# In Defense of the Middleman: Quality Failures in the Generic Pharmaceutical Market*

Catherine Ishitani†

March 20, 2025

Click here for latest version

**Abstract**

This paper studies how intermediary buyers, such as wholesalers and retailers, discipline quality in the market for generic prescription drugs. Consumers are uniquely uninformed about quality in this market, since generic drugs are conventionally assumed to be interchangeable. As in many settings where consumers lack product information, they implicitly rely on intermediaries to observe quality disclosures and select high-quality goods for them. While intermediaries can alleviate adverse selection problems, they may also face incentives to choose inefficiently low-quality goods. Using novel data—the universe of FDA manufacturing quality disclosures from 2000-2022—I first provide evidence that quality failures are pervasive, even in the U.S. market. Over half of manufacturers fail inspections during my study period, and 12% of drugs are recalled, due to their risks to patient health. Next, I show that the disclosure of these failures through recall announcements reduces intermediary purchases of recalled drugs by 60%, with effects persisting for up to a decade—long after recalls are typically resolved. In contrast, the disclosure of inspection failures has little impact on drug purchases, suggesting that intermediaries penalize manufacturers more stringently for signals of supply disruptions than for low quality alone. To isolate the role of intermediaries, I develop a scoring auction model of generic procurement and an estimation technique that takes into account the fact that only winning prices are observed. I find that intermediaries are willing to pay a 2% premium for each 10% reduction in the probability of future recalls, which encourages manufacturers to compete on reliability, rather than just price. Intermediaries enhance the benefits of recall disclosures and ultimately increase the share of high-quality (not recalled) drugs by 27%. Finally, I show that counterfactual pay-for-performance policies, which subsidize high-quality drugs, can improve static welfare but are likely to reduce manufacturer competition in the long-run.

*I am grateful to Alex Olssen, Guy David, Abby Alpert, Atul Gupta, Juan Pablo Atal, Petra Todd, Jeffrey Macher, Alex Azar, Elena Krasnokutskaya, Ashley Swanson, Caitlin Carroll, Genevieve Kanter, Eric Richert, Jacob Dorn, Yihao Yuan, Ruochen Sun, Felipe Barbieri, Jonathan Arnold, Gi Kim, Dragos Potirniche and to participants at the Wharton Healthcare Management and Industrial Organization seminars for helpful guidance and suggestions. Funding from the Mack Institute for Innovation Management and the Leonard David Institute of Health Economics is greatly appreciated. All errors are my own.

†The Wharton School, University of Pennsylvania

# 1 Introduction

Quality disclosure is supposed to improve choice and welfare (Akerlof 1970) but is limited by frictions in consumers' ability to respond to it. Expansive evidence shows that consumers often fail to use disclosed quality information.[1] Instead, in many markets they rely on intermediary buyers to select high-quality goods for them. These firms are pervasive in settings with high degrees of information asymmetry, from healthcare to household goods. For example, the typical consumer trusts that her grocery selects safety-certified produce, that her pharmacy chooses authorized COVID tests, and that her big-box retailer stocks industry-accredited appliances. Despite rich empirical evidence on quality disclosure (Dranove and Jin 2010) and intermediary distortions (Gavazza and Lizzeri 2021) separately, little is known about their combined impact. This paper studies how informed intermediaries[2] discipline quality in the market for generic prescription drugs.

The theoretical benefits of informed intermediaries are straightforward. Compared to consumers, intermediaries are often favorably positioned to observe and react to quality disclosures, due to their scale and experience. Downstream, intermediaries can alleviate adverse selection problems and improve the quality of goods available to consumers. Upstream, they can increase supplier incentives to produce reliable goods. Intermediaries, however, may not fully internalize these social benefits—particularly when quality is difficult to observe and they lack fiduciary responsibility to consumers (Biglaiser 1993).[3] 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 distort *or* enhance the benefits of quality disclosure, with ambiguous implications for equilibrium quality and welfare.

Generic pharmaceuticals[4] are distinguished by an unusually high degree of asymmetric information and intermediation, making them an ideal setting to study informed intermediaries. In contrast to branded drugs, generics are conventionally described as almost perfectly substitutable "commodities" (FTC 2011). Patients and physicians are unlikely to be able to differentiate between generic drugs made by alternative manufacturers, much less between differences in their quality. They implicitly trust intermediaries—group purchasing organizations formed by wholesalers and pharmacies—to select reliable drugs on their behalf. These organizations make purchasing decisions that determine the quality and wholesale prices for nearly all generic drugs in the U.S. market (Fein 2023). For their part, intermediaries promise to deliver high-

1. Evidence from corporate finance, household lending, legal contracts and representation, insurance, automobile fuel economy and healthcare services shows that disclosure routinely fails to protect consumers or improve their decisions (Ben-Shahar and Schneider 2017).

2. The literature on intermediary goods markets refers to these entities as *downstream firms* or *dealers* (Lee, Whinston, and Yurukoglu 2021), focusing on their positions in vertical supply chains. A complementary literature (Gavazza and Lizzeri 2021) on asymmetric information refers to them as *expert intermediaries*, emphasizing their role in reducing information frictions. For simplicity, I refer to them as *intermediaries*.

3. When consumers cannot observe quality, they cannot reward intermediaries with good reputations. In contrast to advisors, the intermediaries in my empirical setting face no legal requirement to act in consumers' best interests, which heightens the degree of their incentive misalignment.

4. A generic drug is a copy of a branded drug whose primary patents have expired. Following the literature, a *drug* or *product* is a molecule-formulation made by a specific manufacturer (e.g., atorvastatin calcium-tablet-Teva).

quality drugs, with slogans like, "It's not just a package, it's a patient" (McKesson 2024).

Most patients perceive that generics are subject to stringent quality regulation, commensurate with their high stakes for health. While generics must meet a minimum bioequivalence standard[5] before entering the market, their *manufacturing quality* ("quality") can deteriorate in the years and decades after. This type of quality assesses whether a drug is pure and performs its approved therapeutic purpose, and is primarily maintained through non-binding disclosures.[6] Specifically, the Food and Drug Administration (FDA) relies on inspection reports and recall announcements to communicate when a drug's quality is potentially injurious or deadly to patient health (FDA 2024b).[7] FDA recall announcements also signal information about a drug's supply reliability, since recalled batches interrupt the flow of goods and are highly predictive of future disruptions.

To study how intermediaries discipline generic quality, this paper does three things. First, it provides evidence of pervasive generic quality failures, even in the U.S. market. Next, it shows that the disclosure of these failures decreases the share of low-quality drugs, despite patients' inability to respond to the disclosed information. Finally, to attribute these effects to intermediaries, it develops a structural model of generic procurement, as well as a new estimation method that requires only winning prices. Using this model, it shows that intermediaries enhance the benefits of quality disclosure and substantially improve the quality of drugs consumed by patients.

I begin by constructing a novel dataset that provides the most comprehensive information on drug manufacturing quality to date. My data include the universe of FDA disclosure actions over 2000-2022, which I obtained through Freedom of Information Act (FOIA) requests. I use drug labels, import manifests and administrative records to identify drug manufacturing locations—historically considered a trade secret (Conti et al. 2020)—and to link drugs to their quality disclosures. Using these data, I show that quality failures are widespread and persistent over 2000-2022. Approximately 57% of manufacturing facilities fail[8] inspections, and 12% of drugs are recalled during this period, demonstrating that quality problems are not isolated to specific firms or market segments. Drugs that experience a quality failure remain twice as likely to repeat that failure for up to a decade, relative to the average drug, implying that quality failures are not random, unavoidable shocks. Taken together, this evidence contradicts the conventional

5. Bioequivalence is typically demonstrated using manufacturer-submitted data and absorption tests in $\leq 60$ patients (Davit et al. 2009). It certifies that a drug has the same active ingredient amounts and dose form as the original.

6. Low manufacturing quality can harm patient health and is responsible, through supply disruptions, "for virtually all" drug shortages (Woodcock and Wosinska 2013).

7. The FDA may escalate its quality enforcement through other actions such as import alerts (bans) and injunctions, but these are very rare in practice. Note that unlike a *permanent* drug withdrawal, recalls only *temporarily* disrupt supply by removing specific batches of defective products.

8. I focus on quality "failures" disclosed by the FDA that indicate that a drug has the potential to cause patient harm. These include class I and II *recalls*, which signal a "reasonable probability" of causing medically reversible injury (II) and irreversible injury or death (I). I also include *inspections* graded "voluntary" or "official action indicated," which assert that a facility is in "objectionable" (VAI) or "unacceptable" (OAI) violation of U.S. law, and that its drugs are being prepared in a way that may make it adulterated or "injurious to health." A full 28% of facilities are awarded OAI inspection grades over my sample period. Since the threshold for patient harm for inspection grades is fuzzy, I replicate all inspection analyses using combined (OAI and VAI) and OAI-only grades. I discuss these disclosure mechanisms in more detail in Section 2.

wisdom about generics: these drugs are not perfect substitutes.

Next, I investigate whether the disclosure of quality failures improves equilibrium market outcomes. This is not obvious *a priori*: since patients face unusually strong barriers to quality information, and intermediaries have no fiduciary responsibility to respond to disclosures, the FDA's disclosure policies might be expected to have limited impact. I quantify causal effects using commercial claims data and a quasi-experimental, generalized difference-in-differences design, which exploits variation in the timing of FDA disclosures within and across markets. I find that recall announcements reduce intermediary purchases of recalled drugs by 60%, substantially improving the quality of drugs available to patients. The effect persists for up to ten years—long 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, the disclosure of even the most serious inspection failures has little effect on intermediary purchases. In other words, manufacturers are penalized more stringently for signals of poor supply reliability than for low quality per se.

To study disclosure's effects on other dimensions of patient welfare, I consider drug prices, accessibility and adherence. While disclosure has limited effects on prices, recalls meaningfully reduce patient access to drugs through shortages. In small markets (with few manufacturers), recalls increase the probability of a national drug shortage by 41%, as substitution to competing manufacturers fails to offset reduced purchases of recalled drugs. Across all markets, recalls reduce standard measures of patient medication adherence by 5%. In sum, even as recalls improve drug quality, they also reduce patient welfare through temporary supply disruptions. As far as I am aware, these results represent the first economic evidence on manufacturing quality, its relationship to drug shortages and its implications for patient welfare.

My quasi-experimental results clarify how disclosure affects available drug quality, but they cannot isolate the equilibrium effects of intermediaries, or the relative importance of supply- and demand-side mechanisms. For example, on the supply-side, disclosure might alter costs and prices, as manufacturers reoptimize their capacity and effort. On the demand-side, it might affect intermediary choices, as they update their expectations about drug supply reliability. To examine these mechanisms, I develop a structural model of generic procurement through scoring auctions. The basic structure of my model is based on evidence from intermediary reference manuals, which describe how they score manufacturers on their historical recalls and supply reliability (CardinalHealth 2022; AmerisourceBergen 2019). I capture intermediaries' stated decision-making processes using a scoring auction, which is multi-dimensional in the sense that buyers consider both price and reliability. Using new data from approximately 68,000 auctions, I estimate the model's primitives: intermediary preferences for reliability and manufacturers' cost and productivity distributions.

My estimation approach extends a recent method from Laffont et al. 2020, where I contribute a new technique that non-parametrically identifies the model primitives when only winning prices are observed. While scoring auctions are widely used in real life (Hortaçsu and Perrigne 2021), current estimation methods require complete data on prices, which is infeasible in many

empirical applications. Identification with incomplete prices is non-standard because the auction model incorporates two sources of asymmetry: bidder characteristics and unobserved buyer preferences. My approach draws on the empirical auction literature (Guerre, Perrigne, and Vuong "GPV" 2000) to identify manufacturer costs, and then applies standard restrictions on bidders' cost distributions to identify intermediary preferences (Athey and Haile 2002).[9] Using Monte Carlo simulations, I show that my algorithm recovers unbiased estimates of the model primitives. Finally, I generalize my approach to multiple scoring auction formats, including linear, optimal and quality-to-price ratio scores.

My structural model shows that intermediaries substantially improve the quality of drugs available to patients, driven by their preferences for reliable suppliers. Intermediaries treat the bids of recalled drugs as if their prices were 13% higher, effectively paying a 2% premium for every 10% reduction in the probability of future recalls. By simulating my model without these preferences, I show that intermediaries increase the share of high-quality (i.e., not recalled) drugs by 27%. In contrast, intermediaries have limited effects on prices and total producer surplus (which each increase by less than 3%). Taken together, my results show that intermediaries improve static quality and welfare, and thus partially overcome the market failures from patient asymmetric information. By encouraging competition on reliability, and not just price, intermediaries may also increase manufacturer incentives to invest in reliability in the long-run.

Next, I explore the welfare implications of counterfactual policies that seek to optimally align intermediary and social incentives. The current administration has proposed two such policies, reflecting its decision—in a period of historic pandemics and drug shortages—to make the pharmaceutical supply chain a "national security" priority (The White House 2021). The respective policies would (1) increase manufacturing quality transparency and (2) provide "pay-for-performance" subsidies for reliable drugs (HHS 2024).[10] My quasi-experimental results suggest that transparency initiatives may have limited impact, since intermediaries already closely track information about drug reliability. On the other hand, my structural model shows that pay-for-performance subsidies may improve welfare by strengthening intermediary incentives to select high-quality drugs. In counterfactual simulations, I show that one version of the policy would increase the share of high-quality drugs by 24% and raise prices by 29%, with moderate effects on producer surplus. The policy's costs, however, are concentrated among low-reliability manufacturers, whose losses indicate that they would likely exit the market. Overall, my analysis suggests that pay-for-performance mechanisms could improve short-run welfare but have ambiguous implications for the market's long-run viability.

Finally, I study the FDA's current quality disclosure policies and decompose their supply-

9. First, I non-parametrically recover the distributions of unobserved prices in symmetric and asymmetric auctions (with identical and non-identical bidder types, respectively), up to a guess of the intermediary's preferences. I then apply GPV to recover the manufacturers' cost distributions from their first order conditions. Finally, using the standard assumption that costs are identically distributed across auctions, I infer intermediary preferences by matching the cost distributions in asymmetric auctions to those in symmetric ones, where preferences are immaterial.

10. The two policies—the Manufacturer Resiliency Assessment Program (MRAP) and Hospital Resilient Supply Program (HRSP)—were proposed by the Department of Health and Human Services (HHS) in April 2024 and build on similar proposals by the FDA and Congress.

and demand-side mechanisms. Consistent with my quasi-experimental results, I find that recall disclosures increase the share of high-quality drugs with little effect on prices, changes that are likely to benefit patients.[11] Producers, in contrast, experience a 14% reduction in surplus, driven by intensified price competition and recall reimbursement costs. This reduction helps explain why—in defiance of theory (Grossman 1981; Milgrom 1981)—manufacturers fail to voluntarily disclose their quality type.[12] Next, I decompose the three mechanisms through which disclosure affects equilibrium outcomes in my model: intermediary preferences, manufacturer costs and strategic pricing responses. I find that intermediary responses explain more than 100% of the policies' benefits: simply updating intermediary preferences while holding the other mechanisms fixed leads to an 18% increase in the share of high-quality drugs after recalls. Supply-side responses—reductions in manufacturer costs and prices—partially offset these benefits and reduce the high-quality drug share by 11%. The decomposition results help rationalize my quasi-experimental findings: specifically, how disclosure improves quality without informing patient demand. This is possible because intermediary responses amplify the effects of disclosure interventions, resulting in meaningful gains for patients.

My paper contributes to two empirical literatures on the economics of quality disclosure and informed intermediaries. The first literature studies the equilibrium effects of quality information revealed through a variety of disclosure mechanisms (Jin and Leslie 2003; Dranove and Sfekas 2008; DellaVigna and Pollet 2009; Higgins, Yan, and Chatterjee 2021; Vatter 2022; Barahona, Otero, and Otero 2023). A key finding from this work is that disclosure can improve consumer choice, but its effectiveness is limited by consumers' abilities to attend to and respond to information. Motivated by this result, I study *intermediary* responses to disclosure, which I find to be an effective but imperfect substitute for consumer ones. The second literature argues that expert intermediaries can help resolve adverse selection problems between buyers and sellers (Biglaiser 1993; Lizzeri 1999). These papers emphasize that intermediaries often face incentives to steer consumers to suboptimal goods, which neither reputational nor competitive mechanisms can fully correct (Benabou and Laroque 1992; Duflo et al. 2013). Much of the empirical evidence focuses expert fiduciaries, such as financial advisers (Gavazza and Lizzeri 2021), realtors (Robles-Garcia 2019) and physicians (Grennan et al. 2024). The intermediaries in this paper are conceptually distinct, in that they make choices for consumers without any legal responsibility to act in their best interests, intensifying their incentive misalignment. My paper connects these two literatures by providing evidence on how intermediary responses can magnify the equilibrium effects of quality disclosure.

Another strand of the disclosure literature focuses on recalls, which are the primary quality regulation tool in many consumer goods markets. Empirical evidence—largely from automobiles, toys, food and appliances—shows that consumer purchases fall after recalls (Jarrell and Peltzman 1985; Salin and Hooker 2001; Seo and Jang 2021), with mixed evidence regarding

11. These benefits may be partially offset by temporary supply disruptions.

12. As previously noted, consumers of generic drugs are uniquely difficult to inform. They may therefore fail to reward high-quality, disclosing manufacturers, which also prohibits the canonical "unraveling" result.

spillovers on competing goods[13] (Toledo and Villas-Boas 2019; Freedman, Kearney, and Lederman 2012; Basker and Kamal 2021)). My innovation is to model intermediary responses to recalls, and to use a structural framework to quantify recall effects on short-run quality and welfare.

While several papers from the literature on pharmaceutical industrial organization study drug quality—focusing on *drug approval* standards (Peltzman 1973; Atal, Cuesta, and Saethre 2022; Kao 2024; Chandra et al. 2024)—I am not aware of any that evaluate quality *after approval*.[14] My paper contributes to this literature by providing the first empirical economic evidence on drug manufacturing quality and its implications for welfare. It also relates to recent papers that study generic drugs, which address shortages (Stomberg 2016; Yurukoglu et al. 2017; Dubois et al. 2023; Galdin 2024), prices (Morton 2000; Ganapati and McKibbin 2023) and collusion (Cuddy 2020; Starc and Wollmann 2022).[15] These papers, however, lack data on manufacturing quality, which is the proximate cause of shortages and has potentially first order implications for prices and competition.

Finally, from a methodological perspective, my paper adds to a growing empirical literature on scoring auctions. While scoring auctions are the rule, not the exception, in many settings (e.g., procurement, private requests-for-proposals), surprisingly little evidence exists regarding their real-world performance. Recent empirical work extends the canonical estimation approach of Guerre, Perrigne and Vuong (2000) to non-parametrically recover bidder costs with exogenous quality, as I do in this paper (Hanazono, Nakabayashi, and Tsuruoka 2013; Andreyanov, Krasikov, and Suzdaltsev 2023; Andreyanov et al. 2024). However, the methods used in these papers require that the scoring rule is known, and that the researcher has complete data on prices, limiting their practical applications.[16] Krasnokutskaya, Song, and Tang 2020 is most similar to my paper, in that the authors recover an unknown scoring rule with incomplete data on quality—but complete prices—in Internet auctions. My paper contributes a new estimation methodology, which builds on the scoring auction model of Laffont et al. 2020 but only requires winning prices. Using this method, I show that scoring auctions increase buyer surplus in generic markets, relative to price-only formats, consistent with theoretical predictions (Asker and Cantillon 2008).

The remainder of the paper proceeds as follows. Section 2 describes quality failures in the U.S. generics market and the empirical setting. Section 3 introduces the data, while Section 4 develops quasi-experimental evidence on the effects of quality disclosure. Section 5 presents

13. A related paper by Cawley and Rizzo 2008 studies *permanent* drug withdrawals, a type of minimum quality standard. The authors find negative spillovers from drug withdrawals onto products in related therapeutic classes.

14. A substantial medical literature compares the efficacy of generic and branded drugs—largely finding them to be similar—but does not consider variation in quality across generic drugs (e.g., Kesselheim et al. 2008).

15. In particular, I rely on the institutional facts established by Cuddy 2020 and Starc and Wollmann 2022. Both of these papers estimate price-only auctions in the generics market, using parametric approximations to accommodate their data, which are aggregated at the drug-market-level. Fortunately, my data are granular enough to accommodate a non-parametric scoring approach.

16. Allen et al. 2024 recover an unknown scoring rule with complete data on bid components. Auctions with endogenous quality, where the scoring rule or bid preference is typically known, are studied in papers by Hanazono et al. 2016, Nakabayashi and Hirose 2016, Sant'Anna 2017 and Takahashi 2018.

the structural model, and Section 6 the counterfactual analyses. Section 7 concludes.

# 2 Background

## 2.1 Generic Quality and Reliability

> "Before I embarked on this project, I had always assumed that a drug is a drug... [that] there isn't necessarily any variation between different generic versions. I was wrong" (Katherine Eban, *Bottle of Lies*).

Generic drugs are widely assumed to be interchangeable, since they are certified as bioequivalent at approval. This principle underlies the basic organization of the U.S. pharmaceutical industry, which relies on generic substitution to make drugs vastly more affordable.[17] However, growing evidence—including the bestselling *Bottle of Lies* exposé—suggests that this assumption is broadly incorrect. In this section, I present evidence of substantial quality variation in the U.S. generics market.

Generic drugs are ubiquitously prescribed and generate tremendous savings for the American patient—benefits that are increasingly threatened by manufacturing quality failures. Two-thirds of American adults are prescribed generics each year, including nearly 90% of the elderly (Cohen and Mykyta 2024). These drugs comprise 91% of prescription volumes and are priced 80-85% lower than branded drugs, collectively saving U.S. patients $450bn per year (AAM 2024).[18] At the same time, generic quality failures are responsible for approximately 1,000 recalled or banned drugs each year, reflecting their significant risks to patient health.[19] Quality failures are also the "foremost" cause of drug shortages, which limit patient access to critical medications and have attracted widespread media and policy attention (Woodcock and Wosinska 2013; Jewett 2023).

For example, in 2023 the FDA banned drugs from two manufacturing facilities owned by Intas Pharmaceuticals. Together, these facilities supplied major shares of specific chemotherapies, anti-epileptics and opioid use disorder treatments in the U.S. market (FDA Import Alert 2023). FDA inspectors had discovered serious manufacturing problems at Intas, describing a "cascade of failure" in quality control and the deliberate destruction[20] of testing records (FDA Warning Letter 2023). The closure of these plants contributed to ongoing shortages in

17. Generics are automatically substituted for each other, without notice to the patient, and are mandatorily substituted for branded versions in 18 states (Rome, Sarpatwari, and Kesselheim 2022). This substitutability is widely credited with fostering intense competition that drives down prices. In fact, American patients pay *less* for generic drugs than their counterparts in other OECD countries (Mulcahy et al. 2021).

18. This figure accounts for direct savings from substitution between branded and generics, but does not consider indirect savings from avoided hospitalizations and other healthcare spending. In general, generics offer extremely high health returns relative to other healthcare spending (Lichtenberg 2007; Stuart et al. 2015).

19. All drug statistics refer to the U.S. market. As examples, four Americans died from using contaminated eye drops in 2023 (later recalled); more than 150 Americans died after being treated with adulterated heparin in 2008 (later banned) (Ables and Cimons 2023; Rosania 2010).

20. The FDA Warning Letter describes Intas employees pouring acid on quality records. The facilities supplied capecitabine, cisplatin, clonazepam and naltrexone, which all later experienced shortages.

each of these drug markets (ASHP 2024). Most prominently, they triggered nationwide rationing of chemotherapy drugs, which has delayed the care of up to 8% of cancer patients since 2023 (ACS CAN 2023).

To contextualize this example—and provide comprehensive evidence on quality in the U.S. market—I plot the distributions of manufacturer quality failures in Figure 1. The figure shows rates of recalls and inspection failures per drug-year, obtained from internal FDA data, for each manufacturer over 2000-2022. Approximately 40% of manufacturers experience recalls or inspection failures with the same frequency as Intas or greater, suggesting it was not an obvious outlier. The average manufacturer (in red) has an economically meaningful chance of being recalled in any given drug-year (2.1%, panel (a)) and fails the majority of its inspections (69%, panel (b)).[21] In panels (c) and (d), I plot the same data aggregated to the market and country of origin levels, respectively. The aggregated distributions tell a similar story: the average market recalls 1.7% of its products each year, and the average country fails 68% of its inspections (on par with the U.S. average of 70%). The similarity in these patterns and the overall broadness of the distributions suggests that quality failures are not isolated to few bad actors—instead, they affect the majority of manufacturers, markets and countries.

Next, I show that quality failures are persistent within drugs over time. In Figure A.4, I plot the average drug's probability of experiencing a quality failure in each year after a reference point (year $t = 0$), conditional on whether it failed (solid line) or did not fail (dashed line) in $t = 0$. The figure shows that failing drugs are much more likely to repeat their failures than the average drug, with elevated probabilities that persist for up to a decade. In each year, recalled drugs are 3.3 times (panel (a))—and inspection failers 1.8 times (b)—more likely to experience quality issues.[22] This suggests that individual drug quality is relatively stable over the duration of my data, consistent with the FDA's narrative that manufacturing processes degrade slowly over decades (Woodcock and Wosinska 2013). Further, pervasive generic quality failures cannot be explained by mere random quality shocks—instead, certain drugs appear to have persistent, lower-quality types. When combined with patients' inability to observe generic quality, these facts provide a strong rationale for targeted policies to improve quality and welfare.

Finally, I present the trends in my data (Figure 1, panel (e)), which show how quality failures have accelerated over time. The drug-level recall rate (in dashed purple) increased from 0.9% to 3.0% over 2000-2022, while the inspection failure rate (solid blue) followed a similar pattern (increasing from 51% to 70%). These trends are temporally correlated with increases in the probability of national drug shortages (Figure A.3, panel (c), dashed green), which grew from 2.2% to 3.9% over the same period.[23] The raw trends present a unified picture of deteriorating

21. The estimated inspection failure rate remains stable (at 70% per drug-year) after conditioning on manufacturers with 10 or more inspections. As noted in the Introduction, my primary definition of a *quality failure* includes all class I/II recalls and inspections that result in OAI or VAI grades. In VAI inspections, the FDA judges that a facility should be able to voluntarily correct its deficiencies, while OAI graded facilities are recommended further regulatory action. In Figure A.3, I present the corresponding distributions for OAI-only inspection grades. The average manufacturing facility in my data receives an OAI inspection grade one fifth of the time, conditional on being inspected.

22. If only OAI inspection failures are considered, the relative rate is 2.6 times.

23. Rates of OAI inspection grades increased from 8% to 13% over 2000-2022 (Figure A.3, (c)). Counts of drug shortages grew from 100 to 250 over the same period and reached an all-time high of 323 in 2024.

generic quality and reliability—a vulnerability the federal government has termed a "national security threat" (Edney and Griffin 2023). A recent flurry of policy proposals—from the White House, Congress and Department of Health and Human Services (HHS)—underscores the depth of regulator concern (The White House 2024; Senate Legislative Counsel 2024; HHS 2024). Despite this, little research has considered the determinants and welfare implications of generic quality and reliability, key inputs to analyzing these policies.

## 2.2 Institutional Framework

To provide context for my empirical approach and structural model, I now describe the salient features of the U.S. generics market. The basic structure and key actors are depicted in Figure A.2. Drug manufacturers make investments in quality, which is monitored and disclosed by the regulator, the FDA. Intermediaries observe the disclosed quality signals and purchase drugs from manufacturers to sell to patients. Importantly, intermediary choices determine the quality of drugs available to patients, who do not observe quality signals. This suggests that intermediaries can improve equilibrium quality and welfare, if their incentives are sufficiently well aligned with the regulator's. I describe each of these actors and their quality information in more detail below.

As previously noted, generic manufacturers must demonstrate bioequivalence to the originator drug before entering the market. Notably, bioequivalence does not require identical manufacturing practices, and a drug's quality can deteriorate as the manufacturer scales up production and its physical capital depreciates. According to a review by the FDA, some generic manufacturing lines have been "running 24 hours a day, 7 days per week... since the 1960s," with "only limited upgrades," leading to increased rates of quality failures (Woodcock and Wosinska 2013). Substandard manufacturing processes can result in bacteria-contaminated drugs or super-potent active ingredients—common defects in my data[24]—with potentially serious adverse effects on patient health.

The FDA is responsible for regulating drug quality in the United States. After a drug's approval, the FDA primarily relies on two non-binding disclosure mechanisms—inspection reports and recall announcements—to maintain a high-quality and reliable supply of drugs.[25] Since these mechanisms generate the key variation in my data, I describe them here briefly. The FDA conducts systematic manufacturing facility *inspections* for general surveillance purposes and to investigate specific quality concerns. FDA officials grade each inspected facility based on its risk to patients, and on the extent to which its manufacturing practices violate the Food, Drug, and Cosmetic (FD&C) Act. In my analysis, I focus on the two most serious grades, which assert that a facility is in "objectionable" (VAI) or "unacceptable" (OAI) violation of U.S. law,

24. For other examples, see Table A.1

25. The FDA also uses other enforcement actions to discipline drug quality, such as import alerts (bans), injunctions, seizures and administrative detention orders. I observe the universe of these actions in my data, where they are relatively rare. The FDA also issues Warning Letters, Untitled Letters, Forms 483 and Regulatory Meetings after inspections, which I observe to be highly correlated with inspection classifications. For additional details about inspections and recalls, please refer to the Appendix.

and that its drugs may therefore be injurious to patient health. *Recalls* may be initiated by the FDA or manufacturers, and are required whenever a batch of drugs is found to violate the quality standards of the FD&C Act. The FDA classifies each recall based the defective drug's potential to cause patient harm; I study class I and II recalls, which have a "reasonable probability" of causing medically reversible injury (II) and irreversible injury or death (I) (FDA 2024b).

The FDA publicly reports inspection and recall outcomes on its website, and in this way discloses information about low drug quality.[26] Recall announcements may also affect market outcomes through a second channel, by signaling that a drug's supply is unreliable. Recalls often cause temporary stockouts, as manufacturers must physically remove defective batches from the market and may need to pause and correct production. While the initial disruption may be resolved in weeks or months,[27] it provides a strong signal of future disruption risks. Recalled drugs are 3.3 times more likely to be recalled over the following decade (as noted in the prior subsection), and they are 1.5 times more likely to exit the market, relative to the average drug.

The effectiveness of the FDA's disclosure policy depends on the extent to which it increases demand for quality. However, patient awareness of generic quality is low for multiple, intractable reasons. Patients face formidable institutional barriers to responding to FDA disclosures: public inspection reports, for example, lack drug names and therefore provide patients with no actionable information. Several sources—internal FDA audits, patient lawsuits and pharmacy recall protocols[28]—indicate that individual recall announcements often fail to reach prescribed patients. Most importantly, U.S. generics are very poorly differentiated: they are typically unbranded and sold without disclosing the product identity (i.e., the manufacturer name).[29] This implies that, even if a patient should observe an FDA disclosure, she would have almost no way to connect this information to her drug purchase.

In contrast, intermediaries (wholesalers and pharmacies) have ample opportunity and incentive to learn about generic quality. These large, sophisticated firms interact with manufacturers over decades and in thousands of drug markets. The largest intermediaries, pictured in Figure A.2, captured 95% of U.S. generic volumes and represented over $20bn in market capitalization in 2020 (Fein 2023). Intermediaries' scale begets monopsony power over manufacturers,

26. From the FDA website, they are often picked up by the industry press. The FDA also announces recalls on X (formerly Twitter) and to healthcare professionals through MedWatch Safety Alerts.

27. The average device recall takes 8 months to complete (to remove all defective products) and another 6 months to be terminated (for the FDA to confirm that quality issues have been resolved). While completion time statistics are not available for drugs, the average reported termination time (2.1 years) suggests that drug recalls operate on a similar timeline, with an implied average completion time of 13 months. Note that drug manufacturers may be able to resupply much faster than they can remove products; the typical lead time for finished dosage form drugs is between several days to a few weeks (The White House 2021; CardinalHealth 2022).

28. Government investigations suggest that manufacturers struggle to contact end users during recalls in medical device and food markets (GAO 2011b). Understaffed pharmacies may lack the resources to inform patients about recalls and are not legally obligated to do so (Fink 2012; NCPA 2022). The only exception is New York state, which began requiring this in 2019.

29. Pharmacies do not typically disclose the generic manufacturer before purchase. After purchase, the manufacturer name may or may not be visible, depending on the pharmacy's packaging. Patients may also struggle to track generic quality for more general reasons: they may fail to attribute new symptoms to defective drugs, and their inattention may induce purchasing inertia, as has been observed with other routine healthcare purchases (Ho, Hogan, and Scott Morton 2017).

to whom they award supply contracts through secretive, competitive auctions.[30] Through these auctions, intermediaries determine the quality and wholesale prices of nearly all drugs in the U.S. market, with important implications for equilibrium welfare.

When choosing suppliers, intermediaries consider both price *and* quality—since the latter can impose significant spillover costs onto them. For example, recalls reduce intermediary revenues through stockouts and require costly tracking of defective inventory (though supply contracts typically stipulate that direct drug costs are reimbursed). Accordingly, intermediaries emphasize that supplier reliability is "critical" to their choices—that "lapses in supply ruin relationship[s]" (FTC 2011; GAO 2011a).[31] Intermediary reference manuals, such as the one presented in Figure A.1, describe how they "score" potential suppliers using reliability criteria, such as historical recalls, product availability and lead times (AmerisourceBergen 2019; CardinalHealth 2022). In contrast, inertia is not a first order concern, since auction winners change rapidly—in some cases, with every batch—as intermediaries continuously seek better terms.

Theoretically, informed intermediaries can increase demand for high-quality drugs, encouraging manufacturer quality investment, and thereby mitigating market failures from asymmetric information. Intermediaries, however, are unlikely to fully internalize these social benefits. This conflict of interest is typical of markets with informed intermediaries (Gavazza and Lizzeri 2021) but is particularly acute in the generics market, for two reasons. First, generic consumers have uniquely limited information and ability to make their own purchasing choices. They cannot easily reward intermediaries for being faithful agents (e.g., with good reputations) and are especially vulnerable to intermediary decisions. Second, the intermediaries in my paper—unlike advisors or physicians—face no legal responsibility to act in patients' best interests. Together, these factors imply that intermediaries may choose socially suboptimal levels of quality, maximizing profits rather than patient welfare.

To summarize, the U.S. generics market is characterized by pervasive quality failures and an unusual degree of asymmetric information, which provide a strong rationale for quality regulation. The FDA relies on disclosure policies—inspection reports and recall announcements—to regulate generic quality and reliability. Since patients cannot respond to quality disclosures, the direct impact of these policies is limited, implying that first-best, full-information welfare is unattainable. In this context, informed intermediaries have substantial scope to improve equilibrium quality—whether they choose to do so, and how close welfare is to first-best, are empirical questions.

30. These institutional details were first established by (Cuddy 2020) and are described in unsealed court records (Connecticut v. Teva 2019). Intermediaries typically award a single contract per pharmacy-market and do not release information about losing bids. In the case of supply disruptions, their contracts may include provisions for secondary suppliers, which I exclude from my data.

31. One industry trade group estimates that the average pharmacy spends $230,000 in labor costs related to recalls per year (Tracelink 2023).

# 3 Data

## 3.1 Data Sources

*FDA quality disclosures*. My analysis relies on three main data sources, from which I construct a panel of drug quality signals, prices and purchases. First, I collect data on drug manufacturing quality measured and disclosed by the FDA. These data include the universe of FDA quality disclosures and enforcement actions over 2000-2022. I obtained inspection and enforcement data through FOIA requests, and I used the Internet Archive to extract recall announcements from Enforcement Reports on the FDA website. To the best of my knowledge, these sources represent the most comprehensive manufacturing quality data studied so far.

Until recently, two factors presented a serious barrier to research on manufacturing quality. First, the majority of FDA quality data is reported at the facility, rather than the drug, level. Second, the FDA considers drugs' manufacturing locations to be trade secrets not subject to FOIA (Conti et al. 2020). Recently, researchers have begun painstakingly tracing drug manufacturing locations, using data extracted from drug labels, trade manifests and OECD administrative records. For example, in 2018 the Center for Infectious Disease Research and Policy (CIDRAP) traced the locations of 156 drugs (CIDRAP 2018). Using machine learning tools, I scale up CIDRAP's methodology to cover all generics in a commercial claims dataset, thereby linking each drug to its quality information.[32]

Recall announcements are reported at the event level, where a recall typically involves one or more drugs produced by a single manufacturer over a short period (up to a few months). The recall data contain the announcement date, product names and National Drug Code (NDC) identifiers, which allow me to follow specific drugs over time. I also observe the recall class, approximate volume of defective drugs and manufacturer-reported recall reasons. Since successive recalls of the same drug are typically extensions of a single event, I count multiple recalls per drug-year as one announcement.

I complement the recall data with FDA inspection outcomes. While recalls can be firm-initiated—implying they could suffer from inconsistent reporting—inspections are FDA-initiated and graded by independent experts. The inspection data are also useful because they reveal relatively little information about a drug's supply reliability, compared to recalls, a factor that I exploit in my quasi-experimental strategy. For each inspection, I observe the quality grade, any citations and whether the inspection was for general surveillance purposes or for cause. For the purposes of quantifying disclosure effects, the FDA data have an additional limitation: they only reveal quality contingent on regulator or manufacturer actions, which are not random. In the next subsection, I describe how I use information about the FDA's facility selection model, as well as plausibly exogenous variation in disclosure timing, to mitigate concerns about this type of selection.

*Prices.* My second main data source contains novel information about drug prices, which I

32. Additional details are provided in the Technical Appendix.

collected from state Medicaid surveys through Open Records Requests. These data report retail pharmacies' average acquisition costs (AACs) for outpatient drugs in eleven states[33] over 2010-2022, which I supplement with national survey data after 2013 (the National Average Drug Acquisition Cost survey, "NADAC"). The prices capture the winning bids of manufacturers in intermediary procurement auctions, averaged at the molecule-formulation-strength-state level.[34] They are calculated by a public accounting firm using monthly pharmacy invoice data, and are updated weekly to reflect pharmacy-submitted pricing changes. The state-level data are useful because they allow me to exploit variation in pharmacy chain presence across states to recover intermediary-specific winning bids.[35] As a result, I am able to estimate my structural model using non-parametric methods and with improved statistical power. To calculate winning bids, I use average prices from the duration of the contract, after dropping the first three "transition" months (Cuddy 2020).

*Patient claims.* My last main data source is administrative claims over 2000–22 from UnitedHealth (Optum), the largest private U.S. health insurer. The data contain approximately 33 million covered individuals with pharmacy benefits each year. The population includes both employer-sponsored (80%) and Medicare Advantage members (20%) and has significant share in all states. I use the claims data for three purposes: (1) to identify intermediary auctions; (2) to estimate the drug volumes associated with each intermediary contract;[36] and (3) to measure patient medication adherence.

I identify intermediary auctions using the methodology developed by (Cuddy 2020), which I outline below. Intermediaries offer a single supply contract per market, allowing me to easily identify their choices (i.e., auction winners) from the drugs stocked by each pharmacy chain. Since new rounds of auctions are usually triggered by the entry and exit of manufacturers, I rely on changes in these patterns (observed in the claims and FDA Orange Book data, described below) and in the composition of winning drugs across pharmacy chains to identify auction timing. Manufacturers typically participate in each auction within a quasi-simultaneous round, allowing me to identify the number of bidders from the manufacturers observed in each market. Finally, I am careful to exclude collusive auctions during my study period, which I observe in unsealed court records (Connecticut v. Teva 2019).

33. Eleven states collected AAC data for at least two years between 2010-2022: LA (2010-2018), OR (2010+), AL (2011+), IA (2013+), TN (2017+), MD (2017+), UT (2021+), CO (2016+), OH (2021+), MT (2019+) and ID (2017+). Under federal rules, state Medicaid programs have flexibility in choosing which cost estimates they use for reimbursement. The data suggest that these states may have chosen AAC-based reimbursement because they acquire drugs at lower prices than the national average.

34. The prices exclude certain volume discounts, as well as wholesaler margins, which are 3.5% on average (Sood et al. 2017). They also exclude manufacturer rebates, which were trivial in the generic market during this time period (Lieberman and Ginsburg 2018).

35. I estimate intermediary-specific winning prices for the largest 5 intermediaries in my data. Since I have up to twelve 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.

36. Note that my measures of intermediary contract volumes capture only Optum enrollees, and are therefore scaled to a fraction of overall U.S. volumes. The Optum claims data are similar to other standard administrative claims datasets: they contain pharmacy IDs, NDC drug codes and purchased volumes for pharmacy encounters, as well as patient demographics such as age and gender.

I estimate annual contract volumes using the total number of prescriptions per pharmacy chain in the claims data. Patient adherence is the 90-day Medication Possession Ratio (MPR), defined as the proportion of days supplied over a specified number of prescribed days. I calculate the MPR at the patient-molecule-form level, allowing any drug within a molecule-form to count towards a patient's adherence level. Following the literature, I use cut-off points—80% and the drug-specific median—to identify non-adherent patients (Andrade et al. 2006).

In addition to these main data sources, I use several supplemental datasets that contain drug and pharmacy characteristics. Two standard datasets from the FDA—the NDC Directory and Orange Book—provide drug ingredients, formulations, strengths, labeling firms and approval dates. I capture drug shortages using the University of Utah Drug Information Service's (UUDIS) data, which are considered the gold standard in the shortage literature (Alpert and Jacobson 2019). Pharmacy characteristics, including location and chains relationships, are from the National Council for Prescription Drug Programs (NCPDP).

### 3.2 Descriptive Statistics

The analysis sample includes generic prescription drugs purchased from retail pharmacies by UnitedHealth beneficiaries over 2000-2022. I focus on unbranded, oligopoly drugs, which are procured through competitive auctions and represent 84% of the total generic sample by volume. Overall, I study the quality characteristics of 10,199 drugs in 1,204 molecule-form markets, and the purchasing responses of approximately 64,000 pharmacies and 63 million patients.

My quasi-experimental approach exploits variation in the timing of quality disclosure through recall announcements and inspection reports, which represent the majority of the FDA's quality regulation actions. To ensure that I observe each drug for at least one year before and after disclosure, I study first recalls and inspections initiated between 2001 and 2021 (observing 1,159 such recalls and 1,483 inspections).

Table 1 describes the key characteristics of the generics in my sample, providing a snapshot of the old, inexpensive products that account for the majority of U.S. drug consumption. Column (1) presents the mean values for all drugs, while columns (2) and (3) summarize the values for drugs that were recalled and never-recalled during the study period, respectively. Each value represents the average over all years in the sample (cols. (1) and (3)) or prior to the first recall (2). The typical drug in my sample was first approved over 19 years[37] ago and is purchased for only $4.55 per unit (e.g., pill or vial), or $0.70 at the median. A cursory comparison of columns (2) and (3) indicates that recalled drugs differ significantly from never-recalled ones on a number of observable dimensions. Notably, they are nearly twice as likely to be injected (13% vs. 7%) and have 4 times greater volumes (56,651 vs. 14,765 prescriptions). They compete in larger markets (with 12.4 vs. 10.7 participants) and are sold at appreciably lower prices ($0.41 vs. $0.76 at the median). These patterns help justify my use of not-yet-recalled drugs—as a more

37. This assumes 12 years of branded exclusivity (Kesselheim, Sinha, and Avorn 2017) plus the average drug's age as of 2011 (7.26), the sample midpoint.

plausible control group than never-recalled ones—for estimating the effects of recall disclosure.

In Table A.2, I present the corresponding descriptive statistics for inspected drugs. Column (1) presents averages for all inspected drugs, which are then divided into OAI or VAI (col. (2)) and passing ones (3). The inspection sample is limited to drugs that are matched to manufacturing locations and FDA inspections. While this sample is smaller (capturing 54% of total drug volume), it has similar observable characteristics to those observed in Table 1. The largest difference is in annual prescription volumes, which are somewhat larger in the inspection sample (28,668 versus 17,211 in the full sample).

# 4 Effects of Quality Disclosure

The theoretical effects of quality disclosure described in Section 2.2 relied on key assumptions about intermediary information and incentives. Specifically, intermediaries must learn about drug quality through FDA disclosures, and they must respond to dynamic incentives to select high-quality drugs. Without these assumptions, patient asymmetric information is likely to negate any benefits of quality disclosure. In this section, I provide quasi-experimental evidence on the equilibrium effects of quality disclosure that is consistent with these assumptions. My results motivate the structural analysis in Section 5, which isolates supply- and demand-side mechanisms to quantify the role of intermediaries.

## 4.1 Empirical Strategy

To estimate the effects of quality disclosure, I use a difference-in-differences framework ("DD," operationalized using two-way fixed effects) to compare outcomes for drugs that experienced quality disclosure to those that did not until later in the study period. To improve exposition, I refer to these drugs as "recalled" and "not-yet-recalled," respectively, though I also apply the framework to study inspection disclosures. 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 of interest for drug $d$ in year $t$.[38] I model the outcome as a function of drug $\alpha_d$ and year fixed effects $\alpha_t$ and unobserved time-varying factors $\varepsilon_{dt}$. $D_{d,t+s}$ is an indicator equal to one $s$ years after the drug's quality is first disclosed. The coefficient of interest $\beta_s$ varies flexibly by year and is normalized to zero in the year before treatment ($\beta_{-1}$ = 0). This coefficient can be interpreted as the difference in outcomes for recalled drugs relative to not-yet-recalled ones in the $s^{th}$ year after disclosure. 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:

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

38. As previously noted, a drug is a molecule-formulation made by a specific manufacturer.

where the key variable $D_{dt}$ is equal to one after disclosure.

My DD 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. It exploits variation in disclosure timing among drugs that eventually experience disclosure to identify causal effects. While the identifying assumption cannot be tested, the absence of differential trends prior to treatment provides evidence consistent with it. I plot the dynamic coefficients $\beta_s$ and find no evidence of pre-existing trends that are correlated with treatment timing. In formal tests of joint significance, I am able to reject the null hypothesis of non-zero pre-trends for all key outcomes.

Since quality disclosure is not random, not-yet-recalled drugs offer a more credible counterfactual than never-recalled ones, which differ in important observable and potentially unobservable ways (described in section 3.2). One source of this selection is the FDA's facility risk model, which is used to prioritize facilities for inspections. The risk model is based on (publicly-announced) drug-specific criteria—such as a drug's "inherent risk," its inspection history and patient exposure—and thus generates non-random variation in inspection and recall probabilities.[39] To control for this variation, I used drug fixed effects to remove inherent risk and unobserved, persistent differences across drugs. In robustness checks, I include inspection grade covariates to control for time-varying differences in drug inspection histories. Finally, I explore heterogeneity in my estimated effects by market and drug size, as proxies for "exposure."[40] While these approaches control for the majority of the risk model's criteria, disclosure timing could still potentially be correlated with the outcomes of interest through other channels. Intermediaries, for instance, may be able to observe other quality signals (e.g., consumer complaints) that allow them to anticipate disclosure, a possibility that I examine using more granular month-level data. Alternatively, the FDA and manufacturers may coordinate disclosure timing to minimize supply disruptions.[41] In these examples, the estimated coefficients are likely to understate the true effects of disclosure, but other non-randomness might have opposite signed effects.

To strengthen my ability to make causal interpretations, I extend the DD approach in two ways. First, I examine *spillover* and *net market effects* of disclosure, using data aggregated at the market-level. I estimate spillover effects for generic bioequivalent competitors of the recalled drug (whose own qualities were not disclosed). This exercise allows me to investigate intermediary substitution patterns within markets and competing manufacturers' responses. In my robustness checks, I explore whether these patterns are sensitive to alternative market definitions, including both smaller and larger markets.[42] I also estimate net market effects, which

39. The FDA has used its Site Selection Model since 2005. The components of the model are publicly announced, and the majority of inputs are captured in my data, as discussed in the main text. To the best of my knowledge, the remaining items—such as foreign government inspection histories—are not releasable under FOIA.

40. Market size is measured using the number of manufacturers per market before disclosure, and drug size using pre-disclosure prescription volumes.

41. When issuing Import Alerts (bans), for example, the FDA often carves out drugs that are at risk of shortage (Brennan 2016).

42. Specifically, I use molecule-formulation-strength and Anatomical Therapeutic Chemical Level 4 (ATC4) markets. While my baseline market—the molecule-formulation—is commonly used in the generics literature, generic

allow me to assess whether disclosure decreases the overall volume of drugs consumed by patients. This evidence would be consistent with an increase in drug shortages, which I formally test using national shortage data from UUDIS.

Second, I compare the effects of two types of disclosure mechanisms, recall announcements and failed inspection reports. While my primary analysis focuses on recalls, failed inspections provide complementary quality signals that allow me to test specific hypotheses about intermediary incentives. Both recall announcements and inspection reports reveal that drug quality is below a threshold that may be injurious to patient health. Recall announcements, however, also signal high risks of future supply disruptions (as discussed in the Section 2). Comparing the effects of recalls and inspections therefore informs whether intermediary purchases respond to signals of drug *reliability* versus *quality* alone.

To study the effects of inspection disclosures, I estimate a triple difference ("DDD") model that compares outcomes for failed and passing drugs, before and after inspections. The DDD specification is identical to Equation (2) but includes an additional term ($\gamma\, D_{dt} \times 1\{\text{failed}_d\}$). This design allows me to difference out potential bias from selection in inspections, as well as any inspection effects that are not caused by negative quality information. It relies on the identifying assumption that passing and failing (OAI or VAI) drugs would have evolved similarly in the absence of inspection (Olden and Møen 2022). Throughout, I focus on surveillance inspections—which are conducted quasi-randomly, and not in response to specific quality concerns—in order to avoid contamination by other disclosure actions.

In my other robustness checks, I report coefficients from the estimator proposed by Callaway and Sant'Anna 2021, which is robust to concerns regarding the consistency of DD estimates in staggered treatment designs. I also test the stability and significance of my estimates to using different outcome measures, control groups, covariates, samples, weighting and clustering.

## 4.2 Quasi-Experimental Evidence

I begin my analysis by describing the causal effects of quality disclosure on intermediary purchases and drug shortages. I consider potential mechanisms for these effects—including signals of drug quality and reliability—by comparing the differential responses to inspections and recalls. Finally, I explore how disclosure affects manufacturer participation and patient welfare, quantifying the policy's unintended effects on prices and medication adherence.

*Effects of recall announcements on intermediary purchases.* I first show that recalls have large, negative effects on intermediary purchases. In Figure 2, I present the event study coefficients from Equation (1) with measures of drug market share and volume as the dependent variables. Drug market share captures intermediaries' purchasing decisions: it is defined as the fraction of retail pharmacies that select a drug in each market (molecule-form).[43] Panel (a) shows the

contracts are typically auctioned off at the smaller molecule-formulation-strength level. ATC-4 groupings are larger and include multiple molecules (therapeutic substitutes), which would typically require new prescriptions to purchase.

43. As previously noted, a market is a molecule-form combination and includes all bioequivalent drugs made

event study for the recalled drug's market share. Prior to the recall, the estimated coefficients are close to zero and statistically insignificant, suggesting no differential pre-trends. (The $p$-value from an $F$-test of joint significance is 0.93). However, the trend for recalled drugs begins to diverge in the disclosure year, with a steady decline in market share over the following five years. Figure A.5 provides additional detail on these dynamics using different cuts of the data. Panel (a) zooms in on the period surrounding the recall, plotting *monthly* effects in the three years before and after disclosure. It shows that recalled and control drugs follow a similar trend up until the last month before disclosure, suggesting anticipatory effects are minimal. Panel (b) zooms out to consider the *long-run* effects of recalls, using data from ten years before and after disclosure. The figure shows that the recalled drug share continues to erode even a decade later, with an estimated coefficient of -13 percentage points (pp) in the tenth year after (-68% of the pre-recall mean of 19%). Finally, Figure 2, panel (b) shows that the recalled drug's prescription volume *levels* also decline, following a very similar pattern to its market share. Relative to its pre-disclosure average, the recalled drug experiences an 89% reduction in prescription volume by the fifth post-disclosure year.[44]

Together, the event studies suggest that recalls strongly penalize low-quality drugs, with effects that last for far longer than the duration of the recall itself. The persistence of these effects cannot be easily explained by short-run supply disruptions or one-time reimbursement costs (discussed in Section 2). They are also unlikely to be caused by intermediary inertia, given the rapid supplier turnover observed in my data. (The average supply contract lasts for less than 1.5 years, and the average supplier is chosen for less than two years in a row.) The persistent effects could, however, be explained by intermediary preferences or long-run changes in manufacturer costs, explanations that I explore using the structural model in Section 5.

Table 2 presents the average effects of recalls, estimated using Equation (2). Consistent with the event study patterns, recalled drugs experience a significant decline in market share of 43% (4.4pp, panel A, col. (1)) and in prescription volumes of 60% (col. (2)), relative to not-yet-recalled drugs. While the estimated coefficients are large, they may actually understate the true recall effects (if drugs with larger effects are more likely to exit and are not observed in the sample).[45] The magnitude of the effects is not surprising. Generic drugs are uniquely undifferentiated—absent quality signals—and are purchased by large, sophisticated intermediaries. Furthermore, their recalls signal quality defects with potentially severe health consequences. Both factors contribute to more aggressive own-substitution patterns than is typically

by different generic manufacturers. This measure of market share captures intermediaries' supplier choices. (It does not capture their "intensive margin" drug volume choices, which cannot be separately identified from consumer purchases.) The market share measure is robust to the fact that the claims data capture only partial U.S. drug volumes, since they do capture an approximate census of U.S. pharmacies. I explore other definitions of market share in my robustness checks, which produce qualitatively similar results.

44. Since prescription volumes are highly skewed, I estimate the effects using logged values. To facilitate comparison, the figure plots the volume effects in levels: specifically, I convert log effects to percentage changes ($\exp(\beta)-1$), which I then multiply by the pre-recall median. Because I capture changes in drug volumes using Optum claims data, which are a subset of total U.S. drug volumes, my measures understate the effects of recalls on volume levels, while overstating their percentage effects.

45. Consistent with this, I estimate significantly smaller effects when excluding recalled drugs that eventually exit the sample (-2.5pp, Table A.3, row 11).

observed for consumer products.[46]

Next, I examine whether quality disclosure increases intermediary substitution to competitor drugs. Figure 2, panels (c) and (d) present the event studies for competitor market shares and volumes, respectively. To facilitate comparison, the competitor effects (in green circles) are overlaid on top of the recalled drug effects (reproduced from panels (a) and (b) in blue diamonds). The figures show that the competitor coefficients are small and insignificant before disclosure ($p$ = 0.25-0.62), consistent with the absence of pre-trends. After disclosure, the effects on competitor and recalled drugs break in strikingly similar but opposite-signed ways—suggesting that intermediaries gradually substitute between the two groups.[47] Table 2 quantifies the average effects: competitor drugs experience an average increase in market share of 6.8pp (9%, col. (3)) and volumes of 12% (col. (4)). These substitution patterns imply that recall disclosures are highly effective: they increase patients' probability of consuming high-quality (i.e., not recalled) drugs by 9% for up to 10 years. These effects are all the more striking given patients' near-total inability to observe the disclosed information.

I test my main results using a variety of robustness checks—including different outcome measures, control groups, samples and specifications, presenting the estimated coefficients in Table A.3. I also repeat my analyses using larger and smaller market definitions in Table A.4. The estimated coefficients remain similar across these tests and are typically within two standard errors of the baseline estimates.[48] Finally, Figure A.6, panel A explores whether my results extend across multiple types of drug dosage forms. This test is motivated by the fact that the drug shortage literature has focused on injectables (Yurukoglu et al. 2017; Dubois et al. 2023; Galdin 2024), which are more likely to experience quality issues. I find that these drugs represent a small share of overall recalls (13%) and actually experience more modest disclosure effects, relative to oral and topical formulations.

*Effects of recalls on drug shortages.* I now investigate whether recalls change *market-level* volumes, which has implications for both producer profits and patient access to drugs. The average effect of disclosure on market volumes is small and insignificant (-4%; Table 2, column (5)).[49] This average, however, masks significant heterogeneity across small and large markets. Conceptually, recalls should be expected to have stronger volume effects in small markets. Here, changes in intermediary purchase decisions require proportionally greater capacity adjustments. The fixed costs of these adjustments, as well as of facility remediation, affect small firms more—

46. Note that, in contrast to drug withdrawals or bans, we should not expect 100% volume reductions following recalls. This is because recalls are only temporary and target specific batches of drugs.

47. The increasing pattern in competitor volumes over time is consistent with the gradual expansion in capacity that I observe when studying manufacturer entry patterns (in the next subsection). It is not driven by effects of subsequent recalls, which are not significant.

48. There are two exceptions: the estimated coefficients for recalled drugs are significantly smaller when drugs that eventually exit the sample are excluded (as discussed in the prior subsection) and when a never-treated control group is used. The latter estimates should be interpreted with caution, since the control group is observably different and has significant pre-trends.

49. The sum of recalled and spillover effects in Table 2 is close to the estimated net market effect: -2,711 + 1,927 $\approx$ -1,065 in cols. (2), (4) and (5), resp. The effects do not precisely match due to estimation error and changes in sample composition caused by the exit of recalled firms.

limiting competitor volume gains while magnifying recalled drug losses. In Figure 2, panel (e), I plot the dynamic effects of disclosure on drug volumes in small and large markets (where small markets are defined as having 10 or fewer participants). Before disclosure, the trends in both markets are close to zero. Small markets (in gray circles), however, experience a significant decline in volume starting in the first year after disclosure. In contrast, large markets (black dots) experience no significant changes. The average effects are significantly different (-7% in small markets versus 15% in large ones; Table 2, panel B, col. (5)). This difference can be attributed to lesser effects on competitors (5% in small markets versus 17% in large ones; col. (4)) and to greater effects on recalled drugs in small markets (-69% versus -45%; col. (2)), consistent with the hypothesized patterns.

The heterogeneity analysis suggests that disclosure effects vary by market size, with small markets at greater risk of experiencing shortages. To formalize this result, I repeat the analysis using data on national drug shortages from UUDIS. In Figure 2, panel (f), I plot the dynamic effects of recalls on shortage probabilities in small and large markets.[50] Shortages are relatively rare events—I observe fewer than 1,200 over the two decades in my data—which causes the effects to be somewhat imprecisely estimated. Nevertheless, the event studies show a clear increasing pattern in the probability of shortage over the five years after disclosure, with effects wholly concentrated in small markets. This is equivalent to an increase in the average market's shortage probability of 33% (2.1pp, significant at $p < 0.05$), and in a small market's probability of 41% (2.5pp; Table 2, col. (6), also significant at $p < 0.05$).[51] These results provide evidence that recalls increase short-run supply disruptions and shortages, and consequently reduce patient access to drugs.

*Effects of inspection reports.* Next, I consider potential mechanisms behind the estimated disclosure effects. Specifically, I explore intermediary incentives to respond to disclosure, by comparing their responses to signals of drug quality versus supply reliability. As previously noted, inspection reports offer a useful counterpoint to recall announcements because they provide quality information without typically disrupting supply. To isolate the effects of inspection reports—and specifically of negative quality information—I compare changes for passing and failing drugs using a DDD design. I begin by plotting separate *DD* event studies for passing and failing drugs (Figure 3, panels (a) and (c)). The DD figures show that passing (purple diamonds) and failing (red dots) drugs follow remarkably similar patterns before disclosure. This remains true even when only OAI grades are counted as "failing," as in panel (c). The closeness of these patterns motivates my use of the DDD design, which estimates the effect of negative quality information while differencing out potential selection bias.

After disclosure, passing and failing drugs continue to track each other closely, implying no significant differences in their effects. The corresponding DD coefficients are presented in Table 3, columns (1)–(4). For each outcome and definition of failure, the passing and failing

50. The corresponding plot for shortages in all markets is Figure A.5, panel (c).

51. These disruptions are typically short-run, as the median shortage in my data is resolved within 1.6 years according to UUDIS data. The average effect in large markets is economically and statistically insignificant, at 0.6pp.

coefficients are remarkably close: for example, drug market share increases by 2.1pp after a passed inspection, and by 2.0pp after a failing one (cols. (1)–(2), panel A).[52] The DDD design formally quantifies the differences between these passing and failing coefficients. Consistent w the closeness in the DD estimates, the DDD event studies (Figure 3, panels (b) and (d)) and average effects (Table 3, columns (5)–(6)) are uniformly insignificant and close to zero.[53]

In sum, the disclosure of even the most severe inspection failures has no measurable effect on intermediary purchases or available drug quality. These effects are starkly different from those of recall disclosures, which substantially reduce the share of low-quality drugs. I interpret the difference in these results as suggestive evidence of intermediary preferences for supply reliability, as signaled by recall announcements, as opposed to quality per se.[54] Alternative explanations—such as intermediaries being uninformed about inspection results, or inspection failures signaling less severe quality problems—are less plausible, for two reasons. First, sophisticated intermediaries should be aware of inspection failures, which are often well-publicized[55] and can jeopardize government contracts (FDA 2024a). Second, while recall announcements may signal slightly different (e.g., more urgent) quality information than the average inspection failure, both reveal potentially "injurious" violations, implying that quality-seeking buyers should respond to them. OAI inspection failures, in particular, should provoke large responses—they signify the FDA's most severe quality concerns and its intention to pursue enforcement action.

Instead, intermediaries may not respond to inspection disclosures simply because they are not costly for them. Recalls, but *not* inspections, cause expensive supply disruptions, as discussed in Section 2. Consistent with this, intermediary reference manuals describe how they penalize manufacturers for recalls—but do not otherwise mention quality (CardinalHealth 2022; AmerisourceBergen 2019). My results suggest that intermediaries may have limited incentives to act in patients' best interests when selecting quality. They also inform policy-makers' design choices, suggesting that the disclosure of drug reliability information—rather than of quality alone—may be most effective for improving patient outcomes.

*Effects of recalls on market structure and patients.* Finally, I study how recall disclosures affect manufacturer participation and patient welfare, focusing on prices and medication adherence. I first estimate how the number of manufacturers per market evolves after disclosure. Figure 4, panel (a) plots the dynamic effects on the recalled drug's exit probability,[56] showing a pro-

52. This estimate uses the broader definition of failure (i.e., VAI or OAI), reported in panel A. Estimates using the stricter definition (OAI only) are also close together and are presented in panel B.

53. The average effects range in size from -3% to 14% of the baseline values, and are precise enough to rule out changes greater than 60% of the reductions observed after recall disclosures.

54. Figure A.6, panel B presents complementary evidence from recalls with different manufacturer-reported reasons. Among the top eight reasons, recalls due to supply-related factors (violations of good manufacturing practices and facility sterility) have the largest, significant responses. Recalls with strong associations for patient health (adverse events and drug contaminations), in contrast, have small, insignificant effects. These patterns are consistent with the idea that intermediaries may prioritize supply reliability over patient-relevant dimensions of quality.

55. See, for example, the dedicated section on Fierce Pharma at https://www.fiercepharma.com/warning-letters.

56. 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).

nounced increase in exit in three years after disclosure. The plot is reproduced (and rescaled) in panel (b), which overlays the effects on competitor participation (in green dots). The figure shows that new drug entry more than offsets recalled drug exit, resulting in a small net gain in market size. The average effects are quantified in Table 4, which shows that recalled drugs become 6.7pp more likely to exit (155% of the average drug's probability, col. (1)), consistent with their estimated market share losses.

Column (3) shows that recall markets gain 0.6 (9%) net participants on average after disclosure. In my robustness checks, I show that this increase is too small to explain the large estimated declines in recalled drug purchases. Specifically, I re-estimate the DD specification using a sample of markets that experience no net participation growth. The estimated effects are slightly smaller but significantly indistinguishable from the baseline ones (-0.029 versus -0.044; Table A.3, row 12 versus Table 2, col. (1)), suggesting that manufacturer participation does not have first order effects on available quality.

Turning to the patient-relevant outcomes, I estimate a modest decline of 11% in the average drug price after recalls (significant only at $p < 0.10$; Table 5, col. 6). This coefficient implies a change in weighted average unit prices of less than \$0.01.[57] It is consistent with prior research, which finds only limited price changes after generic supply disruptions (Dave et al. 2018). The negative sign of the coefficient could be explained in several ways: for example, the disclosure of poor reliability information could increase competition for low-reliability drugs, causing them to lower their prices. Ultimately, since I observe only winning prices, I require a structural model to fully unpack this result.

In my last analysis, I examine how quality disclosure affects patient behavior, focusing on patient adherence to prescribed medications. Patient non-adherence is prevalent among prescription drug users, particularly those with chronic conditions, and is associated with morbidity and mortality costs of approximately \$500bn annually (Watanabe, McInnis, and Hirsch 2018). I present the event study in Figure 4, panel (c), which shows that non-adherence increases rapidly after disclosure, with the largest effects found in the first post-recall year. On average, recalls increase non-adherence by 2.1-3.3pp (4%-5%), depending on the measure used (Table 4, cols. (5)–(6)). This result aligns with previous research that emphasizes how hassle costs—like a stockout at the nearest pharmacy—can nudge patients into non-adherence (McHorney and Spain 2011). In my context, non-adherence may be driven by reduced drug availability, or (theoretically) by negative quality signals, among other factors. While I am agnostic regarding its proximal cause, I note that the estimated non-adherence effects may have large costs in terms of patient health, even relative to benefits of improved drug quality.[58]

*Summary of quasi-experimental evidence.* Disclosure significantly improves the quality of generic drugs available to patients, even when they have limited ability to observe quality themselves. These improvements last for up to ten years—far longer than the duration of a typ-

57. I present price results weighted by pre-recall drug volumes to facilitate comparison with the structural results. The results using unweighted prices are similar in magnitude (-14%) but have significant pre-trends.

58. My claims data are underpowered to detect the health costs of low-quality, recalled drugs (i.e., in terms of hospitalizations).

ical recall—suggesting that they are driven by long-run changes in intermediary preferences or manufacturer costs. Comparing the effects of different disclosure mechanisms implies that manufacturers are more forcefully penalized for signals of low reliability, rather than of quality alone.

Taken together, my results are consistent with a story where intermediaries observe quality disclosures and respond by punishing unreliable manufacturers. An alternative explanation is that manufacturers change their costs and prices after recalls, as they reoptimize their capacity and effort. Since the DD framework estimates changes in equilibrium outcomes, combining both supply- and demand-side channels, I cannot use it to isolate the role of intermediaries, or to predict how available quality would evolve without them. Doing this requires understanding how manufacturer prices would change if intermediaries did not care for drug reliability; this, in turn, presupposes information about intermediary preferences and the distribution of manufacturer costs. Ultimately, I require a structural model to recover these parameters and to estimate the welfare effects of intermediaries and quality disclosure. In the next sections, I develop a model of generic procurement and reliability information that allows me to address these unanswered questions.

# 5 An Empirical Model of Generic Procurement

In this section, I develop and estimate a scoring auction model of generic procurement. This model allows me to recover intermediary preferences for reliability, and to quantify the role of intermediaries in disciplining available drug quality. In Section 6, I use this model to study the welfare effects of counterfactual quality policies.

## 5.1 Model

I first present the scoring auction model, which describes how intermediary preferences for reliability affect equilibrium prices and quality. I reference the auction framework and notation developed by Laffont et al. 2020, which I adapt to a setting with incomplete price data in the next subsection.

Suppose an intermediary[59] 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 price, sealed bid auction with no reservation score. Manufacturer bids are differentiated by two characteristics: the price $b_i$ and the manufacturer's historical reliability $q_i$, which the intermediary weighs using a scoring rule. The intermediary forms expectations about manufacturer reliability—and manufacturers about intermediary scoring—through repeated interactions over years and across thousands of markets. By the start of the auction, drug reliability and the scoring rule are thus assumed to be predetermined and common knowledge. The econometrician, however, does not observe the

59. To improve clarity, I describe the case of a single buyer, but the model may be trivially extended to describe several monopolist buyers.

scoring rule and must estimate it.[60]

Manufacturers are characterized by their reliability type $q_i$ and a random productivity shock $\theta_i$, which jointly determine their cost $C(\theta_i, q_i)$ of producing a drug. While manufacturer reliabilities may be freely correlated—for example, across drugs and over time—their productivity shocks are assumed to be independent of their competitors'. That is, $\theta_i$ is independently distributed conditional on reliability 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 (Milgrom and Weber 1982).

The intermediary receives a payoff of $U(q_i, b_i) = S(q_i) - b_i$ from awarding a drug contract with reliability $q_i$. Given manufacturers' production technologies, Laffont et al. 2020 derive the intermediary's *optimal scoring* mechanism[61]. For each bid, the intermediary assigns the score

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

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.[62]

*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:[63]

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

This equilibrium expression is very similar to the corresponding one from the price-only auction literature (e.g., Proposition 2 in Riley and Samuelson 1981). It differs in two ways, which cause

60. These features of the structural model reflect the institutional characteristics of intermediary procurement described in Section 2. The inclusion of prices and *reliability* in the scoring rule reflects intermediaries' stated decision-making processes, as described in their reference manuals. It is also consistent with the quasi-experimental results, which show large responses to reliability disclosures.

61. This direct mechanism satisfies the manufacturers' incentive and rationality constraints within the first-score auction space.

62. The empirical scoring auction literature (e.g., Krasnokutskaya, Song, and Tang 2020) has focused on linear scoring rules, which are simpler to implement and common in real world applications where the scoring rule is known. As I describe in the next subsection, since I do not observe the scoring rule, I test the robustness of my results to estimating the model with both optimal and linear scores.

63. Equilibrium relies on the following additional technical assumptions: manufacturers are risk neutral and know the number of bidders; costs are monotonically increasing in theta; and high- and low-reliability firms do not lose with certainty.

manufacturers of different reliability types to bid different strategies, even in the same auctions. First, each manufacturer type has a potentially different cost technology $C(\theta_i, q_i)$ and productivity distribution $F(\cdot|q_i)$. Since the optimal bid is strictly increasing in both costs and productivity, these differences in these values generate asymmetric bidding behavior.[64] (This monotonicity of the bidding equation will also be useful for estimation). Second, manufacturers face distinct choice probabilities $P(\theta_i, q_i, q_{-i})$, depending on their type. For example, intermediary preferences for reliability increase the probability that high-reliability types win, which enhances their effective differentiation and tends to increase their optimal price. For low-reliability types, preferences work in the opposite direction and tend to lower prices.

*Equilibrium with reliability disclosure*. Together, the buyer scoring rule (Equation (3)) and manufacturer bidding strategy (Equation (5)) determine the equilibrium prices and quality of drugs available to patients. The basic structure of the model changes in three ways when new information about reliability is disclosed through recalls. First, disclosure may affect bidder variable costs. For example, recalled manufacturers may respond to disclosure by increasing their effort or reducing their capacity (and thereby increasing costs) in order to improve reliability. The first term of optimal bidding function (Equation (5)) implies that increases in manufacturer costs are partly passed on to the intermediary—and ultimately patients—through higher prices. Second, manufacturers' optimal bidding strategies may change: for example, low-reliability types may strategically reduce their prices in order to improve their chances of winning (which enter the second term in Equation (5)).

Third, recalls may allow buyers to update their expectations about bidder reliability. As previously noted, if intermediaries prefer more reliable manufacturers, disclosure decreases the effective competition among high-reliability types and raises their optimal prices. If reliability is costly, disclosure may also reduce productive efficiency, by shifting intermediary purchases to higher-cost bidders. Theoretically, whether disclosure is socially optimal depends on whether the benefits of improved reliability outweigh the potential costs of reduced competition and productive efficiency.

Finally, I note that this analysis is only partial, since I do not model patients' demand responses, or long-run effects on manufacturer participation and quality investment. The quasi-experimental results, however, suggest that these omitted forces are unlikely to have first order consequences in the short- to medium-run. For example, the average market experiences no measurable change in total patient consumption after disclosure. Similarly, small estimated changes in manufacturer participation have no significant effects on intermediary purchases or available drug quality. I leave the study of these long-run forces to future work.

## 5.2 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 $F(\theta_i|q_i)$. By

64. I follow the notation used in Laffont et al. 2020, which implies that costs are increasing in the "inefficiency" $\theta$. For exposition, I refer to $\theta$ as productivity.

adapting the arguments of Guerre, Perrigne, and Vuong "GPV" 2000, these objects can be non-parametrically identified with complete data on prices, qualities and the number of bidders (Laffont et al. 2020). My innovation is to demonstrate how the model can be identified and estimated using incomplete pricing data, as is common in empirical applications.

Identification with incomplete prices is non-standard because the model incorporates two forms of bidder asymmetry, as previously noted: variation in *ex ante* bidder characteristics (costs and productivities) and in buyer preferences for bids. Models with only the first form of asymmetry are identified from data on the winning bid and the winner's identity (Athey and Haile 2002). In this subsection, I show that a model with both forms of asymmetry is also identified from these data (and complete reliabilities), for both optimal and linear scores. My approach is general and can be easily adapted to other scoring formats, such as polynomial scores (e.g., $S(q_i) - b_i^2$) and quality-to-price ratios.

*Unobserved losing prices.* First, I describe the information that can be recovered about the unobserved price distributions from winning price data. The scoring rule (Equation (3)) implies that 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 scoring rule and the composition of bidder types per auction.

In symmetric auctions—where bidders of only one type participate—the unobserved 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 (since bidder types are also identical). This allows me to recover the price distribution using the properties of the first order statistic (Athey and Haile 2002). However, the order statistics method cannot be used to identify the price distribution in asymmetric auctions, where prices are neither independently nor identically distributed. To see this, note that the optimal bidding strategy in Equation (5) depends on $q_i$, and thus implies different and potentially correlated price distributions for each manufacturer type. Furthermore, if buyers prefer reliability, winning prices need not be minimum prices, unless they were submitted by the least reliable bidders.

For most scoring rules, identifying buyer preferences requires the price distributions from at least one type of auction with variation in winner type (i.e., an asymmetric one).[65] I now show that the price distributions from asymmetric auctions can be recovered, up to a guess of the buyer preferences. For exposition, I focus on the case of the *linear score* with two bidder reliability types, $t \in \{H, L\}$. In the Technical Appendix, I show that the identification argument can be extended to other scores 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:

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

65. In the special case of the *optimal score*, the price distributions from symmetric auctions (of each manufacturer type) are sufficient to identify buyer preferences, as I describe later in this subsection.

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 reliability types:

$$(6) \quad Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) = \left[1 - G_L(b_{H,i} - \Delta S)\right]^{N_L} \left[1 - G_H(b_{H,i})\right]^{N_H - 1} g_H(b_{H,i}).$$

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

$$(7) \quad Pr(b_{L,i}|i \text{ wins})Pr(i \text{ wins}) = \left[1 - G_H(b_{L,i} + \Delta S)\right]^{N_H} \left[1 - G_L(b_{L,i})\right]^{N_L - 1} g_L(b_{L,i}).$$

In these equations, the left-hand side is the ratio of two probabilities—the likelihood that the winning price $b_t$ is observed, and that the focal bidder $i$ wins. Both of these objects can be estimated from the observed data,[66] which implies that Equations (6) and (7) define a system of two nonlinear functional equations in the two unknowns that I wish to recover, $g_L(b_L)$ and $g_H(b_H)$. Since the equations include the cumulative distribution functions, the solution at each price in the support depends on the sum of all prior solutions. 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 the preference $\Delta S$. Second, the observed winning price provides a lower bound for the losing one, which allows 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 left-hand side objects using standard non-parametric kernel density estimators. Throughout, I use triweight kernels and rule-of-thumb bandwidths, following Li, Perrigne, and Vuong 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 a defined cutoff, as long as the grid is sufficiently fine.[67] Above this cutoff, one bidder type wins with certainty, implying that the losers' bids are not defined by the model (Hubbard and Paarsch 2014). Monte Carlo simulations presented in Technical Appendix A.10 indicate that this procedure recovers the true bids and preferences without bias.[68]

*Costs and productivities.* With complete information about prices, Laffont et al. 2020 show that manufacturer costs and productivities can be identified by extending the methods of GPV. As noted in the previous subsection, the equilibrium bidding function (5) is strictly monotonic in $\theta_i$. The one-to-one mapping between $\theta_i$ and $b_i$ implies that a firm's probability of winning given $\theta_i$ is equivalent to its probability given $b_i(\theta_i)$; in equations, that is $P(\theta_i, q_i, q_{-i}) =$

66. Specifically, they can be estimated using standard kernel methods, described in the next paragraph.

67. The cutoff is $\overline{b} - \Delta S$ for $b_L$ and $\overline{b} + \Delta S$ for $b_H$. For derivations of (6) and (7), as well as proofs and additional conditions regarding uniqueness, please refer to the Technical Appendix. I also provide additional details regarding the estimation algorithm there.

68. After following the multi-stage estimation algorithm described in this subsection, the standard deviation of the estimated parameter $\Delta S$ is approximately 50% larger than when using the methodology of Laffont et al. 2020 with complete price data.

$H(b_i(\theta_i), q_i, q_{-i})$. The bidding strategy can therefore be rewritten, after differentiating with respect to $\theta_i$, as:

$$b(\theta_i, q_i, q_{-i}) = C(\theta_i, q_i) - \frac{H(b_i, q_i, q_{-i})}{H_b(b_i, q_i, q_{-i})}. \tag{8}$$

To apply Equation (8) with partial price data, I first rewrite $H$ using Bayes' Rule:

$$H(b_i, q_i, q_{-i}) = \frac{Pr(b_i|i \text{ wins}, q_i, q_{-i})Pr(i \text{ wins}|q_i, q_{-i})}{Pr(b_i|q_i, q_{-i}),} \tag{9}$$

where each object on the right-hand side 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 (8) 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 $\theta_i$ and $q_i$. Productivities are then estimated as the residual from projecting logged costs on bidder reliability:

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

*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 $\Delta S$, and then selecting the value that minimizes the distance between the observed and simulated prices. This would set up a nested fixed point algorithm similar to that used in single-agent dynamic games (Rust 1987). However, finding the equilibrium prices in asymmetric auctions (the "inner loop") is computationally demanding and usually requires numerical methods that introduce approximation error (Hubbard and Paarsch 2014).[69] Instead, it is computationally convenient to reformulate the estimation problem from one in *price space* to another in *productivity space*, as I describe below. This approach is valid because each guess of $\Delta S$ is associated with a unique equilibrium price solution (Lebrun 1999).

Conceptually, my approach has similarities to 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

69. For example, the equilibrium prices from a linear score solve the following non-linear system of ordinary differential equations (ODEs):

$$v_L'(b - \Delta S) = \frac{1 - F_L(v_L(b - \Delta S))}{(b - v_H(b))f_L(v_L(b - \Delta S))}$$
$$v_H'(b + \Delta S) = \frac{1 - F_H(v_H(b + \Delta S))}{(b - v_L(b))f_H(v_H(b + \Delta S))},$$

where $N_L, N_H = 1$ and $v = b^{-1}$. This has the following boundary conditions: $b(\overline{\theta}) = \overline{b}$, $b_H(\underline{\theta}) = \underline{b} + \Delta S$ and $b_L(\underline{\theta}) = \underline{b}$, where $\underline{b}$ is not known. Solving this system is challenging because the Lipschitz condition does not hold near $\overline{\theta}$, implying that standard numerical methods for ODEs do not apply. Hubbard and Paarsch 2014 compare various numerical strategies for solving this type of problem, including shooting, projection and fixed point methods. These methods introduce approximation error and are generally not guaranteed to be stable given estimation error in $\hat{f}(\theta)$.

the prices from asymmetric auctions to estimated productivities, which I then compare against the "true" productivities from symmetric auctions in my outer loop. While my approach is tractable and avoids approximation error, it introduces an additional source of estimation error.

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

$$S(q_i) - \hat{\eta}_i \geq S(q_j) - \hat{\eta}_j \quad \forall j \in \{1, ..., N\}, \quad j \neq i, \tag{10}$$

where the distribution of the "random term" $\hat{\eta}_i = \Gamma(\hat{\theta}_i, q_i)$ is a known function of the productivities $\theta_i$ and has thus already been recovered. In this case, Matzkin 1991 shows that the buyer preferences $\Delta S$ are non-parametrically identified and can be estimated $\Delta S$ by maximizing the following likelihood equation:[70]

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

where $A$ is the number of auctions and the indicator $1(i \text{ wins } a)$ is observed in the data. 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 relies on two key assumptions. First, it assumes that bidder productivities are identically distributed across symmetric and asymmetric auctions (conditional on bidder type and auction-level characteristics).[71] This assumption is standard in the empirical auction literature, where it is applied to allow researchers to pool data across auctions during estimation (GPV). Its main threat is that bidders selectively enter symmetric and asymmetric auctions in ways that cause their productivity distributions to differ. In the next subsection, I provide empirical evidence supporting this assumption. Second, it assumes that the buyer's scoring rule only depends on the characteristics of the focal bidder (Equation (3)). This assumption is specific to the optimal score; in general, most scores depend on all bidder characteristics through the price. Together, these two assumptions imply that the distribution of optimal scores in symmetric auctions fully identifies those in asymmetric ones, allowing the researcher to apply Equation (10) for estimation.

Next, I propose a more general method that identifies the buyer's preferences under other scoring formats, including the *linear score*. My approach is computationally tractable and accounts for the fact that the score may depend on all bidder characteristics.[72] To pin down the preferences, it relies on matching the distributions of $\hat{\theta}_i$ from asymmetric auctions—which are conditional on the guess of $\Delta S$—to the $\hat{\theta}_i$ recovered from symmetric auctions—which are not.

70. In contrast to typical applications of maximum likelihood estimation, Matzkin's procedure does *not* involve parametric assumptions on the error term, since their distribution is already recovered. In this sense, the estimation of $\Delta S$ is fully non-parametric.

71. I discuss how I control for auction-level observables and fixed effects later in this subsection.

72. This implies that scores are non-identically distributed in symmetric and asymmetric auctions, and thus the distribution of $\hat{\eta}_i$ in Equation (10) is not known.

As before, the identifying assumption is that the distribution of $\theta_i$ is identical across auctions, conditional on bidder type and auction observables. I then estimate $\Delta S$ using maximum likelihood:

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

*Sample and heterogeneity.* To estimate the structural model, I use bid level data on drug reliability, auction characteristics and winning prices. Descriptive statistics for these data are presented in Table A.5. I measure historical reliability using an indicator for drugs that were recalled within the last five years (which are considered "low-reliability").[73] This definition is consistent with the reduced form evidence, which shows persistent effects on the recalled drug share for up to a decade. It also formalizes the idea that low quality and low reliability are closely related. From the patient's perspective, recently-recalled drugs are low-quality; from the intermediary's, they are also unreliable, because recalls signal a high risk of future supply disruptions.[74] The table shows that, under this definition, unreliability is relatively rare: 27% auctions have at least one unreliable bidder, and 5% have two or more. To take my model to the data, I further assume that intermediaries evaluate reliability using one collective, optimal score, and then explore the robustness of my estimates to other scoring assumptions.

The auctioned supply contracts in my data are observably heterogeneous—differing, for instance, in their realized volumes and drug characteristics. I therefore use the standard method of Haile, Hong, and Shum 2003[75] to semi-parametrically control for auction-level characteristics. The results from the first stage homogenization are presented in Table A.6. My regression includes the realized contract volume and volume squared, the intermediary's identity, the number of bidders and drug characteristics that may affect manufacturing costs, such as the therapeutic area (ATC-1) and extended delivery mechanisms. Since generic prices exhibit strong deflationary trends during my sample period, I also include year fixed effects. The estimated coefficients are consistent with significant and decreasing economies of scale, modest price variation across intermediaries and higher prices for extended-release drugs. I also observe significant price variation across therapeutic areas: for example, drugs that treat neurological conditions (e.g., anti-seizure medications) are priced 12% to 42% higher than cardiovascular and respiratory drugs (e.g., statins and cough suppressants). Overall, the estimated patterns are very similar to those found by Cuddy 2020 in this market.

I plot the distributions of the winning homogenized prices in Figure 5. The figure shows that high-reliability prices (solid blue) are systematically greater than (i.e., nearly stochastically

73. I experiment with other definitions of reliability in my robustness checks, discussed in Section 5.3. These generate similar estimates of buyer preferences.

74. However, this definition does not capture other observable and unobservable characteristics of low quality, such as those disclosed by failing inspection reports.

75. This adjustment assumes that observable heterogeneity enters the prices in a separable fashion. It is valid if the observed characteristics shift the distribution of productivities in the same way for all bidders. In the case of optimal and linear scores, this adjustment preserves equilibrium bidding.

dominate) those of low-reliability types (dashed green). This pattern could potentially be consistent with both sources of bidder asymmetry: first, low-reliability manufacturers might enjoy a lower cost structure, passed through in the form of lower prices. Second, intermediaries might prefer high reliability, causing low-reliability types to only win at lower prices. A key objective of the structural model will be to separately estimate these sources of asymmetry.

For the purposes of the structural model, I define a market as a molecule-formulation-strength combination. (This matches the level at which generic auctions take place (CMS 2024), but is slightly smaller than the standard definition from the literature.) Following Cuddy 2020, I limit my sample to oral generics, which are manufactured using plausibly similar production technologies and represent over 70% of my sample.[76] Finally, I trim the top 5% of homogenized prices and auctions by size (those with more than 8 bidders). My final sample includes 276,495 bids from 67,880 auctions over 2010-2022. From this, I set aside observations from 2022, which I later use as a test sample to evaluate the fit of my model.

### 5.3 Results

Figure 6 presents the key objects recovered from the structural model: the distributions of manufacturer costs (panel (a)) and productivities (b), and intermediaries' choice probabilities (c). Each object is presented in homogenized price terms and separately for high- (solid blue) and low-reliability (dashed green) manufacturers. Panel (a) plots the first object, the cumulative densities of manufacturer unit costs $C(\theta_i, q_i)$. It shows that drug costs are vary widely, with dispersion comparable to that of the winning prices. High-type costs stochastically dominate low-type ones, suggesting that reliable drugs cost significantly more to produce (31% on average).

My cost estimates are credible, in the sense that they imply price-cost margins close to those reported by generic manufacturers. To calculate price-cost 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 then compare the median estimated margin to the operating margins reported by the largest manufacturers in my data. This comparison is only approximate, since the reported margins include certain unrelated line items.[77] Nevertheless, it is reassuring that my estimated margin (25%) falls within the manufacturers' reported range of 18-29%.

The second objects—the estimated productivity distributions—are presented in Figure 6, panel (b). After controlling for average differences in high- and low-type costs, I find that manufacturer productivities are distributed relatively similarly. Above the 80th percentile, low-type manufacturers have slightly larger estimates, or less efficient production capabilities.

The final, and most important objects are the buyer preferences for reliability, which govern

76. Injected and topical drugs are typically more expensive to produce, in part because they require sterile manufacturing conditions and specialized equipment.

77. Reported margins typically include fixed general and administration costs, as well as ex-U.S. profits and specialty drug profits (which may decrease *or* increase the reported margin). I use data from the following six manufacturers with public financial reports: Teva, Zydus, Lupin, Glenmark, Aurobindo, Mylan and Dr. Reddy's. I exclude data from 2013-2015, when several manufacturers participated in a price-fixing cartel.

the extent to which intermediaries discipline quality in my model. Identifying the preferences requires the standard assumption that bidder productivities are identically distributed in symmetric and asymmetric auctions (conditional on type and auction-level characteristics). I now briefly discuss the empirical content of this assumption before providing support for it.[78]

In an asymmetric auction context, at least three potential sources of non-random variation in productivity threaten this assumption. To illustrate each source, I refer to an example of a symmetric, high-reliability auction $S$ (e.g., $N_H = 2, N_L = 0$), which I compare against an asymmetric auction $A$ that differs only in that it has recently experienced a recall ($N_H = 1, N_L = 1$). First, if recalls are not randomly assigned—for example, if they are more likely to affect drugs with small productivity shocks—then the high-reliability shocks in auction $S$ will be smaller than those in $A$, on average (*reclassification bias*). To test for this source of bias, I estimate productivities in *not-yet* symmetric and asymmetric auctions and compare their distributions. This test exploits the same source of plausibly exogenous variation in recall timing as the quasi-experimental strategy.[79] Figure A.7, panel (a) presents the estimated distributions across symmetric, high-reliability (solid blue); symmetric, not-yet low-reliability (solid green) and not-yet asymmetric (dashed purple) auctions. The closeness of these distributions—confirmed using a Wald test—suggests that the identifying assumption holds, at least *before* each auction's first recall. Since I cannot directly compare symmetric and asymmetric productivities *after* recalls,[80] I use a complementary set of tests to explore whether drugs differ observably in this period. I focus on manufacturer size, which has first order implications for productivities through economies of scale. Figure A.7, panels (c)—(h) present the distributions of manufacturer size in the main estimation sample by auction type. The plots show no systematic differences between symmetric and asymmetric auctions, and Wald tests fail to reject the null hypothesis that the means are the same, suggesting that the productivities may also be balanced.

Second, if manufacturers are more likely to enter or exit auctions with recent recalls, this may alter the productivity distributions in symmetric and asymmetric auctions (*selective participation*). To test for this possibility, I estimate the baseline model using a subsample of auctions without changes in bidder participation. These results are presented in Table A.7, row 5, and are qualitatively similar to the baseline estimates (row 1). Finally, it is possible that high- and low-reliability manufacturers may respond to recalls by reducing or increasing their costs (*strategic cost changes*). I am not able to separately identify these changes across high- and low-reliability types, and I maintain the assumption that all estimated changes in costs before and after recalls are driven by the recalled drugs.

I find that intermediaries have substantial preferences for reliability: they are willing to pay up to 13% more to purchase a high-reliability drug (or \$0.75, relative to the mean homogenized

78. This assumption can be justified by a stationary competitive environment with random bidder entry (Athey and Haile 2002).

79. I compare estimated productivities in symmetric, high-reliability (never-recalled) auctions; symmetric, soon-to-be low-reliability (not-yet-recalled) and soon-to-be asymmetric (mixed never-recalled and not-yet low-reliability) ones.

80. Recovering productivities in asymmetric auctions requires knowing the equilibrium asymmetric bidding strategy, which depends on the guess of the buyer preferences.

price of $5.88). Figure 6, panel (c) visualizes how buyer preferences affect their choices in my data, plotting the choice probabilities as functions of bidder type and price. It shows that buyers aggressively penalize low-reliability manufacturers (dashed), whose probability of winning declines faster and farther in prices, relative to high-reliability types (solid). One way to interpret the preference estimate is as intermediaries' expected value from avoiding future supply disruptions. Since recalled drugs are 3.4 times more likely to be recalled again (within 5 years, relative to not recalled ones), this implies that intermediaries are willing to pay a 2% premium for each 10% reduction in their future recall probability.[81]

The preference estimates are interesting because they show that intermediaries prioritize reliability, breaking with standard characterizations of the generics market. The FDA, for example, has recently described the market as fundamentally unable "to observe and reward quality" or to "penalize manufacturers that fail to invest in... facilities to assure a reliable supply" (FDA 2019). On the other hand, these estimates are perfectly consistent with intermediaries' statements about their decision-making processes, which describe reliability as a first order concern (Section 2). They are also broadly similar to preferences for reliability reported from other scoring auction settings, which range from 19%-68%, suggesting my estimates may be on the conservative end.[82]

*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 focus on the model's ability to predict post-recall outcomes *in sample*, for observations used in estimation. I predict the equilibrium outcomes—recalled and competitor prices and market shares—using pre-recall, *ex ante* data, and then compare the predictions against the observed data. The results from this exercise are presented in Figure 7. Panel (a) presents the cumulative winning price distributions, and panel (c) shows the shares of high-reliability drugs after recalls. The close correspondence between the model's predictions (dashed) and the real values (solid) indicates the good fit of the model in sample. Next, I repeat this exercise using post-recall data from the test sample (the year 2022, which was not used in estimation). This sample includes 261 recalls, whose outcomes I predict only once. The results are presented in panels (b) and (d), which show that predicted *out-of-sample* fit is also excellent.[83]

Finally, I assess the robustness of my model to perturbations in the estimation sample and maintained assumptions. Table A.7 presents the estimates of buyer preferences (col. (1)) and average difference between low- and high-type bidder costs (2) recovered with these changes. The

81. Specifically, previously recalled drugs are recalled 15.5% of the time, versus 4.5% for not recalled drugs. Given buyers' 13% preference for high-reliability drugs, this implies an elasticity of $.13/((.045/.155)-1)*.1=2\%$.

82. These include 19% for public utilities in Italy (Spagnolo 2012), 23%-68% for public electricity contracts in the United States (Cameron 2000), 33%-51% in online private auctions (Krasnokutskaya, Song, and Tang 2020; Jin and Kato 2006), < 50% for public contracting in Finland (Jääskeläinen and Tukiainen 2019), and > 60% for public IT service contracts in Italy (Albano, Dini, and Zampino 2009). My estimates may be lower because intermediaries are typically reimbursed for the direct costs of recalled drugs and are not legally liable for associated injuries.

83. I note that confidence intervals are not typically reported in the empirical non-parametric auction literature, since it is not clear whether the estimators converge in a distributional sense. Marmer and Shneyerov (2012) and Ma et al. (2019) have recently proposed pointwise confidence intervals for price-only auctions, but I am not aware of any research that has applied these to scoring auctions.

first row presents the baseline model estimates—assuming that buyers use an optimal score, and that low-reliability drugs are those recalled within the past five years. In the second row, I apply an alternative definition of low-reliability, which includes drugs recalled in the last decade but excludes those recalled in the last year. Intuitively, this definition tests the persistence of buyer preferences while dropping the auctions most likely to be affected by inertia. In the third row, I assess heterogeneity in the estimates by auction size, and present the estimates from a sample of auctions with $N = 2$ bidders. In row four, to isolate changes in recalled drug costs before and after disclosure, I present estimates from a sample of recalled and not-yet-recalled drugs (similar to that used in the quasi-experimental results). In row five, to evaluate the importance of selection in manufacturer participation, I estimate the model using a "balanced" panel of auctions without entry or exit. The final row presents the estimates when buyers are assumed to use a linear score. In each test, I estimate positive intermediary preferences for reliability and greater costs for high-reliability drugs. The directional similarity in these estimates suggests that the key patterns from my model are robust.

# 6 Counterfactual Analyses

In this section, I use the estimated model to study the welfare consequences of informed intermediaries and to evaluate counterfactual quality policies. Throughout, my outcomes of interest are available drug quality, equilibrium prices and manufacturer and intermediary surplus.

To assess the model's implications for patients, I focus on two patient welfare-relevant objects: drug quality and prices. Drug quality affects patient well-being through multiple channels. As previously noted, low-quality (i.e., recently recalled) drugs have manufacturing defects that may cause serious injury or death. Low quality also causes supply disruptions that reduce patient medication adherence (as shown in Section 4.2), which can in turn worsen health outcomes.[84] The second object—wholesale prices—affects patient welfare indirectly, in the form of higher retail ("usual and customary") prices and increased insurance premiums. I do not calculate an explicit measure of patient welfare, since this is challenging when patients pay little to no attention to product characteristics. (In my setting, for example, patients are unable to observe quality and are insulated from price changes through insurance.) I note, however, that drug quality has received extensive media and policy attention, suggesting that it is socially valuable.

My analysis captures changes in manufacturer variable surplus, net of one-time recall costs.[85] I calculate this object using predicted winning prices and unit costs, weighted by realized contract volumes. Since manufacturers are typically required to reimburse intermediaries for the costs of recalled products, I subtract these one-time costs from manufacturer profits.[86] I cal-

84. To the extent that patients can observe quality signals, low quality can also theoretically worsen their perceptions of drug value. While generic aversion is important in many countries (Atal, Cuesta, and Saethre 2022), it does not play a central role in the U.S. market, where generics capture over 90% share. As discussed in Section 2, patients are unlikely to observe specific drugs' quality signals, but may read general media coverage about quality failures.

85. Since I estimate unit costs, I also capture certain fixed costs incurred at the drug contract level (e.g., batch testing costs).

86. I assume that 10% of the recalled revenues are returned, based on the approximate recalled volumes from the

culate intermediary surplus using their payoff function $U(q_i, b_i) = S(q_i) - b_i$, weighted by contract volumes. Since I identify intermediary preferences $S(q_i)$ up to a location normalization, I solve for the location $S(q_i = 0)$ that maximizes intermediary surplus given the other parameter estimates.[87] I measure intermediary surplus using drugs' *ex post* disclosed quality, relying on the assumption that quality is stable over the time horizons that I study (discussed in Section 2). I present all values in non-homogenized, volume-weighted dollar terms.

Finally, I briefly describe how I simulate the counterfactual environments and solve for equilibrium outcomes. As previously discussed, I focus on changes in intermediary choices and manufacturer costs and prices, holding contract volumes, manufacturer participation and quality fixed. I also assume that buyers evaluate bids using the optimal score, which fits my data well and is computationally convenient. In each counterfactual, I simulate 50 draws of productivities $\theta$ from $F(\cdot|q_i)$ per bidder. For each draw of $\theta$, I calculate bidder costs and buyer choice functions, and then integrate over these objects to find bidders' equilibrium pricing strategies (Equation 5). Finally, I use the optimal scoring rule (Equation 3) to identify the winners of each auction.

### 6.1 The Role of Informed Intermediaries

My first counterfactuals study the welfare implications of informed intermediaries, by simulating how available quality would change without them. Since patients lack information about drug quality, intermediaries can theoretically improve welfare by selecting higher-quality, more reliable drugs. Intermediaries also, however, increase the differentiation of high-reliability manufacturers, which may reduce competition and raise prices. Finally, they may decrease allocative efficiency by shifting market share to higher-cost manufacturers.

To quantify these effects, I model a counterfactual environment in which intermediary preferences for reliability (or equivalently, their information about it) are reduced to zero ($\Delta S' = 0$). Table 5, column (1) presents the counterfactual results, which I compare against the status quo in column (2). As expected, informed intermediaries substantially increase the share of high-quality drugs (by 14pp, 27%), relative to the zero-preference counterfactual. Their price effects are also intuitive, though modest: high-reliability manufacturer prices increase by \$0.01 (3%, or 0% at the median), consistent with increased differentiation, while low-reliability prices are essentially unchanged. Taken together, the changes in quality and price imply that patients are almost certainly better off under informed intermediaries. Consistent with these patterns, I find that intermediary responses reward high-reliability manufacturers (whose profits increase by 28%), while penalizing low-reliability ones (whose profits decrease by 37%, but fail to offset the gains for high-types). Intermediary surplus increases slightly, as they make better decisions with more complete information, and total producer (i.e., manufacturer plus intermediary) surplus increases by 1%.

recall data. My main results are not directionally sensitive to this assumption.

87. This value (\$7.40 in homogenized terms) is in the neighborhood of the one obtained through an alternative approach, which calibrates $S(q_i = 0)$ so that intermediary surplus matches their reported profit margins (\$7.85).

My results show that intermediaries improve short-run welfare, relative to a counterfactual where they do not respond to quality disclosure. Intermediaries force manufacturers to compete on reliability, which improves quality without meaningfully changing prices, and slightly increases producer surplus. In the long-run, these actions are likely to encourage manufacturer investment in reliability and may increase the entry of high-reliability drugs.

An alternative interpretation of this counterfactual is that it compares the performance of price-only and scoring auctions. Viewed through this lens, it reproduces a key prediction from the theoretical scoring auction literature: that buyers surplus should increase under scoring rules, as opposed to price-only ones (Asker and Cantillon 2008).[88] Intuitively, this is because price-only auctions limit the space over which buyers and bidders can maximize, consequently reducing their surplus.

*Proposed policies.* In my second counterfactual, I explore whether alternative quality policies—which may improve the alignment between intermediary and social incentives, but are costly to implement—are likely to increase welfare. This exercise is motivated by two recent proposals from the Biden administration that seek to improve pharmaceutical reliability (HHS 2024).[89] The first policy would increase intermediary information about drug reliability by developing and disclosing new reliability measures. The second would subsidize high-reliability drugs, or impose penalties for purchasing low-reliability ones, using a "pay for performance" structure (P4P). Both policies aim to increase demand for reliable drugs, and thereby incentivize reliability investments throughout the supply chain.

My quasi-experimental results suggest that the first policy—targeting transparency—may be redundant in the oral generic market, given intermediaries' already high awareness of drug reliability. Put another way, the "market information failures" that the policy targets are simply less severe in settings where intermediaries are the key decision-makers. The second—targeting intermediary incentives—could theoretically improve welfare if intermediaries undervalue reliability relative to the social planner. To evaluate this possibility, I simulate a counterfactual environment in which intermediary preferences for reliability are increased through a combination of subsidies and penalties. I calculate the social costs of the policy assuming a public cost of funds of 20%, and consider two payment structures: 100% subsidies and a mix of 80% subsidies with 20% penalties.[90]

88. Intermediary surplus is even lower in a closely related counterfactual, with price-only scores and a minimum quality standard. Specifically, if bidders could costlessly "convert" to high-reliability types, buyer surplus would also be lower than in the scoring auction counterfactual (46%). This is because manufacturer costs and prices increase by more than intermediaries are willing to pay for reliability in the status quo scoring auction environment.

89. I refer to the proposed Manufacturer Resiliency Assessment Program (MRAP) and Hospital Resilient Supply Program (HRSP). Related "pay for performance" policies have been proposed by the FDA (Wosińska and Frank 2023) and in the Senate (Senate Legislative Counsel 2024). I note that the HHS white paper focuses on drugs administered in hospitals, in part because this would allow the policy to build on existing Centers for Medicare and Medicaid Services programs without additional statutory authorities (HHS 2024). The white paper states that the policies may be extended to the broader generics market, where supply chain issues are also common, which is where my analysis focuses.

90. The first pays intermediaries a 13% premium for each high-reliability drug purchase. The second assumes that intermediaries bear 20% of the costs for the program, and are alternately taxed (or paid) 13% of the average price for each low- (high-) reliability drug purchase.

The counterfactual policy results are presented in Table 5, column (3). The table reports outcomes from one potential version of the P4P policy, which would double intermediary preferences for reliability (from 13% to 26% of the average price). Relative to the status quo (col. (2)), the P4P policy increases available drug quality by 16pp (24%)—a similar increase to that observed between the no-preference and status quo environments in the first counterfactual (cols. (1)–(2)). Under the P4P policy, however, equilibrium prices grow by a much larger relative amount (by 29% versus 3%). This growth is primarily explained by the adverse effects of the policy on competition: as high-reliability types become more differentiated, they increase their offered prices by an average of 37%. Figure A.8 reproduces these outcomes for other versions of the P4P policy, which alter intermediary preferences to greater or lesser degrees (from the status quo of 13% up to 50% of average prices). Panel (a) plots intermediaries' effective preferences on the $x-$axis against the share of high-quality drugs (solid) and equilibrium prices (dashed) on the $y-$axis. Panel (b) shows the implied elasticity from these changes: the percentage price increase required for each 10% improvement in quality. The figures show that the policy's costs for patients increase gradually around the status quo (a 64% high-quality share), and then sharply above an 80% high-quality share, rising from a 10% elasticity to more than 25%. In other words, while the P4P policy substantially improves quality, it becomes rapidly more expensive as quality increases, with uncertain patient welfare implications.[91]

On the producer side, Table 5 column (3) shows that the policy's costs are primarily concentrated among low-reliability manufacturers. For example, under a policy that doubles intermediary preferences, low-reliability profits decrease by 48%, while high-reliability profits increase by 60%. Total producer surplus (less the social cost of the program) increases by 14-15%, depending on the mix of subsidies and penalties used. Overall, the supply-side effects suggest that the P4P policy is likely to increase manufacturing costs and may drive low-reliability manufacturers from market, which would reduce productive efficiency and competition in the long-run.

## 6.2 Quality Disclosure Mechanisms

In my final counterfactuals, I study the FDA's current recall disclosure policies and decompose their supply- and demand-side mechanisms, linking the structural analysis to the quasi-experimental results. To study the welfare effects of disclosure, I first simulate market outcomes before and after recall announcements (updating $q$ and redrawing $\theta$ to reflect the disclosed quality information).[92] Table 6 reports the results: columns (1) and (2) show the pre- and post-disclosure outcomes, respectively, while column (3) presents the percentage changes. Consistent with my quasi-experimental findings (in Section 4.2), I find that recall disclosure increases the share of high-quality drugs (by 6%) with relatively small effects on prices (-11% on average, or 0% at the median). These results suggest that the disclosure policy benefits pa-

91. Prior research indicates that patient demand for generic drugs is highly inelastic—with estimates ranging from -0.2 to -0.03—suggesting that the increases in quality could still potentially be patient welfare-improving, but this is far from certain (Yeung et al. 2018).

92. For this exercise, I limit my sample to drugs in markets that experience recalls, but no change in manufacturer participation.

tients in the short-run (though they do not consider potential welfare losses from increased drug shortages.)

On the supply side, I find that disclosure makes manufacturers unambiguously worse off (reducing their profits by 28%, on average), by increasing price competition and introducing one-time recall reimbursement costs. These changes help explain why manufacturers do not opt to self-disclose their reliability type, as predicted by the theoretical "unraveling" result of Grossman 1981 and Milgrom 1981. Intermediary surplus increases substantially after recall announcements (approximately doubling), as they benefit from lower prices and improved quality information. Total producer surplus declines moderately (by 14%), suggesting that recall disclosures have ambiguous overall effects on welfare.

Next, I decompose the mechanisms behind these disclosure effects. As discussed in Section 5.1, disclosure may affect market outcomes through three broad channels: intermediary preferences and manufacturer costs and prices. To isolate the role of each, I re-simulate my model while updating one channel at a time. Specifically, for each outcome $y$, I estimate the following equation:

$$\begin{aligned}(11) \qquad \Delta y = &\underbrace{y[R_O(\theta_i, q_i'), C(\theta_i, q_i), b(\theta_i, q)] - y[R_O(\theta_i, q_i), C(\theta_i, q_i), b(\theta_i, q)]}_{\Delta \text{ due to intermediary preferences}} \; + \\ &\underbrace{y[R_O(\theta_i, q_i'), C(\theta_i', q_i'), b(\theta_i, q)] - y[R_O(\theta_i, q_i'), C(\theta_i, q_i), b(\theta_i, q)]}_{\Delta \text{ due to manufacturer costs}} \; + \\ &\underbrace{y[R_O(\theta_i', q_i'), C(\theta_i', q_i'), b(\theta_i', q')] - y[R_O(\theta_i, q_i'), C(\theta_i', q_i'), b(\theta_i, q)]}_{\Delta \text{ due to prices}}.\end{aligned}$$

First, I update *intermediary preferences* while holding the other mechanisms fixed. This is captured by the equation's first right-hand side term, which only updates $q_i$ in the scoring rule. These results are reported in column (4), presented as percentages of the pre-recall values. I find that intermediary responses explain more than 100% of disclosure's quality benefits, and increase the share of high-quality drugs by 18%. By selecting more expensive suppliers, intermediaries also increase costs (24%) and prices (by 22%), implying a modest reduction in productive efficiency.

Manufacturers' supply-side responses partially offset these effects, reducing equilibrium quality, costs and prices. To estimate the role of *manufacturer costs*, I update the model's cost functions while holding prices fixed (captured by the second RHS term in Equation (11)). The results (in col. (5)) show that the average manufacturer's costs decrease by 6% after recalls, driven by changes for recalled manufacturers.[93] These results provide no evidence that manufacturers increase their effort after recalls, but are consistent with a story where recalled manufacturers lower prices to regain share.

93. I note that this result is driven by changes in low-reliability manufacturer costs pre- versus post-recall, and is not imposed by my modeling assumptions. Specifically, I estimate a similar decrease in prices when I limit my sample to recalled drugs before and after disclosure, as in Table A.7 row 4 (discussed in the previous Robustness section). My estimates reflect decreases in variable costs, and do not capture potential changes in fixed costs (such as manufacturing facility investments), which could increase after recalls.

Last, I update manufacturer *prices* (captured by the the third RHS term in Equation (11)). These estimates, in column (6), show that low-reliability manufacturers reduce their prices by 14% after disclosure, passing through their reduced costs and strategically compensating for their dispreferred status. This, in turn, intensifies price competition for high-reliability types, who reduce their prices as well (by 19%). Intermediaries respond to manufacturer price changes by selecting more low-reliability drugs, which decreases the share of high-quality drugs by 11%, and average prices by 32%. The reduction in prices transfers surplus from manufacturers (whose profits fall by 42%) to intermediaries (who gain by 95%), and lowers overall producer surplus by 26%.

In summary, the decomposition results show that quality disclosure benefits patients primarily through intermediary preference channels, as opposed to changes in manufacturer costs or prices. These results help rationalize my quasi-experimental findings: specifically, how disclosure improves quality without changing patient demand. This is possible because intermediary responses magnify the effects of disclosure interventions—which would otherwise be extremely limited—resulting in meaningful benefits for patients.

# 7 Discussion and Conclusion

Quality disclosure can reduce market frictions from asymmetric information—but only when buyers respond to it. In many markets where consumers face high barriers to information—where it is costly to acquire or requires specialized knowledge to use—they instead rely on intermediaries to make informed purchasing decisions for them. This paper studies the welfare implications of quality disclosure and informed intermediaries in the U.S. generic pharmaceuticals market. Quality is of key concern in this setting, given its high stakes for patient health and drug shortages, but patients are uniquely uninformed about it. Together, these factors imply that intermediaries have substantial scope to improve (or potentially distort) equilibrium outcomes.

In this paper, I show that intermediaries discipline quality and enhance the effects of disclosure policies, resulting in substantial welfare gains for patients. I collect novel, comprehensive data on drug manufacturing quality, which I use to document pervasive and persistent quality failures in the U.S. market. I show that the disclosure of these failures significantly improves the quality of drugs consumed by patients, with benefits persisting for up to ten years. To isolate the role of intermediaries, I develop a scoring auction model of generic procurement and a new estimation technique that takes into account the fact that only winning prices are observed. I find that intermediaries are willing to pay a significant premium for more reliable drugs, which encourages manufacturers to compete on reliability and ultimately increases the share of high-quality drugs by 27%.

My results call into question both the conventional wisdom about generics—as having perfectly substitutable quality (FTC 2011)—and typical characterizations of the generic market—as unable "to observe and reward [drug] quality" (FDA 2019). Instead, they show that generic drugs exhibit important and "potentially injurious" variation in quality, which is primarily dis-

ciplined by demand from informed intermediaries. My analyses also have implications for the ongoing design of policies to improve drug quality and reliability. For example, since intermediaries already closely track drug quality signals, additional disclosure interventions may not achieve significant welfare gains in this market. Proposals that increase demand for reliability through targeted subsidies, however, could potentially reduce intermediary distortions and improve equilibrium welfare.

While my paper captures key aspects of how disclosure and intermediaries affect equilibrium outcomes, it does not account for certain long-run economic forces. For example, I do not model changes in manufacturer participation, capacity and quality investment over time, which could lead to important, welfare-relevant changes in drug quality and availability. A natural extension to my analysis would be to measure the long-run effects of disclosure and intermediation on manufacturer quality and patient outcomes.

# References


Ables, Kelsey, and Marlene Cimons. 2023. "What to know about the eyedrop recall linked to 4 deaths and vision loss." *The Washington Post,* https://www.washingtonpost.com/wellness/2023/03/23/eyedrop-recall-2023-pseudomonas-aeruginosa-infection/.

Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential estimation of dynamic discrete games." *Econometrica* 75 (1): 1–53.

Akerlof, George A. 1970. "The market for lemons: quality uncertainty and the market mechanism." *The Quarterly Journal of Economics* 84 (3): 488–500.

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,* 353–363. Springer.

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.

Alpert, Abby, and Mireille Jacobson. 2019. "Impact of oncology drug shortages on chemotherapy treatment." *Clinical Pharmacology & Therapeutics* 106 (2): 415–421.

American Cancer Society Cancer Action Network. 2023. "Survivor Views: Drug Shortages, Telehealth, & Biomarker Testing." Accessed October 29, 2024. https://www.fightcancer.org/sites/default/files/docs/survey_drug_shortages_biomarkers_final_3.19.pdf.

American Society of Health-System Pharmacists. 2024. "Drug Shortage Bulletins." Accessed October 29, 2024. https://www.ashp.org/drug-shortages.

AmerisourceBergen. 2019. "Manufacturer Packaging and Logistics Requirements Guide." Accessed October 29, 2024. https://www.amerisourcebergen.com/-/media/assets/amerisourcebergen/manufacturer/ab-manufacturer-logistics-guideline-v13-updated-20230130.pdf.

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.

Andreyanov, Pasha, Francesco Decarolis, Riccardo Pacini, and Giancarlo Spagnolo. 2024. "Past Performance and Procurement Outcomes." *Available at SSRN 4929595.*

Andreyanov, Pasha, Ilya Krasikov, and Alex Suzdaltsev. 2023. "Scoring and Favoritism in Optimal Procurement Design." *Mimeo.*

Asker, John, and Estelle Cantillon. 2008. "Properties of scoring auctions." *The RAND Journal of Economics* 39 (1): 69–85.

Association for Accessible Medicines. 2024. "2024 U.S. Generic & Biosimilar Medicines Savings Report." Accessed October 29, 2024. https://accessiblemeds.org/sites/default/files/2024-09/AAM-2024-Generic-Biosimilar-Medicines-Savings-Report.pdf.

Atal, Juan Pablo, Jose Ignacio Cuesta, and Morten Saethre. 2022. National Bureau of Economic Research Working Paper 30325. National Bureau of Economic Research. https://doi.org/10.3386/w30325. http://www.nber.org/papers/w30325.

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

Barahona, Nano, Cristóbal Otero, and Sebastián Otero. 2023. "Equilibrium effects of food labeling policies." *Econometrica* 91 (3): 839–868.

Basker, Emek, and Fariha Kamal. 2021. "Recall and response: Relationship adjustments to adverse information shocks." *European Economic Review* 139:103903.

Ben-Shahar, Omri, and Carl E Schneider. 2017. "The failure of mandated disclosure." *Russian Journal of Economics and Law,* no. 4 (44), 146–169.

Benabou, Roland, and Guy Laroque. 1992. "Using privileged information to manipulate markets: Insiders, gurus, and credibility." *The Quarterly Journal of Economics* 107 (3): 921–958.

Biglaiser, Gary. 1993. "Middlemen as experts." *The RAND Journal of Economics,* 212–223.

Brennan, Zachary. 2016. "FDA Allows Banned Chinese Firm to Ship Chemotherapy to US as Shortage Threat Looms." Accessed October 29, 2024. https://www.raps.org/news-and-articles/news-articles/2016/2/fda-allows-banned-chinese-firm-to-ship-chemotherap.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics* 225 (2): 200–230.

Cameron, Lisa J. 2000. "Limiting buyer discretion: effects on performance and price in long-term contracts." *American Economic Review* 91 (1): 265–281.

CardinalHealth. 2022. "Manufacturer Reference Manual." Accessed October 29, 2024. https://www.cardinalhealth.com/content/dam/corp/web/documents/Manual/cardinal-health-pharma-supplier-guidebook-fy22.pdf.

Cawley, John, and John A Rizzo. 2008. "Spillover effects of prescription drug withdrawals." In *Beyond Health Insurance: Public Policy to Improve Health,* 19:119–143. Emerald Group Publishing Limited.

Chandra, Amitabh, Jennifer Kao, Kathleen L Miller, and Ariel D Stern. 2024. "Regulatory Incentives for Innovation: The FDA's Breakthrough Therapy Designation." *Review of Economics and Statistics,* 1–46.

CMS. 2024. "Methodology for Calculating the National Average Drug Acquisition Cost (NADAC) for Medicaid Covered Outpatient Drugs." Accessed October 29, 2024. https://www.medicaid.gov/medicaid-chip-program-information/by-topics/prescription-drugs/ful-nadac-downloads/nadacmethodology.pdf.

Cohen, Robin A, and Laryssa Mykyta. 2024. "Prescription Medication Use, Coverage, and Nonadherence Among Adults Age 65 and Older: United States, 2021–2022." *National Health Statistics Reports.*

Connecticut, State of. 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.

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.*

Cuddy, Emily. 2020. "Competition and collusion in the generic drug market." *Mimeo.*

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.

Davit, Barbara M, Patrick E Nwakama, Gary J Buehler, Dale P Conner, Sam H Haidar, Dewrat T Patel, Yongsheng Yang, Lawrence X Yu, and Janet Woodcock. 2009. "Comparing generic and innovator drugs: a review of 12 years of bioequivalence data from the United States Food and Drug Administration." *Annals of Pharmacotherapy* 43 (10): 1583–1597.

DellaVigna, Stefano, and Joshua M Pollet. 2009. "Investor inattention and Friday earnings announcements." *The Journal of Finance* 64 (2): 709–749.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality disclosure and certification: Theory and practice." *Journal of Economic Literature* 48 (4): 935–963.

Dranove, David, and Andrew Sfekas. 2008. "Start spreading the news: a structural estimate of the effects of New York hospital report cards." *Journal of Health Economics* 27 (5): 1201–1207.

Dubois, Pierre, Gosia Majewska, and Valentina Reig. 2023. "Drug shortages: empirical evidence from France." *TSE Working Paper.*

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India." *The Quarterly Journal of Economics* 128 (4): 1499–1545.

Eban, Katherine. 2019. *Bottle of Lies: The Inside Story of the Generic Drug Boom.* Ecco.

Edney, Anna, and Riley Griffin. 2023. "US Military Is So Worried About Drug Safety It Wants to Test Widely Used Medicines." *Bloomberg,* https://www.bloomberg.com/news/articles/2023-06-07/drug-safety-fears-spur-pentagon-plan-to-test-widely-used-meds.

FDA. 2023. "Import Alert 66-40: Intas Pharmaceuticals Limited." Accessed October 29, 2024. https://www.accessdata.fda.gov/cms_ia/importalert_189.html.

———. 2023. "Warning Letter: Intas Pharmaceuticals Limited 320-23-20." Accessed October 29, 2024. https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/intas-pharmaceuticals-limited-652067-07282023.

———. 2024a. "Pharmaceutical Inspections and Compliance." Accessed October 29, 2024. https://www.fda.gov/drugs/guidance-compliance-regulatory-information/pharmaceutical-inspections-and-compliance.

———. 2024b. "Recalls Background and Definitions." Accessed October 29, 2024. https://www.fda.gov/safety/industry-guidance-recalls/recalls-background-and-definitions.

Fein, Adam J. 2023. "The 2023-2024 Economic Report on Pharmaceutical Wholesalers and Specialty Distributors," Drug Channels Institute.

Fink, Joseph L. 2012. "Legal Duty to Notify Patient of Product Recall?" Accessed October 29, 2024, *Pharmacy Times* 78 (3). https://www.pharmacytimes.com/view/legal-duty-to-notify-patient-of-product-recall.

Force, FDA Drug Shortages Task. 2019. "Drug Shortages: Root Causes and Potential Solutions." Accessed October 29, 2024. https://www.fda.gov/media/131130/download.

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.

FTC. 2011. "Authorized generic drugs: short-term effects and long-term impact," Retrieved October 29, 2024.

Galdin, Anaıs Soledad. 2024. "Trade and Information Provision Across Geographical Space." PhD diss., Princeton University.

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.

Gavazza, Alessandro, and Alessandro Lizzeri. 2021. "Frictions in product markets." In *Handbook of Industrial Organization,* 4:433–484. 1. Elsevier.

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.

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.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong "GPV". 2000. "Optimal nonparametric estimation of first-price auctions." *Econometrica* 68 (3): 525–574.

Haile, Philip, Han Hong, and Matthew Shum. 2003. "Nonparametric tests for common values at first-price sealed-bid auctions." *National Bureau of Economic Research Working Paper 10105.*

Hanazono, Makoto, Yohsuke Hirose, Jun Nakabayashi, and Masanori Tsuruoka. 2016. "Theory, identification, and estimation for scoring auctions." *Mimeo.*

Hanazono, Makoto, Jun Nakabayashi, and Masanori Tsuruoka. 2013. "Procurement auctions with general price-quality evaluation." *KIER Discussion Paper* 845.

Health, Department of, and Human Services. 2024. "White Paper: Policy Considerations to Prevent Drug Shortages and Mitigate Supply Chain Vulnerabilities in the United States." Accessed October 29, 2024. https://aspe.hhs.gov/sites/default/files/documents/3a9df8acf50e7fda2e443f025d51d038/HHS-White-Paper-Preventing-Shortages-Supply-Chain-Vulnerabilities.pdf.

Higgins, Matthew J, Xin Yan, and Chirantan Chatterjee. 2021. "Unpacking the effects of adverse regulatory events: Evidence from pharmaceutical relabeling." *Research Policy* 50 (1): 104126.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton. 2017. "The impact of consumer inattention on insurer pricing in the Medicare Part D program." *The RAND Journal of Economics* 48 (4): 877–905.

Hortaçsu, Ali, and Isabelle Perrigne. 2021. "Empirical perspectives on auctions." In *Handbook of Industrial Organization,* 5:81–175. 1. Elsevier.

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.

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,* 3:37–115. Elsevier.

Jääskeläinen, Jan, and Janne Tukiainen. 2019. "Anatomy of public procurement." *VATT Institute for Economic Research Working Papers* 118.

Jarrell, Gregg, and Sam Peltzman. 1985. "The impact of product recalls on the wealth of sellers." *Journal of Political Economy* 93 (3): 512–536.

Jewett, Christina. 2023. "Drug Shortages Near an All-Time High, Leading to Rationing." Accessed October 29, 2024, *The New York Times,* https://www.nytimes.com/2023/05/17/health/drug-shortages-cancer.html.

Jin, Ginger Zhe, and Andrew Kato. 2006. "Price, quality, and reputation: Evidence from an online field experiment." *The RAND Journal of Economics* 37 (4): 983–1005.

Jin, Ginger Zhe, and Phillip Leslie. 2003. "The effect of information on product quality: Evidence from restaurant hygiene grade cards." *The Quarterly Journal of Economics* 118 (2): 409–451.

Kao, Jennifer. 2024. "Information Disclosure and Competitive Dynamics: Evidence from the Pharmaceutical Industry." *Management Science.*

Kesselheim, Aaron S, Alexander S Misono, Joy L Lee, Margaret R Stedman, M Alan Brookhart, Niteesh K Choudhry, and William H Shrank. 2008. "Clinical equivalence of generic and brand-name drugs used in cardiovascular disease: a systematic review and meta-analysis." *JAMA* 300 (21): 2514–2526.

Kesselheim, Aaron S, Michael S Sinha, and Jerry Avorn. 2017. "Determinants of market exclusivity for prescription drugs in the United States." *JAMA Internal Medicine* 177 (11): 1658–1664.

Krasnokutskaya, Elena, Kyungchul Song, and Xun Tang. 2020. "The role of quality in internet service markets." *Journal of Political Economy* 128 (1): 75–117.

Laffont, Jean-Jacques, Isabelle Perrigne, Michel Simioni, and Quang Vuong. 2020. "Econometrics of scoring auctions." In *Essays in Honor of Cheng Hsiao,* 41:287–322. Emerald Publishing Limited.

Lebrun, Bernard. 1999. "First price auctions in the asymmetric N bidder case." *International Economic Review* 40 (1): 125–142.

Lee, Robin S, Michael D Whinston, and Ali Yurukoglu. 2021. "Structural empirical analysis of contracting in vertical markets." In *Handbook of Industrial Organization,* 4:673–742. 1. Elsevier.

Li, Tong, Isabelle Perrigne, and Quang Vuong. 2000. "Conditionally independent private information in OCS wildcat auctions." *Journal of Econometrics* 98 (1): 129–161.

Lichtenberg, Frank R. 2007. "The impact of new drugs on US longevity and medical expenditure, 1990–2003: evidence from longitudinal, disease-level data." *American Economic Review* 97 (2): 438–443.

Lieberman, Steven M, and Paul B Ginsburg. 2018. "Would price transparency for generic drugs lower costs for payers and patients?" *Brookings Institute Report* 14.

Lizzeri, Alessandro. 1999. "Information revelation and certification intermediaries." *The RAND Journal of Economics,* 214–231.

Matzkin, Rosa L. 1991. "Semiparametric estimation of monotone and concave utility functions for polychotomous choice models." *Econometrica: Journal of the Econometric Society,* 1315–1327.

McHorney, Colleen A, and Charles V Spain. 2011. "Frequency of and reasons for medication non-fulfillment and non-persistence among American adults with chronic disease in 2008." *Health Expectations* 14 (3): 307–320.

McKesson. 2024. "McKesson Pharmaceutical." Accessed October 29, 2024. https://www.mckesson.com/about-mckesson/businesses/mckesson-pharmaceutical/.

Milgrom, Paul R. 1981. "Good news and bad news: Representation theorems and applications." *The Bell Journal of Economics,* 380–391.

Milgrom, Paul R, and Robert J Weber. 1982. "A theory of auctions and competitive bidding." *Econometrica: Journal of the Econometric Society,* 1089–1122.

Minnesota Center for Infectious Disease Research & Policy, University of. 2018. "Resilient Drug Supply: Critical Acute Drugs." Accessed October 29, 2024. https://www.cidrap.umn.edu/resilient-drug-supply/resilient-drug-supply-critical-acute-drugs.

Morton, Fiona M Scott. 2000. "Barriers to entry, brand advertising, and generic entry in the US pharmaceutical industry." *International Journal of Industrial Organization* 18 (7): 1085–1104.

Mulcahy, A, C Whaley, M Tebeka, D Schwarn, N Edemfield, and A Becerra-Oenelas. 2021. "International prescription drug price comparisons." *Santa Monica: RAND Corporation.*

Nakabayashi, Jun, and Yohsuke Hirose. 2016. "Structural estimation of the scoring auction model." *RIETI Working Paper.*

National Community Pharmacists Association. 2022. "Survey: Three-Quarters of Community Pharmacies Report Staff Shortages." Accessed October 29, 2024. https://ncpa.org/newsroom/news-releases/2022/08/11/survey-three-quarters-community-pharmacies-report-staff-shortages.

Olden, Andreas, and Jarle Møen. 2022. "The triple difference estimator." *The Econometrics Journal* 25 (3): 531–553.

Peltzman, Sam. 1973. "An evaluation of consumer protection legislation: the 1962 drug amendments." *Journal of Political Economy* 81 (5): 1049–1091.

Riley, John G, and William F Samuelson. 1981. "Optimal auctions." *The American Economic Review* 71 (3): 381–392.

Robles-Garcia, Claudia. 2019. "Competition and incentives in mortgage markets: The role of brokers." *Mimeo.*

Rome, Benjamin N, Ameet Sarpatwari, and Aaron S Kesselheim. 2022. "State laws and generic substitution in the year after new generic competition." *Value in Health* 25 (10): 1736–1742.

Rosania, Larry. 2010. "Heparin crisis 2008: a tipping point for increased FDA enforcement in the pharma sector." *Food & Drug Law Journal* 65:489.

Rust, John. 1987. "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher." *Econometrica: Journal of the Econometric Society,* 999–1033.

Salin, Victoria, and Neal H Hooker. 2001. "Stock market reaction to food recalls." *Applied Economic Perspectives and Policy* 23 (1): 33–46.

Sant'Anna, Marcelo Castello Branco. 2017. "Empirical analysis of scoring auctions for oil and gas leases." *Mimeo.*

Senate Legislative Counsel. 2024. "Drug Shortage Prevention and Mitigation Act (Draft Copy)," https://www.finance.senate.gov/imo/media/doc/050124_sfc_drug_shortages_discussion_draft_legislative_text.pdf.

Seo, Soobin, and SooCheong Shawn Jang. 2021. "A negative or positive signal? The impact of food recalls on negative word-of-mouth (N-WOM)." *Journal of Hospitality and Tourism Management* 47:150–158.

Sood, Neeraj, Tiffany Shih, Karen Van Nuys, and Dana Goldman. 2017. "The flow of money through the pharmaceutical distribution system." *Health Affairs Blog.*

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

Starc, Amanda, and Thomas G Wollmann. 2022. "Does entry remedy collusion? Evidence from the generic prescription drug cartel." *National Bureau of Economic Research Working Paper 29886.*

Stomberg, Christopher. 2016. "Drug shortages, pricing, and regulatory activity." In *Measuring and modeling health care costs,* 323–348. University of Chicago Press.

Stuart, Bruce C, Mingliang Dai, Jing Xu, Feng-Hua E Loh, and Julia S Dougherty. 2015. "Does good medication adherence really save payers money?" *Medical Care* 53 (6): 517–523.

Takahashi, Hidenori. 2018. "Strategic design under uncertain evaluations: structural analysis of design-build auctions." *The RAND Journal of Economics* 49 (3): 594–618.

The White House. 2021. "Building Resilient Supply Chains, Revitalization American Manufacturing, and Fostering Broad-Based Growth: 100-Day Reviews under Executive Order 14017." Accessed October 29, 2024. https://www.whitehouse.gov/wp-content/uploads/2021/06/100-day-supply-chain-review-report.pdf.

———. 2024. "Biden Cancer Moonshot Announces New Pilot to Mitigate Pediatric Cancer Drug Shortages." Accessed October 29, 2024. https://www.whitehouse.gov/ostp/news-updates/2024/10/28/biden-cancer-moonshot-announces-new-pilot-to-mitigate-pediatric-cancer-drug-shortages/.

Toledo, Chantal, and Sofia Berto Villas-Boas. 2019. "Safe or not? Consumer responses to recalls with traceability." *Applied Economic Perspectives and Policy* 41 (3): 519–541.

Tracelink. 2023. "Pharmacies and Recalls: Understanding the Impact-and Changes to Come." Accessed October 29, 2024. https://www.tracelink.com/resources/resource-center/pharmacies-and-recalls-understanding-the-impact-and-changes-to-come.

United States Government Accountability Office. 2011a. "Drug Shortages: Public Health Threat Continues, Despite Efforts to Help Ensure Product Availability." Accessed October 29, 2024. https://www.gao.gov/assets/gao-14-194.pdf.

United States Government Accountability Office. 2011b. "Medical Devices: FDA Should Enhance Its Oversight of Recalls." Accessed October 29, 2024. https://www.gao.gov/products/gao-11-468.

Vatter, Benjamin. 2022. "Quality disclosure and regulation: Scoring design in medicare advantage." *Available at SSRN 4250361*.

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.

Woodcock, Janet, and Marta Wosinska. 2013. "Economic and technological drivers of generic sterile injectable drug shortages." *Clinical Pharmacology & Therapeutics* 93 (2): 170–176.

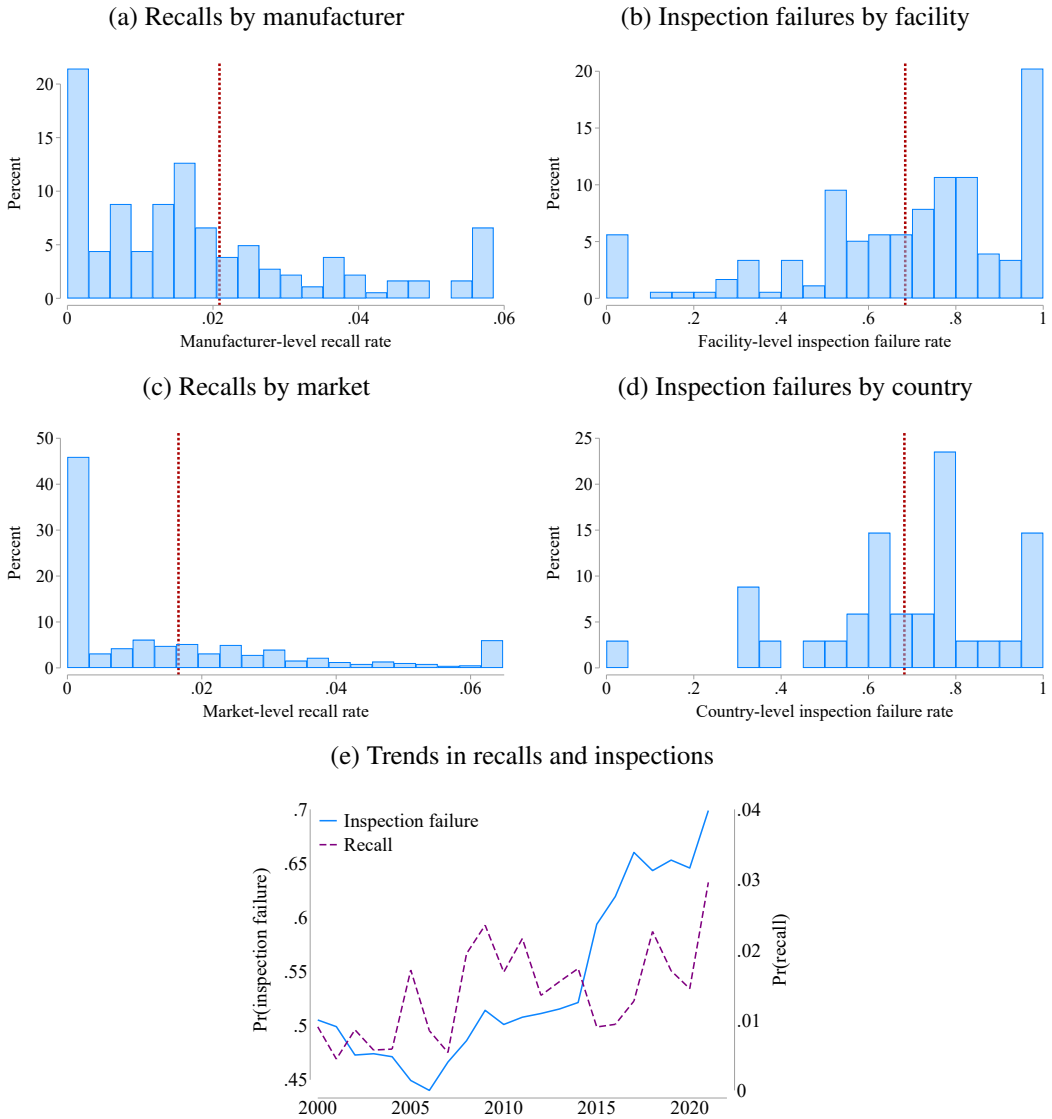
Wosińska, Marta E, and Richard G Frank. 2023. "Federal policies to address persistent generic drug shortages." *Brookings Institution: The Hamilton Project.*

Yeung, Kai, Anirban Basu, Ryan N Hansen, and Sean D Sullivan. 2018. "Price elasticities of pharmaceuticals in a value based-formulary setting." *Health Economics* 27 (11): 1788–1804.

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.

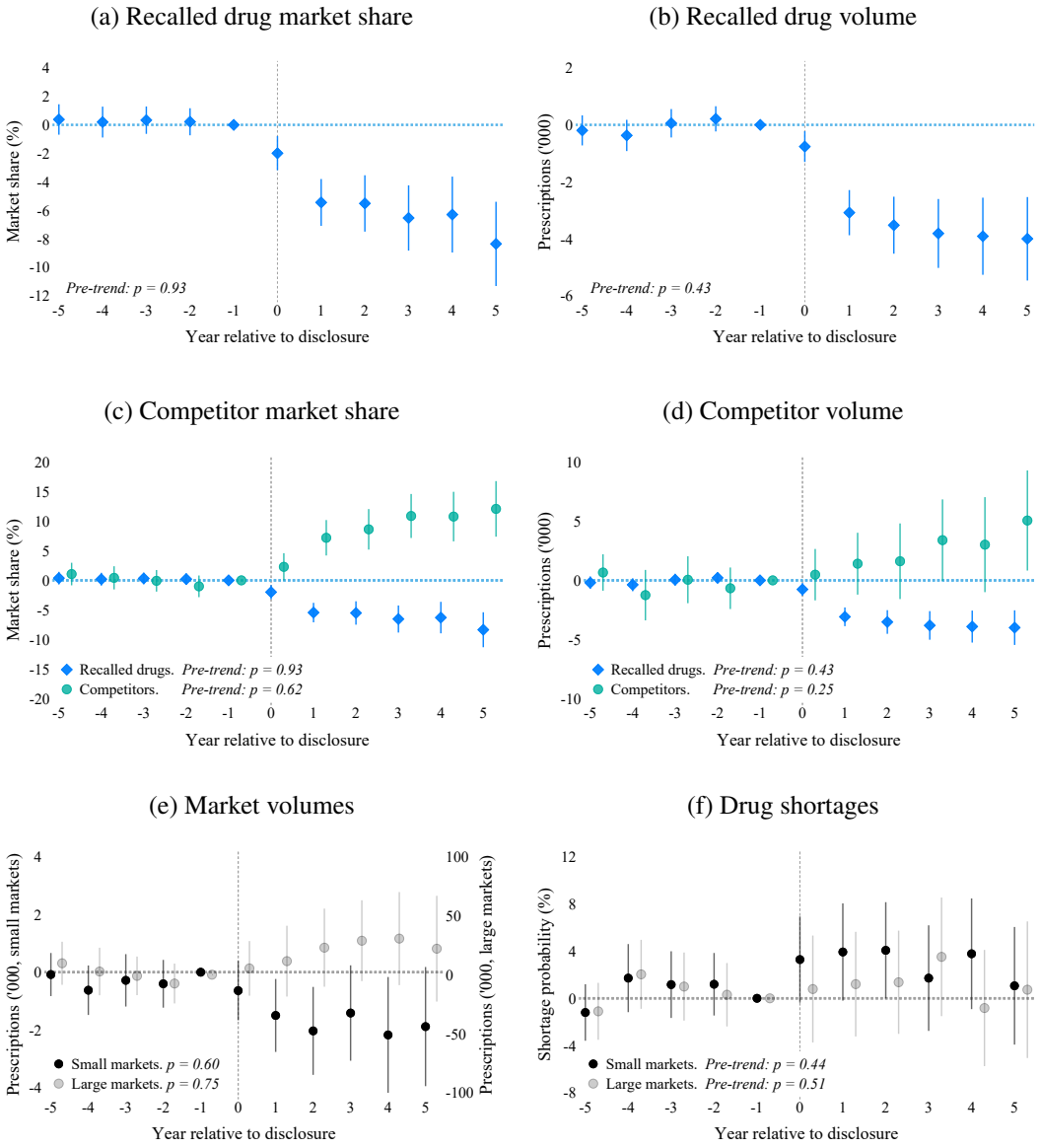
# 8 Figures and Tables

Figure 1: Variation and trends in manufacturing quality




Note: Panels (a)–(c) present distributions of recall and inspection failure rates in the FDA and Optum claims data over 2000-2022. The rates are averaged over drug-years at the manufacturer (panel (a)), facility (b), market (c) and country (d) levels. Panels (a) and (c) show average recall rates, and panels (b) and (d) show inspection failure rates. Each distribution is winsorized at the 95th percentile, and the red dashed lines display the mean rates. Panel (e) presents drug-level recall rates and facility-level inspection failure rates per year in my data. Recalls are observed in FDA Enforcement Reports, while inspections are from internal FDA data obtained through FOIA requests. Failed inspections are those graded "official" or "voluntary action indicated." Corresponding figures using only "official action indicated" inspection grades, as well as drug shortages, are presented in Figure A.3. Each dataset is described in Section 3.

Figure 2: Effects of recall announcements on intermediary purchases

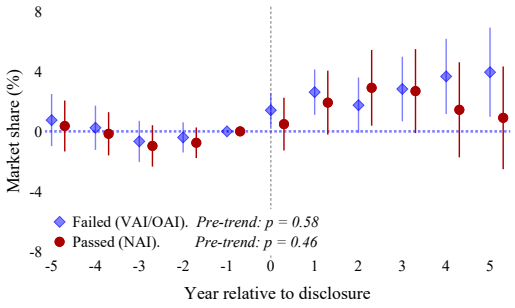



Note: The figure displays the estimated dynamic effects of recall disclosures on intermediary drug purchases. The coefficients were obtained by estimating Equation (1) on drug-year level data constructed from Optum claims. The year prior to the first disclosure event is the omitted reference year. Panels (a) and (b) display the event study results for recalled drugs with drug market share and prescription volumes as the dependent variables, respectively. Drug market share is defined as the fraction of U.S. retail pharmacies supplied by a drug manufacturer, as observed in the Optum data. Panels (c)–(d) present the corresponding results for bioequivalent competitors of the recalled drug (whose qualities were not disclosed), estimated using market-level data. The competitor coefficients are plotted in green dots and overlaid on top of the recalled drug ones, which are in blue diamonds. Panels (e)–(f) plot the total market-level effects of recalls on drug volumes by baseline market structure, where markets are classified as small (N $\leq$10 drugs; black dots) or large (N > 10; gray circles). The coefficients in panel (e) are estimated using drug volumes from the Optum data, while those in panel (f) use an indicator for new national drug shortages in the 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 2 presents the corresponding DD coefficients and mean values.

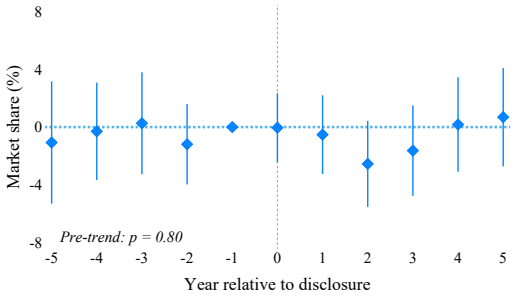
Figure 3: Effects of inspection reports on intermediary purchases

All grades (NAI vs. VAI, OAI)

(a) Inspected drug market share (DD)

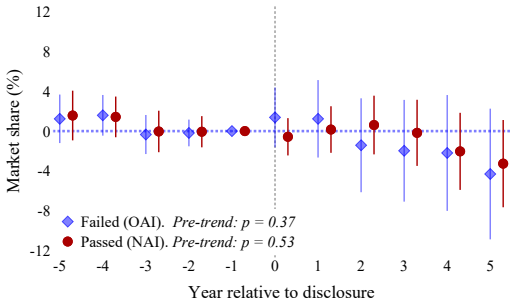



(b) Failed drug market share (DDD)

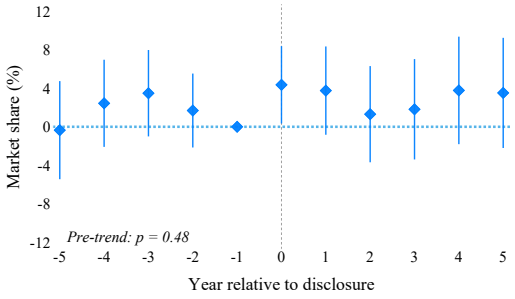



Extreme grades (NAI vs. OAI)

(c) Inspected drug market share (DD)

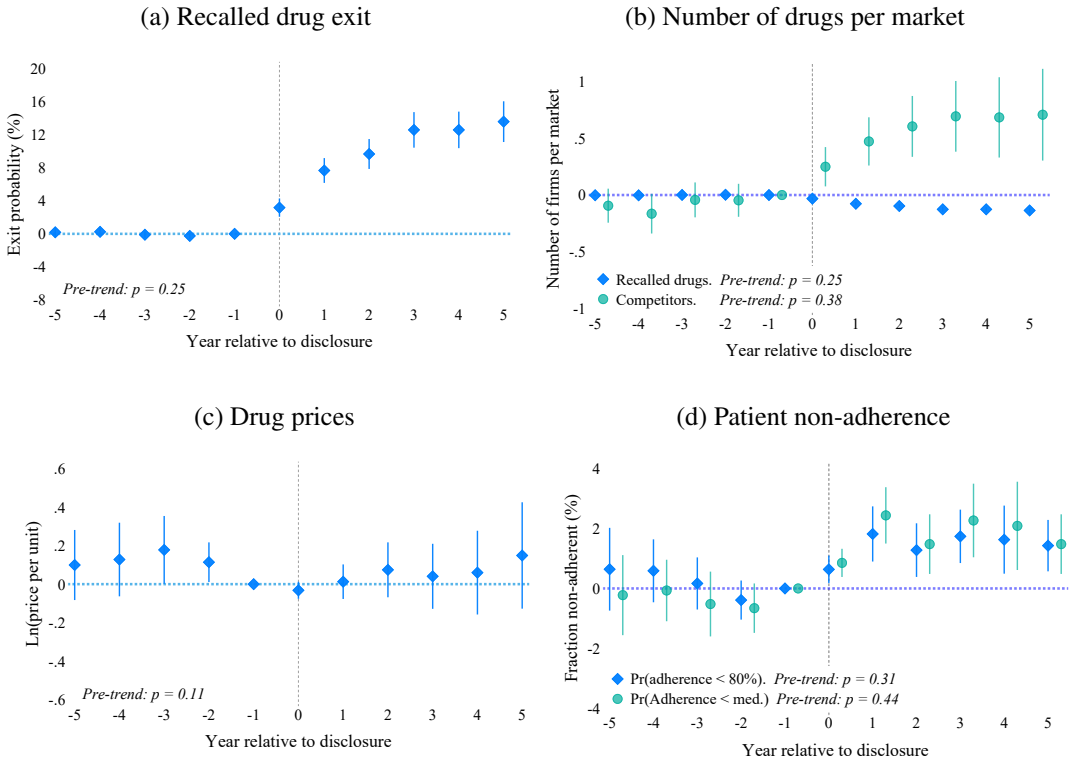



(d) Failed drug market share (DDD)



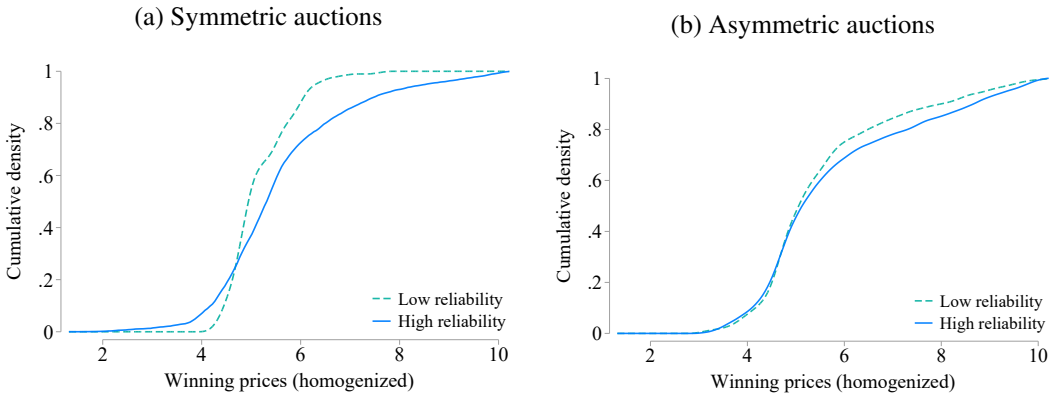

Note: The figure displays the estimated dynamic effects of inspection disclosures on intermediary drug purchases. The coefficients were obtained using drug-year level data constructed from Optum claims. The year prior to the first disclosure event is the omitted reference year. Panels (a) and (c) present the effects of passing (red dots) and failing (purple diamonds) inspection reports, estimated using Equation (1) with drug market share as the dependent variable. Drug market share is defined as the fraction of U.S. retail pharmacies supplied by a drug manufacturer, as observed in the Optum data. Panels (b) and (c) present the dynamic coefficients from a DDD specification, estimated using Equation (1) with an additional term, $(\gamma\, D_{dt} \times 1\{\text{failed}_d\})$. Passing inspections are defined as those graded "No Action Indicated" (NAI) by the FDA. In panels (a)–(b), failing inspections are those graded "Voluntary" (VAI) or "Official Action Indicated" (OAI), while in panels (c)–(d), the definition is restricted to OAI grades only. In each plot, the control group is not-yet-inspected drugs. 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 3 presents the corresponding DD coefficients and mean values.

Figure 4: Effects of recalls on market structure and patient welfare



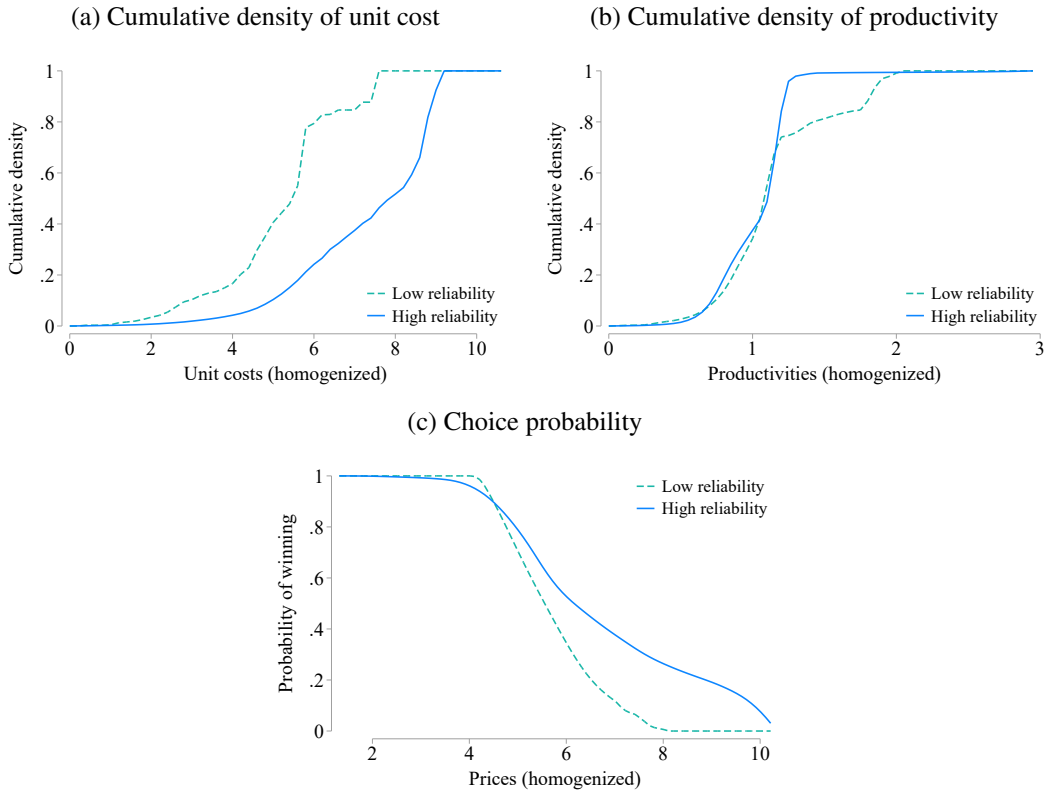

Note: The figure displays the estimated dynamic effects of recalls on market structure, prices and patient drug adherence. The coefficients were obtained by estimating Equation (1) on drug or market-level data constructed from Optum claims. The year prior to the first disclosure event is the omitted reference year. Panel (a) displays the effects on the recalled drug's probability of exit. 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) shows the change in the total count of bioequivalent competitors of the recalled drug, estimated using market-level data. The competitor coefficients are plotted in green dots and overlaid on top of the recalled drug ones, which are in blue diamonds. Panel (c) presents the event study results for log average wholesale prices, weighted by baseline volume. The prices are market-level averages of pharmacy average acquisition costs per unit from state and national surveys (described in Section 3.1) and inflated to 2022 dollar terms. Panel (d) shows the dynamic effects on patient non-adherence. Non-adherence is measured using an indicator variable for patient medication possession ratios ("MPRs") below 80% (blue diamonds) or the drug-specific median (green dots). 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 4 presents the corresponding DD coefficients and mean values.

Figure 5: Cumulative densities of winning prices by auction type



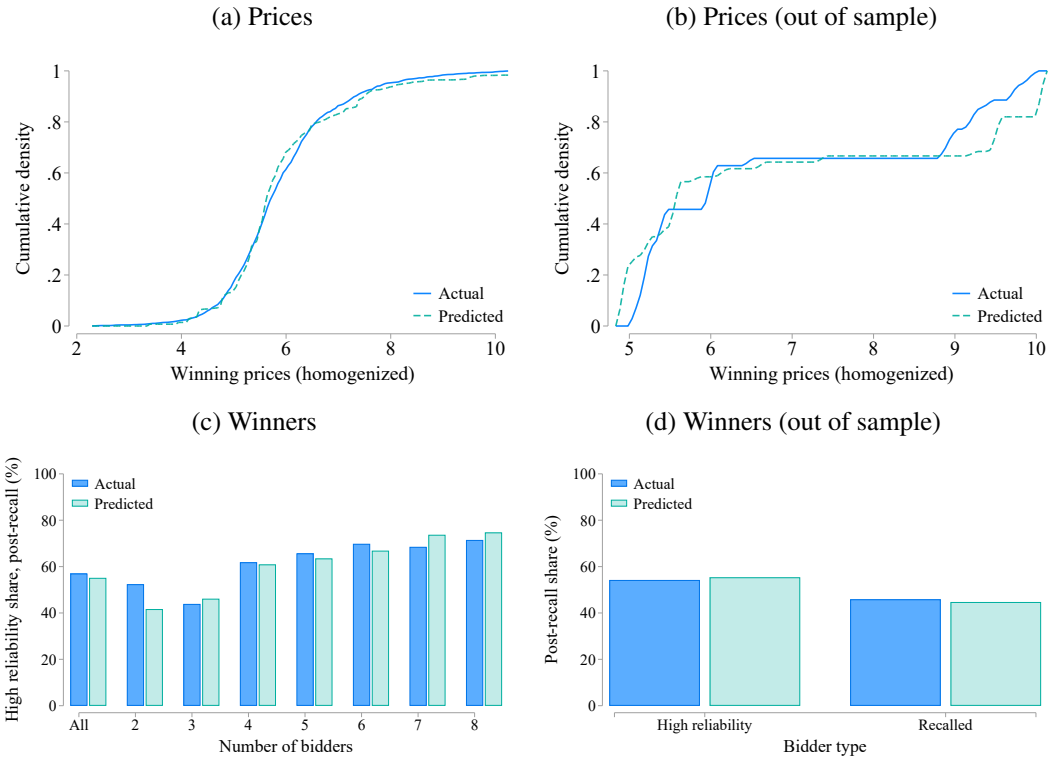

Note: The figures plot the kernel densities of homogenized winning prices from procurement auctions of unbranded, oral generic drugs in the Optum data with $N = 2$ bidders. Panel (a) plots the densities for symmetric auctions—with either high (solid lines) or low reliability (dashedl lines) bidders—while panel (b) plots the corresponding ones for asymmetric auctions. Prices represent intermediary acquisition costs per unit, and are homogenized using the methodology described in Section 5.2. High reliability drugs have are those that have not experienced a recall within the previous five years, as observed in FDA Enforcement Reports.

Figure 6: Scoring auction model estimates




Note: The figures show the key estimated objects from the structural scoring auction model, assuming buyers use an optimal scoring rule. Panel (a) presents the kernel density of the estimated cost per unit of production, while panel (b) presents the density of estimated productivity shocks $\theta_i$. Panel (c) plots the estimated choice probabilities for each bidder type as a function of price. Each panel shows estimates for high- and low-reliability manufacturers separately (in solid and dashed lines, respectively). The model and assumptions are described in Section 5.

Figure 7: Scoring auction model fit: predicted post-recall outcomes




Note: The figure presents the structural model's predictions of market outcomes after recalls, compared against the observed data. The model predictions use only *ex ante*, pre-recall data to predict post-recall outcomes. Throughout, model predictions are plotted in green, while observed outcomes are in blue. Panel (a) shows the kernel density of winning prices of recalled and competitor drugs after recalls. Panel (b) shows the high-reliability drug share across auctions of different sizes. Panels (c) and (d) present the same outcomes but use only out-of-sample data from 2022 (that was not used in estimating the model). Out-of-sample fit was evaluated once. Since the test sample is smaller, panel (d) presents high-reliability and recalled shares pooled across all auction sizes. Model fit is further described in Section 5.3.

Table 1: Descriptive statistics

| | (1) All drugs | (2) Recalled drugs | (3) Never recalled drugs |
|---|---|---|---|
| *Drug characteristics* | | | |
| Years since first generic entry | 18.26 | 18.16 | 18.90 |
| Fraction unit type: | | | |
| Tablet | 0.49 | 0.49 | 0.50 |
| Capsule | 0.15 | 0.10 | 0.15 |
| Oral liquid (ml) | 0.07 | 0.11 | 0.07 |
| Topical (gm) | 0.21 | 0.17 | 0.21 |
| Injection (ml) | 0.07 | 0.13 | 0.07 |
| Extended release | 0.10 | 0.10 | 0.10 |
| *Volume* | | | |
| Mean annual prescriptions | 17,211 | 56,651 | 14,765 |
| | (66,340) | (168,526) | (54,001) |
| Market share of pharmacies | 0.101 | 0.118 | 0.098 |
| | (0.180) | (0.154) | (0.181) |
| *Market structure & prices* | | | |
| Mean drugs per market | 10.73 | 12.44 | 10.66 |
| | (6.77) | (8.16) | (6.69) |
| Mean price per unit | 4.55 | 3.82 | 4.68 |
| Median price per unit | 0.70 | 0.41 | 0.76 |
| | (20.67) | (15.72) | (21.43) |
| N drugs | 10,199 | 1,159 | 9,040 |

Notes: The table presents descriptive statistics for generic drugs in the commercial Optum claims data. It presents all drugs in the sample in column (1), recalled drugs in column (2) and drugs that were not recalled during 2000-22 in column (3). Each value represents the drug-level mean over all years(cols. (1) and (3)) or prior to the first recall (col. (2)). Corresponding standard deviations are reported in parentheses for the key volume, market structure and price outcome variables. Drug characteristics are from the NDC Directory and Orange Book, and years since generic entry is calculated as of 2022. Annual volumes and counts of drugs per market are observed in the claims data. Prices are pharmacy average acquisition costs obtained from state and national surveys and inflated to 2022 dollar values. The datasets are described in Section 3.1. Corresponding descriptive statistics for inspected drugs are presented in Table A.2.

Table 2: Effects of recall announcements on intermediary purchases

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Recalled drug | | Competitor spillovers | | Net market effect | |
| Dependent var. | Market share | Ln(scripts) | Market share | Ln(scripts) | Ln(scripts) | Pr(shortage) |
| *A - Average effects* | | | | | | |
| Recalled * post | -0.044*** | -0.917*** | 0.068*** | 0.115 | -0.037 | 0.021** |
| | (0.007) | (0.082) | (0.014) | (0.083) | (0.060) | (0.010) |
| Recalled * post (%) | -43% | -60% | 9% | 12% | -4% | 33% |
| Recalled * post (level) | – | -2,711 | – | 1,927 | -1,065 | – |
| *B - Heterogeneity by market structure* | | | | | | |
| Low participation: Recalled * post | -0.068*** | -1.174*** | 0.087*** | 0.050 | -0.141** | 0.025** |
| | (0.014) | (0.110) | (0.018) | (0.104) | (0.071) | (0.012) |
| High participation: Recalled * post | -0.018*** | -0.602*** | 0.015 | 0.153 | 0.075 | 0.006 |
| | (0.005) | (0.117) | (0.014) | (0.097) | (0.069) | (0.013) |
| Observations | 12,388 | 12,388 | 6,419 | 6,419 | 6,530 | 6,530 |
| Dependent var. median | 0.103 | 4,516 | 0.738 | 15,810 | 29,333 | 0.063 |

Notes: The table reports the estimated average effects of recall disclosures on intermediary drug purchases. It presents the coefficients obtained by estimating Equation (2) on data constructed from Optum claims. The dependent variables are drug market share (cols. (1) and (3)), logged prescriptions (cols. (2), (4) and (5)) and an indicator for a new national drug shortage (col. (6)). Drug market share is defined as the fraction of U.S. retail pharmacies supplied by a drug manufacturer, as observed in the Optum data. Cols. (1)–(2) present the effects for recalled drugs estimated using drug-year level data. Cols. (3)–(4) present the total spillover effects for bioequivalent competitors of the recalled drug (whose qualities were not disclosed), estimated using market-level data. Col. (5) reports the net effect for recalled and competitors drugs at the market-level, while col. (6) reports market-level shortage probabilities. Panel A presents average effects for all drugs in the analysis sample. To facilitate comparison, the effects are also presented as the percentage of the dependent variable median ("%") and the implied volume level ("level"). Panel B presents results by baseline market structure, where markets are classified as low ($N \leq 10$ drugs) or high ($N > 10$) participation. Standard errors are clustered at the drug- or market-level and reported in parentheses. The dependent variable median is computed for treated drugs in the year prior to the recall; logged medians are exponentiated for clarity. The corresponding event studies are presented in Figure 2. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Effects of inspection reports on intermediary purchases

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Difference-in-difference estimates | | | | Triple-difference estimates | |
| | Market share | | Ln(scripts) | | Market share | Ln(scripts) |
| Dependent var. | Passed | Failed | Passed | Failed | | |
| *A - VAI & OAI effects* | | | | | | |
| Inspection * post | 0.021** | 0.020*** | 0.256** | 0.288*** | 0.027** | 0.502*** |
| | (0.010) | (0.007) | (0.119) | (0.083) | (0.011) | (0.122) |
| Inspection * post * failed | | | | | -0.005 | -0.091 |
| | | | | | (0.012) | (0.134) |
| Observations | 8,051 | 11,131 | 8,051 | 11,131 | 13,425 | 13,425 |
| Dependent var. mean | 0.168 | 0.168 | 1,295 | 1,295 | 0.168 | 1,295 |
| *B - OAI effects* | | | | | | |
| Inspection * post | -0.001 | 0.005 | 0.279** | -0.273* | 0.003 | 0.307** |
| | (0.011) | (0.018) | (0.126) | (0.158) | (0.012) | (0.129) |
| Inspection * post * failed | | | | | 0.024 | 0.001 |
| | | | | | (0.020) | (0.187) |
| Observations | 4,900 | 3,604 | 4,900 | 3,604 | 5,898 | 5,898 |
| Dependent var. mean | 0.177 | 0.177 | 7.186 | 7.186 | 0.177 | 7.186 |

Notes: The table reports the estimated average effects of inspection disclosures on intermediary drug purchases. Cols. (1)–(4) present the DD coefficients obtained by estimating Equation (2) on drug-year level data. Cols. (1) and (3) present the estimates for passing inspections, while cols. (2) and (4) present those for failing ones. Cols. (5)–(6) present the DDD coefficients that quantify the differences between the passing and failing estimates, obtained by estimating Equation (2) (with an additional term $\gamma\, D_{dt} \times 1\{\text{failed}_d\}$). The dependent variables are drug market shares (cols. (1), (2) and (5)) and logged prescriptions (cols. (4), (5) and (6). Drug market share is defined as the fraction of U.S. retail pharmacies supplied by a drug manufacturer, as observed in the Optum data. Throughout, passing inspections are defined as those graded "No Action Indicated" (NAI) by the FDA. In panel A, failed inspections are those graded "Voluntary" (VAI) or "Official Action Indicated" (OAI), while in panel B, the definition is restricted to OAI grades. The control group is not-yet-inspected drugs. Standard errors are clustered at the drug-level and reported in parentheses. The dependent variable mean is computed for treated drugs in the year prior to inspection; logged values are exponentiated for clarity. Corresponding event studies are presented in Figure 3, panel (f). *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Effects of recalls on market structure and patient welfare

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Market structure | | | Patient welfare | | |
| Dependent var. | Recalled drug Pr(exit) | Competitors ln(N drugs) | Net market ln(N drugs) | Ln(price) | Pr(adherence < 80%) | Pr(adherence < drug med.) |
| Recalled * post | 0.067*** | 0.125*** | 0.094*** | -0.111* | 0.021*** | 0.033*** |
| | (0.005) | (0.029) | (0.025) | (0.065) | (0.006) | (0.006) |
| Observations | 11,673 | 6,419 | 6,530 | 7,984 | 20,005 | 20,005 |
| Dependent var. mean | 0.043 | 4.027 | 5.743 | 0.086 | 0.408 | 0.492 |
| Recalled * post (%) | 155% | 13% | 10% | -11% | 5% | 7% |

Notes: The table reports the estimated average effects of recalls on market structure, prices and patient drug adherence. It presents the coefficients obtained by estimating Equation (2) on drug (col. (1)) or market-level data (cols. (2)–(6)). The first three columns show the estimated effects on market structure, where the dependent variables are an indicator for recalled drug exit (col. (1)), and log counts of the number of bioequivalent competitors (col. (2)) and total market participants (col. (3)). The last three columns present outcomes relevant to patient welfare, where the dependent variables are log prices (col. (4)), weighted using baseline volumes), and the fraction of patients with medication adherence (MPR) less than 80% (col. (5)) or the drug median (col. (6)). Exit and the number of firms are observed in the Optum claims data. Log prices are pharmacy average acquisition costs per unit from state and national surveys (described in Section 3.1) and are inflated to 2022 dollar terms. Standard errors are clustered at the drug (cols. 1–2) or market (cols. 3–5) levels and reported in parentheses. The dependent variable mean is computed for treated drugs in the year prior to the recall; logged means are exponentiated for clarity. The corresponding event studies are presented in Figure 4. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Counterfactual policies: price-only and Pay-for-Performance scoring

| | Manufacturer type | (1) Price-only scoring | (2) Baseline scoring | (3) P4P scoring | |
|---|---|---|---|---|---|
| Reliability premium | | 0% | 13% | 26% | |
| Average offered price | All | 1.66 | 1.66 | 2.19 | |
| | Low | 1.12 | 1.12 | 1.25 | |
| | High | 1.85 | 1.85 | 2.53 | |
| Average winning price | All | 0.34 | 0.35 | 0.45 | |
| | Low | 0.35 | 0.35 | 0.42 | |
| | High | 0.34 | 0.35 | 0.46 | |
| *Welfare-relevant effects* | | | | | |
| High reliability share | | 0.51 | 0.64 | 0.80 | |
| Average winning price | | 0.34 | 0.34 | 0.45 | |
| Manufacturer profits | All | 672 | 675 | 846 | |
| | Low | 227 | 143 | 75 | |
| | High | 315 | 402 | 642 | |
| | | | | *20% penalties* | *100% incentives* |
| Program cost | | | | 166 | 207 |
| Intermediary surplus | | 125 | 126 | 239 | 274 |
| Total producer surplus, less program cost | | 797 | 802 | 920 | 913 |

Notes: The table presents results from counterfactual simulations using the estimated structural model. Each simulation changes intermediary preferences for reliability, by varying the parameter $\Delta S$ between 0% (col. (1) and 26% (col. (3)) of average baseline prices. A no-preferences ("price-only") environment is reported in col. (1), the status quo ("baseline) environment in col. (2) and a pay-for-performance ("P4P") policy environment in col. (3). Each counterfactual is described in Section 5. Manufacturer type refers to reliability type. All monetary outcomes are weighted by the realized contract volume and deflated to 2022 dollar terms. Profits and surplus represent average annual values and are presented in millions of dollars.

Table 6: Decomposition of recall effects

| | | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|---|
| | | Total equilibrium effects | | | Decomposition: %Δ due to | | |
| | Segment | Pre-recall | Post-recall | %Δ | Buyer preferences | Bidder costs | Price responses |
| Offered price | All | 2.01 | 1.68 | -16% | – | – | -16% |
| | Recalled | 1.17 | 1.01 | -14% | – | – | -14% |
| | Competitors | 2.31 | 1.87 | -19% | – | – | -19% |
| Unit costs | | 0.17 | 0.18 | 6% | 24% | -6% | -12% |
| *Welfare-relevant effects* | | | | | | | |
| High-quality share | | 0.57 | 0.60 | 6% | 18% | – | -11% |
| Winning price | | 0.37 | 0.33 | -11% | 22% | – | -32% |
| Manufacturer profits | All | 184 | 132 | -28% | 7% | 6% | -42% |
| | Recalled* | 23 | 20* | -12%* | -20% | 47% | 8% |
| | Competitors | 162 | 101 | -37% | 11% | 0% | -49% |
| Intermediary surplus | | 23 | 46 | 100% | 5% | 0% | 95% |
| Total producer surplus | | 208 | 178 | -14% | 7% | 5% | -26% |

Notes: The table presents results from counterfactual simulations of recall disclosures, using the estimated structural model. Cols. (1) and (2) present simulated pre- and post-recall outcomes, while col. (3) reports the percentage changes between them. Cols. (4)–(6) decompose the relative contributions of three mechanisms behind the post-recall changes: changes in intermediary preferences (4), in manufacturer variable costs (5) and in manufacturers' strategic pricing responses (6). Each contribution is presented as a percentage change from the pre-recall baseline, and is described further in Section 5. All values (except the high-quality drug share) are in 2022 dollar terms and weighted by the realized contract volumes. Profits and surplus are expressed as average annual values in millions of dollars. *Post-recall manufacturer profits deduct a one-time recall reimbursement cost, assumed to be 10% of the pre-recall contract value, which is excluded from the last three columns.

# A Appendix

Figure A.1: Intermediary scoring criteria

**Supply Chain Excellence Award scoring**

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

Note: The figure presents as 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 supply disruptions and recalls.

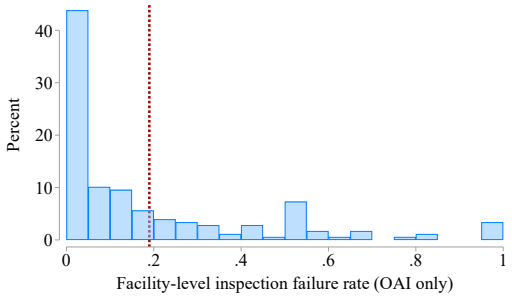
Figure A.2: The structure of the U.S. generic pharmaceutical market



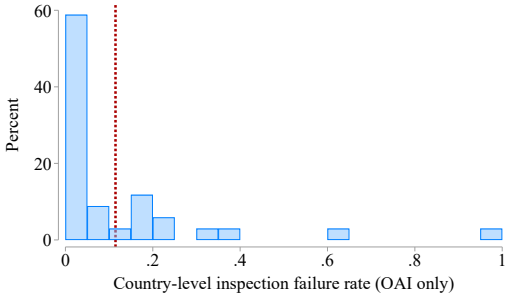

Note: The figure presents the basic structure of the U.S. generic pharmaceutical market over 2010-2022. At the top, it shows competitive generic manufacturers, which sell drugs to monopsonistic intermediaries (purchasing alliances between wholesalers and retail pharmacy chains). Intermediaries, in turn, sell generics to prescribed patients.

Figure A.3: Variation and trends in manufacturing quality: OAI inspections and shortages

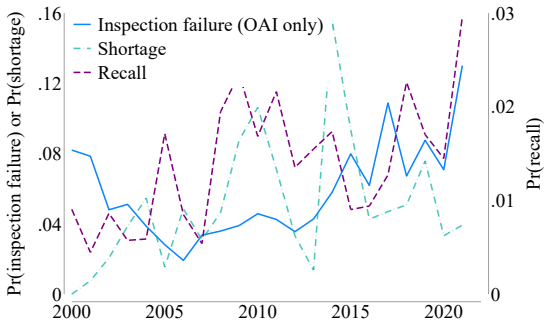
(a) OAI inspection failures by facility




(b) OAI inspection failures by country

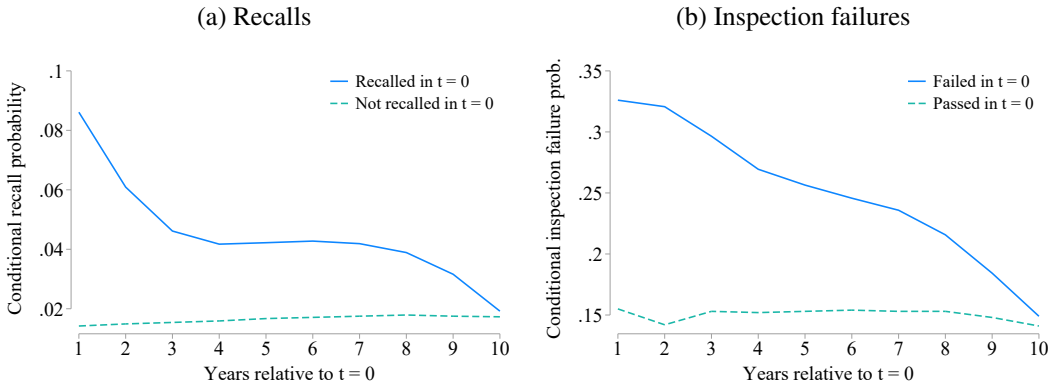



(c) Trends in recalls, OAI inspections and shortages




Note: Panels (a)–(b) present distributions of inspection failure rates in the FDA and Optum claims data over 2000-2022. The rates are averaged over drug-years at the facility (panel (a)) and country (b) levels. The red dashed lines display the mean rates. Panel (c) presents drug-level recall rates, facility-level inspection failure rates, and market-level shortage rates per year in my data. Recalls are observed in FDA Enforcement Reports, while inspections are from internal FDA data obtained through FOIA requests. Failed inspections are those graded "official" or "voluntary action indicated." Corresponding figures using "voluntary" and "official action indicated" inspection grades as are presented in Figure 1. Shortages are from UUDIS data. Each dataset is described in Section 3.

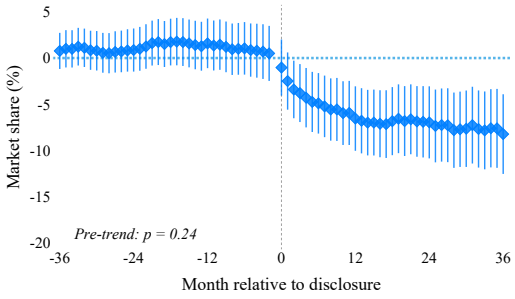
Figure A.4: Persistence in drug manufacturing quality



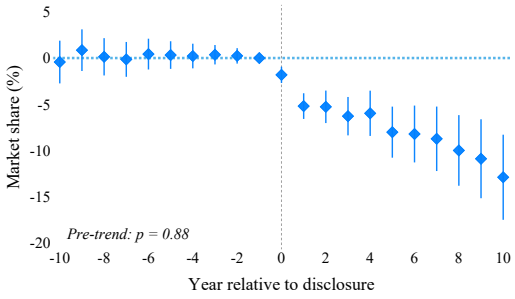

Note: The figure plots persistence in manufacturing quality failures within drugs over time, as observed in the FDA and Optum claims data. It shows the probability of a failing signal in each year for drugs with a previously disclosed failure in year 0 (solid lines), or without such a disclosure (dashed lines). Panel (a) presents the probabilities for recall announcements, and panel (b) the probabilities for inspection failures. 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." Each dataset is described in Section 3.

Figure A.5: Quasi-experimental effects of recall announcements: Timing and shortages

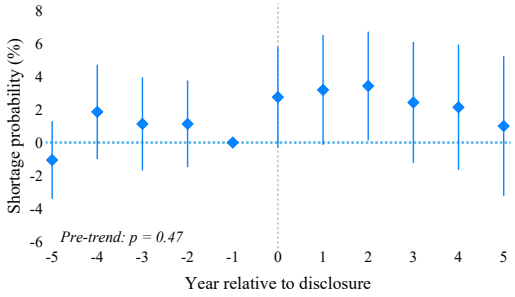
(a) Monthly effects on intermediary purchases




(b) Long-run effects on intermediary purchases

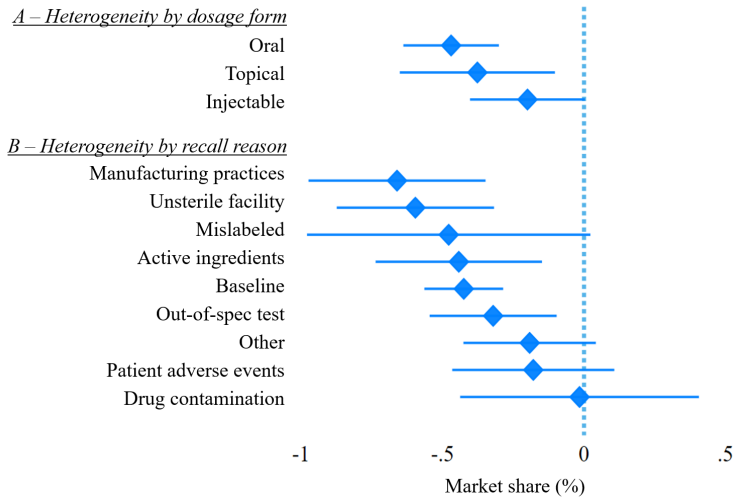



(c) Effects on drug shortages (all markets)



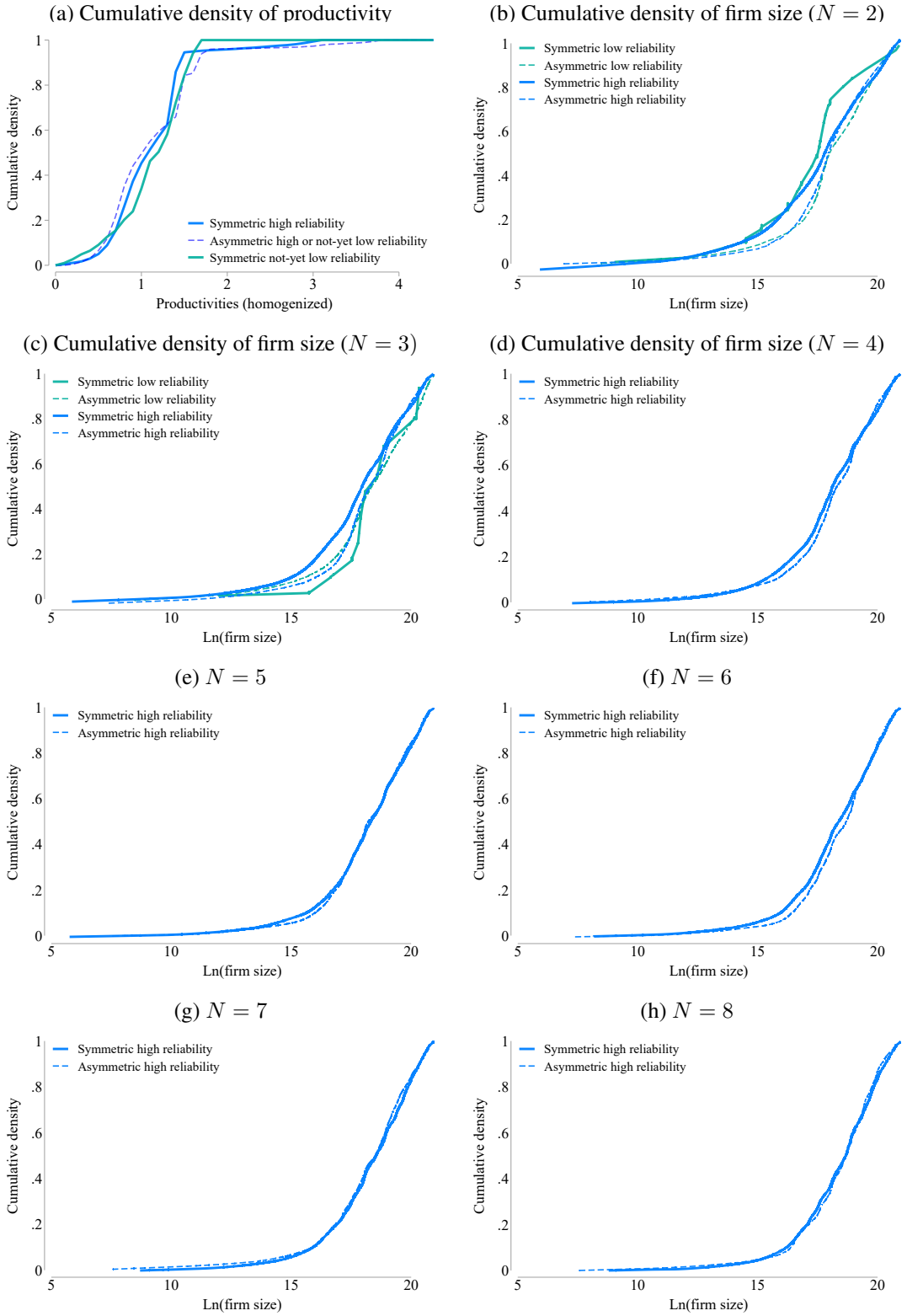

Note: The figure displays the estimated dynamic effects of recall disclosures on intermediary drug purchases and shortages. The coefficients were obtained by estimating Equation (1) on drug-level data constructed from Optum claims. The year or month prior to the first disclosure event is the omitted reference period. Panels (a)–(b) present the dynamic effects from Figure 2, panel (a) in additional detail. Both present the effects for recalled drugs with the market share of pharmacies as the dependent variable. Panel (a) zooms in on the three years surrounding the recall using drug-month level data. Panel (b) presents the long-run effects of recalls, extending the event study to include ten years before and after the recall. Finally, panel (c) shows the estimated effects of recalls on drug shortages, measured using the 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.

Figure A.6: Quasi-experimental effects of recall announcements on intermediary purchases: Heterogeneity by formulation and recall reason



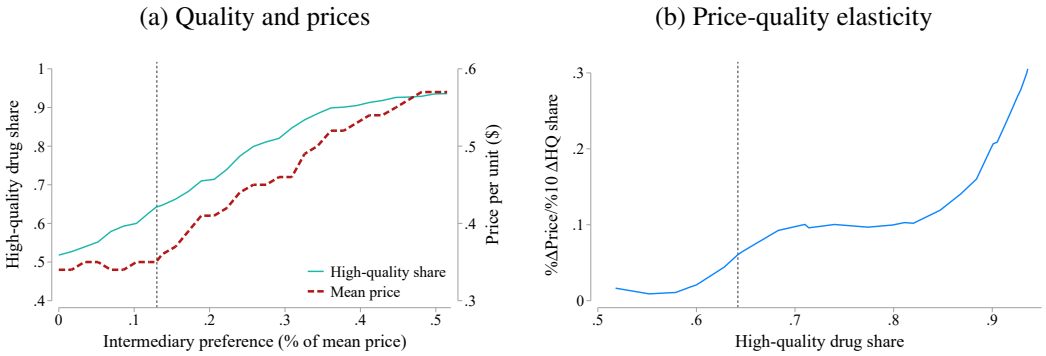

Note: The figure displays the estimated average effects of recall disclosures on intermediary drug purchases. It presents the coefficients obtained by estimating Equation (2) using drug-year level data with the drug market share as the dependent variable. Drug market share is defined as the fraction of U.S. retail pharmacies supplied by a drug manufacturer, as observed in the Optum data. Panel A presents the results by drug formulation, and panel B by manufacturer-reported recall reason. In both panels, the "Baseline" values represent the average effects for all drugs. To facilitate comparison, all values are presented as percentages of the dependent variable mean for treated drugs in each sub-sample in the year prior to the recall. The figure presents 95% confidence intervals with standard errors clustered by drug.

Figure A.7: Scoring auction model:
Balance in productivity and firm size across symmetric & asymmetric auctions




Note: The figure plots the kernel densities of the drug-level estimated productivity shocks $\theta_i$ and firm size in auctions with symmetric (uniform bidder types, solid lines) and asymmetric (mixed types, dashed lines) auctions. Panel (a) presents the cumulative densities of productivities estimated from a subsample of auctions comprised of never recalled (high reliability) and not-yet-recalled (not-yet low reliability) drugs, using the model and algorithm described in Section 5. Reliability type is a binary indicator of whether the focal drug has been recalled in the last five years. Panels (b)–(h) present densities of firm size from the main estimation sample in increasing order of auction size ($N$ bidders). Firm size is measured using annual log volumes observed in Optum claims data.

Figure A.8: Counterfactual policies: quality and prices under alternative preferences




Note: The figures show equilibrium outcomes from counterfactual simulations using the estimated structural model. Each simulation changes intermediary preferences for reliability, by varying the parameter $\Delta S$ (plotted on the $x-$axis) between 0% and 50% of average baseline prices. Panel (a) plots average equilibrium prices (dashed line) and the high-quality drug (solid line) share. Panel (b) plots the same outcomes as a relative elasticity, with the high-quality drug share on the $x$-axis. In each plot, the gray dashed line denotes the status quo environment. All values are weighted by realized contract volumes, and prices are deflated to 2022 dollar terms.

Table A.1: 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 |

Notes: 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. The datasets are described in Section 3.1. 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.

Table A.2: Descriptive statistics: inspections

| | (1) All drugs | (2) Failing drugs | (3) Passing drugs |
|---|---|---|---|
| *Drug characteristics* | | | |
| Years since first generic entry | 19.41 | 19.44 | 19.33 |
| Fraction unit type: | | | |
| Tablet | 0.53 | 0.48 | 0.54 |
| Capsule | 0.14 | 0.14 | 0.11 |
| Oral liquid (ml) | 0.10 | 0.10 | 0.12 |
| Topical (gm) | 0.15 | 0.20 | 0.14 |
| Injection (ml) | 0.07 | 0.07 | 0.09 |
| Extended release | 0.11 | 0.08 | 0.09 |
| *Volume* | | | |
| Mean annual prescriptions | 28,668 | 21,336 | 27,609 |
| | (95,069) | (61,101) | (69,703) |
| Market share of pharmacies | 0.160 | 0.117 | 0.118 |
| | (0.216) | (0.175) | (0.162) |
| Number of drugs | 1,483 | 1,016 | 467 |

Notes: The table presents descriptive statistics for generic drugs in the Optum claims data that are matched to FDA inspections. It presents all inspected drugs in column 1, drugs that receive grades of Voluntary or Official Action Indicated ("failing") in column 2, and drugs are classified as No Action Indicated ("passing") in column 3. Each value represents the drug-level mean over all years prior to the first inspection. 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 market shares are observed in the claims data. The corresponding descriptive statistics for recalled drugs are presented in Table 1.

Table A.3: Quasi-experimental robustness

| | Recalled drug market share |
|---|---|
| 1. Baseline estimate | -0.044*** |
| | (0.007) |
| *Alternative dependent variable* | |
| 2. Share of scripts | -0.065*** |
| | (0.008) |
| 3. Share of first scripts | -0.055*** |
| | (0.007) |
| *Alternative specifications* | |
| 4. Volume weighted | -0.051*** |
| | (0.012) |
| 5. Inspection covariates | -0.045*** |
| | (0.012) |
| 6. Never-recalled control group | -0.020*** |
| | (0.005) |
| 7. Callaway Sant'Anna | -0.031*** |
| | (0.005) |
| 8. Callaway Sant'Anna, | -0.031*** |
| not-yet-treated control group | (0.005) |
| 9. Cluster at firm level | -0.044*** |
| | (0.009) |
| 10. Cluster at market-level | -0.044*** |
| | (0.007) |
| *Alternative samples* | |
| 11. Exclude exiting drugs | -0.025*** |
| | (0.008) |
| 12. Exclude growing markets | -0.029*** |
| | (0.010) |
| 13. Exclude COVID-19 | -0.047*** |
| | (0.008) |
| Observations | 12,388 |
| Dependent var. median | 0.103 |

Notes: The table reports the estimated average effects of recall disclosures on intermediary drug purchases. It presents the coefficients obtained by estimating Equation (2) with different dependent variables, specifications and samples. The first row reproduces the baseline estimate from Table 2, using drug market share as the dependent variable. Rows 2 and 3 vary the dependent variable, using the total share of prescriptions and first prescriptions, respectively. Rows 4 and 5 include pre-recall volume weights and inspection grade covariates, respectively. The sixth row changes the control group to never (as opposed to not-yet) treated drugs. Rows 7–8 include corrections for staggered treatment timing from Callaway and Sant'Anna 2021, while rows 9–10 cluster standard errors at more aggregated levels than the treated unit (the drug). Rows 11—13 refine the estimation sample by excluding treated drugs that exit the market within 5 years; treated drugs in markets that experience net entry; and all observations after the COVID-19 pandemic (2019+), respectively. Unless otherwise noted, standard errors are clustered at the drug-level and reported in parentheses. The dependent variable median is computed for treated drugs in the year prior to the recall. *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Quasi-experimental robustness: alternative market definitions

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Recalled drug | | | Competitor spillovers | | |
| | NADAC market | Molecule formulation | ATC4 | NADAC market | Molecule formulation | ATC4 |
| Recalled * post | -0.043*** | -0.044*** | -0.005** | 0.060*** | 0.068*** | 0.052*** |
| | (0.006) | (0.007) | (0.003) | (0.011) | (0.014) | (0.013) |
| Observations | 27,536 | 12,388 | 23,412 | 13,514 | 6,419 | 3,477 |
| Dependent var. median | 0.233 | 0.103 | 0.047 | 0.596 | 0.738 | 0.756 |
| Recalled * post (%) | -18% | -43% | -12% | 10% | 9% | 7% |

Notes: The table reports the estimated average effects of recall disclosures on intermediary drug purchases using different market definitions. It presents the coefficients obtained by estimating Equation (2) with drug market share as the dependent variable. Cols. (1) and (3) report effects when the market is defined as a molecule-formulation-strength combination, as used by Medicaid in its NADAC survey. Cols. (2) and (4) present effects using the baseline market definition, the molecule-formulation. Cols. (3) and (6) use the larget market definition, the ATC-4 group. Cols. (1)—(2) present the effects for recalled drugs estimated using drug-year level data. Cols. (3)–(6) present the total effects for drugs in the same market that were not recalled ("competitors") using market-level data. Standard errors are clustered at the drug (cols. 1–2) or market (cols. 3–5) level and reported in parentheses. The dependent variable mean is computed for treated drugs in the year prior to the recall. *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: Descriptive statistics: auctions

| Variable | Mean |
|---|---|
| Winning price per unit | 5.88 |
| | (1.18) |
| Mean annual prescriptions ('000,000) | 9.80 |
| Mean bidders per auction | 4.07 |
| Fraction with $\geq$1 recalled participants | 0.273 |
| 1 participant | 0.211 |
| 2 participants | 0.053 |
| 3 participants | 0.006 |
| 4 participants | 0.002 |
| Number of auctions | 67,880 |
| Number of observations | 276,495 |

Notes: The table presents descriptive statistics for generic drug procurement auctions over 2010-2022, the sample used for estimating the structural model. Winning prices per unit represent intermediary acquisition costs and are homogenized using the methodology described in Section 5.2. Annual prescription volumes for each drug-intermediary combination are observed in the Optum claims data. The number of bidders are observed in the Optum claims and FDA Orange Book. Recalls are from FDA Enforcement Reports. The datasets and sample construction are described in Sections 3.1 and 6, respectively.

Table A.6: Scoring auction model price homogenization

| Dependent var. | Price per unit |
|---|---|
| Volume | -0.166*** |
| Volume$^2$ | 0.000*** |
| Extended release | 0.783*** |
| Oral formulation | -5.445*** |
| Topical formulation | -3.383*** |
| Intermediary 2 | -0.147*** |
| Intermediary 3 | -0.154*** |
| Intermediary 4 | -0.173*** |
| Intermediary 5 | -0.168*** |
| *Therapeutic area (ATC-1)* | |
| Blood and blood forming organs | 2.471*** |
| Sensory organs | 1.128*** |
| Antiparasitic products, insecticides and repellents | 0.798*** |
| Genito urinary system and sex hormones | 0.245*** |
| Nervous system | 0.242*** |
| Antineoplastic and immunomodulating agents | 0.139* |
| Systemic hormonal preparations | 0.087 |
| Antiinfectives for systemic use | -0.031 |
| Cardiovascular system | -0.444 |
| Musculo-skeletal system | -0.509*** |
| Alimentary tract and metabolism | -0.599*** |
| Dermatologicals | -0.886*** |
| Respiratory system | -1.683*** |
| Various therapeutic areas | -0.389** |
| N. bidders fixed effects | x |
| Year fixed effects | x |
| Observations | 90,759 |

Notes: The table reports the estimated coefficients from a linear regression of prices on auction characteristics using auction-level data. Annual drug volumes units are presented in millions. Drug volumes and characteristics are from the NDC Directory. Intermediaries are observed in the Optum claims and NCPDP data. Apart from volume and volume$^2$, all regressors are indicator variables. *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Scoring auction model robustness

| | (1) Buyer preferences ($\Delta S$ / mean price) | (2) Bidder costs (high vs. low reliability) |
|---|---|---|
| 1. Optimal score (baseline) | 0.13 | 0.31 |
| 2. Long-run reliability | 0.07 | 0.39 |
| 3. Auctions with N=2 bidders | 0.15 | 0.03 |
| 4. Recalled & not-yet-recalled drugs | 0.37 | 0.47 |
| 5. Balanced panel | 0.05 | 0.24 |
| 6. Linear score | 0.18 | 0.77 |

Notes: The table summarizes the estimates of buyer preferences and bidder costs obtained by changing the estimation sample or various assumptions of the structural scoring auction model. Column (1) presents estimates of buyer willingness to pay for reliable drugs, expressed as a fraction of the mean price. Column (2) reports the average difference between high- and low-reliability drug unit costs, expressed as a fraction of low-reliability costs. In the first row, I present the baseline model estimates, assuming that buyers use an optimal scoring rule, and that "high-reliability" is defined as not being recalled in the past five years. Row 2 reports estimates using an alternate definition of reliability, which includes all drugs recalled within the prior decade, but excludes those that were most recently recalled (within the prior year). Row 3 tests for heterogeneity in buyer preferences by auction size and presents estimates from auctions with only $N = 2$ bidders. Row 4 limits the auction sample to recalled and not-yet-recalled drugs in order to isolate how recalled drug costs change after disclosure. Row 5 reports estimates using a "balanced" panel of auctions that experience no entry or exit. The last row shows the estimates when buyers are assumed to use a linear, rather than an optimal, score.

# B Technical Appendix

## B.1 Estimation Details

**Unobserved prices.** In this subsection, I first derive equations (6)–(7), which form a system of non-linear equations in the unobserved price distributions. Next, I describe the estimation algorithm and the conditions under which a solution to this system is guaranteed to exist.

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

$$Pr(b_{t,i}|i \text{ wins}) = \frac{Pr(i \text{ wins}|b_{t,i})Pr(b_{t,i})}{Pr(i \text{ wins}),} \tag{12}$$

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

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

This is the linear score analogue of Equation (10). 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 (12) and rearranging terms, we have

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

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 (7) can be derived in the same way (after switching $t = H, L$.) For convenience, I reproduce Equation (7) below:

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

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 (6) and (7) are

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

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

*Estimation algorithm.* Together, Equations (6) and (7) define a system of two nonlinear func-

tional 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, Perrigne, and Vuong, 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 $\varepsilon$ 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:

$$Pr(b_{H,i}|i \text{ wins})Pr(i \text{ wins}) = g_H(b_{H,i})\big[1 - G_H(b_{H,i})\big]^{N_H-1}. \tag{13}$$

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

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

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 (13) 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 (6) and (7) as

$$l = g_H(b_{L,i} + \Delta S)\big[k - g_L(b_{L,i})\varepsilon\big]^{N_L}\big[j - g_H(b_{L,i} + \Delta S)\varepsilon\big]^{N_H-1} \tag{14}$$

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

where we observe $N_H, N_L$ and $\varepsilon$, have guessed $\Delta S$, 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.[94] Note that objects $j, k, l$ and $m$ are all positive, since $j,\ k$ are cdfs, and $l,\ m$ are products of pdfs.

*Proofs of existence.* I now show that a solution to the non-linear system defined by Equations (14) and (15) exists at each price, for up to $N = 3$ bidders. For larger auctions, the curse of dimensionality (described in Section 5.2) makes estimating the LHS objects increasingly

94. 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.

(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 (14) and (15) 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

$$0 = k\varepsilon x^2 - (jk - m\varepsilon + l\varepsilon)x + jl \tag{16}$$

$$y = \frac{m}{j - \varepsilon x}. \tag{17}$$

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 (16) is non-negative:

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

This condition is satisfied as 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

$$0 = k\varepsilon^2 x^3 - 2jk\varepsilon x^2 + (-m\varepsilon + j^2 k + l\varepsilon)x - jl \tag{18}$$

$$y = \frac{m}{(j - \varepsilon x)^2} \tag{19}$$

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

$$0 = j\varepsilon^2 y^3 - 2jk\varepsilon y^2 + (m\varepsilon + jk^2 - l\varepsilon)y - km \tag{20}$$

$$x = \frac{l}{(k - \varepsilon y)^2} \tag{21}$$

Since Equations (18) and (20) are odd degree polynomials, they each have at least one real solution. Furthermore, for small values of $\varepsilon$, 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 $y \neq k/\varepsilon$.

**Monte Carlo studies.** To illustrate my estimation algorithm, described in section 5.2, 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 5000 replications of $N$ = 5000 auctions, which corresponds to a realistic auction sample size.[95] 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 repeat this exercise using a range of parameters; for exposition, I show the results from one such experiment in Figure A.9. This uses the following parameter values:

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

Figure A.9: Monte Carlo simulations of prices: Linear score

(a) Unobserved prices (high reliability types) (b) Unobserved prices (low reliability types)

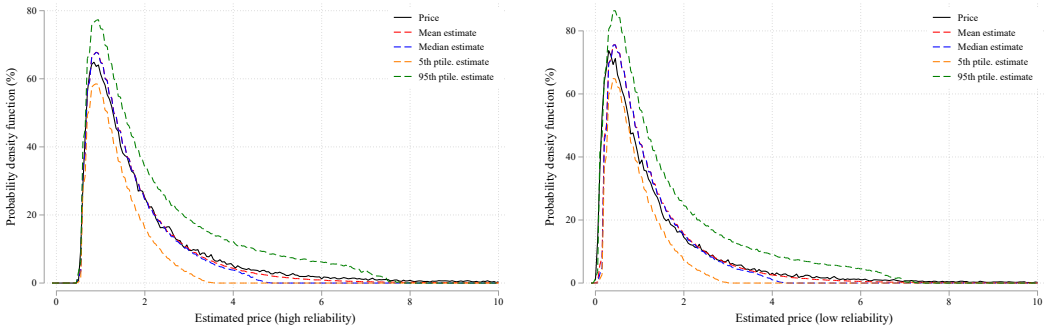



Figure A.9 plots the true and estimated price densities for high and low reliability types (in panels (a) and (b), respectively). The true densities are plotted using solid black lines, while the means, medians, and pointwise 90% confidence intervals of the estimates are plotted using dashed lines. The true values are strikingly similar to the mean and median estimates (conditional on a guess of the true value of $\Delta S$), 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

95. Note, however, that this is an order of magnitude smaller than my data.

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:

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

The results are summarized in Figure A.10, 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 $\theta|q$ (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 $\theta$. 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.

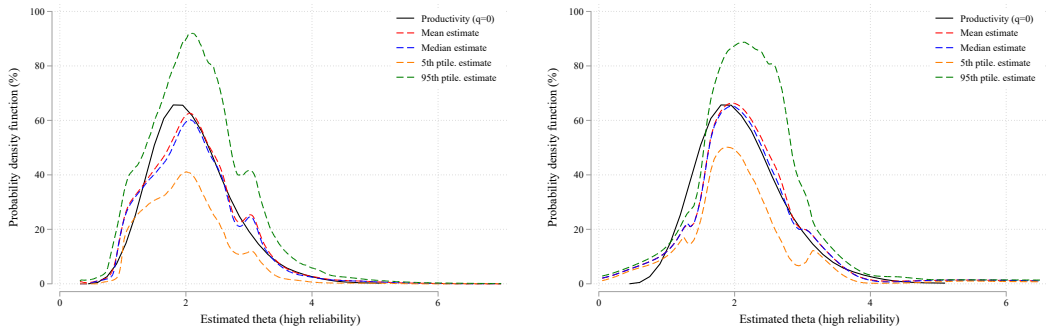
This is consistent with the idea that the distribution of the minimum price is a sufficient statistic for the underlying price distribution, so leaving out non-winning prices (in this case, half of the sample) does not result in significant information loss about the underlying productivity distributions. Comparing the densities of high (panels (c)—(d)) and low (panels (e)—(f)) reliability types provides additional color on this statement; the low reliability 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 $\Delta S$ 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.

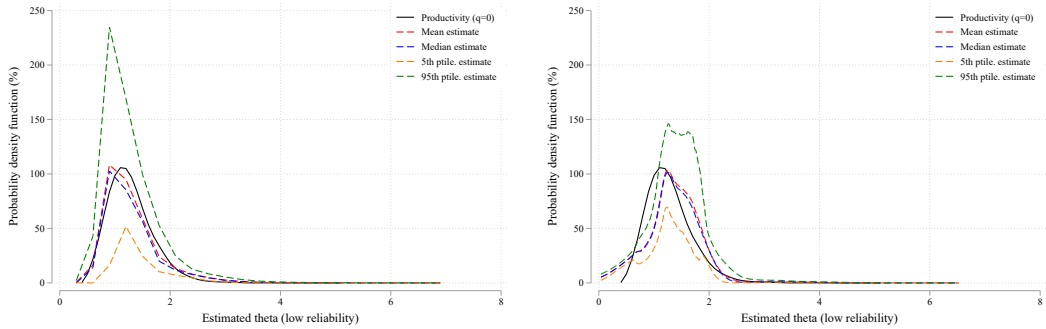
Figure A.10: Monte Carlo simulations of preferences and productivities

Laffont et al. (2020) Ishitani (2024)

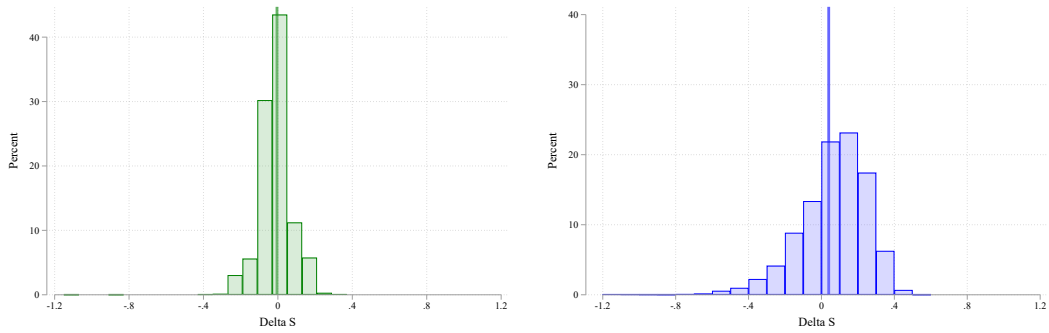
(a)–(b) Productivities (high reliability)




(c)–(d) Productivities (low reliability)




(e)–(f) Buyer preferences




## B.2 Background and Data

**Additional details about recalls:** 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.

While pharmacies are not legally obligated to inform patients about recalls, certain pharmacies have procedures to do so. For example, CVS Caremark and Walgreens state that they do so, 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: 90% of pharmacies reported being understaffed in 2021 (National Community Pharmacists Association, 2021). Anecdotes from lawsuits and the media suggest that information about recalls often fails to reach consumers (Newkirk and Berfield, 2019).

**Data Sources: Manufacturing facility locations.** 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)..."). Finally, I 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.

These datasets were first used to match drugs to facilities by researchers at the University of Minnesota Center for Infectious Disease Research and Policy (CIDRAP, 2018), who used it to identify the locations for 156 drugs. Similar datasets have been used by Joon Noh (2020; using label data) and Anais Galdin (2023; using trade data for injectable drugs). 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 72% of generic NDCs in the Optum sample, representing 92% of generic volumes.

**AAC data.** The national (NADAC) and state pharmacy average acquisition cost data are collected by a single firm, Myers and Stauffer, on behalf of CMS and state Medicaid agencies. Since the state price data have not been used in previous research, I describe them briefly here. Myers and Stauffer conducts monthly, mandatory surveys of random samples of 250-350 independent and chain pharmacies per state, excluding 340B pharmacies. In comparison to the national data, the state data are updated more frequently (weekly versus twice a week). They are also based on larger proportional sampling sizes—NADAC surveys 500-600 pharmacies *nationally*, versus 250-350 pharmacies per *state*—and mandatory rather than voluntary surveys.

The surveys capture invoice data, which represents acquisition costs 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 in the FDA NDC data.