545	-4.34	< .0001

This table reports the result of the testing procedure proposed by Backus et al. (2021). The test statistic is distributed standard normal. The standard error of the difference between $\tilde{Q}(mk^{A1})$ and $\tilde{Q}(mk^{A3})$ is obtained via bootstrapping. See Table VIII for the definitions of A1 and A3 under Tests 1 and 3.