Quality in the Generic Drug Market*

Emily Cuddy[1] Molly Schnell[2,3]

March 27, 2026

[1]Department of Economics, Duke University
[2]Department of Economics, Northwestern University
[3]National Bureau of Economic Research

Abstract

This paper examines whether generic drug quality differs across manufacturers in ways that matter for patient outcomes. We exploit pharmacy-level inventory dynamics as a source of plausibly exogenous variation in manufacturer assignment, comparing outcomes (i) across patients who initiate the same drug but receive different manufacturers and (ii) within patients who switch manufacturers of the same drug when refilling at the same pharmacy. Using administrative medical and pharmacy claims from a large national insurer between 2011 and 2019 and focusing on antidepressants, preliminary evidence suggests economically meaningful differences across manufacturers in early treatment adherence, as well as in subsequent health care utilization and spending. These findings highlight potential limits of current regulatory standards in ensuring substitutability between branded and generic products.

PENDING DISCLOSURE REVIEW; FOR INTERNAL USE ONLY

*We thank Amy Finkelstein, Josh Gottlieb, Matt Notowidigdo, Pëllumb Reshidi, and participants in seminars at the Toulouse School of Economics, the 2024 Duke Applied Micro Jamboree, the 2025 Texas A&M IO Day, and the 2026 Junior(ish) Health Economics Symposium for helpful feedback. We also thank Protima Advani, Oluwayomi Akintola, and Eleena Koep for their assistance in accessing and processing the data. Charles Hanzel, Wenjia Wang, Hellary Zhang, and Yiqing Zheng provided excellent research assistance.

1 Introduction

Generic drugs are fundamental to modern health care systems, underpinning affordability and access for patients worldwide. In the United States alone, generic substitution has generated an estimated $2.9 trillion in cost savings over the past decade, including $408 billion in 2022. The success of this market rests on a core assumption: that generic drugs are therapeutically equivalent to their branded counterparts. To support this premise, the U.S. Food and Drug Administration (FDA) mandates that generics demonstrate bioequivalence at approval and oversees drug safety and quality throughout the product lifecycle. Despite this regulatory framework, growing evidence suggests that generic drugs may differ in clinically meaningful ways. Regulatory recalls point to potential vulnerabilities in the generic supply chain: between 2000 and 2022, approximately 12 percent of drugs experienced at least one recall (Ishitani, 2026), including high-profile cases involving heparin, valsartan, and ophthalmic products (Eban, 2019). While such events are often treated as isolated or transitory, their prevalence raises a fundamental question: whether quality differences that matter for patient outcomes are rare exceptions or a systematic feature of generic drug markets.

This paper studies the prevalence of clinically meaningful quality differences across generic drugs and the market features that generate or amplify them. To answer these questions, we develop two complementary empirical approaches that exploit quasi-experimental variation in the firm responsible for a given generic drug—specifically, the Abbreviated New Drug Application (ANDA) holder—generated by pharmacy procurement processes.[1] In particular, retail pharmacy chains procure generic drugs through competitive bidding at the molecule–strength–form level (which we refer to as a "market"), typically on an annual basis and sometimes more frequently. As a result, patients initiating treatment at a given pharmacy are effectively assigned to the firm selected through this procurement process, and patients refilling the same prescription at the same pharmacy are often switched across firms without their knowledge.

Our two empirical approaches leverage these sources of plausibly exogenous variation in firm exposure to identify firm-level quality differences. We first exploit cross-firm variation at treatment initiation to compare subsequent clinical outcomes across otherwise similar patients starting the same molecule in the same market. In our preferred specification, this initiation design includes patient fixed effects, allowing us to compare outcomes within a

[1]The generic drug supply chain involves three primary actors: the ANDA holder (which owns the FDA approval), the labeler (which markets the drug under a National Drug Code), and the manufacturer (which produces the active ingredient and/or finished dosage form). These roles are often vertically integrated, particularly among large generic firms. Because the ANDA holder is legally responsible for product quality regardless of organizational structure, it forms the basis of our analysis. Throughout, we use "firm" and "manufacturer" interchangeably to refer to the ANDA holder.

patient before and after treatment initiation while allowing the post-initiation change to vary by the manufacturer received. We then leverage within-patient switches across firms induced by changes in pharmacy procurement to estimate market-level regressions in the spirit of Abowd et al. (1999) with patient and firm fixed effects. The initiation design allows us to study impacts on outcomes such as medication adherence, as sample inclusion does not require continued treatment through a manufacturer switch, while the switching design isolates within-patient variation in firm exposure during ongoing therapy. We apply these approaches to antidepressants, one of the most commonly prescribed medication classes in the United States, characterized by high generic penetration, multiple firms per market, and clinically sensitive treatment margins. Using outcomes such as hospitalizations, emergency department visits, and total health care spending, we find evidence of widespread and economically meaningful quality differences across generic drug firms.

There are several reasons to expect quality failures to arise in the generic drug market. First, firms face strong incentives to prioritize cost minimization over quality provision. Quality is difficult to observe and measure, limiting awareness among both buyers and regulators, and generic drugs typically lack brand-level reputational capital that would otherwise discipline quality choices. At the same time, procurement processes are highly price sensitive: buyers can readily switch to a new supplier offering even a marginally lower price, intensifying competitive pressure on costs.[2] Second, regulatory oversight of generic drugs is limited in scope. While the FDA requires firms to demonstrate bioequivalence at approval, these standards are imperfect and permit variation in certain drug characteristics, such as inactive ingredients. Manufacturing sites must also comply with current Good Manufacturing Practice (cGMP) standards and are inspected at approval and periodically thereafter, but inspections are relatively infrequent and typically announced in advance. Post-approval monitoring relies heavily on adverse event reporting through the FDA Adverse Event Reporting System, which requires quality problems to be sufficiently pervasive and actively reported in order to be detected. Together, these features create an environment in which cost pressures are strong and quality failures may arise—and persist undetected.

Our study relies on comprehensive medical and pharmaceutical claims data from a large national commercial insurer covering the period 2011–2019 and more than 65 million unique beneficiaries. The data include all medical and pharmacy claims for covered members, including procedure codes and negotiated prices, as well as limited patient demographics and plan characteristics. For each pharmacy claim, we observe an encrypted patient identifier, the prescription fill date, an encrypted National Provider Identifier (NPI) for the dispensing

[2]Consistent with this logic, Bate et al. (2011) document somewhat lower prices for substandard drugs based on 899 samples drawn from 17 low- and middle-income countries.

pharmacy, and the National Drug Code (NDC) of the dispensed product. We link NDCs to ANDA holders following Cuddy (2020), allowing us to identify the firm legally responsible for the quality of each generic drug. Because claims are observed longitudinally at the patient level and linked across medical and pharmacy files, we can track patients over time both after they initiate a particular product and as they refill the same prescription at the same pharmacy and are switched across generic firms without changing treatment or provider.

We focus our analysis on antidepressants, a drug class that offers at least four advantages for illustrating our empirical design. First, they are among the most commonly prescribed drug classes in the United States: from 2015–2018, 13 percent of adults reported taking an antidepressant in the past month (CDC, 2020). This prevalence is reflected in our data, where 4.2 million beneficiaries—approximately 10 percent of our analysis sample—fill 58 million antidepressant prescriptions over the period 2011–2019. Second, antidepressants are typically used on a chronic basis, with roughly 80 percent of prescriptions in our data filled as refills, allowing us to observe repeated within-patient switching across generic firms over time. Third, the market exhibits high generic penetration—approximately 93 percent in our data—with an average of five firms per market, generating rich variation in suppliers within narrowly defined drug markets. Fourth, antidepressants are clinically sensitive, with relatively narrow tolerability margins, so even modest quality differences are likely to translate into observable changes in patient behavior and health outcomes. While our primary analysis focuses on antidepressants, many of the outcomes we study—including medication adherence, medication adjustments, medical and pharmaceutical spending, and hospitalizations and emergency department visits—are relevant across a broad range of drug classes.

Our first empirical design exploits quasi-experimental variation across patients in generic firm at treatment initiation. We restrict the analysis to the first time a patient fills a given molecule–strength–form and compare patients initiating the same treatment but receiving the product from different firms due to pharmacy procurement decisions. We implement this design using two complementary specifications. First, we estimate patient–week–level regressions with patient fixed effects that compare outcomes within patients before and after treatment initiation, allowing the post-initiation change to vary by the manufacturer received at initiation while controlling for month–by–year fixed effects and flexible relative-time dynamics. Second, we estimate cross-sectional regressions that compare post-initiation outcomes—including measures of medication adherence—across patients receiving different manufacturers, controlling for a rich set of observable patient characteristics and prior medical and pharmaceutical utilization. Because inclusion in this sample does not require patients to remain on treatment beyond their first fill, the initiation design allows us to capture early treatment responses—including adherence, discontinuation, and other short-run

adjustments—as well as downstream clinical outcomes such as emergency department use and total medical spending.

Our second empirical design exploits within-patient changes in firm exposure induced by pharmacy procurement decisions. We focus on patients who return to the same pharmacy to refill the same prescription, including those who are "force switched" to a different firm and "forced stayers" who receive the same firm as on their prior fill.[3] We estimate market-level regressions of weekly patient-level outcomes on patient fixed effects, firm fixed effects, month–by–year fixed effects, and relative-time fixed effects. Patient fixed effects absorb all time-invariant differences in underlying health and preferences, while the relative-time fixed effects net out general refill-related dynamics common to both switchers and stayers. Under this design, the coefficients of interest—the firm fixed effects—are identified by within-patient changes in outcomes as individuals are switched across firms supplying the same drug, relative to patients whose firm exposure remains unchanged. Because patients do not observe or choose the firm supplying their generic drug, changes in firm exposure are unlikely to reflect selection on unobserved health trends.

We first document evidence of manufacturer quality differences at treatment initiation using variation in the generic supplier received when patients begin therapy. Comparing outcomes within patients before and after initiation, we find that the distribution of p-values from market-level tests of equality across manufacturers exhibits substantial excess mass near zero for many outcomes. For measures of health care utilization such as any medical or pharmaceutical spending and mental health evaluations, a large share of markets reject equality across manufacturers at conventional significance levels, far exceeding what would be expected by chance in the absence of quality differences. We find more limited evidence of systematic differences in acute hospital-based outcomes such as emergency department visits and hospitalizations.

Manufacturer differences also manifest at the earliest observable margin of treatment response: whether patients stay on therapy at all. Restricting to patients who filled a 30-day initial prescription, we classify each initiator's treatment status in the six weeks following their first fill into mutually exclusive categories—on-time refill, abandonment, dose or form adjustment, and molecule switch—and test whether these outcomes differ across manufacturers within a market. On-time refill and abandonment together account for approximately 80 percent of initiators and exhibit by far the strongest manufacturer dispersion: 50 percent of markets reject equality across manufacturers at the 5 percent level for on-time refill and 82

[3]Over our sample period, 6 percent of patients refilling an antidepressant are force switched to a different firm. The large majority of refill behavior is stable: 85 percent of patients return to the same pharmacy and receive the same firm as on their prior fill. An additional 9 percent of patients switch pharmacies when refilling their prescription.

percent for abandonment, compared to fewer than 15 percent of markets for dose adjustment and molecule switch. The relative max-gap between the best and worst manufacturer averages 10 percent of the baseline on-time refill rate among rejecting markets, and reaches 44 percent of the baseline abandonment rate in the most extreme case, implying that a patient's probability of never picking up a second fill can differ by more than 12 percentage points depending solely on which manufacturer their pharmacy stocked. These adherence differences are largest among patients with lower attachment to mental health treatment, consistent with quality differences biting hardest where treatment continuation is most discretionary.

We further demonstrate that the switching design can detect quality differences in a setting with (now) well-documented quality failures. In December 2006, the first generic version of Wellbutrin XL 300 mg, a bupropion hydrochloride extended-release tablet, entered the market. Soon thereafter, patients who switched from the branded product to certain generic versions reported reduced efficacy (Associated Press, 2008). Subsequent regulatory reviews conducted several years later confirmed these concerns: in October 2012, the FDA concluded that one widely used generic product, marketed by Teva Pharmaceuticals, was not therapeutically equivalent to the branded drug. A broader FDA review completed in October 2013 further determined that a second generic product, marketed by Watson, also failed to meet bioequivalence standards. Consistent with these findings, we show that patients who are switched to non-equivalent firms—because they refill the same prescription at the same pharmacy but receive a different generic supplier—experience significantly worse outcomes. Specifically, switching to Teva is associated with a 0.5 percentage point increase in emergency department visits or hospitalizations in the subsequent week, an 80 percent increase relative to a mean of 0.6 percentage points, and a 0.4 percentage point increase in mental health evaluations, an 133 percent increase relative to a mean of 0.3 percentage points.

The rest of the paper proceeds as follows. Section 2 provides institutional background on the U.S. generic drug market, including the supply chain, regulation, enforcement, and existing measures of quality. Section 3 introduces our data sources, primary outcome measures, market definition, and sample construction. Our two sets of empirical approaches—the initiation designs and the switching design—are outlined in Section 4, along with a discussion of identification and inference. Section 5 presents preliminary results, and Section 6 concludes.

2 Institutional background

2.1 The U.S. generic drug supply chain

The U.S. generic drug supply chain centers on the Abbreviated New Drug Application (ANDA) holder, the entity that owns FDA approval for a given generic product. As outlined in Section 2.2, the ANDA holder is legally responsible for the product and for compliance with applicable manufacturing standards.

Although the ANDA holder bears regulatory responsibility, the production and commercialization of an approved generic product may involve three distinct actors: the ANDA holder, the labeler (which markets the product under one or more NDCs), and the manufacturer (which produces the active ingredient and/or finished dosage form).[4] These roles are often vertically integrated, especially among large generic firms that both hold ANDAs and manufacture in-house, whereas smaller firms more frequently rely on contract manufacturing organizations. Regardless of organizational structure, the ANDA holder retains ultimate legal responsibility for quality and therefore forms the basis of our analysis.

Multiple ANDA holders may supply products approved as therapeutically interchangeable within a given molecule–strength–dosage form combination (hereafter, a "market") at the same time.[5] Because generics are treated as interchangeable at the pharmacy level, competition occurs primarily through price rather than product differentiation. Pharmacies and wholesalers typically acquire generics through repeated procurement auctions, with manufacturers submitting bids and purchasing intermediaries allocating volume based on price, reliability, and contractual terms (Cuddy, 2020).[6] Because generics within a market are typically reimbursed at a common reference price that does not vary across manufacturers, pharmacy margins depend on the spread between reimbursement and acquisition cost. Pharmacies therefore have strong incentives to source from lower-cost suppliers, and procurement may rotate across manufacturers as entry, exit, or changes in production shift relative prices.

Substitution across generic manufacturers thus occurs largely at the pharmacy level rather than through patient or prescriber choice. Patients typically do not observe manufacturer identity, and prescribers do not specify it. Competition among ANDA holders therefore

[4]The NDC is a unique identifier that encodes the labeler, product, and package configuration. A single ANDA may correspond to multiple NDCs reflecting different packaging configurations or relabeling arrangements. However, each NDC is associated with a single ANDA holder; the FDA does not permit multiple firms to hold approval for the same NDC.

[5]Figure 1 illustrates the hierarchical structure of drug products and the market definition used in our analysis.

[6]Recent work by Ishitani (2026) highlights the importance of reliability in generic drug procurement auctions, documenting that pharmacies and wholesalers shift sourcing away from manufacturers within a market that have previously been subject to recalls.

operates primarily through procurement bids, with limited direct demand-side discipline tied to firm identity. This contrasts with branded pharmaceuticals, where prescribers and patients select products by brand name and manufacturer identity is salient, allowing demand to respond more directly to perceived quality differences. In markets with multiple entrants and compressed margins, the generic procurement structure should reduce returns to quality investments that are costly and not readily observable.

2.2 Regulation in the U.S. generic drug market

Generic drugs are approved in the United States through the Abbreviated New Drug Application (ANDA) pathway. Unlike sponsors submitting new drug applications (NDAs), ANDA applicants are not required to establish clinical efficacy. Instead, they must demonstrate pharmaceutical equivalence and bioequivalence to the reference listed drug. Pharmaceutical equivalence requires the same active ingredient, strength, dosage form, and route of administration, while bioequivalence requires statistically similar delivery of the active ingredient to the bloodstream in terms of rate and extent. Applicants must also submit detailed chemistry and manufacturing controls (CMC) documentation describing production processes, facilities, and quality systems. Approval is product- and site-specific, and manufacturing facilities are subject to pre-approval inspection to verify compliance with current Good Manufacturing Practice (cGMP) requirements. Upon approval, an ANDA holder may market the product under one or more National Drug Codes (NDCs), which identify the labeler, product, and package configuration.

Following approval, oversight continues under the FDA's regulatory framework governing manufacturing quality and post-market compliance. Compliance with cGMP remains central, governing how drugs are produced and monitored. Manufacturers must maintain quality systems, document their processes, and report production problems and adverse events. Oversight includes surveillance inspections, review of CMC documentation and post-approval supplements, and enforcement actions such as inspection observations, warning letters, and recalls (see Section 2.3). Because production frequently occurs in globally dispersed facilities, inspections span both domestic and foreign sites (FDA, 2025). Although regulatory standards are formally uniform, inspection activity is risk-based: timing and frequency reflect assessed risk, prior compliance history, agency resources, and logistical constraints—particularly for foreign facilities requiring international coordination (FDA, 2023). As a result, oversight intensity varies across firms and over time.[7]

[7]During the COVID-19 pandemic, routine inspections—both domestic and foreign—were temporarily postponed while only mission-critical inspections continued (Cuddy et al., 2023).

This institutional structure creates competing incentives for quality. While bioequivalence standards, cGMP requirements, and post-market monitoring in principle impose a regulatory floor on product quality, these safeguards do not guarantee identical performance across manufacturers. Bioequivalence testing permits statistical bounds around pharmacokinetic measures, and firms may differ in inactive ingredients and formulation characteristics that can affect tolerability or side-effect profiles for some patients (FDA, 2001). Although inspections are central to enforcement, they occur intermittently and are often pre-announced for foreign facilities, limiting the extent to which oversight continuously verifies production quality (GAO, 2016; OIG, 2017). At the same time, price-based competition may reduce incentives to invest in process improvements, redundancy, or high-reliability production systems. Because substitution occurs largely at the pharmacy level and prescribers and consumers typically cannot observe or select manufacturers, demand-side forces provide limited discipline for quality (see Section 2.1). Together, these features shape both the scope for quality differentiation and the difficulty of inferring underlying quality from observable regulatory signals such as inspections and recalls, as discussed below.

2.3 Enforcement and existing measures of quality

The regulatory framework described above generates observable enforcement signals—such as recalls, reported shortages, and inspection outcomes—that are often interpreted as indicators of manufacturer quality. While each provides information about compliance and production processes, none directly captures whether process deviations meaningfully affect patient outcomes.

Recalls are frequently interpreted as evidence of manufacturing defects. In practice, however, most recalls are initiated voluntarily by manufacturers following internal quality reviews (Ghijs et al., 2024), with the FDA classifying their severity and overseeing corrective actions. Because recall initiation typically begins within the firm, recall incidence reflects both the underlying rate of defects and the firm's monitoring and disclosure practices. Manufacturers with more stringent internal controls may issue recalls more frequently precisely because they detect and report problems earlier, whereas firms with weaker compliance systems may under-detect or delay disclosure absent regulatory intervention. As a result, recall frequency reflects detection and reporting behavior as much as underlying product quality and does not map cleanly into therapeutic performance.

Drug shortages are another widely cited signal of manufacturing problems. While shortages can arise from production disruptions linked to quality failures, they are also shaped by market structure and firms' supply decisions (Conti and Wosińska, 2025). In competitive

generic markets characterized by thin margins, manufacturers may reduce output or exit markets in response to pricing pressure or procurement dynamics, generating shortages unrelated to manufacturing quality. Conversely, quality problems do not necessarily result in formal shortage declarations if alternative suppliers remain available. Although manufacturers are required to notify the FDA of certain anticipated supply disruptions, shortage listings reflect market-wide availability rather than a direct assessment of therapeutic performance. Shortages therefore conflate production reliability, market incentives, and quality.

Inspection outcomes—such as Form 483 observations and warning letters—provide direct information about compliance with cGMP. These findings are issued by FDA investigators as part of the agency's enforcement authority. However, inspections are risk-based rather than randomly assigned (FDA, 2023). Facilities with prior violations, larger market shares, more complex products, or foreign locations may face different inspection frequencies and levels of scrutiny. Inspection outcomes therefore reflect regulatory targeting and enforcement intensity in addition to underlying manufacturing quality.

Finally, post-market adverse event reports provide another potential signal of product safety. The FDA's Adverse Event Reporting System (FAERS) collects voluntary reports from clinicians, patients, and manufacturers and serves as a pharmacovigilance tool for detecting potential safety concerns. However, adverse event reports are subject to substantial underreporting and reporting heterogeneity, and report incidence depends on product utilization, publicity, and litigation exposure in addition to underlying product performance (Hazell and Shakir, 2006; Bate and Evans, 2009; Alatawi and Hansen, 2017). Moreover, reported events do not establish causality and are rarely designed to distinguish among therapeutically equivalent manufacturers. As a result, while adverse event data are more closely linked to patient health, they provide a noisy and indirect signal of manufacturer-level quality.

Together, these measures provide partial signals of compliance and production stability but do not directly capture differences in patient outcomes across therapeutically equivalent manufacturers. This limitation motivates the outcome-based approach we develop below.

3 Data and sample construction

3.1 Data sources

Our primary data source is administrative medical and pharmacy claims from a large national commercial insurer covering 2011–2019. The data include enrollees in employer-sponsored insurance plans, including both fully insured and self-insured (administrative services only)

arrangements, and cover over 65 million unique lives during the sample period. For each enrollee, we observe an encrypted patient identifier, basic demographic characteristics, and plan enrollment information.

Our empirical design relies on four core features of the pharmacy claims: the National Drug Code (NDC-11), an encrypted pharmacy identifier, the prescription pick-up date, and patient-level linkage to medical claims. Each pharmacy claim records the NDC-11 identifying the exact product dispensed, including manufacturer, formulation, and package size. We map NDCs to their corresponding NDA or ANDA holder—the entity legally responsible for regulatory compliance and product quality (see Section 2.1)—using RxMix, an API provided by the U.S. National Library of Medicine as part of the RxNorm/RxNav system. We supplement these mappings with historical drug label data to resolve cases not identified through RxMix.[8]

For each prescription, we further observe an encrypted National Provider Identifier (NPI) for the dispensing pharmacy, allowing us to identify the exact pharmacy location. Because pharmacies procure drugs through wholesalers and maintain inventories that evolve over time, the same patient filling the same prescription at the same pharmacy may receive different manufacturers across visits as local inventory changes (see Section 2.1). Observing pharmacy identifiers therefore allows us to compare patients within pharmacy over time and to identify situations in which patients remain on the same manufacturer or switch manufacturers due to pharmacy inventory dynamics.

To study how outcomes vary across manufacturers, we leverage the dispensing date and patient-level linkage to medical claims. The dispensing date allows us to track the timing of fill events, and the linked medical claims allow us to examine how health care utilization evolves around those events. The medical claims include place-of-service indicators, procedure codes (CPT/HCPCS), diagnosis codes (ICD-9/10), provider identifiers, and insurer-allowed amounts. We use the diagnosis codes to construct baseline measures of health status, including Elixhauser comorbidity indices, and the provider identifiers to characterize prescribing clinicians. As outlined in the next section, we combine these data with the pharmaceutical claims to construct our primary outcomes.

3.2 Primary outcome measures

We examine several primary outcomes constructed from the medical and pharmaceutical claims. We distinguish between (i) outcomes that can be measured in patient-week data both before and after the focal fill event and therefore used in specifications with patient

[8]For approximately 10% of NDCs (affecting 0.02% of prescriptions), we are unable to assign an NDA or ANDA holder. These are mainly repackaged products, which are rarely dispensed in a retail setting.

fixed effects and relative-time controls and (ii) outcomes that are only defined following the fill event and therefore are analyzed using post-initiation comparisons.

We construct four primary outcomes that can be observed on both sides of the fill event. From the medical claims, we define patient-week indicators for any emergency department visit or hospital admission and for any mental health visit, including psychotherapy and other behavioral health interventions, using place-of-service indicators, diagnosis codes, and procedure codes reported in the claims. We also combine the medical and pharmaceutical claims to construct patient-week measures of total medical and pharmaceutical spending and an indicator for any medical or pharmaceutical spending. We measure spending using insurer-allowed amounts reported in the claims and exclude pharmaceutical spending on the focal prescription so that mechanical cost of the fill event does not directly enter the outcome.[9]

We construct our primary adherence outcomes by classifying each initiator's treatment status in the six weeks following their initial fill. We restrict to patients whose initial prescription was for a 30-day supply so that the six-week window provides approximately two weeks after supply exhaustion for the patient to seek a refill. As shown in Figure 2, we classify each patient into one of five mutually exclusive categories based on the fills observed in this window. Patients with no follow-up antidepressant script of any kind within six weeks are classified as abandoned. Among patients with at least one follow-up script, we identify the last fill by date within the window and classify based on its content: a refill of the same molecule at the same strength and dosage form; a fill of the same molecule at a different strength or dosage form; a fill of a different antidepressant molecule; or a fill of the branded version of the same molecule, identifiable by NDA rather than ANDA holder status in the NDC mapping (branded switch, fewer than 0.1 percent in both samples). When a patient's last fill date contains multiple simultaneous fills, we apply a priority rule: branded switch takes precedence, followed by normal refill, dose switch, and molecule switch, reflecting the principle that continuation on the focal drug is the most conservative classification when the evidence is mixed.[10] Patients whose only fills within the window are reversals of the focal molecule—prescription claims that were processed but subsequently reversed prior to

[9]We adjust all spending measures for inflation and winsorize total spending at the 99.9th percentile of the patient-week distribution.

[10]A potential concern with our classification is that patients assigned to the normal refill category may also receive a new antidepressant molecule within the six-week window, which could reflect augmentation of the original drug rather than continuation alone. However, clinical guidelines recommend augmentation only after an adequate trial of the initial antidepressant, typically defined as six to twelve weeks at therapeutic doses. Augmentation within the six-week window we study would therefore be premature by clinical standards, and any co-occurring new molecule script in this period is more plausibly attributable to a prescriber decision to replace rather than supplement the initial treatment.

pickup—are classified as a reversal. Reversals of non-focal fills, by contrast, are treated as switch attempts and classified according to the molecule of the reversed fill.

3.3 Market definition

We define a drug "market" as a unique combination of molecule, strength, and dosage form (see Figure 1). For example, sertraline 50 mg oral tablets constitute a distinct market from sertraline 100 mg oral tablets or sertraline 50 mg extended-release capsules. This definition aligns with the FDA's Orange Book therapeutic equivalence framework, and thus products within a market are substitutable for purposes of pharmacy dispensing (FDA, 2024).[11]

We focus on antidepressants in the empirical analysis. In our data, there are 182 distinct antidepressant markets. Because our analysis exploits within-market variation across generic manufacturers for identification, we include only markets in which multiple generic manufacturers are active during the sample period. This criterion excludes the 25 markets in which the branded molecule maintained exclusivity through 2019, as well as the 27 markets in which only a single generic manufacturer was active throughout the sample, thereby reducing the sample to 120 markets.

Additional restrictions on the set of included markets depend on the empirical design. As outlined in Section 4, we implement two designs: one comparing outcomes across patients receiving different manufacturers at treatment initiation, and one comparing outcomes among patients who switch manufacturers when refilling at the same pharmacy. We require that markets contain a sufficient number of patients to support both the initiation-based and switching-based analyses. For specifications relying on patient-level switching, we further require sufficient cross-manufacturer switching to ensure connectedness of the manufacturer–patient graph and stable estimation of manufacturer fixed effects. Applying these additional criteria yields a final sample of 44 markets for the initiation-based design and 40 markets for the switching-based design.

Table 1 reports descriptive characteristics of antidepressant markets and compares the full set of markets observed in the data to the subsets used in the initiation and switching analyses. The table summarizes market size, competitive structure, and key drug characteristics. Markets included in the analytic samples are broadly similar to the full set of antidepressant markets in terms of drug class and formulation composition. For example, SSRIs account for 38–43% of markets across all three columns and extended-release formulations account

[11] More precisely, within a molecule–strength–formulation combination, generic products designated as "AB-rated" have demonstrated bioequivalence to the reference listed drug and are legally substitutable. Therapeutic equivalence ratings are assigned relative to a specific dosage form and release mechanism; immediate-release and extended-release formulations are therefore treated as separate markets.

for 23–27%. Consistent with the identification requirements described above, the estimation samples tend to include markets with substantially larger patient counts and prescription volume and more active manufacturers. The average market in the full sample contains 46,100 patients and 321,100 total fills over the sample period, compared to 176,900 patients and 1.2 million fills in the initiator markets and 193,200 patients and 1.4 million fills in the forced-switch markets. Similarly, the mean number of active manufacturers per market-year rises from 3.9 in the full sample to 8.3 and 8.4 in the initiator and forced-switch samples, respectively, reflecting the need for sufficient prescribing activity and, for the switching design, meaningful opportunities for substitution across generic manufacturers within a market.

3.4 Analytic samples

Having defined the eligible markets for each empirical design, we next construct the individual-level analytic samples. We begin with all 65.4M individuals enrolled in the claims data between 2011 and 2019. While the data contain both medical and pharmacy claims, some enrollees have coverage for only one type of service. Because our analysis requires observing both medication utilization and subsequent health care outcomes, we restrict the sample to the 43.8M individuals with both medical and pharmacy benefits during the relevant observation window. We further require continuous enrollment over the outcome measurement period to ensure that prescription activity and health care utilization are fully observed. These restrictions yield a sample of 30.2M individuals.

From this population, we identify individuals with at least one antidepressant prescription during the sample period, resulting in 4.2M individuals and 58.4M antidepressant fills. We then limit the sample to prescriptions in markets that satisfy the design-specific eligibility criteria described above. Within those markets, we construct separate analytic samples for the initiation and switching designs.

Starting with the initiation design, we identify 3.8M individuals with continuous medical and pharmaceutical coverage filling initial antidepressant prescriptions in the 44 included markets. We identify treatment initiators as patients whose first observed fill of a given molecule–strength–form follows at least four months without a prior fill in that market. To ensure that manufacturer assignment at initiation reflects pharmacy inventory rather than patient selection, we further restrict the sample to initiators filling their prescription at a pharmacy they previously used for other medications. After applying these criteria, the analytic sample for the initiation design includes 1.9M individuals.

Turning to the switching design, we observe 4.2M individuals with continuous medical and pharmaceutical coverage filling 57.8M antidepressant prescriptions in the 40 markets

included in that design. From this population, we focus on refill prescriptions of the same molecule–strength–form as the prior fill, yielding 49.5M refill prescriptions among 3.4M patients. We classify these refills according to whether the pharmacy and manufacturer change between fills, yielding four categories: (i) forced switchers (same pharmacy, different manufacturer), (ii) forced stayers (same pharmacy, same manufacturer), (iii) unforced switchers (different pharmacy, different manufacturer), and (iv) unforced stayers (different pharmacy, same manufacturer). To avoid treatment disruptions resulting from or correlated with pharmacy switches, we restrict the switching analysis to same-pharmacy refills. This yields a final analytic sample for the switching design that includes 611,692 patients.

Table 2 reports patient-level averages of plan and demographic characteristics across the populations observed in the data. Column (1) describes all insured beneficiaries with at least one antidepressant prescription during the sample period, column (2) reports characteristics for treatment initiators included in the initiation design (forced takers), and column (3) reports characteristics for patients who refill the same prescription at the same pharmacy but receive a different manufacturer than on their prior fill (forced switchers). These comparisons allow us to assess how representative the initiation and switching samples are relative to the broader population of antidepressant users. The analytic samples look broadly similar to the full antidepressant population: initiators and forced switchers are somewhat younger on average (42.0 and 49.7 years versus 44.3) and more likely to be female (66.1% and 69.3% versus 65.6%), with forced switchers having longer average coverage spells (74 months versus 48), reflecting the requirement that patients remain on treatment long enough to experience an inventory-driven manufacturer change.

Table 3 reports the full distribution of six-week adherence outcomes among initiators, including the cross-market inter-quartile range. Normal refill and abandonment together account for approximately 80 percent of initiators, with the remaining 20 percent distributed across dose switch, molecule switch, and reversal. The low rates of molecule switching and dose adjustment—each below 12 percent—are consistent with clinical guidelines recommending that prescribers allow at least four to eight weeks before concluding that an initial antidepressant has failed and making treatment changes (Boyce et al., 2020; Qaseem et al., 2023). The rarity of branded switching (fewer than 0.1 percent) indicates that patients very rarely pay out-of-pocket to escape to the branded version of their assigned generic, making it an impractical primary adherence outcome in this setting, though we retain it as a conceptually important category in the classification.

One natural question is whether the distribution of treatment outcomes, and any manufacturer differences therein, varies between patients taking antidepressants for an underlying mental health condition versus those receiving them for off-label indications such as chronic

pain, insomnia, or anxiety management. Patients with a confirmed mental health diagnosis on record are arguably those for whom antidepressant treatment is most clearly indicated and for whom abandonment or failure to refill most plausibly reflects a treatment quality problem rather than appropriate discontinuation of an adjunctive therapy.

Table 3 also reports the distribution of outcomes separately for the subset of initiators with a mental health diagnosis on record in the past four months. The two samples are broadly similar: normal refill rates are nearly identical (44.7 versus 44.8 percent), and abandonment is modestly lower among MH-diagnosed initiators (32.0 versus 35.5 percent), consistent with the more active treatment management typical of patients with established mental health care relationships. Molecule switch and dose adjustment rates are correspondingly somewhat higher in the MH sample, suggesting that prescribers are more likely to make active adjustments for these patients rather than allowing them to drift off therapy.

Table 4 reports refill-level summary statistics by refill type among those who refill at least one antidepressant prescription. Across markets, forced events account for the majority of refills, with forced stayers accounting for 44% of refills (column (1)) and forced switches accounting for 21% (column (4)), albeit with significant heterogeneity across markets.[12] Refill timing, prescription characteristics, refill history, and cost-sharing measures are broadly similar across refill types, indicating that forced switches occur in otherwise typical refill circumstances rather than at clinically or economically unusual events. Forced switchers and forced stayers look similar along most dimensions, with comparable days supply, refill timing, and cost-sharing. Refill history is also similar—forced switchers have been on their current drug for an average of 699 days with 10.1 prior fills, compared to 720 days and 10.7 fills for stayers—though the length of this treatment horizon leaves room for survivorship bias to play a role, motivating our initiation design as a complement.

4 Empirical approaches

Our empirical strategies exploit quasi-experimental variation in firm assignment generated by retail pharmacy procurement. As outlined in Section 2.1, retail chains procure generic drugs at the molecule–strength–form ("market") level through periodic competitive bidding. As a result, manufacturer assignment for a given prescription is determined by the pharmacy's inventory at the time of dispensing rather than by patient or prescriber choice.

This procurement structure generates two sources of plausibly exogenous variation. First, patients initiating therapy at their usual pharmacy receive the manufacturer stocked by

[12]Switching frequencies largely reflect differences in the number of active manufacturers and patterns of entry and exit. Markets with fewer manufacturers generate fewer opportunities for switching.

that pharmacy at the time of treatment initiation. Second, continuing patients who refill their prescription at the same pharmacy are switched across manufacturers when inventory changes occur. We leverage both forms of variation to estimate manufacturer-specific effects on patient outcomes, allowing us to examine quality differences both at treatment initiation and during ongoing therapy.

Section 4.1 presents the initiation design, which captures early-treatment responses before selection into continued therapy occurs. Section 4.2 then introduces the switching design, which uses within-patient changes in manufacturer exposure among continuing users to estimate more persistent differences in product performance. Section 4.3 discusses the identifying assumptions underlying the two designs, and Section 4.4 describes our inference procedures.

4.1 Initiation design

We focus on patients initiating therapy in market m who fill their first prescription at a pharmacy they have previously used for other medications. For these patients, firm assignment at initiation is determined by the pharmacy's inventory at the time of the first fill, rather than by active patient or prescriber choice. We refer to these individuals as "forced takers": conditional on selecting their modal pharmacy, they receive the manufacturer stocked by that pharmacy when treatment begins.

For outcomes observed repeatedly over time, we construct an event window around treatment initiation and estimate the following specification separately by market using patient–week-level data:

$$y_{it} = \alpha_i + \Psi_{M(i)} \cdot \mathbf{1}\{t \geq t_{i0}\} + \tau_{my(t)} + \sum_{k \in \kappa, k \neq -1} \gamma_k \cdot \mathbf{1}\{t - t_{i0} = k\} + \epsilon_{it}, \tag{1}$$

where $M(i)$ denotes the manufacturer received at initiation, t_{i0} is patient i's initiation date,[13] α_i denotes patient fixed effects, and $\tau_{my(t)}$ denotes month–by–year fixed effects. The patient fixed effects absorb time-invariant differences across patients, while the month–by–year fixed effects capture common time shocks within markets. By interacting manufacturer indicators with an indicator for periods $t \geq t_{i0}$, we estimate manufacturer-specific shifts in outcomes following initiation relative to the patient's pre-initiation baseline.

We incorporate relative-time effects, γ_k, to flexibly account for common utilization dynamics around treatment initiation. The onset of antidepressant therapy frequently coincides with physician visits, diagnostic evaluation, and adjustments to other medications, generat-

[13] We observe the fill date of the prescription.

ing predictable short-run changes in medical and pharmaceutical spending that are unrelated to firm quality (APA, 2010; Qaseem et al., 2023). Including relative-time indicators ensures that such initiation-related dynamics are not attributed to the firm fixed effects. Identification of Ψ_M therefore comes from cross-firm differences that persist after accounting for common drug initiation dynamics.

We complement this approach with a secondary specification for outcomes that are observed only once following treatment initiation, such as whether a patient discontinues therapy within 30 days, experiences a dose adjustment, or initiates adjunctive pharmacotherapy.[14] Because these outcomes are realized only once per initiation episode, we cannot include relative-time indicators. Instead, we estimate the following patient-level cross-sectional specification, again separately by market:

$$y_i = \Psi_{M(i)} + \tau_{my(i)} + X_i\beta + \epsilon_i, \tag{2}$$

where y_i denotes the outcome measured following treatment initiation, $M(i)$ again denotes the manufacturer received at initiation, and $\tau_{my(i)}$ denotes month–by–year fixed effects corresponding to the calendar month–year of initiation. Because patient fixed effects cannot be included in this specification, we instead control for a rich set of baseline patient characteristics X_i measured prior to treatment initiation. These controls include demographic characteristics (e.g., age), geographic indicators (e.g., ZIP-3), comorbidity measures (e.g., Elixhauser indices), and detailed measures of prior mental health status, including diagnoses, receipt of psychotherapy, prescribing provider specialty, and other indicators of mental health treatment intensity.

The credibility of the cross-sectional specification depends on whether the observable characteristics in X_i adequately control for differences across patients that may be correlated with manufacturer assignment. Such correlation could arise if patients with different underlying health status or treatment histories systematically obtain prescriptions from pharmacies with different inventory patterns. To assess whether the rich controls in X_i adequately capture cross-patient differences, we use the event-time panel design as a validation exercise. Specifically, we compare manufacturer effects estimated from equation (1) when including patient fixed effects with those from an otherwise identical specification that includes the vector of observable characteristics X_i in place of α_i. If the resulting estimates are similar, this suggests that selection on dimensions not captured by X_i is limited and

[14] Adjunctive (or augmentation) therapy refers to the addition of a second medication intended to enhance the effect of the initial treatment. For example, a physician may prescribe a selective serotonin reuptake inhibitor (SSRI) antidepressant and subsequently add bupropion, a norepinephrine–dopamine reuptake inhibitor (NDRI), or aripiprazole, an atypical antipsychotic, to augment the therapeutic response.

supports the use of the conditional cross-sectional design for outcomes measured only once following treatment initiation. This strategy follows a common approach in the empirical literature that compares estimates with and without unit fixed effects to assess whether rich observable controls adequately proxy for unobserved heterogeneity (e.g., Altonji et al., 2005; Oster, 2019).

4.2 Switching design

While the initiation design captures manufacturer-specific shifts in outcomes at the start of treatment, it focuses on early-treatment responses and, for outcomes observed only once, relies on conditioning on observable characteristics. We therefore complement the initiation analysis with a switching design among continuing users that exploits within-patient changes in manufacturer exposure induced by pharmacy procurement changes, allowing us to assess more persistent differences in product performance while accounting for time-invariant unobserved patient heterogeneity.

Among continuing users who refill their prescription at the same pharmacy, manufacturer assignment changes when the pharmacy updates its inventory. We exploit these inventory-driven switches by adapting the Abowd, Kramarz and Margolis (1999) (AKM) framework to the pharmaceutical setting, decomposing outcomes into patient and manufacturer components. The estimation sample includes both "forced switchers," whose manufacturer changes because inventory rotates to a new supplier at the same pharmacy, and "forced stayers," who continue to refill at the same pharmacy but whose manufacturer remains the same.

Whereas AKM decomposes wages into worker and firm components, we decompose health outcomes into patient and firm components within narrowly defined drug markets. Specifically, for each market m, we estimate the following AKM-style model using patient–week-level data:

$$y_{it} = \alpha_i + \Psi_{M(i,t)} + \tau_{my(t)} + \sum_{k \in \kappa, k \neq -1} \gamma_k \cdot \mathbf{1}\{t - t_{is} = k\} + \epsilon_{it}, \tag{3}$$

where $\Psi_{M(i,t)}$ is a firm fixed effect corresponding to the ANDA holder who supplies the product dispensed to patient i at time t, t_{is} denotes the week of refill event s for patient i, and all other variables are defined as in equation (1). Because estimation is conducted separately by market, manufacturer effects are identified only from switching among ANDA holders producing therapeutically equivalent products of the same molecule, strength, and dosage form. Identification comes from within-patient changes in manufacturer exposure over time, with forced stayers helping to identify common refill-related and calendar-time outcome dynamics.

As in equation (1), we include relative-time effects γ_k.[15] Here, these controls flexibly capture common utilization dynamics around refill events, and their inclusion is important for two reasons. First, switching manufacturers may generate short-run adjustment costs that are not indicative of persistent quality differences. Even when products are therapeutically equivalent, changes in pill appearance, packaging, or patient perceptions may induce temporary adherence disruptions or health care utilization responses. Second, prescription refill events themselves are often associated with short-run fluctuations in utilization that are unrelated to manufacturer identity. Refills frequently coincide with physician visits, medication management, or the bundling of multiple prescriptions into a single pharmacy trip, mechanically increasing contemporaneous spending.

Because manufacturer switching in our setting occurs at refill dates, both transition effects and generic refill-related dynamics could be misattributed to manufacturer fixed effects if not explicitly accounted for. The coefficients γ_k absorb these short-run adjustment dynamics and common refill-related fluctuations, ensuring that the firm effects $\Psi_{M(i,t)}$ reflect persistent differences in product performance rather than temporary transition effects. Together, switchers provide within-patient changes in manufacturer exposure, while stayers help identify common refill-related and calendar-time dynamics within the market.

Although forced switching among continuing users provides a clean setting in which to estimate persistent manufacturer effects, it may be less informative about margins that operate through treatment continuation. If lower-quality products lead patients to discontinue therapy, switch to a different molecule, or adjust dose strength shortly after initiation, those responses mechanically reduce the likelihood that affected patients remain on treatment long enough to experience subsequent firm switches within the same market. In this case, the forced-switcher sample is selected toward "survivors" who tolerate or respond to the drug sufficiently to continue refilling, and estimated manufacturer effects would understate quality differences that primarily manifest through early discontinuation or rapid medication adjustment. This concern is particularly salient for antidepressants, where early side effects and tolerability often shape adherence and treatment trajectories. For this reason, the initiation design described above provides an important complement to the switching analysis.

Taken together, the two designs allow us to characterize the profile of manufacturer quality over the course of treatment. The switching analysis isolates persistent firm differences among continuing users, while the initiation analysis captures early-treatment responses and extensive-margin effects before survival into continued therapy is determined. If manufac-

[15] The week immediately prior to the refill event ($k = -1$) is the omitted period. Normalizing to $k = -1$ is convenient because it anchors event time to the patient's pre-refill utilization level while avoiding mechanical spikes in the refill week itself.

turer quality differences primarily reflect persistent differences in efficacy or tolerability, we would expect qualitatively similar patterns in downstream outcomes across both designs. In contrast, divergence between initiation and switching estimates would indicate that quality differences operate through dynamic selection or duration-dependent mechanisms, such as early discontinuation or short-run adjustment costs.

4.3 Identification

Equations (1)–(3) identify manufacturer effects using variation in the generic manufacturer dispensed across pharmacies and over time within a given drug market. In the initiation designs, identification comes from differences in the manufacturer dispensed when treatment begins. In the switching design, identification comes from within-patient changes in the manufacturer dispensed at refill when a pharmacy's inventory rotates to a new supplier. Conditional on calendar-time fixed effects, relative-time controls, and patient fixed effects in equations (1) and (3), or on a rich set of patient covariates in equation (2), identification requires that manufacturer assignment be orthogonal to unobserved determinants of patient outcomes.

Formally, the identifying assumptions underlying equations (1)–(3) can be written as:

$$\begin{aligned} \text{Initiation with patient fixed effects:} \quad & E\left[\epsilon_{it} \mid M(i), \alpha_i, \tau_{my(t)}, t - t_{i0}\right] = 0, \\ \text{Initiation with observed controls:} \quad & E\left[\epsilon_i \mid M(i), \tau_{my(i)}, X_i\right] = 0, \\ \text{Switching:} \quad & E\left[\epsilon_{it} \mid M(i,t), \alpha_i, \tau_{my(t)}, t - t_{is}\right] = 0. \end{aligned}$$

Intuitively, the key distinction across the three designs is the extent to which they rely on within-patient rather than across-patient variation. Equations (1) and (3) are insulated from many standard forms of patient sorting because they include patient fixed effects, while equation (2) instead relies on selection on observables. More broadly, our setting differs from many AKM-style applications in labor and health economics, where identification relies on individuals moving across firms or insurers (Song et al., 2019; Finkelstein et al., 2021; Handel et al., 2025). In those settings, the mobility generating identifying variation may itself respond to shocks to outcomes. In our setting, by contrast, patients typically neither observe nor can request a specific generic manufacturer, and physicians prescribe only the drug molecule, strength, and dosage form, not the firm. As outlined in Section 2.1, firm assignment is instead largely determined by pharmacy procurement and inventory management. Several potential threats to identification are nevertheless worth discussing.

A potential concern is that manufacturer assignment may be correlated with patient risk

through both demand-side and supply-side forces at the pharmacy level. On the demand side, patients with different underlying health risks may sort across pharmacies that tend to carry different generic manufacturers, including in response to time-varying shocks that affect when and where they fill prescriptions. On the supply side, pharmacies may systematically prefer certain manufacturers—for example, if they perceive some suppliers as higher quality or less likely to experience supply disruptions (Ishitani, 2026). If patients with systematically different health profiles fill prescriptions at pharmacies carrying different manufacturers, or if pharmacies serving different patient populations maintain different procurement relationships, comparisons across patients could confound manufacturer effects with differences in patient composition across pharmacies.

These concerns are most directly relevant for the cross-sectional initiation specification in equation (2). They are mitigated in equations (1) and (3), which include patient fixed effects and therefore absorb time-invariant differences in underlying health, preferences, treatment intensity, and pharmacy assignment. However, they remain a concern for equation (1) if time-varying health shocks affect where patients initiate treatment and are correlated with the manufacturer they receive at initiation. In equation (3), these concerns are further reduced because identification comes from within-patient changes in the manufacturer dispensed during ongoing therapy within the same patient–pharmacy pair.[16] Because the pharmacy is fixed, identification does not rely on comparisons across pharmacies, and persistent pharmacy characteristics—including procurement preferences and patient mix—are absorbed by the patient fixed effects. Manufacturer effects are instead identified from inventory-driven substitutions within a given patient–pharmacy relationship.

Finally, manufacturer quality may affect some patients more than others—for example, if certain individuals are more sensitive to formulation differences or inactive ingredients (Reker et al., 2019). Such heterogeneity does not invalidate the design. Rather, our estimates identify average manufacturer effects in the relevant analysis samples: in the initiation designs, they capture the average effect of receiving a given manufacturer at treatment initiation, while in the switching design, they capture the average effect among patients who remain on treatment long enough to experience inventory-driven switches. If quality differences affect only a subset of patients, the estimated manufacturer effects reflect both the magnitude of those effects for affected patients and the share of patients for whom they occur.

[16] A related concern is that short-run changes in patient health could coincide with refill events in ways that are correlated with manufacturer assignment. For example, worsening symptoms could affect refill timing or other treatment behavior. This concern is mitigated by the inclusion of flexible relative-time controls around refill events, so identification comes from deviations from the average outcome path around renewal rather than from routine fluctuations in utilization associated with prescription refilling.

4.4 Inference

Our empirical designs raise three distinct inference considerations: (i) statistical inference for manufacturer fixed effects for a given market and outcome, (ii) multiple hypothesis testing across outcomes within a market, and (iii) multiple hypothesis testing across markets.

4.4.1 Within-market inference

For each market, we estimate manufacturer fixed effects $\Psi_{M(i,t)}$ or $\Psi_{M(i)}$ using within- or across-patient variation induced by pharmacy-level procurement shifts. We test the null hypothesis that manufacturer effects are jointly equal using an F-test of

$$H_0 : \Psi_1 = \Psi_2 = \cdots = \Psi_J,$$

where J denotes the number of manufacturers active in the market.

Because manufacturer switching is driven by pharmacy-level inventory changes, and outcomes may be correlated among patients served by the same pharmacy, we cluster standard errors at the pharmacy level in all specifications. This allows for arbitrary correlation in residuals within pharmacy over time and provides conservative inference given that identification arises from within-pharmacy manufacturer variation.

4.4.2 Multiple testing within markets

Within each market m, we examine multiple outcomes, including total spending, emergency room utilization, discontinuation, and dose changes. Because these outcomes are correlated and may capture related dimensions of treatment response, treating them as independent tests would overstate the effective number of hypotheses.

To address this, we proceed in two ways. First, we report outcome-specific tests of manufacturer equality within market and adjust for multiple testing across outcomes by controlling the false discovery rate. Second, we construct a summary index that aggregates information across outcomes within a market, allowing us to assess whether manufacturers differ along a common dimension of treatment response while reducing the burden of multiple inference.

4.4.3 Multiple testing across markets

We estimate manufacturer dispersion separately in dozens of markets. Even if manufacturer effects were equal in every market, each hypothesis test conducted at the 5 percent level would reject with probability 0.05 purely by chance. Let R denote the number of rejected

null hypotheses across markets. Under the global null that manufacturer effects are equal in all markets,

$$R \sim \text{Binomial}(45, 0.05).$$

The expected number of rejections is therefore

$$E[R] = 45 \times 0.05 = 2.25,$$

with variance

$$\text{Var}(R) = 45 \times 0.05 \times 0.95 = 2.1375.$$

We compare the observed number of rejections to this distribution and report the probability of observing at least as many rejections as in the data under the global null, $\Pr(R \geq r^{\text{obs}})$, where r^{obs} denotes the observed number of significant markets. Observing substantially more rejections than predicted by this distribution provides evidence that manufacturer-level dispersion exceeds what would arise from sampling variation alone.

We next ask which specific markets exhibit evidence of manufacturer dispersion. Because this requires simultaneous inference across many hypothesis tests, we adjust for multiple testing to control the family-wise error rate. This ensures that the probability of falsely classifying one or more markets as exhibiting significant manufacturer dispersion remains bounded at the chosen significance level.

5 Results

5.1 Treatment initiation

Figure 3 summarizes evidence from the initiation design on differential manufacturer effects on health care utilization. As outlined in Section 4.1, we estimate equation (1) separately for each market, comparing outcomes within a patient before and after treatment initiation while allowing the post-initiation change to vary by the manufacturer received at initiation. The specification includes patient fixed effects, flexible relative-time controls around initiation, and month–by–year fixed effects, so the manufacturer effects capture differential changes in outcomes after treatment begins relative to the same patient's pre-initiation baseline, net of common utilization dynamics and calendar-time shocks. For each market and outcome, we test the null hypothesis that the manufacturer effects are jointly equal using an F-test and plot the empirical cumulative distribution function (eCDF) of the resulting p-values.

The figure shows substantial evidence of manufacturer differences for several outcomes. For the indicator for any medical or pharmaceutical spending (Figure 3(a)), the eCDF lies

well above the 45-degree line, indicating far more small p-values than would be expected by chance in the absence of manufacturer differences: 33 percent of markets reject at the 1 percent level and 58 percent reject at the 10 percent level. There is also clear, though less pronounced, excess mass near zero for total medical and pharmaceutical spending (Figure 3(b)), with 7 percent of markets rejecting at the 1 percent level and 29 percent at the 10 percent level. By contrast, the eCDF for any emergency department (ED) visit or hospital admission (Figure 3(c)) lies much closer to the 45-degree line, suggesting limited evidence that manufacturers systematically differ in their effects on acute hospital-based utilization. Mental health evaluations (Figure 3(d)) again show substantial excess mass near zero, with 22 percent of markets rejecting at the 1 percent level and 53 percent at the 10 percent level. Taken together, these patterns suggest that manufacturer differences are common for several dimensions of treatment response, particularly extensive-margin utilization and mental health care, but are less pervasive for acute hospital-based outcomes.

A natural extension of this analysis is to examine whether manufacturer differences also manifest in early treatment adherence—specifically, whether patients assigned to different manufacturers at initiation differ in their likelihood of picking up a second fill within six weeks. Unlike the panel outcomes examined above, however, adherence measures are defined only once following the initial fill and do not permit a patient fixed effects specification. We therefore estimate adherence outcomes using the observables specification, which replaces patient fixed effects with the rich set of baseline controls described in Section 4.1.

To assess whether this substitution materially affects the results, we compare the patient fixed effects and observables specifications for the panel outcomes where both are feasible. Two findings from this comparison support the validity of the observables approach for adherence. First, Appendix Figure A1 shows that the empirical CDFs are broadly similar across all five specifications for any spending and MH evaluation—the two outcomes with the strongest evidence of manufacturer dispersion—and that adding successive layers of patient controls generally shifts the eCDF closer to the diagonal rather than further from it, indicating that richer controls produce more conservative rather than more permissive inference. Second, Appendix Figure A2 shows that the max-gap statistics from the richest observables specification and the patient fixed effects specification are highly correlated across markets. Moreover, for any spending and MH evaluation, the dots tend to cluster on or below the 45-degree line, indicating that the observables specification produces max-gap estimates that are at typically smaller than the patient fixed effects estimates. We therefore interpret the adherence results that follow as conservative estimates of the true manufacturer differences in early treatment continuation.

Figure 4 presents evidence on manufacturer differences in early treatment adherence.

As described in Section 3.4 and Table 3, the six-week window following a 30-day initial prescription is dominated by two outcomes: on-time refill and abandonment together account for approximately 80 percent of initiators across markets, with dosage or form adjustment and molecule switch each occurring in fewer than 12 percent of cases.

For on-time refill (panel a) and abandonment (panel b), the empirical CDFs lie well above the diagonal in both the full initiator sample and the MH-diagnosed subsample, indicating far more small p-values than would be expected under the null of no manufacturer differences. The evidence is particularly strong for abandonment: 70 percent of markets reject the null at the 1 percent level and 82 percent at the 5 percent level among all initiators, with nearly identical rejection rates in the MH-diagnosed subsample (67 and 79 percent respectively).

For on-time refill, rejection rates are somewhat lower but still substantial, with 40 percent of markets rejecting at the 1 percent level and 50 percent at the 5 percent level among all initiators. By contrast, the eCDFs for dosage or form adjustment (panel c) and molecule switch (panel d) track much closer to the diagonal, with limited evidence of systematic manufacturer dispersion: only 10 percent of markets reject at the 5 percent level for dosage adjustment and 15 percent for molecule switch. Given the six-week window, most prescriber-initiated modifications have not yet had the opportunity to occur, and these outcomes are therefore unlikely to be sensitive to manufacturer quality differences even if such differences exist. The absence of manufacturer dispersion for dosage and molecule adjustment is therefore uninformative about quality, reflecting the clinical timeline rather than the absence of underlying quality differences.

Moreover, the two samples—all initiators and initiators with a mental health diagnosis on record—tell broadly similar stories, suggesting that the pattern of manufacturer dispersion is not an artifact of sample composition or off-label use. The most notable divergence between samples occurs in panel (a) for on-time refill, where the full initiator sample exhibits somewhat stronger excess mass near zero relative to the MH-diagnosed subsample, with rejection rates approximately 14 percentage points higher at both the 1 and 5 percent levels. For abandonment (panel b), the two samples are nearly indistinguishable.

Abandonment represents complete disengagement from treatment, an extreme enough response that it occurs at similar rates across both populations regardless of clinical attachment. On-time refill is a more discretionary margin: among patients embedded in ongoing mental health care, prescriber monitoring may sustain prompt refill behavior even when the product experience is suboptimal, attenuating manufacturer differences on this margin but not on the dropout margin. That quality differences are harder to detect among high-need patients does not mean they are absent—it reflects that clinical inertia sustains treatment continuation regardless of manufacturer quality, making the data less informative precisely

where the welfare stakes may be highest. A patient who abandons a suboptimal product at least escapes continued exposure to it; a patient who remains on treatment due to prescriber inertia bears the full cost of any quality shortfall.

In this vein, Figure 4 characterizes the magnitude of these differences among markets with statistically significant manufacturer dispersion, providing a sense of the stakes involved. The relative max-gap—the difference between the highest- and lowest-performing manufacturer expressed as a share of the market's mean outcome rate—varies substantially across markets, reflecting meaningful heterogeneity in the scope of manufacturer quality differences.

For on-time refill, among the 12 markets rejecting at the 5 percent level, the relative max-gap ranges from 7 to 15 percent of the baseline refill rate, with a mean of 10 percent. At a typical baseline on-time refill rate of 48 percent, this corresponds to an absolute difference of roughly 3 to 6 percentage points between the best and worst manufacturer in the same market. For abandonment, the heterogeneity is considerably larger: among the 26 rejecting markets, the relative max-gap ranges from 6 to 44 percent of the baseline abandonment rate, with a mean of 21 percent. Baseline abandonment rates in these markets average 31 percent and range from 18 to 44 percent, so the implied absolute differences span from roughly 2 to 12 percentage points, meaning that in the most extreme markets, a patient's probability of never picking up a second fill varies by more than 12 percentage points depending solely on which manufacturer their pharmacy happened to stock.

The distribution of baseline abandonment rates across rejecting markets is also informative in light of the earlier finding that the MH-diagnosed and full initiator samples show similar rejection rates for abandonment. Markets where abandonment is highest are precisely the markets most likely to include lower-attachment patients for whom treatment continuation is more discretionary. That these markets show the largest relative max-gaps is consistent with the demand elasticity interpretation: quality differences bite hardest where attachment is weakest, and the markets with the most elastic patient populations exhibit the widest manufacturer gaps. The similarity in rejection rates between the two samples for abandonment thus does not conflict with this interpretation: it reflects the fact that abandonment, as an extreme dropout response, is sensitive to manufacturer quality across both high- and low-attachment populations, while the magnitude of the gap is largest in markets dominated by lower-attachment patients.

5.2 Ongoing treatment

Figure 6 provides a proof of concept for the switching design by focusing on the bupropion XL 300 mg market, where product-quality problems are now well known. In October

2012, the FDA concluded that the Teva-marketed version of bupropion XL 300 mg was not therapeutically equivalent to the branded product, and a broader FDA review completed in October 2013 further found that the Watson-marketed product also failed to meet bioequivalence standards. This market therefore provides a useful setting in which to assess whether the switching design detects manufacturer-specific differences in patient outcomes in a case where such differences are known to have existed.

To do so, we estimate equation (3) in the bupropion XL 300 mg market and plot the resulting manufacturer fixed effects, $\Psi_{M(i,t)}$, for each firm. As outlined in Section 4.2, the switching specification compares outcomes within a patient over time as the dispensed manufacturer changes during ongoing therapy, while controlling for patient fixed effects, flexible relative-time effects around refill events, and month–by–year fixed effects. The estimated manufacturer effects therefore capture differences in outcomes across firms within this market, net of time-invariant patient heterogeneity, common calendar-time shocks, and common utilization dynamics around refill. Figure Figure 6 reports these estimated firm effects for the same four weekly outcomes considered above.

The figure shows that patients switched to the product marketed by Teva experienced sharply elevated adverse outcomes, particularly for ED visits or hospital admissions (Figure Figure 6(c)) and for mental health evaluations (Figure 6(d)). The estimated Teva fixed effect implies that patients switched to the Teva product were 0.5 percentage points more likely to experience an ED visit or hospital admission in a given week, relative to a weekly mean of 0.6 percentage points, and 0.4 percentage points more likely to receive a mental health evaluation, relative to a weekly mean of 0.3 percentage points. These are economically large effects. Watson shows somewhat elevated spending outcomes in Figures 6(a) and 6(b), but the effects are small in magnitude. The p-values on the market-level F-tests are around 0.05 for each outcome, consistent with detectable manufacturer differences in a market known to have bioequivalence failures. The bupropion evidence therefore provides a useful proof of concept for the switching design: in a setting where quality problems are known to have occurred, the design recovers economically meaningful manufacturer-specific differences in patient outcomes.

5.3 Policy implications

The results above establish that manufacturer quality differences are widespread, economically meaningful, and detectable using patient-level claims data. We now turn to two sets of policy implications: how our outcome-based quality measures relate to existing regulatory signals, and how a regulator with access only to claims data could implement a feasible

monitoring approach. [*Note: The empirical implementation in this subsection is ongoing; hence, we present the conceptual framework and planned analyses here.*]

5.3.1 Comparison to existing quality metrics

Existing assessments of generic drug manufacturer quality rely primarily on three observable signals: product recalls, reported shortages, and regulatory inspection outcomes. Recalls and inspection findings reflect compliance with manufacturing standards while shortages capture disruptions in supply. However, as discussed above, these measures are indirect proxies for realized product performance and may reflect monitoring intensity, enforcement targeting, or strategic supply decisions in addition to underlying quality.

Having constructed outcome-based manufacturer quality measures using health claims data, we next examine how closely these align with existing regulatory signals. Conceptually, if recalls, shortages, and inspection outcomes are informative proxies for true product quality, manufacturers identified as lower quality in our outcome-based analysis should also exhibit worse performance on these traditional metrics. Conversely, weak correspondence would suggest that existing signals provide limited information about realized patient outcomes within therapeutically equivalent markets.

To assess alignment, we plan to link our manufacturer-level quality estimates to publicly available data on recall activity, shortage history, and inspection findings, and examine cross-sectional correlations between our estimated manufacturer effects and each regulatory metric. We will also evaluate whether manufacturers flagged as lower quality in our outcome-based ranking are disproportionately represented among firms with prior recalls, documented shortages, or adverse inspection outcomes. Because regulatory metrics are measured at the firm or facility level rather than at the molecule–strength–form market level, we will aggregate our manufacturer effects to the corresponding firm level where necessary, weighting by patient exposure within markets.

5.4 Utility for policymakers

Consider a regulator who only observes longitudinal health claims outcomes and NDCs on pharmacy claims, allowing each fill to be mapped to an ANDA holder, but who does not observe pharmacy identifiers or procurement regimes. Without pharmacy information, the regulator cannot condition on the inventory process that drives quasi-experimental switching across manufacturers and therefore cannot implement the within-pharmacy, within-patient design that underlies our baseline AKM estimates.

In this information environment, the regulator's natural approach is to compare outcomes

across manufacturers using only claims data. A simple implementation is to estimate, within each molecule–strength–dosage form market m, a "claims-only" model of the form

$$y_{it} = \sum_{j \in \mathcal{J}_m} \tilde{\Psi}_{jm} \, \mathbf{1}\{M(i,t) = j\} + \tau_{my(t)} + u_{it}, \tag{4}$$

possibly augmenting (4) with observable risk adjusters X_{it}, e.g., age, diagnosis, lagged spending, etc. The key policy question is: how far can such claims-only comparisons recover the manufacturer quality rankings identified by our procurement-based design, and where do they fail?

The conceptual issue is that manufacturer assignment is not random in claims-only data. Even if patients do not directly choose manufacturers, they choose pharmacies, plans, and refill timing, and these choices affect which manufacturer they receive. Moreover, if lower-quality products induce discontinuation, dose adjustment, or rapid switching to alternative therapies, then outcomes among continuing users may be selected. Both channels can cause a claims-only estimator to mis-rank manufacturers or attenuate true quality dispersion.

We formalize these concerns using a simplified version of the forced-switcher AKM outcome model within market m from Section 4.2 (*reprinted below*):

$$y_{it} = \alpha_i + \Psi_{M(i,t)} + \tau_{my(t)} + \varepsilon_{it}, \tag{5}$$

where α_i captures time-invariant patient health and utilization propensity and Ψ_M captures persistent manufacturer quality differences.[17] Our baseline design restricts attention to forced switches within the same pharmacy location, so that Ψ_M is identified from within-patient changes in manufacturer induced by pharmacy inventory shifts.

To see why claims-only comparisons can be biased, consider the population conditional expectation implied by Equation 5:

$$E[y_{it} \mid M(i,t) = j, my(t)] = \tau_{my(t)} + \Psi_j + E[\alpha_i \mid M(i,t) = j, my(t)]. \tag{6}$$

Thus, relative to the causal object Ψ_j, the claims-only estimand in (4) absorbs a selection term reflecting differences in baseline patient health across manufacturers. Abstracting from sampling variation and assuming correct time controls, we can write

$$\tilde{\Psi}_j \approx \Psi_j + \left(E[\alpha_i \mid M(i,t) = j] - E[\alpha_i]\right), \tag{7}$$

so that the discrepancy between claims-only and procurement-based estimates is driven by

[17] For brevity, we omit the relative-time fixed effects.

sorting of patients into manufacturers.

This selection component is directly related to the covariance between patient baseline health and manufacturer exposure. Let

$$D_{ijt} \equiv \mathbf{1}\{M(i,t) = j\}$$

denote exposure to manufacturer j at time t. Then

$$\text{Cov}(\alpha_i, D_{ijt}) = E[\alpha_i \mid D_{ijt} = 1] \Pr(D_{ijt} = 1) - E[\alpha_i] \Pr(D_{ijt} = 1), \tag{8}$$

so that $\text{Cov}(\alpha_i, D_{ijt})$ provides a natural diagnostic for the magnitude of sorting embedded in claims-only comparisons. When this covariance is close to zero for all j, cross-sectional claims-based manufacturer differences are less likely to be distorted by patient composition; when it is large, naive rankings may be substantially misleading.

We will use our baseline AKM estimates $\hat{\Psi}_j^{AKM}$ as a benchmark for manufacturer quality and evaluate how closely regulator-feasible claims-only approaches approximate these rankings. Specifically, within each market we construct three alternative estimators: (i) raw outcome differences by manufacturer, (ii) risk-adjusted differences controlling for patient observables and month-by-year fixed effects, and (iii) within-patient fixed effects estimates that do not condition on pharmacy-level procurement regimes. This comparison will quantify the extent to which existing monitoring approaches, which necessarily rely on claims data without pharmacy procurement information, can recover the true distribution of manufacturer quality, and where they systematically fail.

6 Conclusion

The generic drug market is predicated on the assumption that branded and generic products are therapeutically equivalent and therefore interchangeable. Under this view, differences across generic manufacturers should not translate into meaningful differences in patient outcomes. However, a growing body of evidence, including regulatory actions and post-market surveillance, suggests that this assumption may not always hold in practice (Eban, 2019).

In this paper, we develop and implement empirical approaches that use administrative claims data to measure variation in drug quality across generic manufacturers. By leveraging quasi-experimental variation in firm exposure generated by pharmacy procurement and inventory dynamics, we identify manufacturer-specific effects on a range of clinically relevant outcomes. The initiation design allows us to capture early treatment responses,

including adherence, discontinuation, and other short-run adjustments, while the switching design isolates persistent differences in outcomes within patients over time.

Our results provide evidence of economically meaningful quality differences across generic drug manufacturers in the antidepressant market. In the initiation design, we document widespread rejection of equality across manufacturers for several measures of health care utilization, particularly for outcomes reflecting treatment stability and mental health care use. In the switching design, we show that within-patient changes in manufacturer exposure similarly lead to systematic differences in patient outcomes.

These findings have several implications. First, they suggest that current regulatory standards and post-market monitoring may not fully capture clinically relevant variation in product quality. Second, they highlight the importance of supply-side factors—such as manufacturing practices and procurement relationships—in shaping patient outcomes, even in settings where products are formally deemed interchangeable. Finally, they point to the value of administrative claims data as a tool for real-time monitoring of drug quality at scale. Ongoing work builds on our preliminary findings to better understand the mechanisms underlying these quality differences, including the roles of market forces, regulatory oversight, and supply chain dynamics, and to assess the extent to which similar patterns arise in other therapeutic classes.

References

Abowd, J., F. Kramarz, and D. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Alatawi, Y. and R. Hansen, "Empirical estimation of under-reporting in adverse event reporting systems: a systematic review," *Drug Safety*, 2017, *40* (9), 761–767.

Altonji, J., T. Elder, and C. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

American Psychiatric Association, "Practice guideline for the treatment of patients with major depressive disorder," *American Journal of Psychiatry*, 2010, *167*, 1.

Associated Press, "FDA Says Generic Wellbutrin OK Despite Reports," *NBC News*, April 2008.

Bate, A. and S. Evans, "Quantitative signal detection using spontaneous ADR reporting," *Pharmacoepidemiology and Drug Safety*, 2009, *18* (6), 427–436.

Bate, R., G. Z. Jin, and A. Mathur, "Does Price Reveal Poor-Quality Drugs? Evidence from 17 Countries," *Journal of Health Economics*, 2011, *30* (6), 1150–1163.

Boyce, Philip, Malcolm Hopwood, Grace Morris, Amber Hamilton, Darryl Bassett, Bernhard T Baune, Roger Mulder, Richard Porter, Gordon Parker, Ajeet B Singh et al., "Switching antidepressants in the treatment of major depression: When, how and what to switch to?," *Journal of affective disorders*, 2020, *261*, 160–163.

Brody, D. and Q. Gu, "Antidepressant Use Among Adults: United States, 2015–2018," Technical Report Data Brief No. 377, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD 2020.

Conti, R. and M. Wosińska, "The economics of generic drug shortages: The limits of competition," *Journal of Economic Perspectives*, 2025, *39* (2), 79–102.

Cuddy, E., "Competition and collusion in the generic drug market," *PhD dissertation, Princeton University*, 2020.

_ , Y.P. Lu, and D. Ridley, "FDA Global Drug Inspections: Surveillance Of Manufacturing Establishments Remains Well Below Pre-COVID-19 Levels: Study examines FDA

surveillance inspections of manufacturing establishments around the world," *Health Affairs*, 2023, *42* (12), 1758–1766.

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

Finkelstein, A., M. Gentzkow, and H. Williams, "Place-based drivers of mortality: Evidence from migration," *American Economic Review*, 2021, *111* (8), 2697–2735.

Ghijs, S., E. Wynendaele, and B. De Spiegeleer, "The continuing challenge of drug recalls: Insights from a ten-year FDA data analysis," *Journal of Pharmaceutical and Biomedical Analysis*, 2024, *249*, 116349.

Handel, B., J. Holmes, J. Kolstad, and K. Lavetti, "The Role of Insurers in Health Care Spending and Production: Evidence from Utah," Technical Report, National Bureau of Economic Research 2025.

Hazell, L. and S. Shakir, "Under-reporting of adverse drug reactions: a systematic review," *Drug Safety*, 2006, *29* (5), 385–396.

Ishitani, C., "In Defense of the Middleman: Quality Failures in the Generic Pharmaceutical Market," Working Paper, Yale University 2026.

Oster, E., "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Qaseem, A., D. Owens, I. Etxeandia-Ikobaltzeta, J. Tufte, J.T. Cross Jr, T. Wilt, Clinical Guidelines Committee of the American College of Physicians, C. Crandall, E. Balk, T. Cooney et al., "Nonpharmacologic and pharmacologic treatments of adults in the acute phase of major depressive disorder: a living clinical guideline from the American College of Physicians," *Annals of internal medicine*, 2023, *176* (2), 239–252.

Reker, D., S. Blum, C. Steiger, K. Anger, J.M. Sommer, J. Fanikos, G. Traverso, and A.S. Kesselheim, "Ïnactiveïngredients in oral medications," *Science Translational Medicine*, 2019, *11* (483), eaau6753.

Song, J., D. Price, F. Guvenen, N. Bloom, and T. Von Wachter, "Firming up inequality," *Quarterly journal of economics*, 2019, *134* (1), 1–50.

U.S. Department of Health and Human Services, Office of Inspector General, "The FDA Has Taken Steps to Improve the Foreign Drug Inspection Program, but Challenges Remain," 2017.

U.S. Food and Drug Administration, "Statistical Approaches to Establishing Bioequivalence," 2001. Guidance for Industry. Updated 2022.

_ , "MAPP 5014.1: Understanding Center for Drug Evaluation and Research's Risk-Based Site Selection Model," Manual of Policies and Procedures, U.S. Food and Drug Administration, Silver Spring, MD 2023.

_ , "Approved Drug Products with Therapeutic Equivalence Evaluations (Orange Book)," Technical Report, U.S. Food and Drug Administration 2024.

_ , "Report on the State of Pharmaceutical Quality: Fiscal Year 2024," Technical Report, U.S. Food and Drug Administration 2025.

U.S. Government Accountability Office, "FDA: Improvements Needed in Foreign Drug Inspection Program," 2016.

7 Tables

Table 1: Characteristics of Antidepressant Markets

	All Antidepressant Markets	Initiator Markets	Forced Switch Markets
	(1)	(2)	(3)
A. Market Size (2011–2019)			
Number of patients per market (thousands)	46.1	176.9	193.2
Total fills per market (thousands)	321.1	1231.6	1355.5
B. Market Competition			
Mean active manufacturers per market-year	3.9	8.3	8.4
Total manufacturers ever observed in market	5.4	11.8	12.0
C. Antidepressant Class			
Selective serotonin reuptake inhibitors (SSRIs) (%)	38	41	43
Serotonin–norepinephrine reuptake inhibitors (SNRIs) (%)	19	18	18
Tricyclic antidepressants (TCAs) (%)	17	18	12
Other antidepressants (%)	25	23	27
D. Formulation Characteristics			
Extended-release formulation (%)	26	23	27
Capsule dosage form (%)	27	30	25
Number of markets	172	44	40

Notes: The table reports characteristics of antidepressant markets observed in the claims data over the period 2011–2019. Column (1) includes all antidepressant markets in which at least one generic manufacturer was active during the sample period. Column (2) restricts to markets included in the initiation analysis. Column (3) restricts to markets included in the forced-switch analysis. Market size and fill counts are constructed from pharmacy claims for all enrollees with continuous medical and pharmacy coverage during the relevant observation window.

Table 2: Descriptive Statistics

	All Insurees on AD	Initiators *(Forced Takers)*	Same Pharmacy Different Manufacturer *(Forced Switchers)*
	(1)	(2)	(3)
A. Insurance Coverage			
Months with coverage (mean)	48	72	74
B. Demographics			
Age (mean)	44.3	42.0	49.7
Female (%)	65.6	66.1	69.3
Number of patients (thousands)	4,230	1,992	611
Number of prescriptions (millions)	58.44	1.99	1.22

The table reports patient-level descriptive statistics for the analytic samples used in the initiation and switching designs. Column (1) includes all insured beneficiaries with at least one antidepressant prescription during the sample period 2011–2019. Column (2) includes treatment initiators—patients whose first observed fill of a given molecule–strength–form follows at least four months without a prior fill in that market—who fill their prescription at their modal pharmacy (forced takers). Column (3) includes patients who refill the same prescription at the same pharmacy but receive a different manufacturer than on their prior fill (forced switchers). Panel A reports insurance coverage characteristics. Panel B reports demographic characteristics. All statistics are measured at the time of the relevant fill event. The number of prescriptions in column (1) reflects all antidepressant fills observed during the sample period; columns (2) and (3) reflect fills within the analytic sample for each design.

Table 3: Drug adherence at six weeks following initiation

	All initiators			MH diagnosis only		
	Mean	P25	P75	Mean	P25	P75
Normal refill	44.7	36.9	51.9	44.8	36.9	52.1
Abandonment	35.5	29.6	41.0	32.0	27.6	37.2
Molecule switch	9.0	6.3	11.9	11.3	7.2	13.9
Dose/form adjustment	7.8	4.2	11.0	9.0	4.5	12.8
Reversal	3.0	2.6	3.4	2.9	2.4	3.3
Total	100.0			100.0		

Notes: The table reports the distribution of treatment status at six weeks following initiation among patients who filled a 30-day initial prescription of a generic antidepressant. Each patient is assigned to exactly one mutually exclusive category based on the last non-reversed fill of the focal molecule within the six-week window, with a priority rule applied when multiple fills share the same date (see Section 3.2 and Figure 2 for details). Statistics are computed at the market (molecule-strength-form) level and then averaged across markets; P25 and P75 denote the 25th and 75th percentiles of the cross-market distribution. Branded switch is omitted as it accounts for fewer than 0.1 percent of initiators in all markets. The MH diagnosis column restricts to initiators with a mental health diagnosis on record in the 120 days preceding treatment initiation.

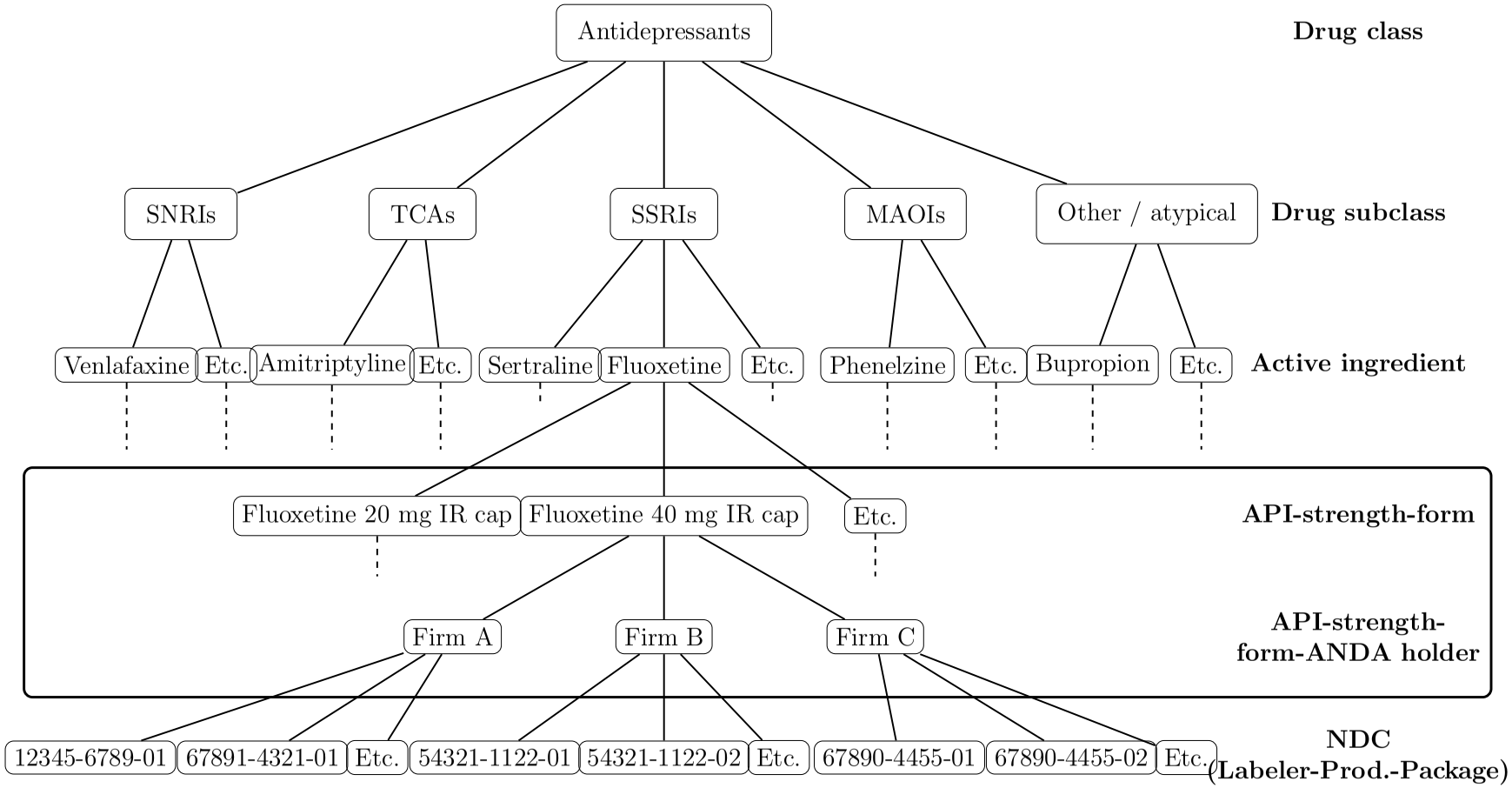
Table 4: Characteristics of Refill Events by Refill Type

	Refill Classification			
	Same Pharmacy Same Manufacturer	Different Pharmacy Same Manufacturer	Different Pharmacy Different Manufacturer	Same Pharmacy Different Manufacturer *(Forced Switch)*
	(1)	(2)	(3)	(4)
A. Refill Timing and Event Characteristics				
Days supply (mean)	78.8	82.5	82.1	79.2
Days since previous fill (mean)	84.1	81.1	82.5	92.9
Refill before prior supply exhausted (%)	52.3	56.5	61.8	51.8
Refill after gap in therapy (%)	2.2	3.2	5.5	4.5
B. Treatment Characteristics				
Branded drug (%)	9.8	16.9	5.7	4.6
C. Refill History				
Days on current drug at refill date	719.9	638.7	564.1	699.0
Number of prior fills for current drug	10.7	9.4	8.6	10.1
Number of manufacturer versions used previously	1.9	1.8	2.2	2.2
D. Cost and Formulary Status				
Patient out-of-pocket cost ($)	32.3	41.0	31.3	26.9
Total allowable cost ($)	109.4	151.1	108.1	96.5
Number of refill events	1,917,529	718,222	856,829	901,263

Notes: The table reports refill-level summary statistics separately by refill type for patients refilling antidepressant prescriptions. Column (1) includes refills at the same pharmacy from the same manufacturer as the prior fill (forced stayers). Columns (2) and (3) include refills at a different pharmacy from the same or different manufacturer, respectively. Column (4) includes refills at the same pharmacy from a different manufacturer (forced switchers), who form the primary analytic sample for the switching design. Panel A reports characteristics of the refill event itself, including days supply, days elapsed since the prior fill, and indicators for early refills and gaps in therapy. Panel B reports treatment characteristics at the time of the refill. Panel C reports refill history measures cumulated through the refill date. Panel D reports cost and formulary characteristics of the dispensed prescription.

8 Figures

Figure 1: Drug product hierarchy and market definition



Notes: The above figure illustrates the hierarchical structure of drug products used to define markets in the analysis. Drugs are organized from broad therapeutic classes to subclasses, active ingredients, and finally to the market level defined as a specific molecule–strength–dosage form. Within each market, variation arises across ANDA holders, each of which may market products under one or more National Drug Codes (NDCs). The boxed region highlights the level at which markets are defined and the level at which manufacturer variation is measured. Dashed lines indicate omitted categories within each level.

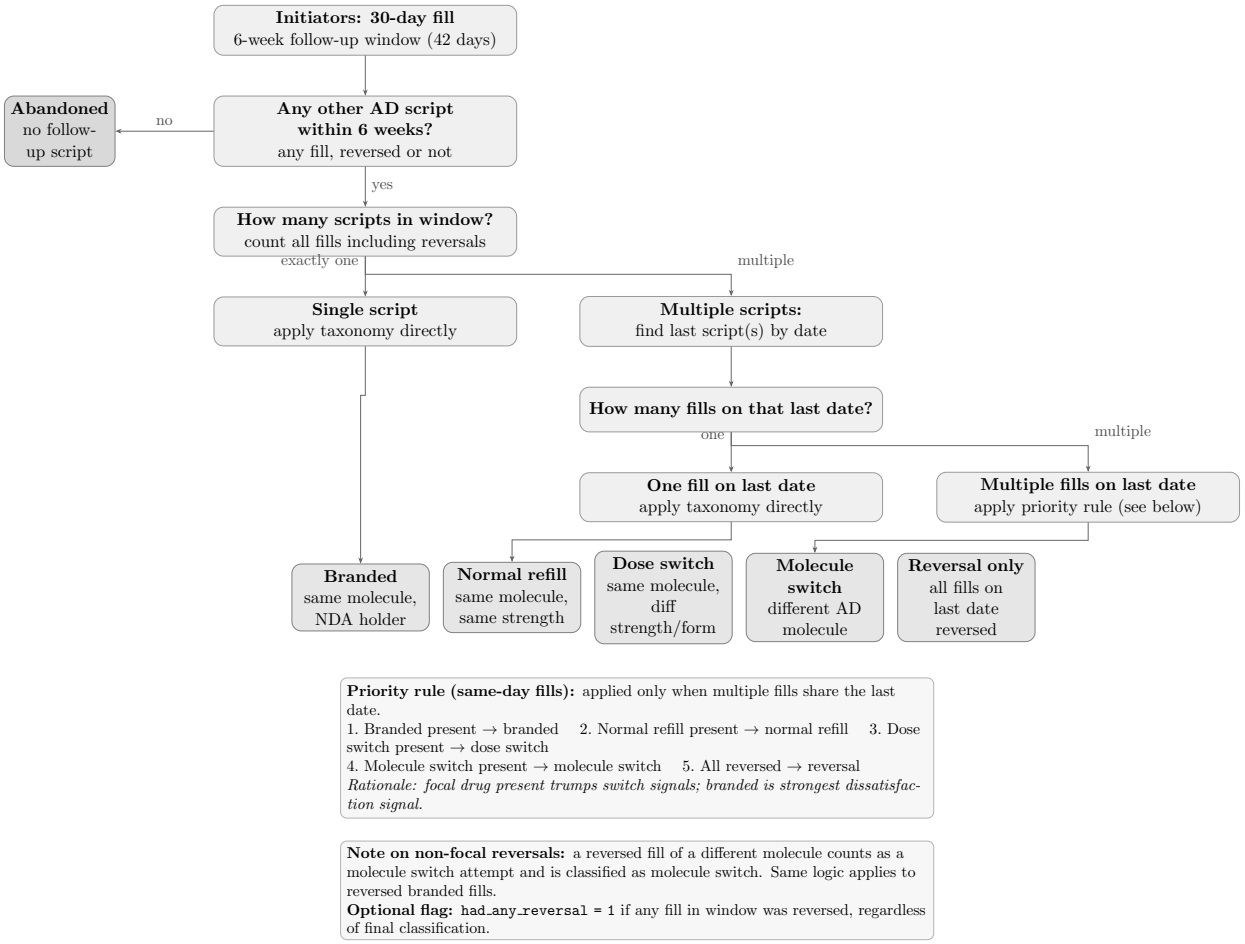
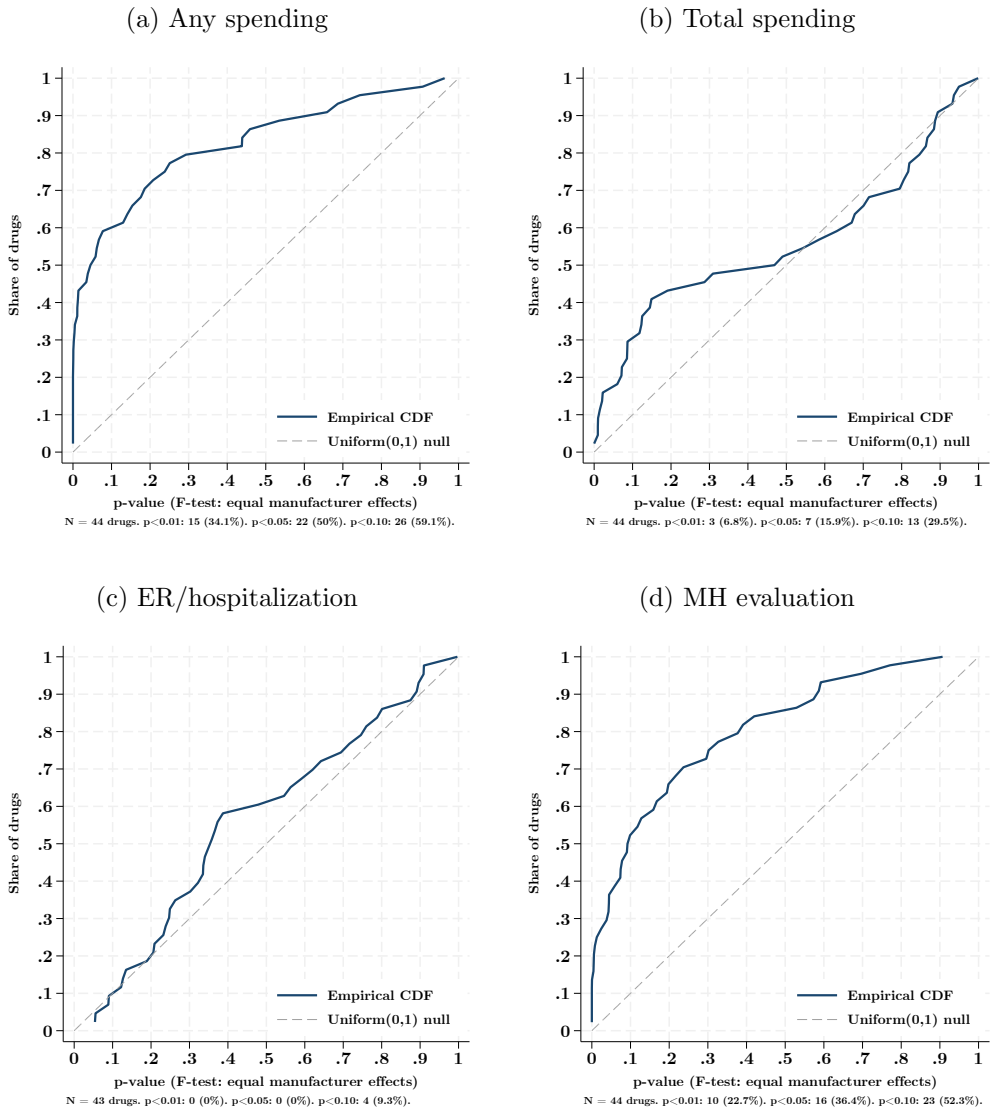


Figure 2: Classification of follow-up scripts within 6-week window

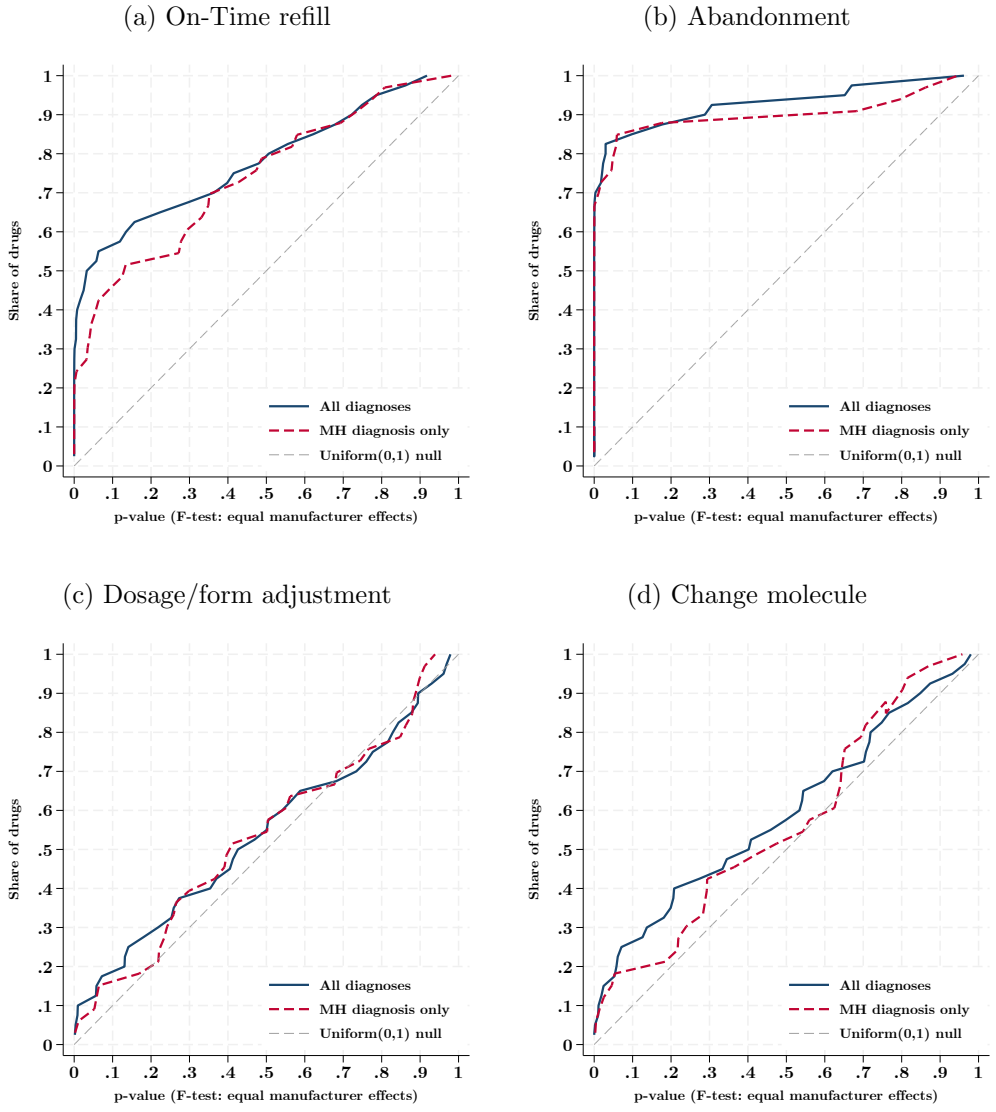
Notes: The figure illustrates the algorithm used to classify treatment status at six weeks following initiation for patients who filled a 30-day initial prescription of a generic antidepressant.

Figure 3: Initiation design: Empirical CDF of p-values from market-level F-tests of equality across manufacturers



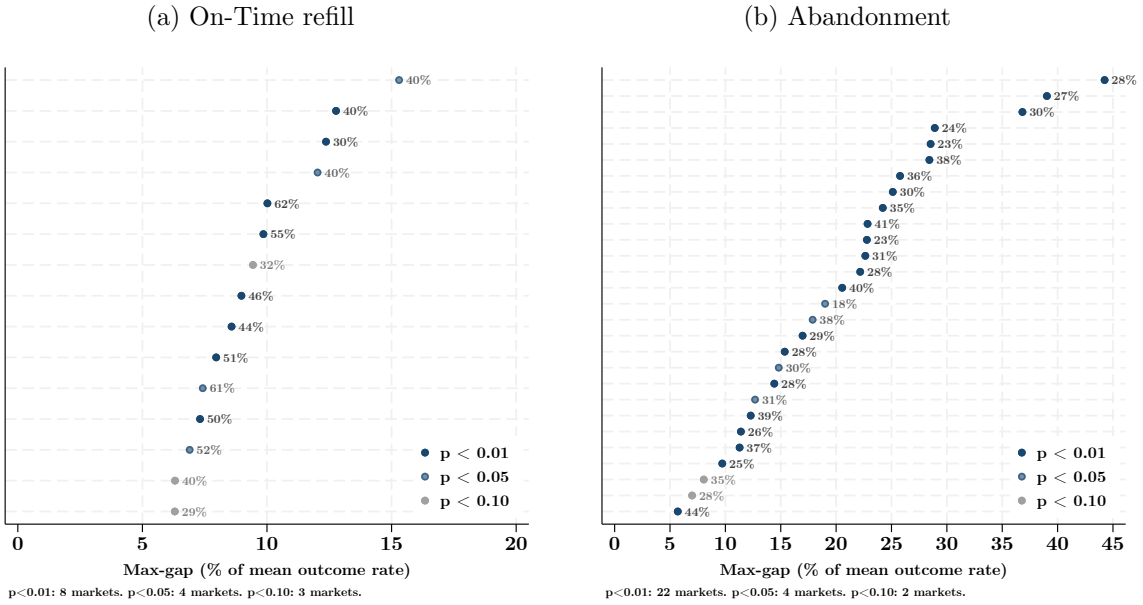
Notes: The above figures plot the empirical cumulative distribution function of p-values from market-level F-tests of the null hypothesis that the manufacturer effects $\Psi_{M(i)}$ are jointly equal within a market, as estimated from equation (1). The outcome variable in subfigure (a) is an indicator for any medical or pharmaceutical spending, excluding pharmaceutical spending on the focal prescription; in subfigure (b), total medical and pharmaceutical spending, again excluding pharmaceutical spending on the focal prescription; in subfigure (c), an indicator for any emergency department (ED) visit or hospital admission; and in subfigure (d), an indicator for a mental health evaluation. All regressions include patient, month–by–year, and relative-time fixed effects; standard errors are clustered by pharmacy.

Figure 4: Drug adherence



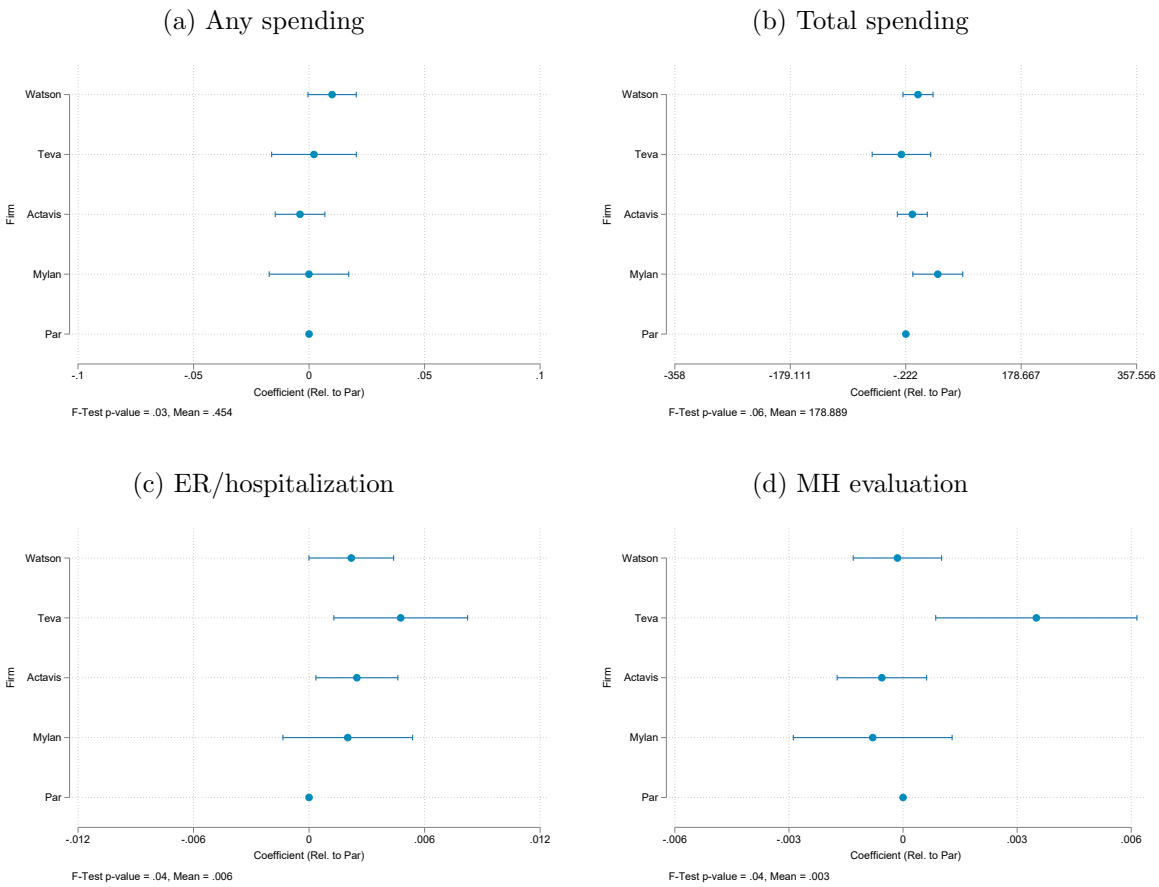
Notes: The above figures plot the empirical cumulative distribution function of p-values from market-level F-tests of the null hypothesis that the manufacturer effects $\Psi_{M(i)}$ are jointly equal within a market, as estimated from equation (2). Panel (a) plots results for on-time refill, panel (b) for abandonment, panel (c) for dosage or form adjustment within the same molecule, and panel (d) for molecule switch. The specification includes patient demographics, general health controls measured over the 90 days prior to initiation, mental health controls, and indicators for the specialty of the prescribing provider while retaining month–by–year fixed effects. Each figure shows two empirical CDFs: one for the full initiator sample (solid navy line) and one restricted to initiators with a mental health diagnosis on record (dashed cranberry line). Standard errors are clustered by pharmacy in all specifications. The dashed diagonal line represents the Uniform(0,1) distribution expected under the null hypothesis of no manufacturer differences.

Figure 5: Manufacturer differences in drug adherence



Notes: The above figures plot the distribution of max-gap statistics across antidepressant markets with statistically significant manufacturer differences, separately for on-time refill (panel a) and abandonment (panel b). Each dot represents one market. The x-axis shows the max-gap—the difference between the highest- and lowest-performing manufacturer fixed effects within a market—expressed as a percentage of the mean outcome rate in that market, estimated from equation (2). Labels next to each dot report the mean outcome rate for that market; multiplying the x-axis value by the label recovers the absolute max-gap in percentage points. Dot color indicates the strength of the F-test rejection: dark navy indicates $p < 0.01$, medium navy indicates $p < 0.05$, and gray indicates $p < 0.10$. The sample is restricted to initiators with a mental health diagnosis on record who filled a 30-day initial prescription. All specifications include patient demographics (age, sex, insurance plan type, and ZIP-3 fixed effects), general health controls measured over the 120 days prior to initiation (deciles of total spending, indicators for any prescription drug spending, any emergency department or hospital visit, and any record of self-harm), mental health controls (indicators for psychiatrist and therapist visits, mental health evaluations, and diagnosis indicators for depression, anxiety, bipolar disorder, OCD, substance use disorder, adjustment disorders, ADHD/autism spectrum disorder, eating disorders, or no mental health diagnosis), and indicators for the specialty of the prescribing provider (primary care physician, therapist, psychiatrist, nurse practitioner or physician assistant, and other medical specialty). Standard errors are clustered by pharmacy.

Figure 6: Switching design: Manufacturer fixed effects for Bupropion XL 300 mg

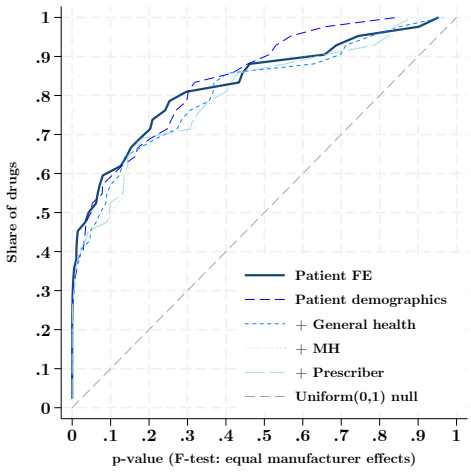


Notes: The above figures plot the firm fixed effects $\Psi_{M(i,t)}$ estimated from equation (3) using patient–week–level data from 2011 to 2014 for the bupropion XL 300 mg market. The outcome variable in subfigure (a) is an indicator for any medical or pharmaceutical spending, excluding pharmaceutical spending on the focal prescription; in subfigure (b), total medical and pharmaceutical spending, again excluding pharmaceutical spending on the focal prescription; in subfigure (c), an indicator for any emergency department (ED) visit or hospital admission; and in subfigure (d), an indicator for a mental health evaluation. All regressions include patient, month–by–year, and relative-time fixed effects; standard errors are clustered by pharmacy. In October 2012, the FDA concluded that the Teva-marketed product was not therapeutically equivalent to the branded drug. A broader FDA review completed in October 2013 further found that the Watson-marketed product also failed to meet bioequivalence standards.

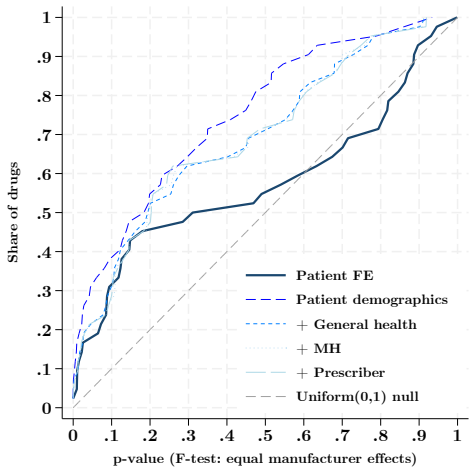
A Appendix Figures and Tables

Figure A1: Robustness to selection on observables: Empirical CDF of p-values from market-level F-tests of equality across manufacturers

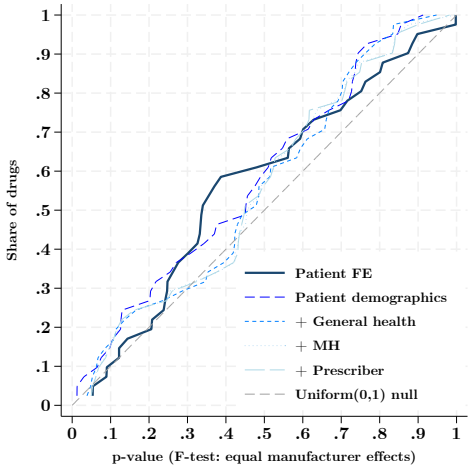
(a) Any spending



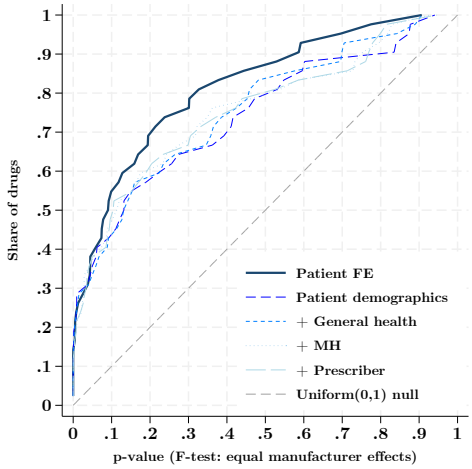
(b) Total spending



(c) ER/hospitalization

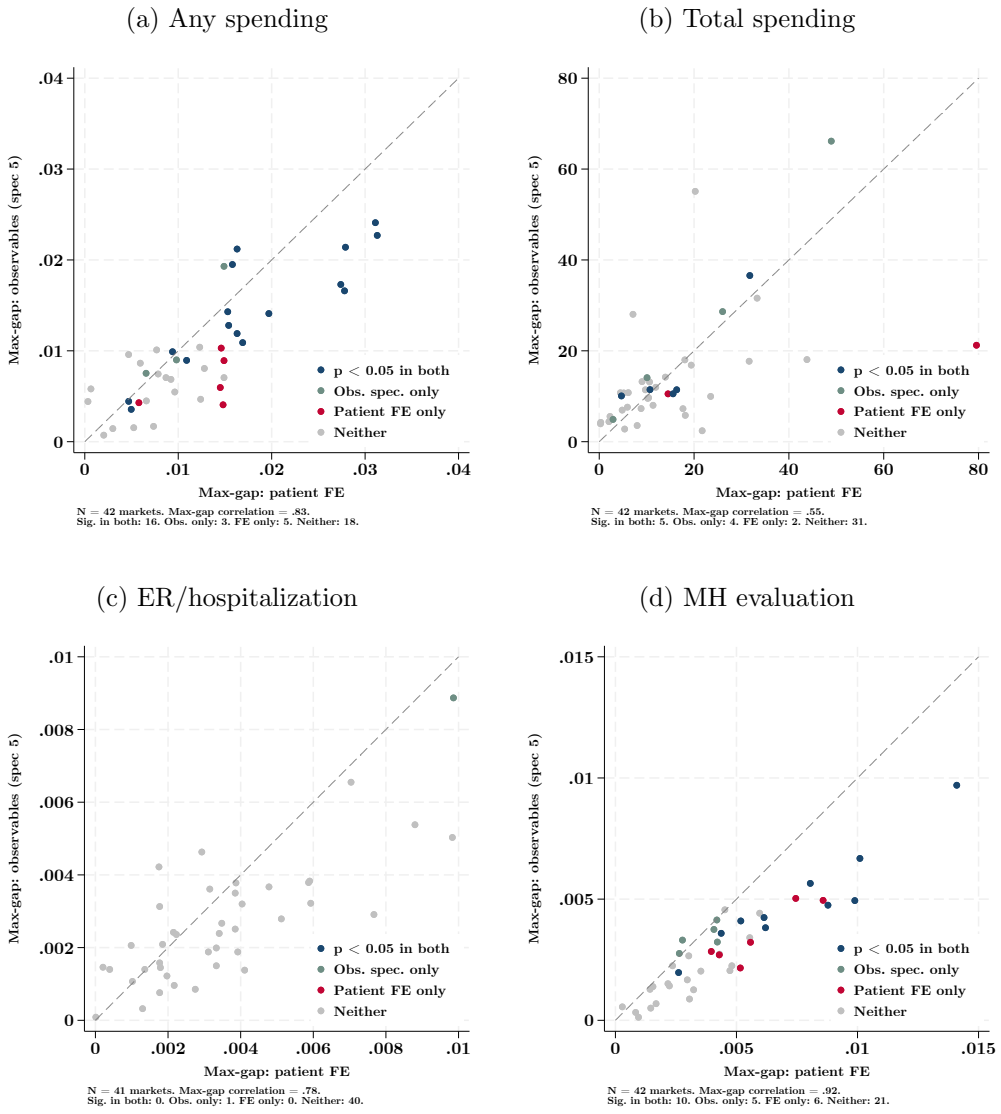


(d) MH evaluation



Notes: The above figures plot the empirical cumulative distribution function of p-values from market-level F-tests of the null hypothesis that the manufacturer effects $\Psi_{M(i)}$ are jointly equal within a market, as estimated from equation (1). The baseline ((1) includes patient, month-by-year, and relative-time fixed effects. Specifications 2-5 exclude patient fixed effects. Specification 2 includes patient demographics. Specification 3 further adds general health controls measured over the 90 days prior to initiation. Specification 4 further adds MH controls. Specification 5 further adds indicators for the specialty of the prescribing provider. Standard errors are clustered by pharmacy in all specifications. The dashed diagonal line represents the Uniform(0,1) distribution expected under the null hypothesis of no manufacturer differences.

Figure A2: Initiation design: Robustness to selection on observables: Max-gap statistics across specifications



Notes: The above figures plot the max-gap statistic from the richest observables specification against the max-gap statistic from the patient fixed effects specification, separately for each outcome. Each dot represents one market. The max-gap is defined as the difference between the highest- and lowest-performing manufacturer fixed effects within a market, estimated from equation (1). The patient fixed effects specification includes patient, month–by–year, and relative-time fixed effects. The observables specification replaces patient fixed effects with patient demographics, general health controls measured over the 90 days prior to initiation, mental health controls, and indicators for the specialty of the prescribing provider while retaining month–by–year and relative-time fixed effects. Dot color indicates agreement in statistical significance across specifications: dark navy indicates markets where both specifications reject the null of equal manufacturer effects at $p < 0.05$; teal indicates markets where only the observables specification rejects; cranberry indicates markets where only the patient FE specification rejects; and gray indicates markets where neither specification rejects. The dashed diagonal line represents the 45-degree line along which the two specifications produce identical max-gap estimates. Standard errors are clustered by pharmacy in all specifications.