IN DEFENSE OF THE MIDDLEMAN: QUALITY FAILURES IN THE GENERIC PHARMACEUTICAL MARKET

CATHERINE ISHITANI
Yale University

This paper studies how intermediary buyers discipline quality in the generic pharmaceutical market. Consumers are uninformed about generic quality and implicitly rely on intermediaries, such as wholesalers and pharmacies, to select quality for them. Using the universe of FDA manufacturing quality disclosures from 2000-2022, I document pervasive quality failures—showing that generics are not interchangeable, as is conventionally assumed. Disclosing quality failures through recalls reduces purchases of low-quality drugs by 60% for a decade. To isolate the role of intermediaries, I develop a scoring auction model of generic procurement that is identified with incomplete price data. I find that intermediaries are willing to pay a 2% premium for every 10% reduction in recalls. Their preferences, rather than manufacturer responses, drive the benefits of disclosure and increase equilibrium quality by 34%. Modest subsidies can improve the alignment between intermediary and social objectives, yielding greater welfare gains than interventions that limit the supply of low-quality drugs.

KEYWORDS: Healthcare, quality regulation, disclosure, intermediation.

1. INTRODUCTION

Quality disclosure is supposed to improve choice and welfare. However, its impact is often limited by consumer responses (Akerlof, 1970). Expansive evidence shows that consumers frequently overlook or ignore disclosures (Ben-Shahar and Schneider, 2017). Instead, in many markets they rely on intermediary buyers to demand quality on their behalf. These firms are pervasive in settings with severe information asymmetry, from healthcare to

Catherine Ishitani: catherine.ishitani@yale.edu

I thank Alex Olssen, Guy David, and Abby Alpert for their invaluable guidance. I also thank Rena Conti, Zack Cooper, Jacob Dorn, Atul Gupta, Phil Haile, Yunan Ji, Elena Krasnokutskaya, Eric Richert, John Romley, Fiona Scott Morton, Natalia Serna, Maria Polyakova, Yihao Yuan, and seminar participants at several institutions for helpful suggestions. I gratefully acknowledge financial support from the Mack and Leonard David Institutes.

household goods. For example, the typical consumer trusts that her pharmacy selects legitimate COVID tests and that her grocery stocks safety-certified produce. This paper studies how informed intermediaries discipline quality in the market for generic prescription drugs.

The theoretical benefits of informed intermediaries are straightforward. Relative to consumers, intermediaries are often better positioned to observe and react to quality disclosures, given their scale and experience. Downstream, intermediaries can alleviate adverse selection and improve the quality of goods available to consumers. Upstream, they can incentivize suppliers to invest in higher quality. Intermediaries, however, may not fully internalize these social benefits, particularly when they lack reputational or legal consequences. This creates the conditions for a conflict of interest, in which intermediaries select goods that benefit themselves rather than consumers. In this context, intermediaries can enhance *or* distort the benefits of disclosure, with ambiguous implications for equilibrium welfare.

The generic pharmaceutical market is distinguished by an unusual degree of asymmetric information and intermediation, making it an ideal setting to study informed intermediaries.[1] In contrast to branded drugs, generics are conventionally described as almost perfectly substitutable "commodities" (FTC, 2011). Patients and physicians typically cannot differentiate between generics made by alternative manufacturers, much less between differences in their quality. Instead, they implicitly trust intermediaries, such as wholesalers and pharmacies, to select high-quality drugs for them. These firms make purchasing decisions that determine the quality and wholesale prices of nearly all generics in the U.S. market. For their part, intermediaries promise to deliver safe and effective drugs, with slogans like, "It's not just a package, it's a patient" (McKesson, 2024).

Most patients perceive generics to be tightly regulated, commensurate with their high stakes for health. Although generics must meet a minimum clinical standard at launch, their *manufacturing quality* can deteriorate in the years and decades after. This type of quality reflects a drug's purity, strength, and ability to perform its therapeutic purpose—capturing, for example, if it is contaminated with bacteria or has insufficient active ingredients.[2] The

[1] A generic drug is a copy of a branded drug whose primary patents have expired; two-thirds of Americans consume generics each year (Cohen and Mykyta, 2023). A *drug* is a molecule-formulation made by a specific manufacturer (e.g., atorvastatin-tablet-Teva Pharmaceuticals).

[2] These are representative examples of quality failures, which are summarized in Appendix Table V.

Food and Drug Administration (FDA) primarily maintains manufacturing quality through recall and inspection failure announcements. These disclose that a drug poses a "reasonable probability" of serious harm, but do not permanently remove it from the market (FDA, 2024b). Recalls also require manufacturers to retrieve defective batches, which can temporarily disrupt the drug supply.

I begin by providing evidence of pervasive quality failures in the U.S. generics market. Next, I show that disclosing these failures sharply reduces purchases of low-quality drugs. To isolate the role of intermediaries, I develop a structural model of generic procurement. I show that intermediaries enhance the effects of disclosure and substantially improve quality. Finally, I evaluate leading proposals to improve drug quality and supply reliability.

First, I construct a new dataset that provides the most comprehensive information on drug manufacturing quality to date. The data include the universe of FDA disclosure actions from 2000-2022, obtained through Freedom of Information Act (FOIA) requests. I use drug labels, import manifests, and administrative records to identify drug manufacturing locations—historically considered trade secrets—and to link drugs to quality disclosures (Conti et al., 2020). These data reveal that quality failures are pervasive and persistent: 57% of facilities fail inspections, and 12% of drugs are recalled during the study period, demonstrating that failures are not limited to a few bad actors. Moreover, a failing drug is much more likely to fail again (relative to the average), suggesting that quality resembles a persistent type. Taken together, these results contradict the conventional wisdom: generic drugs are not perfect substitutes.

Next, I explore whether disclosing quality failures improves equilibrium outcomes. This is not obvious *a priori*. Since patients face strong barriers to using disclosures, the FDA's policy could be expected to have limited impact. I quantify the causal effects of disclosure using commercial claims data and a generalized difference-in-differences design that exploits variation in disclosure timing. I find that recall announcements reduce intermediary purchases of low-quality, recalled drugs by 60%. This effect persists for at least ten years—far after recalls are typically resolved—suggesting it is driven by long-run changes in intermediary preferences or manufacturer costs, rather than short-run supply disruptions. In contrast, disclosing even the most serious inspection failures has little effect on intermediary purchases. It thus appears that manufacturers are penalized more stringently for poor supply reliability than for low quality alone. Finally, I study other determinants of

patient welfare, including drug prices and accessibility. While recalls have limited impact on prices, they increase the probability of a national drug shortage by 33% and reduce patient medication adherence by 5%. Thus, even as recalls improve quality, they also reduce patient welfare by disrupting the drug supply.

The quasi-experimental evidence demonstrates that disclosure improves drug quality, but it cannot disentangle supply- and demand-side mechanisms. For example, disclosure might alter manufacturer *costs* and *prices*, as they reoptimize capacity and positioning. On the demand-side, it might alter *intermediary preferences*, as they update expectations about supply reliability. To separate these mechanisms, I develop a structural model of intermediary purchases through scoring auctions. The model is based on intermediary reference manuals, which specify their criteria for scoring potential suppliers, including historical recalls and fill rates. I capture this process using a scoring auction framework, in which buyers weigh both price and quality, and estimate the model using data from 78,000 auctions.

My estimation approach extends a recent method from Laffont et al. (2020) and introduces a new technique that non-parametrically identifies the model primitives when only winning prices are observed. While scoring auctions are common in practice, current estimation methods require complete data on prices, which is infeasible in most empirical applications. Identification with incomplete prices is non-standard because the model incorporates two sources of asymmetry: bidder characteristics and unobserved buyer preferences. My approach draws on the empirical auction literature to identify manufacturer costs (Guerre et al., 2000), and then imposes standard restrictions on bidder cost distributions to identify intermediary preferences (Athey and Haile, 2002). I provide empirical support for these restrictions by exploiting variation in recall timing and showing balance in uncontrolled drug characteristics. Using Monte Carlo simulations, I show that my algorithm recovers unbiased estimates of the model primitives. Finally, I generalize my approach to other scoring auction formats, including linear, optimal, and quality-to-price ratio scores.

The estimated model shows that intermediary responses are central to the equilibrium effects of disclosure. Intermediaries prefer high-quality drugs and are willing to pay a 13% premium for them (roughly 2% more for each 10% reduction in recall risk). Their preferences account for over 100% of the quality improvements observed after recalls and increase the share of high-quality drugs by 8%. Manufacturer responses offset half of these gains, lowering quality by 4 percentage points (pp). On net, disclosure benefits patients but

reduces producer surplus by 9%, as losses for recalled manufacturers more than offset gains for their competitors and intermediaries. These losses help explain why—counter to economic theory—manufacturers do not voluntarily disclose their quality types, and regulation is required in the first place (Grossman, 1981, Milgrom, 1981).

Next, I quantify the equilibrium role of intermediaries. Simulating a counterfactual environment without quality preferences shows that intermediaries increase the share of high-quality drugs by 34% and producer surplus by 16%. They have modest effects on prices, which increase by $0.12, or less than one cent at the median. Overall, intermediaries likely improve short-run quality and welfare, and thus partially correct the market failures from patient asymmetric information. By encouraging competition on quality, and not just price, they may also strengthen manufacturer incentives to invest in quality in the long-run.

In my final counterfactuals, I study policies that seek to better align intermediary and social incentives, or to alter the composition of manufacturers in the U.S. market. During a period of historic pandemics and drug shortages, the federal government proposed four such policies: expanding quality disclosures, subsidizing high-quality drugs, raising minimum quality standards, and reshoring drug manufacturing to the U.S. (HHS, 2024, GAO, 2024, The White House, 2025). The quasi-experimental evidence shows that intermediaries already strongly respond to quality signals, suggesting that more salient disclosures may yield limited gains. In contrast, counterfactual simulations show that modest purchase subsidies could increase consumer welfare, by reinforcing intermediary incentives to select high-quality drugs. Improving quality, however, is increasingly expensive, and reaching an 80% high-quality drug supply would require price increases of more than 90%. In the long-run, large subsidies are also likely to drive low-quality manufacturers to exit, heightening supply risks for drugs already prone to shortage.

Higher quality standards could approximately double equilibrium drug quality, but at the cost of nearly 3.7 times greater prices. Reshoring production produces the least efficient quality gains, as it increases prices by a factor of 2.7 in return for only 50% higher quality. This reflects more expensive domestic production, sharply reduced competition, and poor targeting, as baseline U.S. quality only slightly exceeds the rest of the world's. Finally, to consider the welfare consequences of each proposal, I apply laboratory evidence on the value of quality from the FDA. Recent recalled drugs increased the expected incidence of cancer by 1/8000, implying that higher quality may be worth at least 20% more

at standard values of Quality-Adjusted Life Years (QALYs). Taken together, this suggests that moderate subsidies could improve welfare, and are likely to yield larger gains than proposals that limit the supply of low-quality drugs.

Although this paper's focus on drug manufacturing quality is new, research on quality disclosure is well-established (Akerlof, 1970). Prior work documents its effects in settings such as restaurants, packaged foods, toys, and insurance (Dranove and Jin, 2010, Freedman et al., 2012, Barahona et al., 2023, Vatter, 2022). This work finds that disclosure can improve consumer choice and quality, but its impact is often attenuated by frictions in consumer responses. Motivated by this finding, I study how *wholesalers and pharmacies* respond to quality information, connecting evidence on disclosure to research on expert intermediaries (Grennan et al., 2024, Gavazza and Lizzeri, 2021). I show that intermediaries can partially resolve adverse selection problems between buyers and sellers and substantially enhance the equilibrium effects of quality disclosure.

My paper complements a large literature on pharmaceutical quality, providing the first economic evidence on quality *after approval* and its welfare implications (Peltzman, 1973, Atal et al., 2022). It also adds to prior work on generics, which has addressed shortages (Yurukoglu et al., 2017, Dubois et al., 2023, Galdin, 2024), pricing (Scott Morton, 2000, Ganapati and McKibbin, 2023), and collusion (Cuddy, 2020, Starc and Wollmann, 2022). I build on this work by introducing data on manufacturing quality, which I show has first order implications for supply reliability and competition. Finally, this paper contributes methodologically to work on empirical scoring auctions. This literature has largely focused on settings with complete bid data and known scoring rules (Marra et al., 2024, Allen et al., 2024). Krasnokutskaya et al. (2020) is most similar, in that the authors recover an unknown scoring rule using incomplete data (on quality, but complete prices). My paper extends the canonical estimation methods of Guerre et al. (2000) to a broader set of empirical applications with incomplete price and scoring data.

The paper proceeds as follows. Section 2 describes the empirical setting and data. Section 3 presents the quasi-experimental evidence, and Section 4 introduces the structural model. Section 5 discusses the counterfactual analyses, and Section 6 concludes.

2. INSTITUTIONAL SETTING AND DATA

2.1. *Generic Pharmaceuticals and Quality Disclosure*

Generic interchangeability underpins the basic organization of the U.S. pharmaceutical market, where it generates intense price competition. As a result, U.S. generics are priced 80% lower than their branded counterparts and capture 90% of prescription volumes (Cohen and Mykyta, 2023). However, growing evidence suggests that generic drugs may not be perfect substitutes. Generic quality failures lead to approximately 1,000 recalled or banned drugs each year and are the "foremost" cause of drug shortages (Woodcock and Wosinska, 2013, Jewett, 2023). For example, the FDA's discovery of a "cascade of failure" in manufacturing at Intas Pharmaceuticals caused widespread shortages of anti-epileptics and chemotherapies, delaying the care of up to 10% of U.S. cancer patients in 2023 (FDA, 2023, American Cancer Society Cancer Action Network, 2023). While regulators describe pharmaceutical supply as a matter of "national security," little research has examined the determinants and welfare implications of manufacturing quality, key inputs to evaluating policy proposals (The White House, 2024).

Internal FDA data indicate that quality failures are pervasive and persistent. Over 2000-2022, 57% of manufacturing facilities failed inspections, and 12% of drugs were recalled for health risks.[3] Quality failure rates have also broadly increased over time (Figure 2.1 panel (a)). The distributions of failures show that they affect the majority of facilities, manufacturers, and countries of origin, and thus cannot be explained by a few bad actors (panels (b)–(c)). Nor can failures be explained by random shocks: a recalled drug remains 3.3 times more likely to be recalled for the next decade (relative to the average), suggesting that drugs are characterized by persistent quality types (Figure 6). Together, these patterns align with the FDA's description of slowly deteriorating generic manufacturing capital, driven by continuous use and minimal upgrades over decades (Woodcock and Wosinska, 2013).

After a drug's approval, the FDA primarily regulates quality through public disclosures, such as inspection reports and recall announcements. Inspection reports document violations of good manufacturing practices observed during routine inspections, while recalls

[3] A full 28% of facilities received the worst inspection grade (OAI) over the same period. Complete results using only these grades are presented in the Appendix.

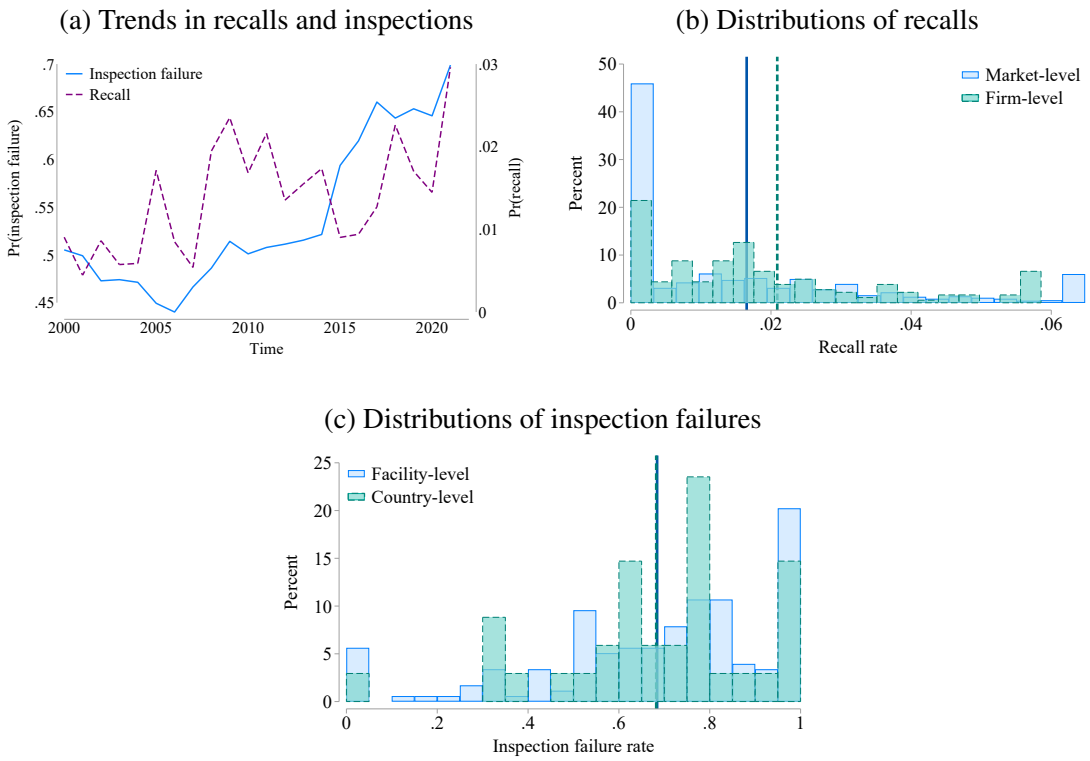


FIGURE 1.—Trends and variation in quality failures

Note: Panel (a) presents annual drug recall rates and facility inspection failure rates over time. Panels (b) and (c) present distributions of recall and inspection failure rates averaged over drug-years at the manufacturer, facility, market, and country levels. Distributions are winsorized at the 95th percentile, and vertical lines display mean rates.

must be initiated when a drug fails to meet the safety standards of the Food, Drug, and Cosmetic Act. I measure quality using the most serious types of disclosures: VAI and OAI inspection grades and class I and II recalls, which apply when a drug poses a "reasonable probability" of causing serious injury or death (FDA, 2024b). Unlike inspection failures, recalls also require firms to remove defective drug batches, which often causes temporary stockouts. This creates a second channel—information about supply reliability—through which recalls may affect market outcomes.

The impact of the FDA's disclosure policies depends on the extent to which they increase demand for quality. However, patient awareness of quality is very low, for multiple reasons. Disclosures often fail to reach patients or omit actionable information, such as the drug's name (GAO, 2011). Most importantly, U.S. generics are poorly differentiated: they are typically unbranded and sold without identifying the exact product (i.e., the manufacturer).

As a result, even if a patient should observe a disclosure, she would have almost no way to apply that information to her drug purchase.

In contrast, intermediaries have strong incentives and latitude to monitor quality disclosures. Intermediaries purchase drugs from manufacturers to sell to patients, with the five largest purchasers capturing 95% of U.S. generic volumes (Fein, 2023). They procure generics through auctions, awarding supply contracts to a single manufacturer per market (Cuddy, 2020, State of Connecticut, 2019). When selecting manufacturers, intermediaries weigh both price and quality, as quality failures impose substantial costs on them. Shortages reduce drug sales, and recalls require inventory tracking, with estimated labor costs of approximately $230,000 per store annually (Tracelink, 2023). Consistent with this, intermediaries emphasize that "lapses in supply ruin relationship[s]," and publicly specify their quality criteria for "scoring" manufacturers (e.g., Figure 7) (FTC, 2011, AmerisourceBergen, 2019, CardinalHealth, 2022).[4]

Since patients do not observe quality, intermediary choices determine the equilibrium quality and wholesale prices of U.S. generics. In theory, this gives intermediaries significant scope to improve quality and reduce welfare losses from asymmetric information. Intermediaries, however, face a classic agency problem and may choose socially suboptimal levels of quality, maximizing profits rather than total welfare (Gavazza and Lizzeri, 2021). This agency problem is particularly acute in the generics market, where intermediaries (unlike physicians or financial advisors) face no legal responsibility to act in patients' interests, and where patients cannot monitor quality choices. In this context, whether intermediaries improve quality, and how close welfare is to first-best, are empirical questions.

2.2. *Data*

This paper combines three main data sources. The first is data on drug quality, including the universe of FDA quality disclosures and enforcement actions from 2000-2022. Inspection and enforcement records were obtained through FOIA requests, while recall data are from archived FDA Enforcement Reports. Linking drugs to quality disclosures typically requires information on drug manufacturing locations, which is not releasable under FOIA

[4]Auction winners change rapidly—in certain markets, with every batch—and intermediaries do not release losing bids, implying limited a role for inertia or common information.

TABLE I

DESCRIPTIVE STATISTICS

	(1) All drugs	(2) Recalled drugs	(3) Never recalled drugs
Drug characteristics			
Years since first generic entry	18.46	18.90	18.37
Fraction unit type:			
Oral pill	0.74	0.71	0.75
Injection	0.07	0.13	0.07
Topical & other	0.19	0.17	0.19
Extended release	0.10	0.10	0.10
Market shares & prices			
Share of pharmacies	0.16	0.19	0.16
	(0.22)	(0.23)	(0.22)
Mean annual prescriptions	21,056	56,668	18,998
	(80,379)	(168,562)	(72,604)
Mean price per unit	0.97	0.94	0.97
Median price per unit	0.32	0.29	0.32
	(1.53)	(1.59)	(1.53)
Mean drugs per market	11.55	12.44	11.51
N drugs	11,045	1,158	9,887

Note: The table presents descriptive statistics for multisource generic drugs in Optum claims data. Column (1) shows statistics for all drugs in the sample, col. (2) shows recalled drugs, and col. (3) shows drugs that were not recalled during 2000-2022. Statistics are drug-level means over all years (cols. (1) and (3)) or prior to the first recall (col. (2)). Corresponding standard deviations are reported in parentheses for key market share and price outcomes. Drug characteristics are from the FDA NDC Directory and Orange Book, and years since generic entry is as of 2022. Shares of pharmacies, annual volumes, and counts of drugs per market are observed in claims data. Prices are pharmacy acquisition costs in 2022 dollars from Medicaid surveys.

(Conti et al., 2020). To overcome this, I use drug labels, trade manifests, and OECD administrative records to recover manufacturing locations.[5] For recalls, I observe announcement dates, classes, reasons, and National Drug Code (NDC) identifiers. For inspections, I observe dates, grades, citations, and whether the inspection was routine or "for cause."

The second source is data on drug prices, which capture weekly average acquisition costs for retail pharmacies over 2010-2022 (the "AAC" or "NADAC"). These data are from

[5]This method is described in the Appendix and was first validated by CIDRAP for 156 drugs (Center for Infectious Disease Research & Policy, 2018). Inspections and recalls represent the vast majority of FDA actions, which also include regulatory meetings, import alerts, injunctions, etc.

surveys conducted by fourteen state and federal Medicaid agencies and were collected through Open Records Requests and the Medicaid.gov website. To recover intermediary-specific winning bids, I exploit variation in pharmacy chain presence and average prices across states.

The third source is administrative claims data from UnitedHealth, the largest private U.S. health insurer. The data cover nearly all U.S. retail pharmacies, as well as over 30 million individuals with pharmacy benefits per year from 2000-2022, including both employer-sponsored and Medicare Advantage enrollees. I use these data to identify intermediary auctions and to measure patient medication adherence. I identify auctions following Cuddy (2020), observing winning drugs stocked by each pharmacy, and auction timing from changes in manufacturer participation and drug composition across pharmacies. Manufacturers typically participate in each auction within a round, allowing me to identify the number of bidders from the manufacturers observed in each market.[6] Supply contract volumes are estimated from annual prescription counts per pharmacy. To measure adherence, I use the Medication Possession Ratio (MPR), defined as the ratio of days filled (by any manufacturer) to days prescribed. I also use supplemental data on drug, pharmacy, and shortage characteristics from the FDA NDC Directory and Orange Book, the National Council for Prescription Drug Programs, and the University of Utah Drug Information Service (UUDIS), respectively.

My analysis sample consists of generic prescription drugs purchased by retail pharmacies over 2000-2022. I focus on unbranded, oligopoly drugs, which account for 84% of generic volume, and first recall and inspection announcements over 2001-2021 (1,158 and 6,633 events, respectively).[7] In total, I study the quality characteristics of 11,045 drugs across 1,209 molecule-form markets, and the purchasing responses of approximately 64,000 pharmacies and 63 million patients.

Table I summarizes the main characteristics of my sample. Column (1) reports values for all drugs, while columns (2) and (3) summarize drugs that were recalled and never-recalled during the study period, respectively. Each value represents the average over the

[6]My sample excludes collusive auctions, observed in unsealed court records (State of Connecticut, 2019).

[7]I focus on surveillance inspections, which are conducted quasi-randomly conditional on facility risk (and not for cause), to avoid contamination by other disclosure actions.

full sample period (cols. (1) and (3)) or prior to the first recall (2). The typical drug is old and inexpensive: its first generics were approved 18 years earlier, and it costs just $0.97 per unit (e.g., pill). Columns (2) and (3) show that recalled drugs differ systematically from never-recalled ones, with larger volumes (56,668 vs. 18,998 prescriptions) and more competitive markets (with 12.4 vs. 11.5 participants). These differences motivate my use of not-yet-recalled drugs—rather than never-recalled ones—as a more appropriate control group for estimating the effects of recall disclosure.[8]

3. EVIDENCE ON THE EFFECTS OF QUALITY DISCLOSURE

The theoretical benefits of quality disclosure described in Section 2.1 rely on two assumptions: intermediaries must observe quality through disclosures and must respond to incentives to select high-quality drugs. In this section, I provide evidence on the effects of recall and inspection disclosures that is consistent with these assumptions. To estimate causal effects, I use a difference-in-differences framework ("DD") that compares outcomes for drugs that experienced disclosure to those that did not until later in the study period.[9] My primary event study specification is:

$$Y_{dt} = \alpha_d + \alpha_t + \sum_{s \neq -1} \beta_s \, D_{d,t+s} + \varepsilon_{dt}, \tag{1}$$

where Y_{dt} denotes the outcome for drug d in year t. A drug is a molecule-formulation made by a specific manufacturer, and a market includes all generic bioequivalent drugs. I model the outcome as a function of drug α_d and year fixed effects α_t and unobserved time-varying factors ε_{dt}. $D_{d,t+s}$ is an indicator equal to one s years after the drug's quality is first disclosed. Its coefficient β_s varies flexibly by year and is normalized to zero in the year before treatment. I estimate the model using drug-year-level data and cluster standard errors at the drug-level (the level at which treatment occurs). I also report the average effects on treated drugs by estimating the DD specification below:

[8] Table VI presents the corresponding statistics for inspected drugs. This sample includes drugs that are matched to manufacturing locations and inspection data and represents 60% of drug volumes. It is observably similar to the sample in Table I, with slightly younger drugs (17.3 vs 18.5) and smaller volumes (9,786 vs. 21,056).

[9] For exposition, I refer to these drugs as "recalled" and "not-yet-recalled," respectively, though I also apply the framework to study inspection disclosures.

$$Y_{dt} = \alpha_d + \alpha_t + \beta\, D_{dt} + \varepsilon_{dt}, \tag{2}$$

where D_{dt} is equal to one after disclosure.

This design relies on the identifying assumption that the outcomes of recalled and not-yet-recalled drugs would have evolved in parallel in the absence of disclosure, and it exploits variation in disclosure timing to identify causal effects. While this assumption cannot be tested, the absence of differential trends prior to treatment provides evidence consistent with it. I plot the dynamic coefficients β_s and find no evidence of pre-existing trends that are correlated with disclosure timing. In formal tests of joint significance, I reject the null hypothesis of non-zero pre-trends for all key outcomes.

Disclosure is not random, and one source of selection into treatment is the FDA's risk-based inspection model. The model prioritizes facilities based on drug characteristics such as "inherent risk," inspection history, and patient exposure. To control for the model's criteria, I include drug fixed effects to remove "inherent," unobserved, and persistent differences across drugs. In robustness checks, I add inspection grade covariates to control for time-varying differences in inspection histories, and I explore heterogeneity in the estimated effects by market size as a proxy for patient exposure. Disclosure timing could also correlate with outcomes through other channels. Intermediaries, for instance, might observe private quality signals (e.g., consumer complaints), allowing them to anticipate disclosure, a possibility that I examine using more granular monthly data. Alternatively, the FDA might relax quality enforcement for drugs at high risk of shortage (Brennan, 2016). In these cases, the estimates are likely to understate the true effects of disclosure, but other non-randomness may have opposite signed effects.

3.1. *Effects of Recall Announcements on Intermediary Purchases*

I first show that recall announcements have large, negative, and persistent effects on intermediary purchases. Figure 3 panel (a) presents the event study coefficients from Equation (1) with each drug's share of pharmacies as the dependent variable.[10] Prior to disclosure, the estimated coefficients are insignificant and close to zero, suggesting no differential pre-trends ($p = 0.93$). In the disclosure year, the trend for recalled drugs diverges sharply, with

[10]This measure captures intermediaries' "extensive margin" supplier choices, rather than choices over volume.

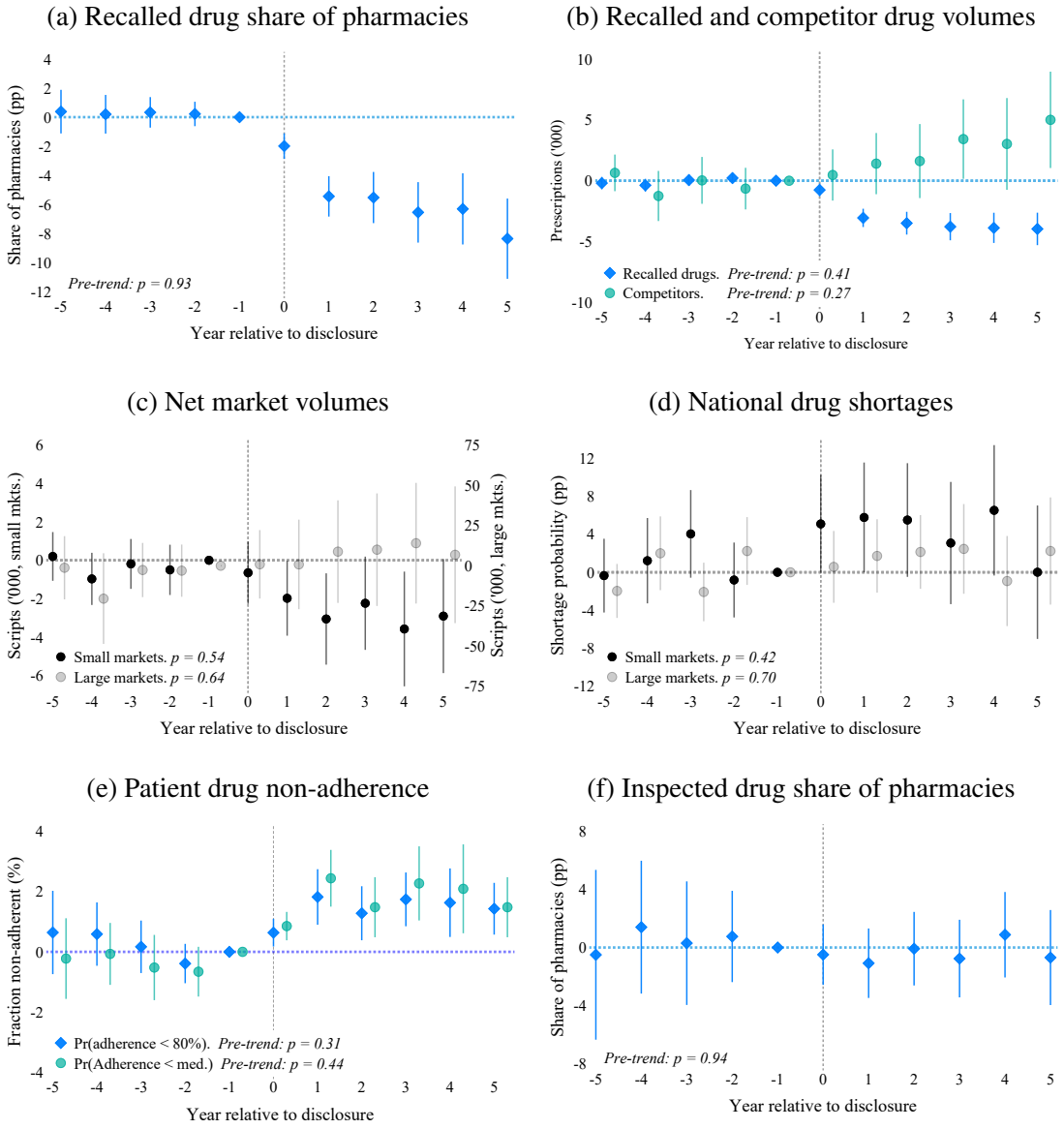


FIGURE 2.—Effects of quality disclosure on drug purchases

Note: The figure displays the estimated dynamic effects of quality disclosures, obtained using Equation (1). The sample includes multisource generic drugs in Optum claims, and the unit of observation is a drug- or market-year. The period prior to the first disclosure event is the omitted reference. The figures present the effects of recall disclosures on recalled drugs (panels (a)—(b)), bioequivalent competitors of the recalled drug (b), and net market effects ((c)—(e)). Panel (f) shows the effect of inspection failure disclosures on inspected drugs, obtained using a DDD specification (Equation (1) with an additional term, $\gamma\, D_{dt} \times 1\{\text{failed}_d\}$.) The dependent variables are each drug's share of pharmacies ((a), (f)) and prescriptions ((b), (c)), the probability of a new drug shortage, (d) and patient drug adherence (e). The figures present 95% confidence intervals with standard errors clustered by drug or market, and p-values from F-tests of the joint significance of pre-treatment coefficients.

their share declining steadily over the next five years. Figure 8 provides additional detail on these dynamics using finer and coarser cuts of the data. Panel (a) zooms in on the *months* surrounding recalls. It shows that recalled and control drugs follow similar trends up until the last month before disclosure, suggesting anticipatory effects are minimal. Panel (b) zooms out to consider the *long-run* effects of recalls, showing that the recalled drug continues to lose share even a decade later.

Table II presents the average effects, estimated using Equation (2). Recalled drugs experience a 43% decline in pharmacy share (4.4pp, panel A, col. (1)), relative to not-yet-recalled drugs. While large, this estimate may actually understate the true magnitude, if drugs with stronger responses are more likely to exit and are not observed in the sample.[11] In another sense, the magnitude is not surprising. Generics are wholly undifferentiated—apart from quality, which has potentially severe health consequences—implying more aggressive substitution patterns than is observed for most consumer products.[12]

These results show that recall announcements sharply reduce pharmacy purchases of low-quality drugs, with effects that last far longer than the recall itself or most disclosure interventions (Ben-Shahar and Schneider, 2017). The persistence of these effects cannot be easily explained by supply disruptions—which are typically resolved within months—or by purchasing inertia, given the rapid supplier turnover observed in the data (GAO, 2011). (The average supply contract in my data lasts for 1.5 years, and the average supplier is chosen for 1.9 years in a row.) Similarly, I do not find significant effects for subsequent recalls. Persistent effects could, however, be explained by long-run changes in intermediary preferences or manufacturer costs, possibilities I explore in Section 4.

Next, I consider intermediary substitution patterns and how competing manufacturers respond to disclosure. Figure 3 panel (b) compares the event studies for recalled and competitor drugs, using drug volumes as the dependent variables.[13] After disclosure, the coefficients for recalled and competitor drugs break in similar but opposite-signed ways, con-

[11] Consistent with this, excluding recalled drugs that eventually exit the sample yields significantly smaller estimated effects (-28%, Table IX row 11).

[12] Volume should not be expected to fall by 100%, as recalled and inspected drugs need not exit the market.

[13] Since prescription volumes are highly skewed, I estimate effects using logged values. Competitor and market level effects are estimated using data at the market-year level. My data capture a subset of U.S. drug volumes, implying the estimates understate effects on volume levels, while overstating percentage effects.

sistent with intermediaries substituting from low- to high-quality drugs. Table II shows that competitor volumes increase by 12% on average, while recalled drug volumes decline by 60% (cols. (4), (2)). As a result, the average patient's probability of receiving a high-quality drug increases by 9% for 10 years, implying that recall announcements are strikingly effective, even when patients do not observe them.

In my robustness checks, I report coefficients from the estimators proposed by Callaway and Sant'Anna (2021) and Borusyak et al. (2024), which address concerns about DD estimates in staggered treatment designs. I also test the stability and significance of my estimates to using different outcome measures, market definitions, control groups, covariates, samples, weighting and clustering. The results, shown in Table IX, are stable and typically fall within two standard errors of the baseline estimates.[14]

3.2. *Effects of Recall Announcements on Supply Reliability and Patients*

I now explore how disclosure affects the drug supply, using national shortages as a proxy for supply disruptions. Shortages are relatively rare—I observe fewer than 1,200 in my data—causing the effects to be imprecisely estimated. Nevertheless, I find that recalls increase the probability of a national drug shortage by 33% (significant at $p < 0.05$, Table II, col. (6)). Conceptually, one might expect these effects to be concentrated in small markets, where substituting between manufacturers requires proportionally greater capacity adjustments and fixed costs. Consistent with this, Figure 3 panel (d) shows that the estimated effects are driven by small markets (with fewer than 7 manufacturers), which experience a 53% increase in shortages (significant at $p < 0.05$), while large markets experience no significant change. The claims data confirm these patterns: panel (c) shows that small markets experience greater declines in volume than large markets do (-15% versus 4% on average, Table II panel B, col. (5)).[15] Overall, these results suggest that intermediary transitions

[14]The exceptions use a never-treated control group or exclude drugs that exit the sample (footnote 11). The former should be interpreted cautiously, as the controls have observably different characteristics and pre-trends.

[15]The difference in small markets stems from greater declines for recalled drugs and lesser increases for competitors (panel B, col. (2),(4)). The effects in small and large markets largely offset, so that the average market experiences no significant change in volume (panel A, col. (5)).

TABLE II: Effects of quality disclosures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Recalled drug		Competitor spillovers		Net market effect				Inspected drug	
Dependent var.	Share of pharms.	Log scripts	Share of pharms.	Log scripts	Log scripts	Pr(short -age)	Pr(non-adherence)	Log price	Share of pharms.	Log scripts
A - Average effects										
Disclosed * post	-0.044	-0.918	0.068	0.115	-0.037	0.021	0.021	-0.036	-0.008	0.040
	(0.007)	(0.082)	(0.014)	(0.083)	(0.060)	(0.010)	(0.006)	(0.036)	(0.011)	(0.123)
Disclosed * post (%)	-43%	-60%	9%	12%	-4%	33%	5%	-4%	-10%	4%
Disclosed * post (level)	–	-2,683	–	1,927	-1,071	–	–	-0.010	–	-57
B - Heterogeneity by market structure										
Low participation: Disclosed * post	-0.077	-1.054	0.108	0.028	-0.160	0.042	0.031	-0.160	-0.036	-0.102
	(0.022)	(0.140)	(0.025)	(0.136)	(0.088)	(0.018)	(0.008)	(0.031)	(0.028)	(0.194)
High participation: Disclosed * post	-0.026	-0.779	0.039	0.142	0.037	0.009	0.012	-0.023	0.002	0.097
	(0.005)	(0.100)	(0.012)	(0.076)	(0.058)	(0.013)	(0.011)	(0.035)	(0.012)	(0.157)
Observations	12,380	12,380	6,420	6,420	6,531	6,531	20,005	3,984	14,080	14,080
Dependent var. median	0.103	4,466	0.738	15,810	29,333	0.063	0.408	0.290	0.081	1,389

Note: The table reports the estimated average effects of disclosures, obtained using Equation (2). It presents the effects of recall disclosures on recalled drugs (columns (1)—(2)), bioequivalent competitors of the recalled drug ((3)—(4)), and net market effects ((5)—(8)). Cols. (9)—(10) present effects of inspection failure disclosures on inspected drugs, obtained using a DDD specification (Equation (2) with $\gamma\, D_{dt} \times 1\{\text{failed}_d\}$.) The dependent variables are each drug's share of pharmacies and logged prescriptions, the probability of a new drug shortage, patient drug adherence, and logged volume-weighted average wholesale prices. To facilitate comparison, effects are presented in levels and as percentages of the dependent variable medians ("%"). Panel A presents average effects for all drugs in the sample, while Panel B presents results by baseline market structure, for low (N $\leq$6 drugs) and high participation (N > 6) markets. Standard errors are clustered at the drug- or market-level and reported in parentheses.

between suppliers after recalls are not seamless and may generate short-run supply disruptions, particularly in concentrated markets.

To consider whether supply disruptions affect patient access to drugs, I estimate changes in patient medication adherence. Following the literature, I consider patients adherent if their MPR exceeds 80% (Andrade et al., 2006). Baseline pre-recall adherence is low: approximately half of patients take their medications regularly, in line with prior research (Watanabe et al., 2018). Following a recall, adherence declines rapidly, with the largest effects observed in the first post-recall year (Figure 3 panel (e)). On average, recalls reduce adherence by 2.1pp (5%), consistent with evidence that modest hassle costs, such as a temporary pharmacy stockout, can persistently reduce adherence (De Avila et al., 2021). Non-adherence is associated with large morbidity and mortality costs in the U.S. (up to $500bn annually), suggesting that recalls may impose substantial welfare losses, which must be weighed against the benefits from improved quality.

In contrast, I find little evidence that disclosure affects patients through drug prices. Recalls do not significantly affect wholesale prices (which fall 4% on average; Table II, col. (8)), consistent with prior evidence from supply disruptions (Dave et al., 2018).[16] I also examine whether disclosure affects supply through manufacturer participation. Table VII col. (3) shows that markets gain 0.6 manufacturers on average after recalls (an increase of 10%). This increase, however, cannot explain the estimated changes in recalled drug purchases, as effects in markets without net entry are statistically indistinguishable from the main estimate (-33% versus -43%; Table IX row 12 versus 1).

3.3. *Mechanisms and Effects of Inspection Reports*

Next, I explore potential mechanisms behind the disclosure effects. To do this, I exploit differences in the information content between recall and inspection failure disclosures. Both reveal that quality falls below an acceptable threshold, but recalls also signal a heightened risk of supply disruption. Comparing the two effects therefore informs whether intermediary purchases respond to signals of drug *supply reliability* versus *quality* alone. To isolate the effects of negative quality information from inspections, I estimate a triple

[16] Since the AAC is a type of volume-weighted average, I present price results weighted by pre-recall drug volumes. Results using unweighted prices are similar (-5%).

difference ("DDD") specification that compares outcomes for failing and passing drugs, before and after inspections. The specification is identical to Equation (2) but includes an additional term $(\gamma\, D_{dt} \times 1\{\text{failed}_d\})$, allowing me to difference out potential bias from selection into inspections. This strategy assumes that passing and failing drugs would have evolved similarly in the absence of inspections.[17] The coefficients from the event study (Figure 3 panel (b)) are uniformly close to zero and insignificant, the average effects are precise enough to rule out changes greater than 60% of the recall effects (Table II, cols. (9)–(10)). I find similar results even when restricting failures to the lowest possible inspection grade (OAI), which signals the most serious quality violations and the FDA's intent to escalate enforcement (Table VII, panel B).

In sum, disclosing even severe inspection failures has no detectable effect on intermediary purchases or drug quality. Intermediaries are almost certainly aware of inspection disclosures, which they claim to closely monitor, are often publicized in industry press, and can jeopardize future drug launches and government contracts (FDA, 2024a). And while recalls and inspection failures capture slightly different quality dimensions, both signal violations that may be "injurious" to patient health, implying that quality-seeking buyers should respond to them. Thus, one interpretation of the differential effects of recall and inspection disclosures is that the market responds more strongly to information about supply reliability than quality per se. Intermediary contracts and reference manuals accord with this explanation, describing recall and failure-to-supply penalties, but making no other mention of quality (CardinalHealth, 2022, AmerisourceBergen, 2019). From a policy design perspective, this suggests that intermediaries face limited incentives to act in patients' best interests when selecting quality. Furthermore, disclosures that reveal supply reliability information—rather than quality alone—may produce greater quality improvements.

3.4. *Summary of Disclosure Effects*

The quasi-experimental evidence shows that disclosure significantly improves the quality of generics available to patients, even when they cannot observe quality themselves. These improvements last for at least a decade—far beyond the duration of a typical recall—

[17] DD event studies for passing and failing drugs show very similar patterns before disclosure, supporting the use of the DDD design (Figure 9).

suggesting they are driven by persistent changes in intermediary preferences or manufacturer costs. Finally, manufacturers are penalized more stringently for signals of low supply reliability than low quality alone. Together, these patterns could be explained by changes in intermediary expectations. For example, intermediaries might observe quality disclosures, update their expectations about supply reliability, and respond by punishing low-quality manufacturers. Alternatively, manufacturers might adjust their costs and prices after disclosures, as they reoptimize capacity and positioning, responses that could also rationalize the patterns. Since the DD framework identifies changes in equilibrium outcomes, it ultimately cannot disentangle these channels or predict how quality would evolve without disclosure. Doing so requires understanding how prices would evolve without intermediary and manufacturer responses; this, in turn, presupposes information about their preferences and costs. In the next section, I develop a structural model to recover these primitives and study the welfare effects of intermediaries.

4. AN EMPIRICAL MODEL OF GENERIC PROCUREMENT

In this section, I develop and estimate a scoring auction model of generic procurement. The model characterizes how intermediary preferences and manufacturer costs jointly determine equilibrium prices and quality. I reference the framework and notation developed by Laffont et al. (2020), which I adapt to an empirical setting with incomplete price data.

Suppose an intermediary wishes to procure a quantity of a generic drug. I assume it auctions off a single supply contract to $N \geq 2$ manufacturers in a first score, sealed bid auction with no reservation score. Manufacturer bids are differentiated by two features: the price b_i and the drug's historical quality q_i, which the intermediary weighs using a scoring rule. The intermediary observes manufacturer quality signals—and manufacturers observe intermediary scoring—through repeated interactions over years and across thousands of markets. By the start of the auction, I therefore assume that quality and the scoring rule are predetermined and common knowledge. The econometrician, however, does not observe the scoring rule and must estimate it.

Manufacturers are characterized by their quality type q_i and a random productivity shock θ_i, which jointly determine their cost $C(\theta_i, q_i)$ of producing a drug. Qualities may be mutually dependent and non-identically distributed, but manufacturer productivity shocks are assumed to be conditionally independent of their competitors'. That is, θ_i is independently

distributed conditional on quality as $F(\cdot|q_i)$; furthermore, this distribution (but not its realization) is commonly known. Together, these assumptions situate the model within the conditional independent private values framework.

The intermediary receives a payoff of $U(q_i, b_i) = S(q_i) - b_i$ from awarding a drug contract with quality q_i. Given manufacturers' production technologies, Laffont et al. (2020) derive the intermediary's *optimal scoring* mechanism.[18] For each bid, the intermediary assigns the score

$$R_O(\theta_i, q_i) = S(q_i) - C(\theta_i, q_i) - \frac{F(\theta_i|q_i)}{f(\theta_i|q_i)} C_\theta(\theta_i, q_i), \tag{3}$$

and selects the manufacturer with the highest R_O. In contrast to a *linear-in-price* score, $R_L(\theta_i, q_i) = S(q_i) - b_i$, the optimal score exploits the buyer's monopoly power to induce lower prices from low-cost manufacturers.

4.1. *Equilibrium*

Under the optimal score, Laffont et al. (2020) show that each manufacturer's probability of winning is determined by

$$P(\theta_i, q_i, q_{-i}) = \prod_{j \neq i} \left[1 - F\Big(\Gamma^{-1}(S(q_j) - S(q_i) + \Gamma(\theta_i, q_i), q_j) | q_j \Big) \right], \tag{4}$$

where $\Gamma(\theta_i, q_i) = C(\theta_i, q_i) + \frac{F(\theta_i|q_i)}{f(\theta_i|q_i)} C_\theta(\theta_i, q_i)$. Manufacturers choose prices to maximize their expected profits $[b_i(\theta_i, q) - C(\theta_i, q_i)] P(\theta_i, q_i, q_{-i})$. The first order condition yields the following type-symmetric Bayesian Nash equilibrium bidding strategy:

$$b_i = \sigma(\theta_i, q_i, q_{-i}) = C(\theta_i, q_i) + \frac{\int_{\theta_i}^{\bar{\theta}} C_\theta(u, q_i) P(u, q_i, q_{-i}) du}{P(\theta_i, q_i, q_{-i})}.^{19} \tag{5}$$

[18] The preference for quality $S(q_i)$ could reflect altruism or expected supply disruption costs. The optimal direct mechanism satisfies manufacturers' incentive and rationality constraints within the space of first-score auctions. Since I do not observe the scoring rule, I check whether my results are robust to using optimal and linear scores.

[19] For exposition, I refer to the cost residual or inefficiency θ as "productivity." Equilibrium relies on the following technical assumptions: manufacturers are risk neutral and know the number of bidders; costs and payoffs

This equilibrium expression is similar to the analogue from price-only auctions (e.g., Proposition 2 in Riley and Samuelson (1981)). It departs in two ways, which cause manufacturers of different quality types to bid different strategies in the same auctions. First, each manufacturer type is characterized by a potentially distinct cost technology $C(\theta_i, q_i)$ and productivity distribution $F(\cdot|q_i)$. Since the optimal bid is strictly increasing in cost and productivity, differences in these values generate asymmetric bidding behavior. Second, each type faces different choice probabilities $P(\theta_i, q_i, q_{-i})$ that reflect intermediary preferences. Stronger preferences for quality raise the probability that high-quality types win, increasing their differentiation and optimal prices. For low-quality types, quality preferences work in the opposite direction and tend to lower prices.

4.2. *Equilibrium with quality disclosure*

Equilibrium prices and qualities are jointly determined by the buyer's scoring rule (Equation (3)) and manufacturers' bidding strategies (Equation (5)). The basic structure of the model changes in three ways when new information about quality is disclosed. First, disclosure may alter intermediary preferences by allowing them to update their expectations about manufacturer reliability. Stronger preferences for high-quality (and in expectation, more reliable) manufacturers may reduce the auction's competitiveness and increase prices. If quality is costly, disclosure may also lower the auction's productive efficiency by reallocating intermediary purchases toward higher-cost manufacturers. Second, disclosure may affect manufacturer costs. For example, recalled manufacturers may increase their effort or reduce their capacity to improve their quality, increasing variable costs. In Equation (5), increased costs are partly passed through to the intermediary—and ultimately to patients—in higher prices. Third, manufacturers may strategically adjust their prices in response to new information about quality and costs. For example, low-quality manufacturers may reduce their prices to rehabilitate their chances of winning. Whether disclosure is socially optimal depends on whether the gains from increased quality outweigh the potential costs from reduced competition and productive efficiency.

This analysis is only partial, since I do not model patient demand responses, or disclosure's long-run effects on manufacturer entry and investment. The quasi-experimental

---

from quality are positive; costs increase monotonically in θ; $f(\cdot|q_i) > 0$ and $F(\cdot|q_i)/f(\cdot|q_i)$ is non-decreasing over the support of θ; and high- and low-quality firms do not lose with certainty.

evidence, however, suggests that these forces are unlikely to have first order effects in the short-run. For example, the average market shows no detectable change in patient demand after disclosure, and changes in manufacturer participation are uncorrelated with intermediary purchases and drug quality. I leave the study of these long-run forces to future work.

4.3. *Estimation*

The key primitives of the scoring auction model are buyer preferences for quality $S(q)$, the cost function $C(\theta_i, q_i)$, and the distributions of manufacturer productivities $F(\theta_i|q_i)$. With complete data on prices, qualities, and the number of bidders, these objects are non-parametrically identified following the arguments of Guerre et al. (2000) and Laffont et al. (2020). My contribution is to show how the model can be identified using incomplete price data, which is common in empirical applications.

Identification with incomplete prices is non-standard because the model incorporates two forms of asymmetry: variation in *ex ante* bidder characteristics (costs and productivities) and in buyer preferences. Models with only the first form of asymmetry are identified from data on winning bids and winner identities (Athey and Haile, 2002). I show that a model with both forms of asymmetry is also identified from these data (and complete data on qualities), for both optimal and linear scores. My approach is general and can be easily adapted to other scoring formats, such as polynomial scores and quality-to-price ratios.[20]

4.3.1. *Unobserved Prices*

First, I describe what information about the unobserved price distributions can be recovered from winning prices. Winning prices are only observed when the bidder has the highest score, which in turn depends on all bidder types. This suggests that the information that can be recovered will depend on the particular scoring rule and the composition of bidder types.

In homogeneous auctions—where all bidders share the same type—the price distributions can be fully recovered from the winning prices. Here, the winning price is necessarily the minimum from each auction, and prices are independently and identically distributed (as bidder types are also identical). This allows me to recover the price distribution using

[20]Derivations, proofs, and additional details on the estimation algorithm are in the Supplementary Materials.

the properties of the first order statistic (Athey and Haile, 2002). However, this method cannot be used to identify the price distribution in heterogeneous auctions, where prices are neither independently nor identically distributed. To see this, note that equilibrium prices depend on q_i (Equation (5)), implying different and potentially dependent distributions across manufacturer types. Moreover, if buyers value quality, winning prices need not be minimum prices, unless they were submitted by the lowest-quality type.

For most scoring rules, identifying buyer preferences requires the price distributions from at least one heterogeneous auction (i.e., one with variation in winner type). I now show that these price distributions can be non-parametrically recovered, up to a guess of the buyer preferences. For exposition, I focus on a *linear score* with two bidder types, $t \in \{H, L\}$. In the Supplementary Materials, I show the identification argument can be extended to other scoring rules and an arbitrary, finite number of types. Let N_t refer to the number of type t bidders in an auction, and let prices be distributed as $b_t \sim G_t(\cdot|q_i = t, q_{-i})$. I first rewrite the observed winning price distributions using Bayes' Rule:

$$Pr(b_{t,i}|i \text{ wins}) = \frac{Pr(i \text{ wins}|b_{t,i})Pr(b_{t,i})}{Pr(i \text{ wins}),}$$

where I suppress conditioning on $q_i = t, q_{-i}$ throughout. After substituting in the linear scoring rule and rearranging terms, the following holds for high-quality types:

$$Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) = \big[1 - G_L(b_{H,i} - \Delta S)\big]^{N_L}\big[1 - G_H(b_{H,i})\big]^{N_H - 1} g_H(b_{H,i}).$$

Note that the buyer preferences for quality have been normalized relative to the lowest quality type, so that $\Delta S = S(q_H) - S(q_L)$. The preferences are thus identified up to a location, as in discrete choice models. Similarly, for low-quality types, I can write:

$$Pr(b_{L,i}|i \text{ wins})Pr(i \text{ wins}) = \big[1 - G_H(b_{L,i} + \Delta S)\big]^{N_H}\big[1 - G_L(b_{L,i})\big]^{N_L - 1} g_L(b_{L,i}).$$

In these equations, the left-hand side (LHS) is the product of two probabilities: the likelihood that the winning price b_t is observed, and that the focal bidder i wins. Both objects can be estimated from the data using standard kernel methods. This implies that these two equations define a system of two nonlinear functional equations in the two unknowns I wish to recover, $g_L(b_L)$ and $g_H(b_H)$. Intuitively, this identification strategy exploits two pieces

of information from each auction. First, knowing the type of the winning bidder tells me about the relative order of the lowest b^L and b^H, up to the guess of ΔS. Second, knowing the winning price provides a lower bound for the losing one, allowing me to carve out the marginal probability of each price.

To estimate $g_L(b_L)$ and $g_H(b_H)$, I first recover the LHS objects using standard non-parametric kernel density estimators (with triweight kernels and rule-of-thumb bandwidths, Li et al. (2000)). I then discretize the prices and solve for $g_L(b_L)$ and $g_H(b_L + \Delta S)$ at each point in the price grid, starting with the lowest price $\underline{b}$. A solution to this system exists at each price, up to defined cutoffs, as long as the grid is sufficiently fine.[21] Above these cutoffs, one bidder type wins with certainty, implying the losers' bids are not defined by the model. Monte Carlo simulations (presented in Figures 14–15) indicate that this procedure recovers the true bids and preferences without bias.

4.3.2. *Costs and Productivities*

With complete price data, manufacturer costs and productivities can be identified by extending the arguments of GPV (Laffont et al. (2020)). Since the bidding function (5) is strictly monotonic in θ_i, a firm's probability of winning given θ_i is equivalent to its probability given $b_i(\theta_i)$; in equations, that is $P(\theta_i, q_i, q_{-i}) = H(b_i(\theta_i), q_i, q_{-i})$. The bidding strategy can therefore be rewritten, after differentiating with respect to θ_i, as:

$$b(\theta_i, q_i, q_{-i}) = C(\theta_i, q_i) - \frac{H(b_i, q_i, q_{-i})}{H_b(b_i, q_i, q_{-i})}. \tag{6}$$

To apply Equation (6) with partial price data, I first rewrite H using Bayes' Rule:

$$H(b_i, q_i, q_{-i}) = \frac{Pr(b_i | i \text{ wins}, q_i, q_{-i}) Pr(i \text{ wins} | q_i, q_{-i})}{Pr(b_i | q_i, q_{-i})}, \tag{7}$$

where each object on the right-hand side (RHS) was already recovered when estimating the price distributions. Using $\hat{H}(b_i, q_i, q_{-i})$ and prices simulated from $\hat{G}(b|q_i, q_{-i})$, I then apply Equation (6) to recover the distribution of costs for each bidder type. To recover productivities, I follow Laffont et al. (2020) in assuming that the cost function is multiplicatively separable in θ_i and q_i. Productivities are then estimated as the residual from projecting

[21] The cutoffs are $\overline{b} - \Delta S$ for b_L and $\overline{b} + \Delta S$ for b_H.

logged costs on bidder quality:

$$\log \hat{C}(\theta_i, q_i) = \log C_0(q_i) + \log \theta_i.$$

4.3.3. *Intermediary Preferences*

In principle, having recovered the cost functions and productivity distributions, buyer preferences could be estimated by simulating winning prices for each guess of ΔS, and then selecting the value that minimizes the distance between the observed and simulated prices (yielding a nested fixed point algorithm (Rust, 1987)). In practice, solving for equilibrium prices in heterogeneous auctions (the algorithm's "inner loop") is computationally demanding and typically requires numerical methods that introduce approximation error (Hubbard and Paarsch, 2014).[22]

Instead, it is computationally convenient to reformulate the estimation problem from one in *price space* to another in *productivity space*. This approach is valid because each guess of ΔS maps to a unique equilibrium price solution (Lebrun, 1999). Conceptually, my approach parallels two-step methods for estimating dynamic games, which approximate a challenging inner loop by mapping it to probability space and forward simulating it (Hotz et al., 1994, Aguirregabiria and Mira, 2007). In my inner loop, I map prices to estimated productivities (from heterogeneous auctions), which I then compare to the "true" productivities (from homogeneous auctions) in my outer loop. While my approach is tractable and avoids approximation error, it introduces an additional source of estimation error that decreases in the sample size.

For the *optimal score*, I follow Laffont et al. (2020) and repose the estimation problem as a discrete choice one. I rewrite the optimal scoring rule in Equation (3) as follows: a buyer selects bid i if and only if

$$S(q_i) - \hat{\eta}_i \geq S(q_j) - \hat{\eta}_j \quad \forall j \in \{1, ..., N\}, \;\; j \neq i, \tag{8}$$

[22] For example, the prices from a linear score solve a non-linear system of ordinary differential equations, where the Lipschitz condition does not hold near $\overline{\theta}$, implying standard numerical methods cannot be used. Hubbard and Paarsch (2014) compare numerical strategies for solving this type of system, which introduce approximation error and are not guaranteed to be stable given estimation error in $\hat{f}(\theta)$ (Kastl et al., 2025).

where the distribution of the "random term" $\hat{\eta}_i = \Gamma(\hat{\theta}_i, q_i)$ is a known function of the productivities θ_i and has thus already been recovered. In this case, Matzkin (1991) shows that the buyer preferences ΔS are non-parametrically identified and can be estimated by maximizing the following likelihood equation:

$$\hat{\Delta S} = \text{argmax } \mathcal{L}(\Delta S) = \sum_{a=1}^{A} \sum_{i=1}^{N} 1(i \text{ wins } a) \log \hat{P}(i \text{ wins} | q_i, q_{-i}, \Delta S), \tag{9}$$

where A is the number of auctions, and the indicator $1(i \text{ wins } a)$ is observed in the data.[23] The final term, the *ex ante* choice probability $P(i \text{ wins} | q_i, q_{-i}, \Delta S)$, is equal to $\int_{\underline{\theta}}^{\overline{\theta}} \hat{P}(\theta_i, q_i, q_{-i}) \hat{f}(\theta_i | q_i) d\theta_i$, where $P(\theta_i, q_i, q_{-i})$ is defined in Equation (4).

This identification strategy rests on two assumptions. First, bidder productivities must be identically distributed across homogeneous and heterogeneous auctions, conditional on the bidder type and auction-level characteristics. This assumption is standard in the empirical auction literature, where it is used to pool data across auctions (Guerre et al., 2000), and I discuss evidence for it in Section 4.4.1. Second, the buyer's scoring rule must depend only on the focal bidder's characteristics. This assumption only pertains to the optimal score, as most scoring rules depend on all bidder characteristics through the price. Together, these assumptions imply that the distribution of optimal scores in homogeneous auctions identifies those in heterogeneous ones, allowing estimation via Equation (8).

Next, I propose a general method that identifies buyer preferences under alternative scoring rules, including the *linear score*, which depend on all bidder characteristics.[24] To pin down intermediary preferences, I match the distributions of $\hat{\theta}_i$ from heterogeneous auctions—which depend on the guess of ΔS—to the $\hat{\theta}_i$ recovered from homogeneous auctions—which do not. As before, the identifying assumption is that the distribution of θ_i is conditionally identical across auctions. I then estimate ΔS using maximum likelihood:

$$\hat{\Delta S} = \text{argmax } \mathcal{L}(\Delta S) = \sum_{a=1}^{A} \sum_{i=1}^{N} \log \hat{f}_{\text{asym}}(\hat{\theta}_{\text{sym},i} | q_i, \Delta S).$$

[23] Matzkin's maximum likelihood procedure does *not* involve parametric assumptions on the error term, since its distribution has already recovered. In this sense, estimation of ΔS is fully non-parametric.

[24] This implies that the distribution of scores differs across homogeneous and heterogeneous auctions, so the distribution of $\hat{\eta}_i$ in Equation (8) is not known.

I estimate the model using bid-level data on drug quality, auction characteristics, and winning prices. The sample is limited to oral generic drugs, which are manufactured using similar production technologies and represent over 70% of volumes, and to auctions held by the five largest intermediaries. I begin by assuming that intermediaries evaluate quality using one collective, optimal score, and that historical quality is an indicator of whether a drug was recalled in the last five years.[25] In Section 4.4, I assess robustness to each of these assumptions by using alternative scoring rules, samples, and definitions of quality.

The auctioned supply contracts in my data are observably heterogeneous—differing, for instance, in their realized volumes and drug characteristics. I follow Haile et al. (2003) to semi-parametrically control for auction-level heterogeneity from realized volumes, intermediary identities, the number of bidders, and drug characteristics that may affect costs, such as the therapeutic area and delivery mechanism. Year fixed effects control for strong deflationary price trends.[26] Table X presents the estimated coefficients, which are consistent with significant and decreasing economies of scale, modest price variation across intermediaries, and higher prices for extended-release drugs. Overall, the estimated patterns are very similar to those found by Cuddy (2020) in this market. After trimming the top 5% of homogenized prices and auctions by size ($N > 9$ bidders), my final sample includes 313,800 bids from 78,263 auctions over 2010-2022. From this sample, I set aside observations from 2022, which I use to evaluate model fit.

Table XI presents descriptive statistics for this sample. Low quality is relatively common: 25% auctions have at least one low-quality bidder, and 4% have two or more. Figure 11 plots the distributions of winning homogenized prices. The prices exhibit high variance (with a coefficient of deviation of 0.55), and high-quality prices exceed low-quality ones on average. These patterns could reflect multiple sources of bidder asymmetry: low-quality manufacturers might have lower costs, or they might set smaller markups if intermediaries prefer quality. A key objective of the structural model is to disentangle these sources.

[25] This definition is consistent with the quasi-experimental evidence, which shows persistent effects from first recalls and limited effects from subsequent ones. For the structural model, I define a market as a molecule-formulation-strength combination, reflecting the level at which generic auctions are held. This is narrower than the standard market definition (molecule-formulation), used to study substitution patterns in Section 3.

[26] This method assumes that auction-level heterogeneity enters prices in a separable fashion. The adjustment preserves equilibrium bidding for optimal and linear scores.

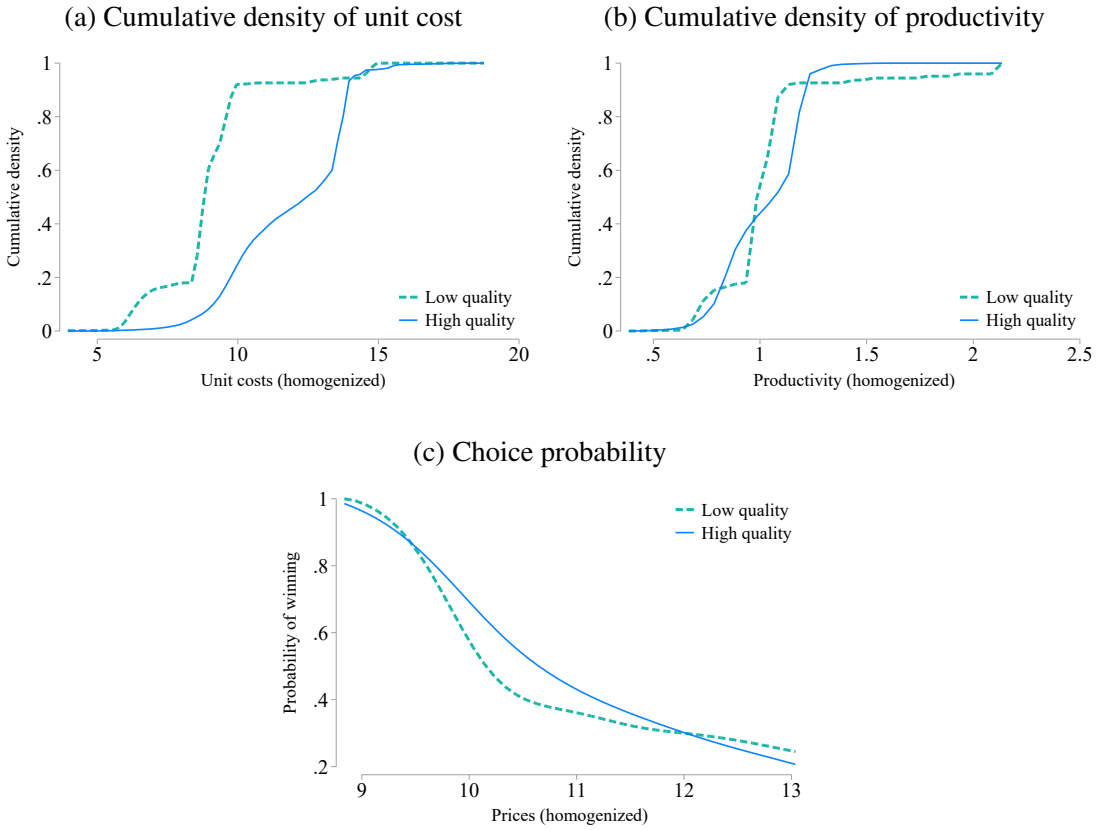


FIGURE 3.—Scoring auction model estimates.

Note: The figure presents the key estimated objects from the optimal scoring auction model in homogenized 2022 dollars. Panel (a) presents the kernel density of the cost per unit, while panel (b) presents the density of productivities. Panel (c) plots the choice probabilities for each bidder type as a function of the price.

4.4. *Results*

Figure 4.3.3 presents the estimated primitives from the structural model: the distributions of manufacturer costs and productivities, as well as intermediaries' preferences for quality. Each object is shown in homogenized price terms for high- and low-quality manufacturers. Panel (a) shows the first primitive, the distribution of manufacturer unit costs $C(\theta_i, q_i)$. On average, high-quality drugs cost 25% more to produce, with costs that stochastically dominate those of low-quality drugs. The cost estimates are credible, in the sense that they imply price-cost margins close to those reported by generic manufacturers. This comparison is only approximate, as the reported margins include unrelated line items (e.g., specialty drug profits). Nevertheless, it is reassuring that the estimated margin (23%) falls within the

manufacturers' reported range of 18-29%.[27] Panel (b) plots the second primitive, the estimated productivity distributions. After controlling for average cost differences, high- and low-quality manufacturers exhibit similar productivities.

The third and most important primitives are the buyer preferences, which govern the extent to which intermediaries discipline quality in my model. I estimate that intermediaries are willing to pay 13% more to purchase high-quality drugs (1.63 relative to the mean homogenized price of 12.34). Panel (c) illustrates how these preferences shape intermediary choices, showing that low-quality manufacturers' probability of winning declines farther and faster in prices (relative to high-quality types). One way to interpret this estimate is as intermediaries' expected value from avoiding future supply disruptions. Since recalled drugs are 3.4 times more likely to be recalled within five years, this implies that intermediaries are willing to pay a 2% premium for each 10% reduction in their future recall risk.[28]

My estimates imply that intermediaries place substantial weight on quality, contrasting with standard accounts of the generics market. For example, the FDA has described the market as "fundamentally" unable "to observe and reward quality" or to "penalize manufacturers that fail to invest in... a reliable supply" (FDA, 2019). On the other hand, the estimates closely align with intermediaries' own statements about sourcing, which describe reliability as a first order concern (Section 2). They are also broadly similar to quality premia reported from other procurement auctions, which range from 19%-68% (Spagnolo, 2012, Krasnokutskaya et al., 2020, Jääskeläinen and Tukiainen, 2019, Albano et al., 2009). My estimates may be more conservative because generic intermediaries are relatively insulated from the costs of low quality: they face limited reputational risk, are reimbursed for direct recall costs, and are not liable for patient injuries (Fink, 2012).

4.4.1. *Tests of the Identifying Assumption*

Identifying buyer preferences requires the standard assumption that bidder productivities θ_i are conditionally identically distributed across homogeneous and heterogeneous

[27] To calculate margins, I take the ratio of the estimated winning prices and costs and weight them by the realized volume of each contract (Starc and Wollmann, 2022). I compare the median estimated margin to the operating margins reported by the seven largest public manufacturers in my data, excluding a cartel period (2013-2015).

[28] Previously recalled drugs are recalled each year with probability 15.5%, versus 4.5% for not recalled drugs, implying an elasticity of $.13/((.045/.155)-1)*.1=2\%$.

auctions. I now discuss the empirical content of this assumption before providing evidence for it. For exposition, I compare a heterogeneous auction with two bidders of different types $A := \{N_H = 1, N_L = 1\}$ to a homogeneous, high-quality auction $S := \{N_H = 2, N_L = 0\}$, which differs in that it has not yet experienced a recall.

At least three sources of non-random variation threaten the identifying assumption. First, recalls may not be randomly assigned. For example, if low-cost (and low-θ_L) manufacturers are disproportionately recalled, then the productivities in auction A will systematically differ from those in S. To test for this source of *selection into treatment,* I compare the distributions of productivities in (not yet) homogeneous and (not yet) heterogeneous auctions, estimated using data before each market's first recall. Intuitively, this test exploits the same source of variation as the quasi-experimental design: if recall timing is random, the closeness of the distributions in Figure 13 panel (a) suggests that the identifying assumption holds. Using two one-sided t-tests, I able to reject differences in the mean productivities greater than 1% of the overall mean at $p < 0.01$.[29]

Since I cannot directly compare the productivities in homogeneous and heterogeneous auctions after recalls, I use a complementary "balance test" to assess the identifying assumption for these auctions.[30] I compare firms' observable characteristics across auctions, focusing on features that are correlated with productivity but are not controlled for in the first stage homogenization. For example, panels (c)–(e) show the distributions of manufacturer sizes, which affect productivity through economies of scale. I find no systematic differences in manufacturer sizes across homogeneous and heterogeneous auctions, suggesting that productivities may also be balanced after recalls. (I am able to jointly reject differences in means greater than 0.7% of the overall sample, or greater than 4.0% for each auction size, at $p < 0.01$.)

A second threat could arise from manufacturers' *selective participation* in auctions after recalls—for example, if recalls reduce competition and induce the entry of high-cost types. To examine this possibility, I re-estimate the model using auctions that experience no change in bidder participation. The results are similar to the baseline estimates (Table

[29]For optimal scores, the distributions imply differences in P in the estimating Equation (9) of less than 0.001.

[30]Estimating productivities in heterogeneous auctions requires knowing buyer preferences, which in turn requires the identifying assumption.

XII row 5 versus row 1), suggesting that selective participation is not a major threat to identification. A final concern is that high-quality manufacturers may *strategically adjust their costs* in response to a competitor's recall. Since I cannot separately identify changes made by high- and low-quality types, I assume that any differences in costs after recalls are attributed to low-quality manufacturers. Using this assumption and a sample of eventually recalled drugs, I find that low-quality manufacturers sharply reduce their costs after recalls (by 24%, Table XII row 4). This is comparable in magnitude to the difference between high- and low-quality costs from the main sample (25%) and suggests that productivity differences between types are driven by recalls, rather than selection into treatment.

4.4.2. *Model Fit and Robustness*

In addition to the cost and preference benchmarks discussed above, I assess the fit of my model in two ways. First, I consider the model's ability to predict post-recall outcomes in sample. Using pre-recall data, I predict equilibrium prices by type, which I compare against the observed data (Figure 12 panel (a)). Second, I repeat this exercise using post-recall data from 2022 (which was not used in estimation, (b)). The close correspondence between the predicted and observed data indicates the good fit of the model in and out of sample.[31]

Finally, I assess the robustness of my results to perturbations in the sample and maintained assumptions. Table XII presents estimates of intermediary preferences (col. (1)) and average differences in high- and low-quality costs (col. (2)) obtained with these perturbations. The baseline model assumes optimal scoring and that "low-quality" is defined as being recalled in the last five years (row 1). Alternative specifications use stricter definitions of quality (2); fix the auction size (3), drug quality types (5), or bidder participation (5); or assume that intermediaries use linear scoring (6). Across all specifications, I estimate positive preferences for quality and greater costs for high-quality drugs. The consistency of these patterns suggests that the model's key insights are robust.

5. COUNTERFACTUAL ANALYSES

In this section, I use the estimated model to decompose the equilibrium effects of quality disclosure and to study the welfare consequences of informed intermediaries. I then evalu-

[31] Confidence intervals are not typically reported in the empirical scoring auction literature, since it is not clear that non-parametric estimators converge in a distributional sense (Ma et al., 2019).

ate federal policy proposals that seek to improve drug quality and reliability. Throughout, my outcomes of interest are equilibrium quality, prices, and producer surplus. To assess patient welfare, I focus on quality and prices. Quality affects patient well-being directly—by definition, low quality can cause serious injury—and indirectly, by increasing the risk of supply disruptions and non-adherence. To approximate patients' value of quality, I rely on laboratory estimates of the direct health effects of recent recalls. I assume that drug wholesale prices affect patients through higher retail prices.

To measure producer surplus, I consider changes in manufacturer and intermediary variable profits, net of one-time recall penalties.[32] Intermediary surplus is computed from their payoffs $U(q_i, b_i) = S(q_i) - b_i$, which are a function of *ex post* revealed quality (as quality is fixed). Since intermediary preferences are identified up to a location normalization, I solve for the location $S(q_i = 0)$ that maximizes surplus under the estimated parameters. All outcomes are expressed in 2022 dollars.

For each counterfactual environment, I simulate 50 draws of productivities per bidder from $F(\cdot|q_i)$. I calculate bidder costs and intermediary choice probabilities, and then integrate over these objects to find equilibrium prices. Winners are determined by the optimal scoring rule, which fits the data well and is computationally convenient. Finally, I compute manufacturer and intermediary surplus, holding patient demand, manufacturer participation, and drug quality fixed.

5.1. *Quality Disclosure Mechanisms*

I begin by studying the FDA's status quo disclosure policy and decomposing its supply- and demand-side mechanisms. Section 3 showed that disclosure improves quality without significantly changing prices. To quantify how it affects producer surplus, I simulate equilibria before and after recall announcements, updating q and redrawing θ to reflect the new quality information.[33]

[32] Intermediary contracts typically stipulate "failure-to-supply" clauses that require manufacturers to reimburse the direct costs of recalled drugs. I assume that 10% of recalled drug volumes are reimbursed, based on volumes in the recall data; my results are not directionally sensitive to this assumption.

[33] To isolate the recall mechanisms and facilitate comparison with the quasi-experimental results, I use a sample of auctions with recalled and not-yet recalled drugs.

Consistent with the quasi-experimental estimates, Table 5.1 shows that disclosure increases the share of high-quality drugs (by 4%), with limited effects on prices (which increase 4% on average, but decrease slightly in weighted terms, column (3)). On the supply-side, disclosure lowers realized costs by 10%, implying a modest improvement in productive efficiency. Manufacturer profits fall by 11%, as disclosure increases price competition and failure-to-supply penalties. These losses help explain why manufacturers do not voluntarily disclose their quality types, as predicted by the "unraveling" result of Grossman (1981) and Milgrom (1981), and quality regulation is needed in the first place. Intermediary surplus increases by 2%, as they benefit from lower prices and better information. In total, producer surplus declines by 9%, suggesting that recall announcements benefit patients but have ambiguous welfare effects.

To decompose these effects, I next simulate the model while updating one mechanism—intermediary preferences, manufacturer costs, and prices—at a time. For each outcome y, I implement the following decomposition:

$$\begin{aligned}\Delta y = &\underbrace{y[R_O(\theta_i, q_i'), C(\theta_i, q_i), b(\theta_i, q)] - y[R_O(\theta_i, q_i), C(\theta_i, q_i), b(\theta_i, q)]}_{\Delta \text{ due to intermediary preferences}} + \\ &\underbrace{y[R_O(\theta_i, q_i'), C(\theta_i', q_i'), b(\theta_i, q)] - y[R_O(\theta_i, q_i'), C(\theta_i, q_i), b(\theta_i, q)]}_{\Delta \text{ due to manufacturer costs}} + \\ &\underbrace{y[R_O(\theta_i', q_i'), C(\theta_i', q_i'), b(\theta_i', q')] - y[R_O(\theta_i, q_i'), C(\theta_i', q_i'), b(\theta_i, q)]}_{\Delta \text{ due to prices}}.\end{aligned} \tag{10}$$

I first update intermediary preferences (the q_i in the first RHS term), holding the other mechanisms fixed. The results suggest that intermediary responses explain more than 100% of policy's quality benefits, increasing the share of high-quality drugs by 8% (col. (4)). By selecting more expensive suppliers, intermediaries also modestly increase costs (by 5%) and prices (by 6%).

Manufacturer responses partially offset these effects, reducing equilibrium quality, costs, and prices. Updating the model's cost function (the second RHS term) while holding prices fixed, I find that average costs fall by 1% after disclosure (col. (5)), driven by lower costs for recalled manufacturers. Finally, I allow manufacturers to re-optimize *prices* (the third RHS term). Recalled manufacturers lower prices by 36% (col. (6)), fully passing through cost savings and strategically compensating for their lower-quality type. High-quality man-

ufacturers, by contrast, raise prices by 40% as demand for their drugs increases. On net, average prices decline by 2%, which transfers surplus from manufacturers (whose profits decrease 12%) to intermediaries (whose profits increase 17%). Intermediaries respond to these price changes by shifting purchases toward low-quality drugs, reducing the share of high-quality drugs by 4% and undoing half of the policy's benefit.

In summary, disclosure benefits patients and intermediaries but penalizes manufacturers. The policy's estimated quality benefits stem wholly from intermediary, rather than supply-side, responses. Furthermore, I find no evidence that disclosure increases manufacturer effort to produce quality.[34] Taken together, these results highlight how intermediaries reshape the equilibrium effects of disclosure in a setting where patients cannot observe quality.

TABLE III

DECOMPOSITION OF RECALL EFFECTS

		(1)	(2)	(3)	(4)	(5)	(6)
		Total equilibrium effects			Decomposition: %Δ due to		
	Manufacturer type	Pre-recall	Post-recall	%Δ	Buyer preferences	Bidder costs	Price responses
Bid price	All	3.13	4.07	30%	–	–	30%
	Recalled	2.23	1.43	-36%	–	–	-36%
	Competitors	3.35	4.69	40%	–	–	40%
Unit costs		0.51	0.45	-0.10	5%	-1%	-14%
Welfare-relevant effects							
High-quality share		0.55	0.57	4%	8%	–	-4%
Winning price	All	0.94	0.98	4%	6%	–	-2%
Manufacturer profits	All	167	149	-11%	-7%	9%	-12%
	Recalled*	51	21*	-58%*	-18%	30%	-66%
	Competitors	117	128	10%	-2%	0%	11%
Intermediary surplus		28	29	2%	-15%	0%	17%
Total producer surplus		195	178	-9%	-8%	8%	-8%

Note: The table presents the simulated effects and mechanisms of recall disclosures. Columns (1) and (2) present pre- and post-recall outcomes, and col. (3) the percentage changes. Cols. (4)–(6) decompose the relative contributions of three mechanisms: changes in intermediary preferences (4), manufacturer variable costs (5), and manufacturers' strategic pricing responses (6). All monetary values are in 2022 dollars; profits and surplus are aggregated across drugs and expressed in millions of dollars per year. *Post-recall manufacturer profits deduct one-time recall penalties, which are excluded from the last three columns.

[34] My estimates capture variable costs and do not account for potential changes in manufacturer investment.

5.2. *The Role of Informed Intermediaries*

Next, I formally quantify the role of informed intermediaries, by asking how equilibrium outcomes would change if they did not care for (or equivalently, had no information about) quality. In theory, informed intermediaries can improve equilibrium quality relative to the no-information benchmark. However, they may also increase manufacturer differentiation and prices, or decrease allocative efficiency by shifting demand to high-cost manufacturers. To quantify these effects, I simulate a counterfactual environment in which intermediary preferences for quality are reduced to zero ($\Delta S' = 0$). Table IV presents the results (col. (1)) relative to the status quo (col. (2)). I find that informed intermediaries substantially increase the share of high-quality drugs (by 14pp or 34%). They have limited effects on offered prices, which change by less than 1% for both quality types. In equilibrium, informed intermediaries purchase more high-quality drugs, which increases average prices by \$0.12 (20%), or less than one cent (3%) at the median. These responses reward high-quality manufacturers (whose profits increase by 74%) and penalize low-quality ones (whose profits decrease by 22%, but fail to offset gains for high-types). With better information, intermediaries capture more surplus, and total producer surplus increases by 16%.

Overall, these results suggest that informed intermediaries are likely to improve static welfare. By forcing manufacturers to compete on quality, and not just price, intermediaries may also encourage manufacturer investment in quality and the entry of high-quality drugs in the long-run. Alternatively, this counterfactual can be interpreted as a comparison between price-only and scoring auctions. Viewed through this lens, it reproduces a key prediction from the theoretical scoring auction literature: buyers should achieve greater surplus in scoring auctions than in price-only ones (Asker and Cantillon, 2008).[35] Intuitively, this is because price-only auctions limit the space over which buyers and bidders can optimize, thereby reducing the potential surplus.

[35]The surplus in the scoring auction also exceeds that from a price-only auction with a minimum quality standard (discussed in subsection 5.3). This holds true even if bidders are allowed to endogenously "convert" to high-quality types, as the surplus gains from higher quality fail to offset losses from higher prices.

TABLE IV

COUNTERFACTUAL POLICIES: PRICE-ONLY AND PAY-FOR-PERFORMANCE SCORES

		(1)	(2)	(3)	
	Manufacturer type	Price-only scoring	Baseline scoring	P4P scoring	
Quality premium		0%	13%	26%	
Bid price	All	2.97	2.95	3.00	
	Low	0.91	0.90	0.90	
	High	3.57	3.54	3.61	
Winning price	All	0.61	0.74	0.91	
	Low	0.51	0.52	0.44	
	High	0.76	0.90	1.14	
Welfare-relevant effects					
High-quality share		0.42	0.56	0.66	
Winning price	All	0.61	0.74	0.91	
Manufacturer profits	All	132	152	164	
	Low	80	62	52	
	High	52	90	112	
				20% penalties	*100% incentives*
Program cost				6	7
Intermediary surplus		19	22	26	28
Producer surplus less program cost		150	174	185	184

Note: The table presents equilibrium outcomes from counterfactual simulations that vary intermediary preferences for quality. Column (1) presents a "price-only" environment ($\Delta S = 0\%$ of average prices), column (2) presents the status quo ($\Delta S = 13\%$), and column (3) presents preferences under quality subsidies ($\Delta S = 26\%$). All monetary values are in 2022 dollars; profits and surplus are aggregated across drugs and expressed in millions of dollars per year.

5.3. *Policy Proposals*

In my final counterfactuals, I study the welfare implications of recent federal proposals that intend to improve drug quality and supply reliability in the U.S. These proposals fall into two broad classes. The first class seeks to increase *demand* for high-quality drugs and incentivize quality investment throughout the supply chain (HHS, 2024). For example, the Manufacturer Resiliency Assessment Program would develop new manufacturing quality measures to increase information about quality. Another proposal, the Hospital Resilient Supply Program, would subsidize purchases of high-quality drugs (or penalize low-quality purchases), providing an instrument to align intermediary preferences with social objec-

tives. The second class of policies would instead alter the *supply* of drugs, by raising minimum quality standards or reshoring production to the U.S. (GAO, 2024, The White House, 2025). Since I do not explicitly model capacity constraints, which may bind under many of the simulated policies, these counterfactuals are best interpreted applying to a limited set of drugs (e.g., essential medicines).[36]

The quasi-experimental evidence suggests that the first policy, targeting transparency, may be redundant in the generics market, where intermediaries already closely monitor quality. Put another way, the "market information failures" that the policy targets are less severe when sophisticated intermediaries are the primary decision-makers. The second policy, targeting intermediary preferences, could improve welfare if intermediaries undervalue quality relative to the social planner. To assess this possibility, I simulate counterfactual environments in which intermediary preferences are perturbed through varying combinations of subsidies and penalties. Table IV column (3) presents results for one version of the policy, which subsidizes high-quality drugs at 13% of the average price, effectively doubling intermediary preferences for quality.[37] Relative to the status quo, the policy increases the share of high-quality drugs by 10pp (18%), relatively less than the intermediary effect in cols. (1)–(2). However, prices increase by relatively more (by \$0.17, or 23%), as greater differentiation weakens price competition. On the supply-side, producer surplus rises by approximately 6%, depending on the mix of subsidies and penalties used.

Figure 4 presents equilibrium outcomes under many alternative subsidy levels, comparing them against proposals to raise quality standards and reshore manufacturing. I begin by simulating intermediary preferences at subsidies between 0% and 70% of average prices, plotting the equilibrium share of high-quality drugs (on the $x-$axis) against prices (on the $y-$axis). Panel (a) shows that improving quality through subsidies is increasingly costly. While prices increase gradually around the status quo, reaching an 80% high-quality supply would require price increases of over 90%.

---

[36]The HHS proposal focuses on physician-administered drugs, building on existing Centers for Medicare and Medicaid Services authorities. It states that the policies may be extended to the broader generics market, where supply issues are also pervasive and my analysis is directed. While other motivations to reshore manufacturing may exist, I focus exclusively on its potential effects on short-run prices, costs, and quality.

[37]I calculate the policy's social costs assuming a public cost of funds of 20% and consider two designs: 100% subsidies and a mix of 80% subsidies with 20% penalties.

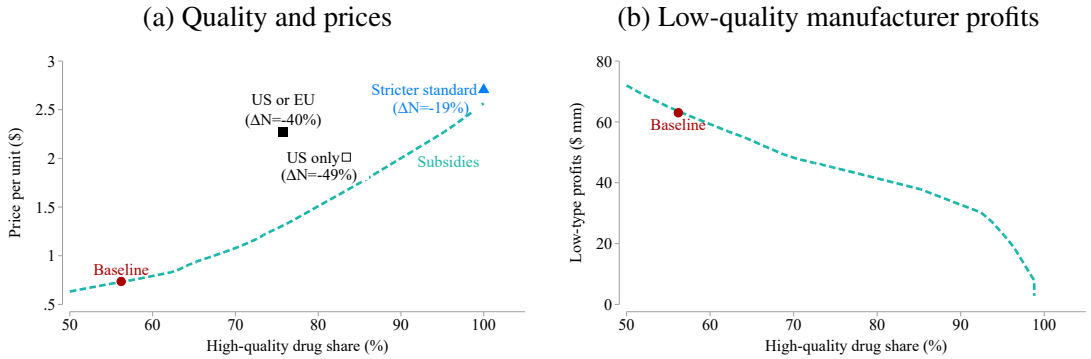


FIGURE 4.—Quality, prices, and profits under counterfactual policies

Note: The figures show equilibrium outcomes from counterfactual simulations of a range of quality policies. Panel (a) plots average equilibrium prices against the share of high-quality drugs, simulated under various levels of quality subsidies, increased minimum quality standareds, and reshored manufacturing. Panel (b) plots average profits for low-quality manufacturers against quality under subsidies. All monetary values are in 2022 dollars.

On the supply side, subsidies also impose substantial costs on low-quality manufacturers. Panel (b) plots their profits, which decrease steadily in quality (falling 35% at an 80% high-quality supply). While this analysis is static, the magnitude of these losses suggests that low-quality manufacturers are likely to exit the market at high subsidy levels. This, in turn, has the potential to increase supplier churn, reduce manufacturing capacity, and worsen drug shortages, a primary target of the policy.

The third proposal, which would raise minimum quality standards, nearly doubles the share of high-quality drugs but sharply reduces competition (lowering the average number of bidders by 19%; panel (a)). As a result, equilibrium prices increase by a factor of 3.7 (slightly more than comparable quality improvements achieved through subsidies). The last policy, reshoring manufacturing, produces the least efficient quality gains. Reshoring improves quality by only 27pp (reaching an 80% high-quality share), but requires 2.7 times higher prices (implying an arc elasticity of 3.6, versus 3.4 for quality standards). A more conservative policy, reshoring to the U.S. *and* Europe, produces qualitatively similar results. The structural model helps decompose why this policy is relatively inefficient: domestic production entails 9% higher marginal costs, on average, and reduces bidder participation by nearly one-half. The intervention is also poorly targeted, as domestic drug quality only slightly exceeds foreign quality (by less than 1pp pre-policy). Consequently, equilibrium quality gains are only possible with large price increases.

A final question concerns whether these interventions are likely to improve patient welfare. Estimating patients' willingness-to-pay for quality is challenging, as they do not observe changes in quality (or even prices, if insured). Furthermore, no systematic evidence exists on the social value of manufacturing quality or its health impacts. Instead, I approximate the value of quality using laboratory evidence from recent recalls. For example, the FDA estimated that 145 recently recalled drugs increased the incidence of cancer by 1 in 8000, roughly twelve times the agency's accepted standard (FDA, 2020). At typical QALY values, this implies that higher-quality drug are worth at least 20% more.[38] This value is likely conservative, as it reflects estimates from less severe (class II) recalls, and patients may be willing to pay more to avoid cancer. Nevertheless, it suggests that only policies with small price increases, such as moderate subsidies, are likely to improve consumer welfare at even generous valuations of quality.

6. DISCUSSION AND CONCLUSION

This paper studies how informed intermediaries and regulatory disclosure discipline quality in the U.S. generics market—a setting where quality has high stakes but is not observed by patients. Using comprehensive data on manufacturing quality over 2000-2022, I show that quality failures are widespread, and that disclosing these failures meaningfully improves the quality of drugs consumed by patients for at least a decade. I then formalize generic procurement as a scoring auction, showing that intermediaries prefer high-quality, reliable drugs, which in turn encourages manufacturers to compete on quality. In equilibrium, intermediaries increase the share of high-quality drugs by a third and magnify the effects of FDA disclosure, generating substantial welfare gains for patients.

These findings challenge the conventional wisdom that generics are perfectly substitutable and that the market is "unable to observe and reward quality" (FTC, 2011, FDA, 2019). Instead, they reveal large and potentially injurious variation in generic quality, which is primarily disciplined by intermediaries. The results also suggest that the design of quality policies should explicitly account for intermediary responses. In particular, since intermediaries already closely monitor quality signals, additional disclosures may yield limited

[38]This estimate reflects the expected increase in cancer incidence from four years of daily use, assuming 10 QALYs lost per cancer diagnosis, and $150,000 per QALY.

welfare gains. Modest quality subsidies, however, could reduce intermediary distortions and improve equilibrium quality, generating greater welfare gains than policies that limit the supply of low-quality drugs.

While this paper captures key mechanisms through which disclosure and intermediaries shape equilibrium outcomes, it abstracts away from manufacturer participation, capacity adjustment, and quality investment. Incorporating these long-run forces would inform the design of quality policy, efficient scoring auctions, and intermediary antitrust enforcement (FTC, 2024). Another natural extension would be to link quality failures directly to patient health outcomes, providing a fuller account of the welfare consequences of drug quality.

REFERENCES

AGUIRREGABIRIA, VICTOR AND PEDRO MIRA (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75 (1), 1–53. [26]

AKERLOF, GEORGE A (1970): "The market for lemons: quality uncertainty and the market mechanism," *The Quarterly Journal of Economics*, 84 (3), 488–500. [1, 6]

ALBANO, GIAN LUIGI, FEDERICO DINI, AND ROBERTO ZAMPINO (2009): "Bidding for complex projects: Evidence from Italian government's acquisitions of IT services," in *Electronic Government: 8th International Conference, EGOV 2009, Linz, Austria, August 31-September 3, 2009. Proceedings 8*, Springer, 353–363. [30]

ALLEN, JASON, ROBERT CLARK, BRENT HICKMAN, AND ERIC RICHERT (2024): "Resolving Failed Banks: Uncertainty, Multiple Bidding and Auction Design," *Review of Economic Studies*, 91 (3), 1201–1242. [6]

AMERISOURCEBERGEN (2019): "Manufacturer Packaging and Logistics Requirements Guide," https://www.amerisourcebergen.com/-/media/assets/amerisourcebergen/manufacturer/ab-manufacturer-logistics-guideline-v13-updated-20230130.pdf. [9, 19]

ANDRADE, SUSAN E, KRISTIJAN H KAHLER, FERIDE FRECH, AND K ARNOLD CHAN (2006): "Methods for evaluation of medication adherence and persistence using automated databases," *Pharmacoepidemiology and Drug Safety*, 15 (8), 565–574. [18]

ASKER, JOHN AND ESTELLE CANTILLON (2008): "Properties of scoring auctions," *The RAND Journal of Economics*, 39 (1), 69–85. [36]

ATAL, JUAN PABLO, JOSE IGNACIO CUESTA, AND MORTEN SAETHRE (2022): "Quality Regulation and Competition: Evidence from Pharmaceutical Markets," *NBER Working Paper 30325*. [6]

ATHEY, SUSAN AND PHILIP A HAILE (2002): "Identification of standard auction models," *Econometrica*, 70 (6), 2107–2140. [4, 23, 24]

BARAHONA, NANO, CRISTÓBAL OTERO, AND SEBASTIÁN OTERO (2023): "Equilibrium effects of food labeling policies," *Econometrica*, 91 (3), 839–868. [6]

BEN-SHAHAR, OMRI AND CARL E SCHNEIDER (2017): "The Failure of Mandated Disclosure," *Russian Journal of Economics and Law*, 11 (4), 146–169. [1, 15]

BORUSYAK, KIRILL, XAVIER JARAVEL, AND JANN SPIESS (2024): "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, 91 (6), 3253–3285. [16]

BRENNAN, ZACHARY (2016): "FDA Allows Banned Chinese Firm to Ship Chemotherapy to US as Shortage Threat Looms," https://www.raps.org/news-and-articles/news-articles/2016/2/fda-allows-banned-chinese-firm-to-ship-chemotherap. [13]

CALLAWAY, BRANTLY AND PEDRO HC SANT'ANNA (2021): "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 225 (2), 200–230. [16]

CARDINALHEALTH (2022): "Manufacturer Reference Manual," https://www.cardinalhealth.com/content/dam/corp/web/documents/Manual/cardinal-health-pharma-supplier-guidebook-fy22.pdf. [9, 19]

COHEN, ROBIN A AND LARYSSA MYKYTA (2023): "QuickStats: Percentage of Adults Aged $\geq$ 18 Years Who Took Prescription Medication During the Past 12 Months, by Sex and Age Group - National Health Interview Survey, United States, 2021," *MMWR Morb Mortal Wkly Rep*. [2, 7]

CONTI, RENA M, ERNST R BERNDT, NERIMAN BESTE KAYGISIZ, AND YASHNA SHIVDASANI (2020): "We still don't know who makes this drug," *Health Affairs Forefront*. [3, 10]

CUDDY, EMILY (2020): "Competition and collusion in the generic drug market," *Mimeo*. [6, 9, 11, 28]

DAVE, CHINTAN V, AJINKYA PAWAR, ERIN R FOX, GREGORY BRILL, AND AARON S KESSELHEIM (2018): "Predictors of drug shortages and association with generic drug prices: a retrospective cohort study," *Value in Health*, 21 (11), 1286–1290. [18]

DE AVILA, JORGE L, DAVID O MELTZER, AND JAMES X ZHANG (2021): "Prevalence and persistence of cost-related medication nonadherence among Medicare beneficiaries at high risk of hospitalization," *JAMA Network Open*, 4 (3), e210498–e210498. [18]

DRANOVE, DAVID AND GINGER ZHE JIN (2010): "Quality disclosure and certification: Theory and practice," *Journal of Economic Literature*, 48 (4), 935–963. [6]

DUBOIS, PIERRE, GOSIA MAJEWSKA, AND VALENTINA REIG (2023): "Drug shortages: empirical evidence from France," *TSE Working Paper*. [6]

FDA (2019): "Drug Shortages: Root Causes and Potential Solutions," https://www.fda.gov/media/131130/download. [30, 40]

——— (2020): "Recalls of Angiotensin II Receptor Blockers (ARBs) including Valsartan, Losartan and Irbesartan," https://www.fda.gov/drugs/drug-safety-and-availability/recalls-angiotensin-ii-receptor-blockers-arbs-including-valsartan-losartan-and-irbesartan. [40]

——— (2023): "Import Alert 66-40: Intas Pharmaceuticals Limited," https://www.accessdata.fda.gov/cms$_ia/importalert_189.html$. [7]

——— (2024a): "Pharmaceutical Inspections and Compliance," https://www.fda.gov/drugs/guidance-compliance-regulatory-information/pharmaceutical-inspections-and-compliance. [19]

——— (2024b): "Recalls Background and Definitions," https://www.fda.gov/safety/industry-guidance-recalls/recalls-background-and-definitions. [3, 8]

FEIN, ADAM J (2023): "The 2023-2024 Economic Report on Pharmaceutical Wholesalers and Specialty Distributors," . [9]

FINK, JOSEPH L. (2012): "Legal Duty to Notify Patient of Product Recall?" *Pharmacy Times*, 78 (3). [30]

FREEDMAN, SETH, MELISSA KEARNEY, AND MARA LEDERMAN (2012): "Product recalls, imperfect information, and spillover effects: Lessons from the consumer response to the 2007 toy recalls," *Review of Economics and Statistics*, 94 (2), 499–516. [6]

FTC (2011): "Authorized generic drugs: short-term effects and long-term impact," https://www.ftc.gov/sites/default/files/documents/reports/authorized-generic-drugs-short-term-effects-and-long-term-impact-report-federal-trade-commission/authorized-generic-drugs-short-term-effects-and-long-term-impact-report-federal-trade-commission.pdf. [2, 9, 40]

——— (2024): "FTC, HHS Seek Public Comment on Generic Drug Shortages and Competition Amongst Powerful Middlemen," . [41]

GALDIN, ANAÏS SOLEDAD (2024): "Trade and Information Provision Across Geographical Space," Ph.D. thesis, Princeton University. [6]

GANAPATI, SHARAT AND REBECCA MCKIBBIN (2023): "Markups and fixed costs in generic and off-patent pharmaceutical markets," *Review of Economics and Statistics*, 105 (6), 1606–1614. [6]

GAO (2011): "Medical Devices: FDA Should Enhance Its Oversight of Recalls," https://www.gao.gov/products/gao-11-468. [8, 15]

——— (2024): "FDA Has Faced Persistent Challenges Overseeing Foreign Drug Manufacturing," https://www.gao.gov/assets/d24107359.pdf. [5, 38]

GAVAZZA, ALESSANDRO AND ALESSANDRO LIZZERI (2021): "Frictions in product markets," in *Handbook of Industrial Organization*, Elsevier, vol. 4, 433–484. [6, 9]

GRENNAN, MATTHEW, KYLE MYERS, ASHLEY SWANSON, AND AARON CHATTERJI (2024): "No free lunch? welfare analysis of firms selling through expert intermediaries," *Review of Economic Studies*, rdae090. [6]

GROSSMAN, SANFORD J (1981): "The informational role of warranties and private disclosure about product quality," *The Journal of Law and Economics*, 24 (3), 461–483. [5, 34]

GUERRE, EMMANUEL, ISABELLE PERRIGNE, AND QUANG VUONG (2000): "Optimal nonparametric estimation of first-price auctions," *Econometrica*, 68 (3), 525–574. [4, 6, 23, 27]

HAILE, PHILIP, HAN HONG, AND MATTHEW SHUM (2003): "Nonparametric tests for common values at first-price sealed-bid auctions," *NBER Working Paper 10105*. [28]

HHS (2024): "White Paper: Policy Considerations to Prevent Drug Shortages and Mitigate Supply Chain Vulnerabilities in the United States," https://aspe.hhs.gov/sites/default/files/documents/3a9df8acf50e7fda2e443f025d51d038/HHS-White-Paper-Preventing-Shortages-Supply-Chain-Vulnerabilities.pdf. [5, 37]

HOTZ, V JOSEPH, ROBERT A MILLER, SETH SANDERS, AND JEFFREY SMITH (1994): "A simulation estimator for dynamic models of discrete choice," *The Review of Economic Studies*, 61 (2), 265–289. [26]

HUBBARD, TIMOTHY P AND HARRY J PAARSCH (2014): "On the numerical solution of equilibria in auction models with asymmetries within the private-values paradigm," in *Handbook of Computational Economics*, Elsevier, vol. 3, 37–115. [26]

JÄÄSKELÄINEN, JAN AND JANNE TUKIAINEN (2019): "Anatomy of public procurement," *VATT Institute for Economic Research Working Papers*, 118. [30]

JEWETT, CHRISTINA (2023): "Drug Shortages Near an All-Time High, Leading to Rationing," *The New York Times*. [7]

KASTL, JAKUB, ERIC RICHERT, AND JESPER RUDIGER (2025): "Efficient Grid Search to Solve Static Games with Private Information," *Kilts Center at Chicago Booth Marketing Data Center Paper*, available at http://dx.doi.org/10.2139/ssrn.5187301. [26]

KRASNOKUTSKAYA, ELENA, KYUNGCHUL SONG, AND XUN TANG (2020): "The role of quality in internet service markets," *Journal of Political Economy*, 128 (1), 75–117. [6, 30]

LAFFONT, JEAN-JACQUES, ISABELLE PERRIGNE, MICHEL SIMIONI, AND QUANG VUONG (2020): "Econometrics of scoring auctions," in *Essays in Honor of Cheng Hsiao*, Emerald Publishing Limited, vol. 41, 287–322. [4, 20, 21, 23, 25, 26]

LEBRUN, BERNARD (1999): "First price auctions in the asymmetric N bidder case," *International Economic Review*, 40 (1), 125–142. [26]

LI, TONG, ISABELLE PERRIGNE, AND QUANG VUONG (2000): "Conditionally independent private information in OCS wildcat auctions," *Journal of Econometrics*, 98 (1), 129–161. [25]

MA, JUN, VADIM MARMER, AND ARTYOM SHNEYEROV (2019): "Inference for first-price auctions with Guerre, Perrigne, and Vuong's estimator," *Journal of Econometrics*, 211 (2), 507–538. [32]

MARRA, MARLEEN R, NATHAN H MILLER, AND GRETCHEN SILEO (2024): "Phoning Home: The Procurement of Telecommunications for Incarcerated Individuals in the United States," *NBER Working Paper 33292*. [6]

MATZKIN, ROSA L (1991): "Semiparametric estimation of monotone and concave utility functions for polychotomous choice models," *Econometrica: Journal of the Econometric Society*, 1315–1327. [27]

AMERICAN CANCER SOCIETY CANCER ACTION NETWORK (2023): "Survivor Views: Drug Shortages, Telehealth, & Biomarker Testing," https://www.fightcancer.org/sites/default/files/docs/survey_drug_shortages_biomarkers_final_3.19.pdf. [7]

CENTER FOR INFECTIOUS DISEASE RESEARCH & POLICY (2018): "Resilient Drug Supply: Critical Acute Drugs," https://www.cidrap.umn.edu/resilient-drug-supply/resilient-drug-supply-critical-acute-drugs. [10]

STATE OF CONNECTICUT (2019): "Connecticut et al. v. Teva Pharmaceuticals USA, Inc. et al. (Unredacted complaint)," *U.S. District Court for the District of Pennsylvania*, 2:19-cv-02407. [9, 11]

THE WHITE HOUSE (2024): "Biden Cancer Moonshot Announces New Pilot to Mitigate Pediatric Cancer Drug Shortages," https://www.whitehouse.gov/ostp/news-updates/2024/10/28/biden-cancer-moonshot-announces-new-pilot-to-mitigate-pediatric-cancer-drug-shortages/. [7]

——— (2025): "Executive Order on Regulatory Relief to Promote Domestic Production of Critical Medicines," https://www.whitehouse.gov/presidential-actions/2025/05/regulatory-relief-to-promote-domestic-production-of-critical-medicines/. [5, 38]

MCKESSON (2024): "McKesson Pharmaceutical Distribution," https://www.mckesson.com/pharmaceutical-distribution/. [2]

MILGROM, PAUL R (1981): "Good news and bad news: Representation theorems and applications," *The Bell Journal of Economics*, 380–391. [5, 34]

PELTZMAN, SAM (1973): "An evaluation of consumer protection legislation: the 1962 drug amendments," *Journal of Political Economy*, 81 (5), 1049–1091. [6]

RILEY, JOHN G AND WILLIAM F SAMUELSON (1981): "Optimal auctions," *The American Economic Review*, 71 (3), 381–392. [22]

RUST, JOHN (1987): "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, 999–1033. [26]

SCOTT MORTON, FIONA (2000): "Barriers to entry, brand advertising, and generic entry in the US pharmaceutical industry," *International Journal of Industrial Organization*, 18 (7), 1085–1104. [6]

SPAGNOLO, GIANCARLO (2012): "Reputation, competition, and entry in procurement," *International Journal of Industrial Organization*, 30 (3), 291–296. [30]

STARC, AMANDA AND THOMAS G WOLLMANN (2022): "Does entry remedy collusion? Evidence from the generic prescription drug cartel," *NBER Working Paper 29886*. [6, 30]

TRACELINK (2023): "Pharmacies and Recalls: Understanding the Impact-and Changes to Come," https://www.tracelink.com/resources/resource-center/pharmacies-and-recalls-understanding-the-impact-and-changes-to-come. [9]

VATTER, BENJAMIN (2022): "Quality disclosure and regulation: Scoring design in medicare advantage," *Available at SSRN 4250361*. [6]

WATANABE, JONATHAN H, TERRY MCINNIS, AND JAN D HIRSCH (2018): "Cost of prescription drug–related morbidity and mortality," *Annals of Pharmacotherapy*, 52 (9), 829–837. [18]

WOODCOCK, JANET AND MARTA WOSINSKA (2013): "Economic and technological drivers of generic sterile injectable drug shortages," *Clinical Pharmacology & Therapeutics*, 93 (2), 170–176. [7]

YURUKOGLU, ALI, ELI LIEBMAN, AND DAVID B RIDLEY (2017): "The role of government reimbursement in drug shortages," *American Economic Journal: Economic Policy*, 9 (2), 348–382. [6]

SUPPLEMENT TO "IN DEFENSE OF THE MIDDLEMAN: QUALITY FAILURES IN THE GENERIC PHARMACEUTICAL MARKET"

CATHERINE ISHITANI
Yale University

1. DATA CONSTRUCTION

1.1. *Quality*

I use four secondary data sources to construct the manufacturing facility dataset. First, I obtain the names and locations of manufacturing facilities that produce drugs for the U.S. market from the FDA by FOIA. Second, I use drug labels scraped from the Drugs@FDA and National Library of Medicine's DailyMed websites. Manufacturers are required to list the manufacturer name and country of origin on imported drug labels under The Tariff Act of 1930 and The Food, Drug, and Cosmetic Act of 1938. In practice, they often include additional information regarding drug manufacturers, packagers and distributors and their locations. The FDA began publishing electronic copies of drug labels on these websites in 2006. Third, I obtain detailed customs data from ImportGenius for all countries that export drugs to the U.S. market from 2006-2022. The data contain the names of shipping and consigning firms, their locations and a detailed product description (e.g., "nifedipine extended-release tablets(procxl), usp 60mg (300ct) metoprolol succinate extended-release tablets, usp, 25mg (500ct) nifedipine extended-release, tablets(procxl), usp 90mg (100ct) metoprolol succinate extended-release tablets, usp, 50mg (1000ct)..."). I also use information reported by Medsafe, the New Zealand Health Authority, which includes the drug names, sponsoring and manufacturing firms and their locations, for all drugs currently marketed in New Zealand. I confirm that drugs produced for the U.S. and New Zealand markets are manufactured in the same facilities using the label and import data. Finally, I use drug location data from ProPublica's Rx Inspector website, which includes a limited number of records obtained from ProPublica's lawsuit against the FDA filed in November 2024.

The first three datasets were first used to match drugs to facilities by researchers at the University of Minnesota Center for Infectious Disease Research and Policy, who used them to identify the locations for 156 drugs (Center for Infectious Disease Research & Policy, 2018). Similar datasets have been used by Joon Noh (using label data) and Anais Galdin (using trade data for injectable drugs) (Noh et al., 2025, Galdin, 2024). The main advantages of using multiple data sources to identify locations are that they increase coverage and allow me to cross-check my matches.

To scale up the CIDRAP methodology, I rely on several machine learning packages. To convert label images to text, I use Optimal Character Recognition from Tesseract. To fuzzy match drug and firm names, I use string cosine distance measures based on NLP Term Frequency-Inverse Document Frequency and the Universal Sentence Encoder from TensorFlow. Lastly, to match locations, I use the Geocoding API from Google. Using this method, I am able to recover at least one facility location for 73% of generic NDCs in the Optum sample, representing 95% of generic volumes.

TABLE V

TOP REASONS FOR RECALLS AND INSPECTION CITATIONS

A - Recall reason	*N*
1. Failure to comply with Good Manufacturing Practices	2,353
2. Cross contamination: chemical or drug	1,497
3. Lack of sterility	916
4. Active ingredient: sub or super potent	856
5. Lack of stability	764
6. Drug mixup or mislabeling	720
7. Drug tests out of specification	654
8. Foreign substance in drug	413
9. Improper storage	392
10. Defective packaging	347
11. Counterfeit or marketed without approval	291
12. FDA requested	201
13. Supplier manufacturing failure	190
14. Consumer complaints	186
15. Cross contamination: microbial	148

B - Inspection citation reason	*N*
1. Failure to investigate out-of-specification results and batch discrepancies (211.192)	2,174
2. Inadequate validation of sterilization process (211.113(b))	1,586
3. Lack of scientifically sound laboratory controls (211.160(b))	1,128
4. No written procedures for process controls to assure drugs conform to standards of identity, strength, quality & purity (211.100(a))	925
5. Equipment is not cleaned at appropriate intervals to prevent contamination (211.67(a))	529
6. Records do not meet retention and audit trail requirements to ensure they are trustworthy (211.68(b))	503
7. Quality control unit lacks authority to review production records to investigate errors (211.22(a))	498
8. Equipment is not of appropriate design for its intended use, cleaning and maintenance (211.63)	461
9. Laboratory records do not include data from all tests necessary to assure compliance with standards (211.194(a))	368
10. Test methods have not established accuracy, sensitivity, specificity and reproducibility (211.165(e))	356
11. Sterile processing areas have deficient monitoring systems (211.42(c)(10)(iv))	332
12. Drug is released without appropriate laboratory testing to determine conformance to standards (211.165(a))	308
13. Deficient control systems to prevent contamination and mix-ups (211.42(c))	308
14. Lack of established time limits for production to assure drug quality (211.111)	258
15. Sterile processing areas have deficient cleaning systems (211.42(c)(10)(v))	232

Note: The table presents the 15 most common reasons for recalls (panel A) and inspection citations (panel B) in the Optum data and the corresponding number of drug-year observations. Recalls and inspections may have multiple reasons and citations, respectively. Recall reasons are manufacturer-reported and observed in FDA Enforcement Reports. Citations and descriptions are inspector-reported and from internal FDA data obtained through FOIA requests. To capture citations associated with failing inspections, the citation sample is limited to those with above-median shares of failing versus non-failing inspections. Citation numbers from the Code of Federal Regulations, Title 21 are presented in parentheses.

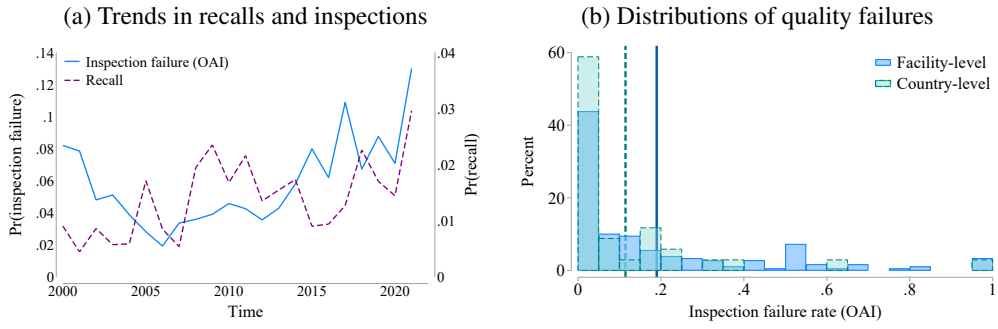


FIGURE 5.—Trends and variation in quality failures

Note: Panel (a) presents annual drug recall rates and facility inspection failure rates, where inspection failures are limited to "Official Action Indicated" (OAI) grades. Panel (b) presents distributions of inspection failure rates averaged over drug-years at the facility and country levels. Distributions are winsorized at the 95th percentile, and vertical lines display mean rates.

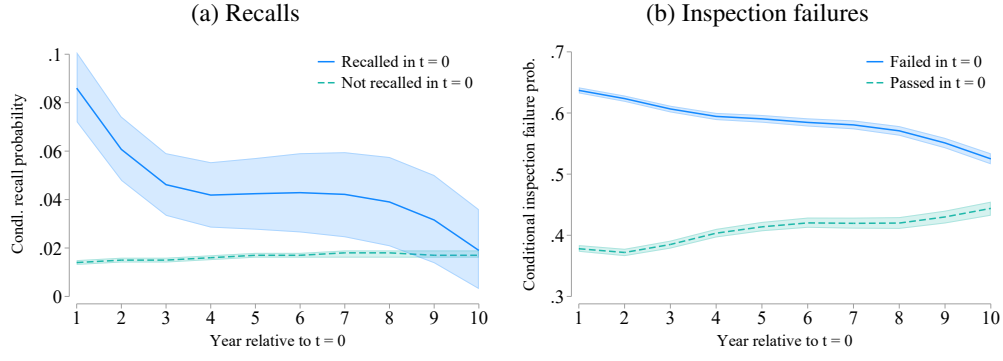


FIGURE 6.—Persistence in drug manufacturing quality

Note: The figure plots persistence in manufacturing quality failures within drugs over time, as observed in the FDA and Optum claims data. It shows the probability of observing a failing signal for drugs with a previously disclosed failure in year 0 (solid lines), or without such a disclosure (dashed lines). Panel (a) presents the probabilities for recall announcements, and panel (b) the probabilities for inspection failures, along with 95% confidence intervals. Recalls are observed in FDA Enforcement Reports, and inspection grades are from internal FDA data obtained through FOIA requests. Failed inspections are those graded "official" or "voluntary action indicated."

1.2. *Prices*

The national and state pharmacy average acquisition cost data are used by state and federal Medicaid agencies to set generic reimbursement levels for publicly-insured individuals. Since the state price data have not been used in previous research, I describe them briefly here. The data have been collected by two firms, Myers and Stauffer and Mercer, since 2010. The firms conduct monthly surveys of random samples of 250-350 independent and chain pharmacies per state (and 500-600 pharmacies nationally), excluding 340B pharmacies. The state data are updated twice a week, while the national data are updated weekly. The surveys capture invoice data, which represents acquisition costs for all patients. These costs are net of rebates paid by manufacturers to plans and off-invoice volume discounts. The former are typically very small ($< 5\%$) for generic drugs, and I am not aware of a research dataset that contains estimates of the latter. Since AAC data do not contain NDCs, I fuzzy match drug names, formulations, and strengths to those found in the FDA NDC data.

Thirteen states (and the federal government) collected AAC data for at least two years between 2010-2022: AL, AR, CO, IA, ID, LA, MD, MT, OH, OR, SC, TN, and UT. I collected average prices for each state over 2010-2022 through Open Records Requests, which I supplement with national data from the Medicaid.gov website after 2013. For six states where I received partial data in early years, I fill in missing prices using data from Medicaid State Drug Utilization Files, adjusted to reflect state-year-specific dispensing fees as described in the Medicaid Covered Outpatient Prescription Drug Reimbursement Information Files. Finally, I also use the Medicare Federal Upper Limit (FUL) as an additional moment after 2016. The FUL provides a volume-weighted average of the actual manufacturer price (AMP). My results are not sensitive to the inclusion of the Medicaid State Drug Utilization and FUL data.

I recover intermediary-specific winning prices for the largest 5 intermediaries in my data. Since I have up to 15 price observations from each auction, this implies that I solve an (often overidentified) system of linear equations in state-level prices and pharmacy shares to estimate each intermediary's winning price.

2. INSTITUTIONAL DETAILS

2.1. *Quality Regulation*

In a typical inspection, officials observe production for 2-5 days, interview employees and review the facility's quality management systems and historical records. At the start of a recall, the FDA convenes an expert committee to classify the defective drug's risks. While recalls, in principle, are voluntary, the FDA can file a court order if the firm should refuse a recall request. In practice, this sort of escalation is extremely rare, and fewer than 70 drug court orders have been granted since 2000.

FDA enforcement of recalls is described in two reports from the GAO and OIG, respectively: "Medical Devices: FDA Should Enhance Its Oversight of Recalls" (2011) and "The FDA's Food-Recall Process did not Always Ensure the Safety of the Nation's Food Supply" (2017). These reports identify gaps in the FDA's process to ensure timely and effective recalls of devices and food (no such report is available for drug recalls). The investigators found that manufacturers often could not locate and remove all defective products, even in high-risk recalls (manufacturers were successful in removing all products in just 52% of completed recalls). Inconsistent with its procedures, the FDA frequently did not conduct post-recall inspections to verify that the recall was complete and that sufficient actions had been taken to prevent a recurrence of the underlying quality problems (53% of completed recalls.)

While pharmacies are not legally obligated to inform patients about recalls, certain pharmacies have written procedures to do so. For example, CVS Caremark and Walgreens state that they may notify patients, depending on the recall classification. Other pharmacies' written procedures do not include steps to inform patients, e.g., UC San Diego Health System. Even pharmacies with formal outreach procedures may struggle to find the resources to do so: 76% of pharmacies reported being understaffed in 2022 (National Community Pharmacists Association, 2022). Anecdotes from lawsuits and the media suggest that information about recalls often fails to reach consumers (Newkirk and Berfield, 2019).

2.2. *Generic Procurement Auctions*

Supply Chain Excellence Award scoring

SCE	Maximum points	Score structure					
Manufacturer fill rate	30 pts	97-100% = 30 95-96% = 15 <95% = 0					
Year-end fill rate	10 pts	97-100% = 10 95-96% = 5 <95% = 0					
Item level lead time	10 pts	**Consumer health:** 0-10 days = 10 11-14 days = 5 >15 days = 0	**Brand:** 0-4 days = 10 5-7 days = 5 >8 days = 0	**Private label:** 0-21 days = 10 22-28 days = 5 >29 days = 0	**Generic:** 0-6 days = 10 7-10 days = 5 >11 days = 0	**SPD (oncology):** 0-4 days = 10 5-7 days = 5 >8 days = 0	**SPD (non-oncology):** 0-15 days = 10 16-20 days = 5 >21 days = 0
Sourcing performance	15 pts	**Consumer health: Account management:** 15 • Sales = 5 • Inventory efficiency = 5 • DSA and market support = 5 Best Practices and Collaboration: +5 (bonus)		**Brand and generics: Account management:** 15 Best Practice and Collaboration: +5 (bonus)		**SPD: Account management:** 15 • Responsiveness = 3 • Dispute resolution = 3 • Pipeline engagement = 3 • Service fee value = 3 • Timely WAC change = 1 • Permit off cycle order planning = 1 • Provide compensation for enhances distribution services = 1	
Product Availability Report (PAR)	10 pts	9.5-10 average pts = 10 8.5-9.4 average pts = 9 7.5-8.4 average pts = 8 6.5-7.4 average pts = 7		5.5-6.4 average pts = 6 4.5-5.4 average pts = 5 3.5-4.4 average pts = 4 2.5-3.4 average pts = 3		1.5-2.4 average pts = 2 0.5-1.4 average pts = 1 <0.4 average pts = 0	
• • •							
Supply chain excellence quality	20 pts	**Deductions based on:** Class 3 recall/withdraw: -2 Class 2 recall: -4 Class 1 recall: -6		Incomplete recall notices: -2 Concealed shortage on control substances or list chemicals: -2 Complete, accurate and prompt recall notices: +1 (bonus)			

FIGURE 7.—Intermediary scoring criteria

Note: The figure presents an example of an intermediary's supplier scoring criteria (Cardinal Health, 2020). For brevity, it is shortened at the ellipsis. Approximately 60% of the possible points are related to the manufacturer's historical recalls and supply disruptions.

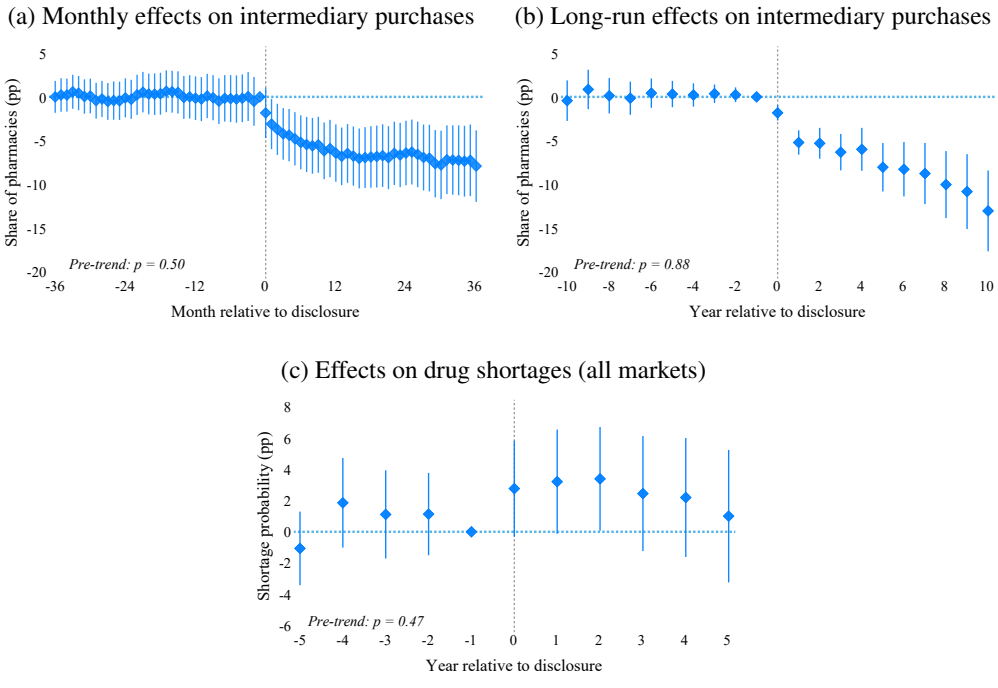


FIGURE 8.—Quasi-experimental effects of recall announcements: Timing and shortages

Note: The figure displays the estimated dynamic effects of recall disclosures on intermediary drug purchases and shortages, obtained using Equation (1). The sample includes multisource generic drugs in Optum claims, and the unit of observation is a drug- or market-period. The year or month prior to the first recall is the omitted reference. Panels (a)–(b) present the effects from Figure 2 panel (a) in additional detail, with the share of pharmacies as the dependent variables. Panel (a) zooms in on the three years before and after the recall using monthly data. Panel (b) extends the event study to ten years before and after the recall. Finally, panel (c) shows the estimated effects of recalls on drug shortages, measured using UUDIS data. The figures present 95% confidence intervals with standard errors clustered by drug or market, and p-values from F-tests of the joint significance of pre-treatment coefficients.

TABLE VI

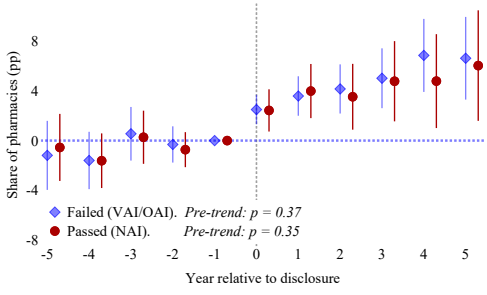
DESCRIPTIVE STATISTICS: INSPECTIONS

	(1) All drugs	(2) Failing drugs	(3) Passing drugs
Drug characteristics			
Years since first generic entry	17.31	17.15	17.74
Fraction unit type:			
Oral pill	0.74	0.75	0.73
Injection	0.07	0.07	0.08
Topical & other	0.19	0.18	0.20
Extended release	0.11	0.10	0.12
Market share			
Share of pharmacies	0.101	0.100	0.103
	(0.197)	(0.196)	(0.199)
Mean annual prescriptions	9,786	10,405	8,155
	(36,550)	(38,497)	(30,791)
Mean drugs per market	11.11	10.97	11.49
N drugs	6,633	4,940	1,693

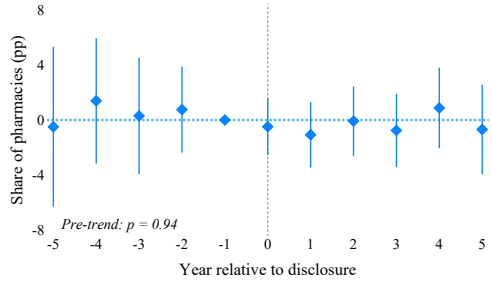
Note: The table presents descriptive statistics for multisource generic drugs in Optum claims data during 2000-2022. Column (1) shows statistics for all inspected drugs, column (2) for drugs with grades of Voluntary or Official Action Indicated ("failing"), and column (3) for drugs with a grade of No Action Indicated ("passing"). Statistics are drug-level means, and corresponding standard deviations are reported in parentheses for the primary volume outcome variables. Years since generic entry is calculated as of 2022. Annual volumes and shares of pharmacies are observed in the claims data.

All grades (NAI vs. VAI, OAI)

(a) Inspected drug share of pharmacies (DD)

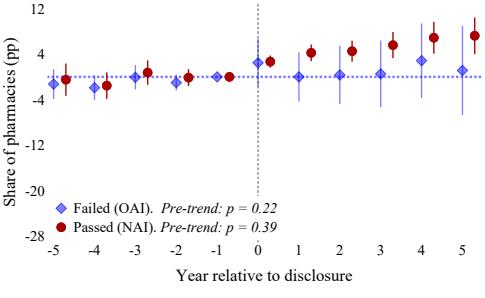


(b) Failed drug share of pharmacies (DDD)



Extreme grades (NAI vs. OAI)

(c) Inspected drug share of pharmacies (DD)



(d) Failed drug share of pharmacies (DDD)

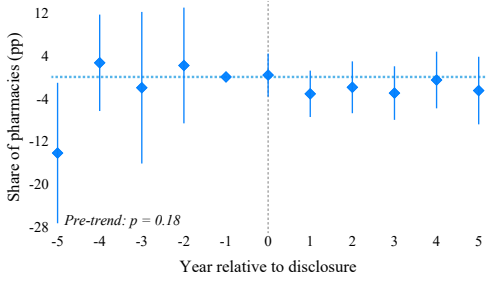


FIGURE 9.—Effects of inspection reports on intermediary purchases

Note: The figure displays the estimated dynamic effects of inspection disclosures on intermediary drug purchases. The sample includes multisource generic drugs in Optum claims, and the unit of observation is a drug-year. The year prior to the first inspection is the omitted reference. Panels (a) and (c) present the effects of disclosing passing and failing inspections, obtained using Equation (1) with the share of pharmacies as the dependent variables. Panels (b) and (d) present the differential effects of failing inspections, obtained using using a DDD specification (Equation (1) with $\gamma D_{dt} \times 1\{\texttt{failed}_d\}$). Passing is defined as receiving a "No Action Indicated" (NAI) grade from the FDA. In panels (a)–(b), failing is defined as receiving a "Voluntary" (VAI) or "Official Action Indicated" (OAI) grade. In panels (c)–(d), this definition is restricted to OAI grades only. The figures present 95% confidence intervals with standard errors clustered by drug, and p-values from F-tests of the joint significance of pre-treatment coefficients.

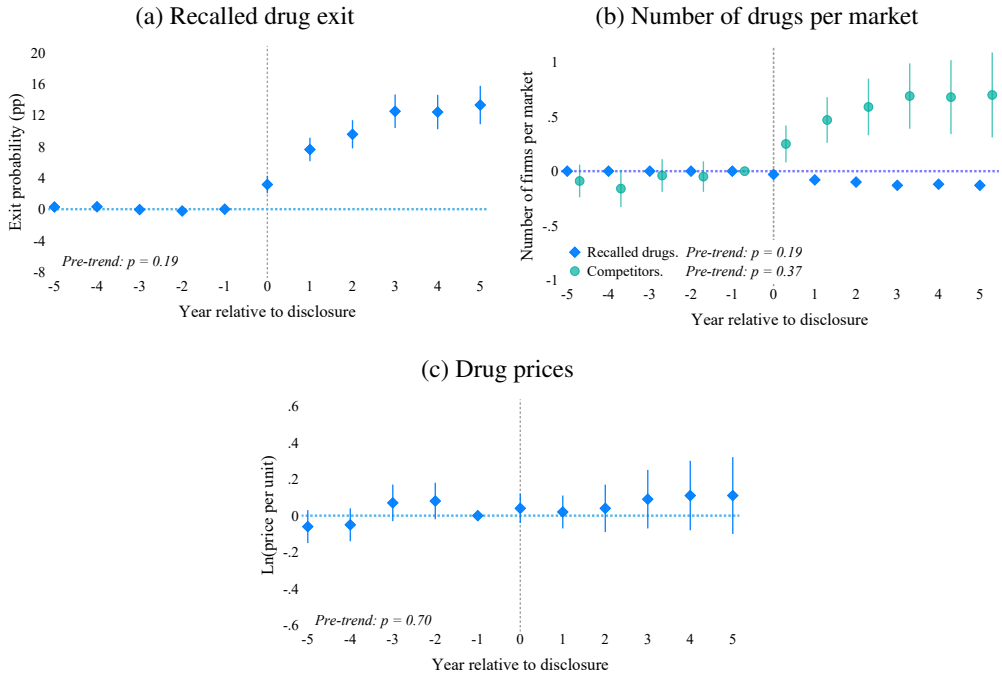


FIGURE 10.—Effects of recalls on market structure and prices

Note: The figure displays the estimated dynamic effects of recalls on market structure and prices, obtained using Equation (1). The sample includes multisource generic drugs in Optum claims, and the unit of observation is a drug- or market-year. The year prior to the first inspection is the omitted reference. Panel (a) presents the effects on the recalled drug's probability of exiting the market, where the dependent variable is an indicator equal to one in the last year before a drug disappears from the Optum data (for at least two consecutive years). Panel (b) presents the effects on the total counts of recalled drugs and bioequivalent competitors. Panel (c) displays the effects on log average wholesale prices, weighted by pre-recall prescription volumes. Prices are market-level averages of pharmacy acquisition costs per drug unit from state and national surveys and inflated to 2022 dollar terms. The figures present 95% confidence intervals with standard errors clustered by drug or market, and p-values from F-tests of the joint significance of pre-treatment coefficients.

TABLE VII

EFFECTS OF INSPECTION REPORTS ON INTERMEDIARY PURCHASES

	(1)	(2)	(3)	(4)	(5)	(6)
	Difference-in-difference estimates				Triple-difference estimates	
	Share of pharmacies		Ln(scripts)		Share of	Ln(scripts)
Dependent var.	Passed	Failed	Passed	Failed	pharmacies	
A - VAI & OAI effects						
Inspection * post	0.033	0.026	0.841	1.013	0.033	1.093
	(0.010)	(0.007)	(0.102)	(0.069)	(0.009)	(0.102)
Inspection * post * failed					-0.008	0.040
					(0.011)	(0.123)
Observations	7,071	11,116	7,071	11,116	14,080	14,080
Dependent var. mean	0.067	0.078	1,075	1,318	0.081	1,389
B - OAI effects						
Inspection * post	0.027	0.014	1.085	0.664	0.031	1.140
	(0.006)	(0.023)	(0.059)	(0.226)	(0.009)	(0.098)
Inspection * post * failed					-0.018	0.266
					(0.023)	(0.229)
Observations	13,341	4,846	13,341	4,774	5,281	5,281
Dependent var. mean	0.080	0.054	1,362	862	0.081	1,389

Note: The table reports the estimated average effects of inspection disclosures on intermediary drug purchases. Cols. (1)–(4) present the effects of disclosing passing (odd cols.) and failing inspections (even cols.), obtained using Equation (2) on drug-year level data. Cols. (5)–(6) present the differential effects of failing inspections, obtained using using a DDD specification (Equation (2) with $\gamma\, D_{dt} \times 1\{\texttt{failed}_d\}$). The dependent variables are each drug's share of pharmacies (cols. (1), (2) and (5)) and logged prescriptions (cols. (4), (5) and (6)). Passing is defined as receiving a "No Action Indicated" (NAI) grade from the FDA. In panel A, failing is defined as receiving a "Voluntary" (VAI) or "Official Action Indicated" (OAI) grade. In panel B, this definition is restricted to OAI grades only. Standard errors are clustered at the drug- or market-level and reported in parentheses. Logged dependent variable means are exponentiated for clarity.

TABLE VIII

EFFECTS OF RECALLS ON MARKET STRUCTURE AND PATIENT DRUG ADHERENCE

	(1)	(2)	(3)	(4)	(5)
		Market structure		Patient drug adherence	
Dependent var.	Recalled drug Pr(exit)	Competitors ln(N drugs)	Net market ln(N drugs)	Pr(adherence > 80%)	Pr(adherence > drug med.)
Recalled * post	0.068	0.124	0.094	0.021	0.033
	(0.005)	(0.029)	(0.025)	(0.006)	(0.006)
Observations	12,380	6,420	6,531	20,005	20,005
Dependent var. mean	0.043	4.027	5.743	0.408	0.492
Recalled * post (%)	157%	13%	10%	5%	7%

Note: The table reports the estimated average effects of recall disclosures on market structure and patient drug adherence, obtained by estimating Equation (2) on drug- or market-level data. Cols. (1)–(3) present the effects on market structure, including recalled drug exit (col. (1)), log counts of bioequivalent competitors (col. (2)), and log counts of market participants (col. (3)). Cols. (4)–(5) present effects on patient adherence, including the fraction of patients with MPRs less than 80% (col. (5)) or the drug median (col. (6)). Standard errors are clustered at the drug (cols. 1–2) or market (cols. 3–5) levels and reported in parentheses. Logged dependent variable means are exponentiated for clarity.

TABLE IX

QUASI-EXPERIMENTAL ROBUSTNESS

	Recalled drug share of pharmacies (%)
1. Baseline estimate	-42.7
	(7.1)
Alternative dependent variable	
2. Share of scripts	-63.1
	(7.7)
3. Share of first scripts	-53.4
	(7.1)
Alternative specifications	
4. Volume weighted	-49.5
	(11.4)
5. Inspection covariates	-43.7
	(11.2)
6. Never-recalled control group	-19.4
	(5.0)
7. Borusyak Jaravel Spiess	-50.5
	(8.7)
8. Callaway Sant'Anna	-29.7
	(5.0)
9. Cluster at market level	-42.7
	(7.2)
10. Cluster at firm level	-42.7
	(9.0)
Alternative samples	
11. Exclude exiting drugs	-27.5
	(8.3)
12. Exclude growing markets	-33.2
	(11.3)
13. Balanced panel	-20.4
	(5.8)
14. Exclude COVID-19	-44.7
	(7.3)
15. Oral tablets	-36.9
	(6.8)
Alternative market definitions	
16. NADAC markets	-38.6
	(4.7)
17. ATC-4 markets	-175.4
	(76.6)
Observations	12,380
Dependent var. median	0.103

Note: The table reports the estimated average effects of recall disclosures on intermediary drug purchases, obtained using Equation (2) with a variety of dependent variables, specifications, samples, and market definitions. The first row reproduces the baseline estimate from Table II col. (1), using each drug's share of pharmacies as the dependent variable. Rows 2 and 3 alter the dependent variable, using shares of prescriptions and first prescriptions, respectively. Rows 4 and 5 are obtained using specifications with pre-treatment volume weights and inspection grade covariates, respectively. Row 6 uses never treated drugs as the control group. Rows 7–8 include corrections for staggered treatment timing from Borusyak et al. (2024) and Callaway and Sant'Anna (2021), while rows 9–10 cluster standard errors at more aggregated levels than the treated unit (the drug). Rows 11–15 alter the estimation sample, by excluding drugs that exit the market within 5 years of treatment (11) and drugs in markets that experience net entry (12). Row 13 uses treated drugs with exactly three observations before and after treatment and control drugs with at least seven observations. Rows 14–15 exclude observations from the COVID-19 pandemic (2019+) and non-oral drug formulations, respectively. Rows 16–17 narrow and broaden the definition of a drug market, using molecule-formulation-strength combinations (the "NADAC" definition) and ATC-4 groups. To facilitate comparison, all effects are presented as percentages of the dependent variable medians. Unless otherwise noted, standard errors are clustered at the drug-level and reported in parentheses.

4. ADDITIONAL EVIDENCE ON THE SCORING AUCTION MODEL

TABLE X

PRICE HOMOGENIZATION

Dependent var.	Price per unit
Volume	-0.100
Volume^2	0.000
Extended release	0.964
Intermediary 2	-0.209
Intermediary 3	-0.287
Intermediary 4	-0.327
Intermediary 5	-0.208
Therapeutic area (ATC-1)	
Antiparasitics and insecticides	0.999
Antiinfectives	0.810
Sensory organs	0.310
Nervous system	0.149
Antineoplastic and immunomodulating agents	0.135
Dermatologicals	0.040
Alimentary tract and metabolism	-0.156
Systemic hormonal preparations	-0.166
Genito urinary system and sex hormones	-0.174
Respiratory system	-0.189
Musculo-skeletal system	-0.195
Cardiovascular system	-0.342
Blood and blood forming organs	-0.600
Other therapeutic areas	0.911
N. bidders fixed effects	x
Year fixed effects	x
Observations	84,403
Dependent var. mean	1.04

Note: The table reports the estimated coefficients from a linear regression of drug prices on auction characteristics. The sample includes multisource, oral generic drugs in Optum claims, and the unit of observation is an auction. The dependent variable is the intermediary-specific winning price per unit in 2022 dollars. Annual drug volumes are in millions of units and are observed in Optum claims data. Therapeutic areas (ATC-1 groups) are from the National Library of Medicine. All other drug characteristics are from the NDC Directory. Robust standard errors are reported in parentheses.

TABLE XI

DESCRIPTIVE STATISTICS: AUCTIONS

Variable	Mean
Winning price per unit	0.63
	(0.92)
Annual prescriptions (mm)	12.09
	38.91
N bidders per auction	4.62
Fraction with $\geq$ 1 recalled bidders	0.25
with 1 recalled bidder	0.19
with 2 recalled bidders	0.04
with 3+ recalled bidders	0.01
N Auctions	78,263
N Observations	313,800

Note: The table presents descriptive statistics for generic drug procurement auctions during 2010-2022. It presents means and standard deviations (in parentheses) using auction-level data. Prices are intermediary-specific winning prices per unit in 2022 dollars, trimmed at the 95th percentile. Annual drug volumes are in millions of units and are observed in Optum claims data. Numbers of bidders are from Optum claims and the FDA Orange Book. Recalled bidders include all drugs recalled within the last five years.

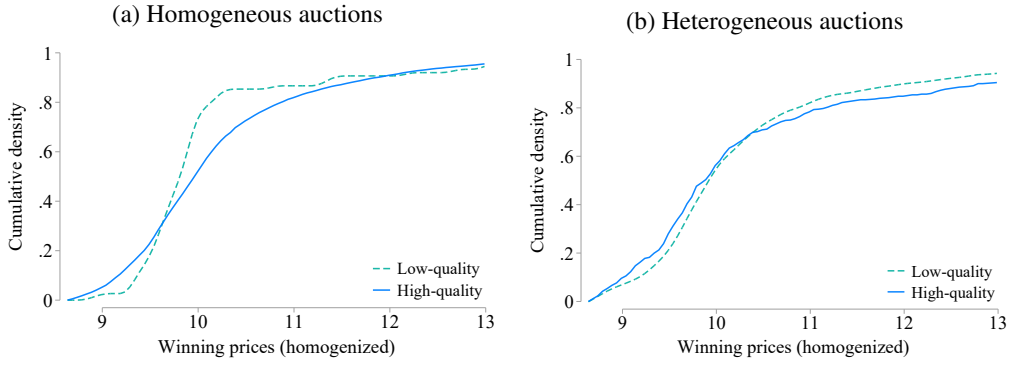


FIGURE 11.—Cumulative densities of winning prices by auction type

Note: The figures present the kernel densities of homogenized winning prices from generic procurement auctions with $N = 2$ bidders. Panel (a) plots the densities for homogeneous auctions—with uniform high- (solid lines) or low-quality (dashed lines) bidders—while panel (b) plots the corresponding densities for heterogeneous auctions. Prices are intermediary-specific winning prices per unit in 2022 dollars, trimmed at the 95th percentile and homogenized using the methodology described in Section 4.3. "Low-quality" drugs have experienced a recall within the last five years.

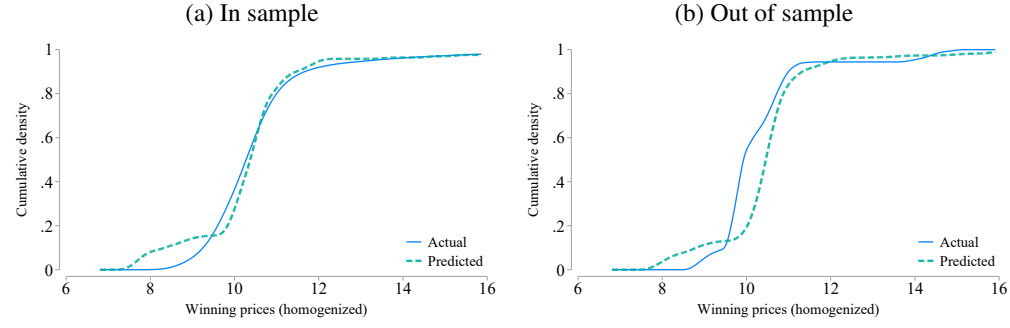


FIGURE 12.—Model fit: predicted post-recall prices

Note: The figure presents the kernel densities of winning prices after recalls, comparing the structural model's predictions against the observed data. The model uses *ex ante*, pre-recall data to predict post-recall outcomes. Panel (a) presents densities using the estimation sample, while panel (b) uses out of sample data from 2022.

TABLE XII

SCORING AUCTION MODEL ROBUSTNESS

	(1) Buyer preferences (fraction of mean price)	(2) Bidder costs (high vs. low quality)
1. Optimal score (baseline)	0.13	0.25
2. Long-run quality	0.12	0.21
3. Auctions with $N = 2$ bidders	0.10	0.11
4. Recalled & not-yet-recalled drugs	0.18	0.24
5. Balanced panel	0.25	0.27
6. Linear score	0.34	0.75

Note: The table presents estimates of buyer preferences and bidder costs obtained after perturbing the estimation sample and assumptions of the scoring auction model. Col. (1) presents estimates of buyer willingness to pay for quality drugs, expressed as a fraction of the mean price. Col. (2) reports the average difference in high- and low-quality drug costs per unit, expressed as a fraction of low-quality costs. Row 1 presents the baseline model estimates, assuming that buyers use an optimal scoring rule, and that "low-quality" is defined as being recalled in the last five years. Row 2 reports estimates using a stricter definition of quality, including drugs recalled within the last decade. Row 3 presents estimates from auctions with only $N = 2$ bidders. To isolate how recalled drug costs change after disclosure, row 4 limits the estimation sample to recalled and not-yet-recalled drugs. Row 5 reports estimates using a "balanced" panel of auctions that experience no entry or exit. The final row presents estimates assuming that buyers use a linear score.

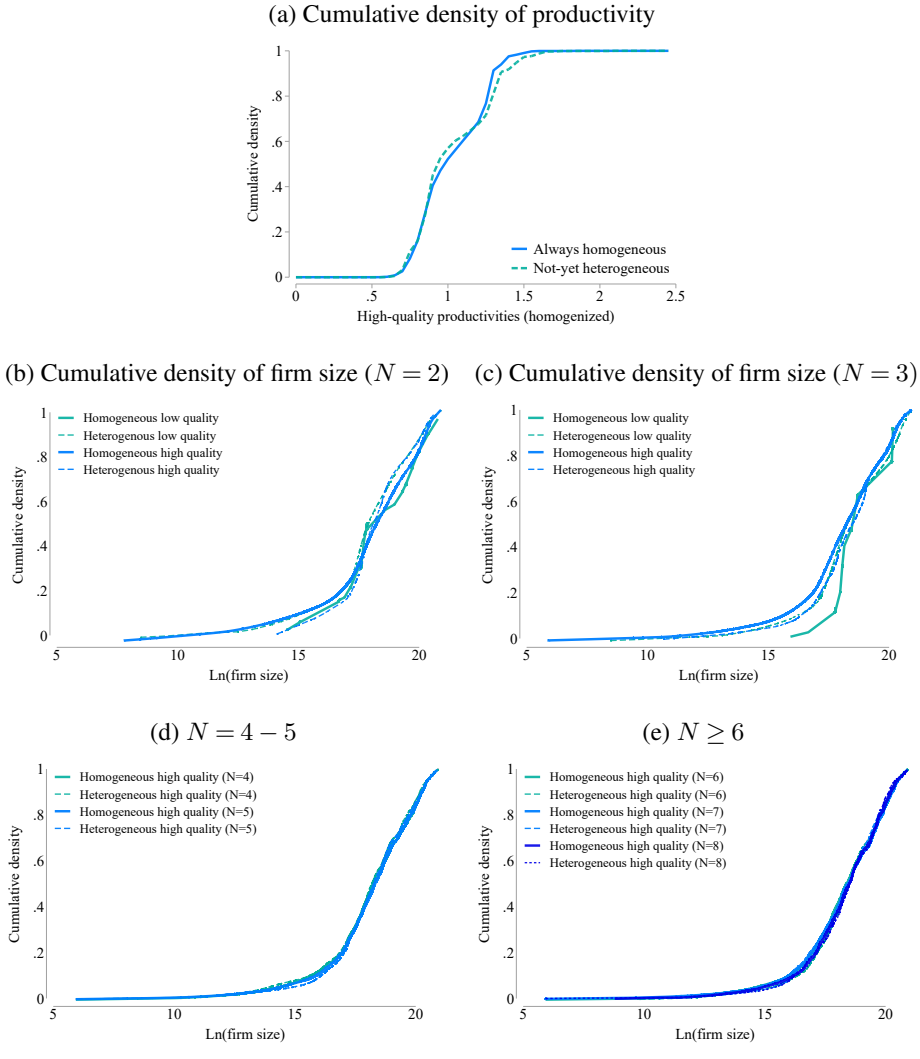


FIGURE 13.—Balance in productivity and firm size in homogeneous & heterogeneous auctions

Note: The figure provides evidence to support the identifying assumption of the scoring auction model: that productivities are identically distributed in auctions with homogeneous (uniform quality) and heterogeneous (mixed quality) bidders. It presents the kernel densities of estimated productivities and firm sizes in each auction type. Panel (a) presents estimated productivities for high-quality drugs, obtained using homogeneous and heterogeneous auctions before the market's first recall ("always homogeneous" and "not yet heterogeneous," respectively). These densities imply a difference in ex ante choice probabilities in estimating Equation (9) of less than 0.001. I am able to reject differences in mean productivities greater than 1% of the overall mean using two one-sided t-tests. Panels (b)–(c) present firm sizes in homogeneous and heterogeneous auctions in the main estimation sample, ordered by auction size (N bidders). I am able to jointly reject differences in means greater than 0.7% of the overall sample (or greater than 4.0% for each individual N) at $p < 0.01$. "Low-quality" drugs have experienced a recall within the last five years. Firm size is measured using annual log volumes observed in the Optum claims.

5. TESTING THE IDENTIFYING ASSUMPTION OF THE SCORING AUCTION MODEL

6. SCORING AUCTION ESTIMATION ALGORITHM

6.1. *Unobserved prices.*

I first derive the system of equations that identifies the unobserved prices in Section 4.3.1. For convenience, these equations are reproduced below. Next, I describe the estimation algorithm

and the conditions under which a solution to this system is guaranteed to exist.

$$Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) = \left[1 - G_L(b_{H,i} - \Delta S)\right]^{N_L} \left[1 - G_H(b_{H,i})\right]^{N_H - 1} g_H(b_{H,i}) \quad (11)$$

$$Pr(b_{L,i}|i \text{ wins})Pr(i \text{ wins}) = \left[1 - G_H(b_{L,i} + \Delta S)\right]^{N_H} \left[1 - G_L(b_{L,i})\right]^{N_L - 1} g_L(b_{L,i}) \quad (12)$$

Derivation. I rewrite the distribution of the observed winning prices using Bayes' Rule:

$$Pr(b_{t,i}|i \text{ wins}) = \frac{Pr(i \text{ wins}|b_{t,i})Pr(b_{t,i})}{Pr(i \text{ wins}),} \quad (13)$$

where I suppress conditioning on $q_i = t, q_{-i}$ throughout. Under the linear scoring rule, bidder i wins i.f.f.

$$S(q_i) - b_i \geq S(q_j) - b_j \;\; \forall j \in \{1, ..., N\}, \;\; j \neq i.$$

Without loss of generality, I focus on the case of $N = 2$ types, $t \in \{H, L\}$. After substituting the definition of the linear score into Equation (13) and rearranging terms, we have

$$\begin{aligned} Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) &= Pr(b_{H,i})\left[Pr(S(q_i) - b_i \geq S(q_j) - b_j \;\; \forall j \in \{1, ..., N\}, \;\; j \neq i \mid b_{H,i})\right] \\ &= Pr(b_{H,i})\left[Pr(b_L \geq b_{H,i} - \Delta S \mid b_{H,i})\right]^{N_L} \left[Pr(b_H \geq b_{H,i} \mid b_{H,i})\right]^{N_H - 1} \\ &= g_H(b_{H,i})\left[1 - G_L(b_{H,i} - \Delta S))\right]^{N_L} \left[1 - G_H(b_{H,i})\right]^{N_H - 1}. \end{aligned}$$

The second line holds because b_i is independently and identically distributed (conditional on q_i, q_{-i}). Note that b_L is a random variable corresponding to the price of another bidder of type L; b_H is similarly defined. The last line substitutes in the distributions of $b_t \sim g_t(\cdot)$. Equation (12) can be derived in the same way (after switching $t = H, L$.)

Note that similar equations can be easily derived for quality-to-price scores, another common scoring format ($V = S(q_i)/b_i$) and non-linear scores (e.g., $V = S(q_i) - b_i^2$). For example, using a quality-to-price score, the analogues to Equation (11) and (12) are

$$\begin{aligned} Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) &= g_H(b_{H,i})\left[1 - G_L(b_{H,i}r))\right]^{N_L} \left[1 - G_H(b_{H,i})\right]^{N_H - 1} \\ Pr(b_{L,i}|i \text{ wins})Pr(i \text{ wins}) &= g_L(b_{L,i})\left[1 - G_H(b_{L,i}/r))\right]^{N_H} \left[1 - G_L(b_{L,i})\right]^{N_L - 1}, \end{aligned}$$

where the normalized quality preference is $r = \frac{S(q_L)}{S(q_H)}$.

6.2. *Estimation algorithm.*

Together, Equations (11) and (12) define a system of two nonlinear functional equations in two unknowns, $g_L(b_L)$ and $g_H(b_H)$, at each point price in the support. To solve this system, I first estimate the LHS objects using the observed data and standard non-parametric kernel density estimators (Li et al., 2000). I then discretize the prices b and solve for $g_L(b_L)$ and $g_H(b_H)$ at each point in the price grid, starting with the lowest price $\underline{b}$. Let ε refer to the difference between two consecutive points in the discretized price grid.

For all $b_H < \underline{b} + \Delta S$, type H wins with certainty, implying that the system reduces to a single equation:

$$Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) = g_H(b_{H,i})\left[1 - G_H(b_{H,i})\right]^{N_H - 1}. \quad (14)$$

At the lowest price $b_H = \underline{b}$, we further have

$$Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) = g_H(\underline{b})\left[1 - g_H(\underline{b})\right]^{N_H - 1},$$

where I have used the fact that the empirical cdf in the leftmost bin is equal to the pdf in that bin. We can solve this equation for $g_H(\underline{b})$, as the root of an N_H^{th} degree polynomial in $g_H(\underline{b})$. We can then solve for subsequent $g_H(b_H)$ using Equation (14) and approximating $G_H(b_H)$ with the empirical cdf, namely the Riemann sum of the prior solutions (for $g_H(b_H)$ up to $b_H < \underline{b} + \Delta S$). Note that the boundary condition for $g_L(\underline{b} - \Delta S)$ is not yet defined. I now explain how to iteratively construct $g_H(b_H)$ and $g_L(b_H - \Delta S)$.

More generally, for all b, we can rewrite Equations (11) and (12) as

$$l = g_H(b_{L,i} + \Delta S)\left[k - g_L(b_{L,i})\varepsilon\right]^{N_L}\left[j - g_H(b_{L,i} + \Delta S)\varepsilon\right]^{N_H - 1} \tag{15}$$

$$m = g_L(b_{L,i})\left[j - g_H(b_{L,i} + \Delta S)\varepsilon\right]^{N_H}\left[k - g_L(b_{L,i})\varepsilon\right]^{N_L - 1}, \tag{16}$$

where we observe N_H, N_L and ε, have guessed ΔS, and have already estimated j, k, l and m (as defined below).

$$l = Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) \quad k = 1 - \sum_{x=\underline{b}}^{b_L - \varepsilon} g_L(x)\varepsilon$$

$$m = Pr(b_{L,i}|i \text{ wins})Pr(i \text{ wins}) \quad j = 1 - \sum_{x=\underline{b}}^{b_L + \Delta S - \varepsilon} g_H(x)\varepsilon.$$

This leaves two equations in two unknowns ($g_H(b_H = b_L + \Delta S)$ and $g_L(b_L)$), which can be solved for each price where bidding behavior is defined.[1] Note that objects j, k, l and m are all positive, since j and k are cdfs, and l and m are products of pdfs.

6.3. *Proofs of existence.*

I now show that a solution to the non-linear system defined by Equations (15) and (16) exists at each price, for up to $N = 3$ bidders. For larger auctions, the curse of dimensionality makes estimating the LHS objects increasingly (exponentially) impractical. To avoid this problem, Laffont et al. (2020) recommend using the estimates from only the smallest auction size, which is typically be $N = 2$. For $N = 2$ auctions, I also describe the conditions under which the solution is unique and bounded in $[0, 1]$.

Equations (15) and (16) can be rewritten, after some algebra, as (1) a polynomial of order N in the first unknown, and (2) an expression that defines the second unknown in terms of the first. From here, proving existence relies on well-known properties of polynomial functions. For example, for $N = 2$ ($N_H = 1, N_L = 1$), we can write

$$0 = k\varepsilon x^2 - (jk - m\varepsilon + l\varepsilon)x + jl \tag{17}$$

$$y = \frac{m}{j - \varepsilon x}. \tag{18}$$

[1] Bidding behavior is not defined for prices $b_L > \overline{b} - \Delta S$ and $b_H > \overline{b} + \Delta S$, where one bidder wins or loses with certainty.

For brevity, I have replaced $x = g_H(b_L + \Delta S)$ and $y = g_L(b_L)$. A real solution exists as long as $x \neq j/\varepsilon$ and the discriminant of the quadratic equation (17) is non-negative:

$$(jk - m\varepsilon + l\varepsilon)^2 - 4k\varepsilon jl \geq 0.$$

This condition is satisfied as the grid of prices becomes increasingly fine (as $\varepsilon \to 0$). Taking the limit of the Quadratic Equation as $\varepsilon \to 0$ implies that the two roots of x approach $\{\frac{l}{k}, \infty\}$, and $y = \frac{m}{j}$. Each root is positive since all of the quantities involved (j, k, l, m) are positive. Furthermore, x has a unique solution in $[0, 1]$ if $l \leq k$, or

$$Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) \leq 1 - G_L(b_L - \varepsilon).$$

Likewise, y is in $[0, 1]$ if $m \leq j$, or

$$Pr(b_{L,i}|i \text{ wins})Pr(i \text{ wins}) \leq 1 - G_H(b_L + \Delta S - \varepsilon).$$

In sum, a solution for $N = 2$ exists as long as the grid of prices is sufficiently fine, and a unique positive solution is bounded in $[0, 1]$ if $l \leq k$ and $m \leq j$.

For $N = 3$ ($N_H = 2, N_L = 1$), we can write

$$0 = k\varepsilon^2 x^3 - 2jk\varepsilon x^2 + (-m\varepsilon + j^2 k + l\varepsilon)x - jl \tag{19}$$

$$y = \frac{m}{(j - \varepsilon x)^2} \tag{20}$$

Similarly, for ($N_H = 1, N_L = 2$),

$$0 = j\varepsilon^2 y^3 - 2jk\varepsilon y^2 + (m\varepsilon + jk^2 - l\varepsilon)y - km \tag{21}$$

$$x = \frac{l}{(k - \varepsilon y)^2} \tag{22}$$

Since Equations (19) and (21) are odd degree polynomials, they each have at least one real solution. Furthermore, for small values of ε, they each have three sign changes, implying that they must have at least one positive root each (by Descartes's Rule of Signs). Equations (20) and (22) have positive solutions as long as $x \neq j/\varepsilon$ and $y \neq k/\varepsilon$.

6.4. *Monte Carlo studies.*

To illustrate the estimation algorithm, described in Section 4.3, I conduct a range of Monte Carlo studies. I first show that my algorithm for linear scores recovers the unobserved price distributions. Next, I compare the performance of my full estimation procedure (which uses only winning prices) to Laffont et al.'s (using complete data).

Each Monte Carlo experiment consists of 5,000 replications of N = 5,000 auctions, which corresponds to a realistic auction sample size.[2] To show that my estimation routine recovers the unobserved price densities for linear scores, I first draw prices from log-normal distributions conditional on bidder type. I then identify the winning bids using the linear scoring rule, remove all non-winning prices, and apply my estimation algorithm to recover the price densities. I

[2]Note, however, that this is an order of magnitude smaller than my data.

repeat this exercise using a range of parameters; for exposition, I show the results from one such experiment in Figure 6.4. This uses the following parameter values:

$$N = 2,\ \ (q_i, q_j) = (1, 0),\ \ b|q = 0 \sim \text{LogN}(\text{shape} = 1.0, \text{loc} = 0.0, \text{scale} = 0.9),$$
$$b|q = 1 \sim \text{LogN}(\text{shape} = 1.0, \text{loc} = 0.5, \text{scale} = 1.0),\ \ \Delta S = 0.7.$$

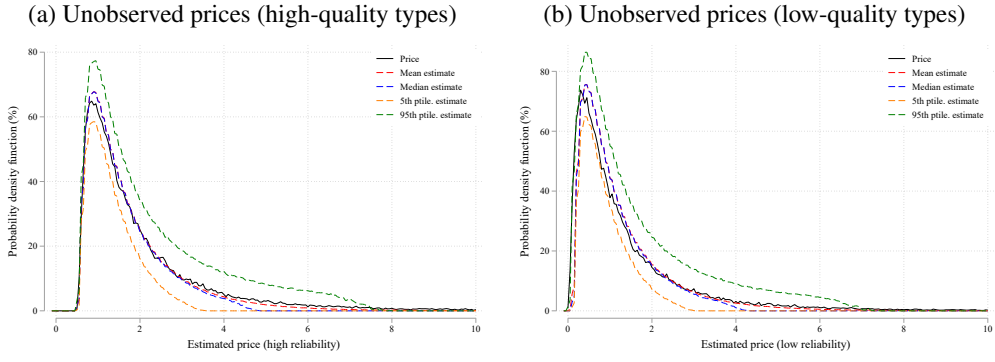


FIGURE 14.—Monte Carlo simulations of prices: Linear score

Figure 6.4 plots the true and estimated price densities for high- and low-quality types (in panels (a) and (b), respectively). The true densities are plotted using solid black lines, while the means, medians and pointwise 90% confidence intervals of the estimates are plotted using dashed lines. The true values are strikingly similar to the mean and median estimates (conditional on a guess of the true value of ΔS), and they appear to fall between the confidence intervals over the whole support.

To show that my full estimation routine recovers the model primitives, and to compare its performance against Laffont et al.'s, I conduct an additional set of experiments. For each replication, I draw productivities from log-normal distributions conditional on bidder type. I then numerically compute the corresponding bids using Equation (5) and apply Laffont et al.'s estimation algorithm. Finally, I leave out all non-winning prices—reducing the sample size by at least half—and apply my algorithm. Again, I repeat the experiment for a range of parameter values and auction sizes, which produces qualitatively similar results. I present the results from one such experiment, with the following parameter values:

$$N = 2,\ \ q \sim \text{Bernoulli}(0.7),\ \ \theta|q = 0 \sim \text{LogN}(0.2, 0.3),\ \ \theta|q = 1 \sim \text{LogN}(0.7, 0.3),\ \ \Delta S = 0.7.$$

The results are summarized in Figure 6.4, with estimates from Laffont et al.'s procedure on the left, and from mine on the right. Panels (a)—(d) report the true and estimated conditional densities of the productivities (in solid black and dashed lines, respectively). The figures show that the estimation routines recover the shape, scale and location of the underlying distributions: the true values are very close to the mean and median estimates, and lie in between the confidence intervals for all but small values of θ. These patterns are similar to those from GPV's Monte Carlo exercise, where even after truncation, extreme values are more difficult to capture due to the boundary bias associated with non-parametric estimators. Comparing estimates from Laffont et al.'s algorithm (on the left) and mine (right) suggests that my estimation procedure recovers the productivity distributions with comparable fidelity as the full-data routine.

This is consistent with the idea that the distribution of the minimum price is a sufficient statistic for the underlying price distribution, so leaving out non-winning prices (in this case, half of the sample) does not result in significant information loss about the underlying productivity distributions. Comparing the densities of high (panels (c)—(d)) and low (panels (e)—(f)) quality types provides additional color on this statement; the low-quality densities are somewhat more noisily estimated, but they represent just 4–8% of the original sample. My algorithm, however, does a poorer job than Laffont et al.'s of fitting the left tail of the distributions. This may be because, in limiting my sample to winning bids, I differentially discard the right tail of the distribution (high prices). Since each pdf must each sum to one, this may cause my results to overestimate sample in the left tails.

Panels (e)—(f) plot the distributions of the estimated ΔS parameters less the true value, with a vertical line at the mean estimate. Both distributions are centered around zero and approximately unbiased; the estimates from my algorithm are less precise, consistent with using a smaller sample size. The corresponding means (and standard deviations) are -0.005 (0.091) for Laffont et al. and 0.040 (0.189) for my algorithm.

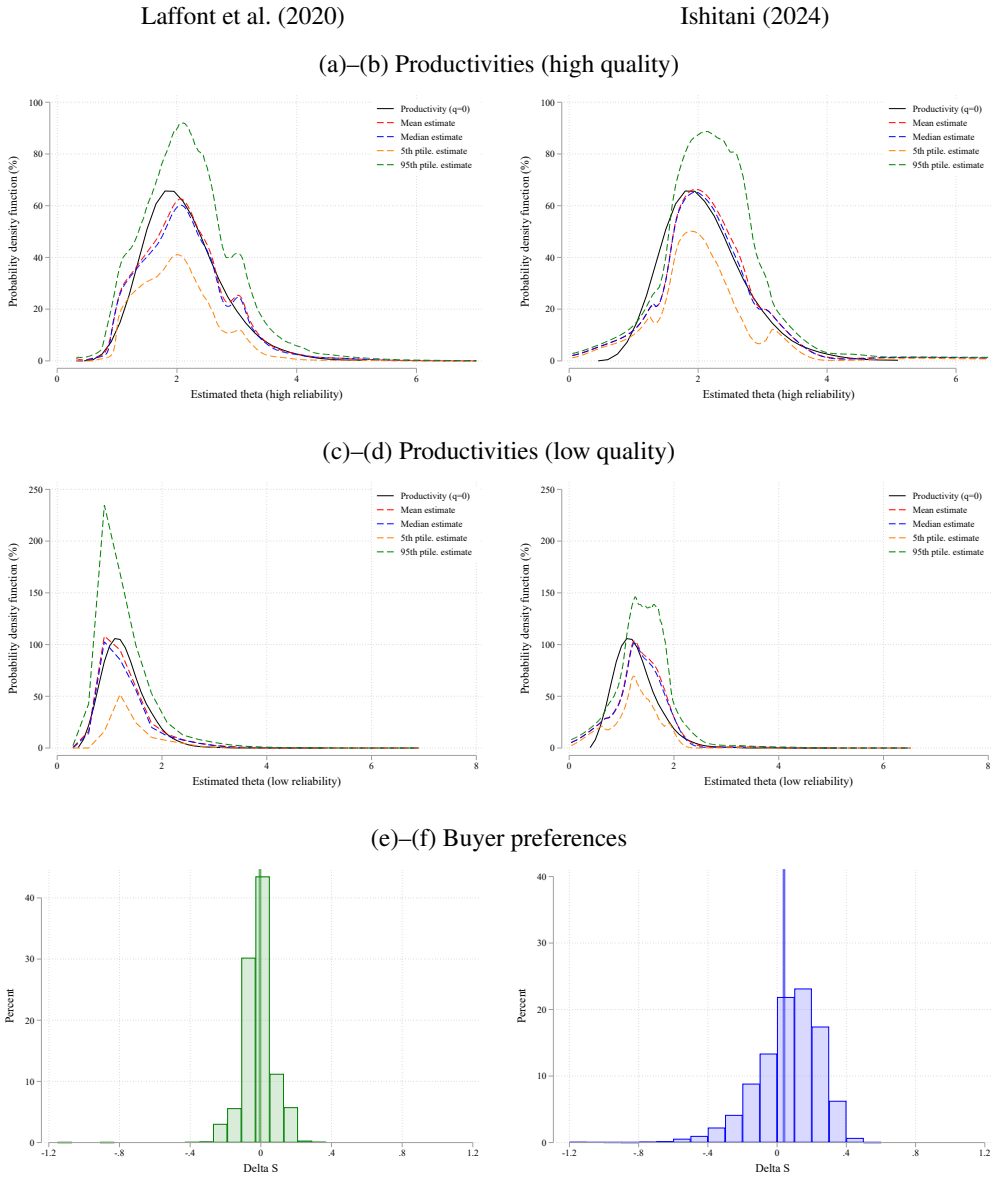


FIGURE 15.—Monte Carlo simulations of preferences and productivities

REFERENCES

BORUSYAK, KIRILL, XAVIER JARAVEL, AND JANN SPIESS (2024): "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, 91 (6), 3253–3285. [11]

CALLAWAY, BRANTLY AND PEDRO HC SANT'ANNA (2021): "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 225 (2), 200–230. [11]

GALDIN, ANAÏS SOLEDAD (2024): "Trade and Information Provision Across Geographical Space," Ph.D. thesis, Princeton University. [1]

LAFFONT, JEAN-JACQUES, ISABELLE PERRIGNE, MICHEL SIMIONI, AND QUANG VUONG (2020): "Econometrics of scoring auctions," in *Essays in Honor of Cheng Hsiao*, Emerald Publishing Limited, vol. 41, 287–322. [17]

LI, TONG, ISABELLE PERRIGNE, AND QUANG VUONG (2000): "Conditionally independent private information in OCS wildcat auctions," *Journal of Econometrics*, 98 (1), 129–161. [16]

CENTER FOR INFECTIOUS DISEASE RESEARCH & POLICY (2018): "Resilient Drug Supply: Critical Acute Drugs," https://www.cidrap.umn.edu/resilient-drug-supply/resilient-drug-supply-critical-acute-drugs. [1]

NATIONAL COMMUNITY PHARMACISTS ASSOCIATION (2022): "Survey of Community Pharmacy Economic Pressures: 2022 Report," https://ncpa.org/sites/default/files/2022-08/Pharmacy [4]

NEWKIRK, MARGARET AND SUSAN BERFIELD (2019): "The FDA Drug Recall System is Voluntary, Haphazard, and Broken," *Bloomberg Businessweek*. [4]

NOH, IN JOON, JOHN GRAY, GEORGE BALL, ZACHARY WRIGHT, AND HYUNWOO PARK (2025): "Are all generic drugs created equal? An empirical analysis of generic drug manufacturing location and serious drug adverse events," *Production and Operations Management*, 10591478251319691. [1]