The Generic Drug Trilemma

Daniel J. Hemel
New York University School of Law

Lisa Larrimore Ouellette
Stanford Law School

July 5, 2022

Abstract:

More than 90 percent of prescriptions dispensed in the United States each year are for off-patent drugs. Yet the bulk of scholarship on prescription drug policy focuses on patented drugs. Discussions of prescription drugs are typically oriented around the “innovation-access dilemma”—the tradeoff between stronger patent-based incentives for innovators and higher prices for purchasers of patented products. But for drugs in the “patent afterlife”—the period after patent protection and other forms of market exclusivity have expired—the innovation-access dilemma is not the fundamental policy tradeoff. Higher prices do not necessarily redound to the benefit of innovators, and price is not the only significant impediment to access: drug shortages—often triggered by safety concerns—also prevent patients from obtaining the medicines they need.

This chapter seeks to provide scholars and policymakers with a unifying framework for analyzing the variegated challenges of the pharmaceutical patent afterlife. We argue that the key tradeoffs in off-patent drug policy take the form of a trilemma—the three corners of which are price, quantity, and quality (i.e., safety and efficacy). The ideal for off-patent drug policy is to facilitate (1) low prices and (2) sufficient quantities of drugs that are (3) equivalent or similar to brand-name drugs approved by the Food and Drug Administration. But policies that improve outcomes along one or two of those dimensions typically entail some sacrifice along the third.

The trilemma framing yields several analytical payoffs. First, it sheds light on the root causes of puzzling problems plaguing some off-patent drug markets, such as sudden price spikes and persistent mismatches between supply and demand. Second, it draws attention to the non-innovation costs of recent drug pricing reform proposals—costs that are often overlooked amid the emphasis on innovation versus access. Finally, it motivates a search for solutions that can transcend the trilemma and optimize along all three dimensions of off-patent drug policy. That search brings us, full circle, back to innovation and access—the concerns that dominate discussions of pharmaceutical policy at the beginning of the patent life. But this time, the focus is on innovation—and, specifically, manufacturing innovation—in service of rather than in tension with access.
# Introduction

The innovation-access dilemma is the dominant theme in discussions of patent law and policy (e.g., Landes and Posner 2003; Barnes 2010; Sampat and Williams 2019). Patent protection incentivizes innovation by offering a time-limited monopoly for novel and nonobvious inventions, but the high prices facilitated by patent monopolies limit access to knowledge goods. The stakes on both sides of the innovation-access tradeoff are particularly high in the prescription drug context: pharmaceutical innovation has fueled significant health and longevity gains in recent decades (Lichtenberg 2019), but high prices prevent millions of patients in and beyond the United States from accessing medicines they need (WHO 2016; Kearney et al. 2021).

In prior work, we have argued that the innovation-access dilemma is, in fact, less of a dilemma than the conventional framing lets on. First, policymakers can avoid the tradeoff by replacing patents with non-patent rewards, including grants, tax credits, and prizes (Hemel and Ouellette 2013). Second, even without eschewing intellectual property, policymakers can sidestep the innovation-access tradeoff by “matching” patent-based rewards with non-patent allocation mechanisms such as government procurement (Hemel and Ouellette 2019). The Covid-19 experience offers an illustration of “matching” in action: while preserving patent protection on the incentive side, the federal government provided all Americans with free access to vaccines through procurement contracts with Pfizer, Moderna, and Johnson & Johnson. That is, the price paid to innovators can be set separately from the price paid by patients.

But whether one accepts or rejects the claim that the innovation-access dilemma is the fundamental policy problem in patent law generally or pharmaceutical patents specifically, it is clearly not the fundamental policy problem in the market for generic drugs, which constitute 90 percent of prescription drugs dispensed in the United States (Association for Accessible Medicines 2021). Once drugs lose the protection of patents and other forms of market exclusivity—in the period we call the “patent afterlife”—many of the profits flow to firms that had no role in the development of the drug in question. While patients still often pay high prices for off-patent drugs, high prices paid to firms that did not develop the relevant drug are not part of society’s bargain for more innovation.

To be sure, drug policy in the patent afterlife is not free from tradeoffs. The tradeoffs, though, are different from—and potentially more difficult than—the innovation-access tradeoff in patent law. And while these tradeoffs have garnered much less attention than the innovation-access dilemma, they have significant implications for access to medicines in the United States.

---

1 “Off-patent” drugs constitute even more than 90 percent of prescription drugs dispensed. The category of off-patent drugs includes not only generics, but also biosimilar versions of biologic products as well as branded drugs that are no longer patent-protected.
This chapter argues that the key tradeoffs in generic drug policy take the form of a trilemma—the three corners of which are price, quantity, and quality (encompassing efficacy as well as safety). As with other trilemmas in economics—e.g., the classic macroeconomic trilemma (Fleming and Mundell 1964) and the political trilemma of the world economy (Rodrik 2000)—the trilemma framing captures the impossibility (or at least extreme difficulty) of achieving all three goals at the same time (here, low prices, sufficient quantities, and a safety and efficacy profile that mirrors the profile of originator drugs).

Policy solutions that satisfy one—or even two—of the objectives identified by the trilemma are feasible, but solutions that adequately address all three remain elusive for many drugs. We can have a cheap and plentiful supply of generic drugs, a plentiful supply of high-quality generic drugs, or a cheap supply of high-quality generics, but the abiding challenge of generic drug policy is to solve for all three variables simultaneously.

To see how the goals of generic drug policy come into conflict, start with the corner of the trilemma that has recently attracted the most policy attention: price. Policies that reduce prices paid to manufacturers, such as price caps, also reduce incentives for generic entry, increase incentives for exit, and discourage manufacturers that remain in the market from building up excess supply. As a result, the probability of shortages will rise, negatively impacting the second corner: adequate quantities. (As discussed below, the U.S. and Canadian experiences largely bear out this prediction.) Reciprocally, policymakers could minimize the risk of shortages by offering to reimburse off-patent drug manufacturers at high rates, but those generous reimbursements would (of course) come at the expense of the first corner of the trilemma: low prices.

The third corner of the trilemma—quality—implies a similar quandary. With respect to small-molecule drugs, the U.S. Food and Drug Administration (FDA) requires generic manufacturers to demonstrate “bioequivalence” with the brand-name product. Firms seeking to market the equivalent of generics for off-patent biologics must satisfy even more demanding requirements. The FDA also requires firms to adhere to rigorous manufacturing protocols. These prerequisites for approval raise entry barriers that reduce competition. That lack of competition, in turn, leaves the market for off-patent drugs more vulnerable to sudden price hikes and to persistent mismatches between supply and demand.

At least in theory, policymakers still can achieve any two of the three objectives straightforwardly. For example, policymakers can achieve both low prices and sufficient quantities by reducing entry barriers for off-patent drugmakers. But those entry barriers exist to protect safety and efficacy, so reduced regulatory scrutiny will entail a tradeoff with drug quality. Similarly, policymakers can satisfy the price and quality corners of the trilemma by combining price caps with rigorous safety and efficacy evaluations. But the combination of low prices and high entry barriers will exacerbate the risk of shortages. To satisfy both quantity and quality objectives, policymakers can combine thorough regulatory scrutiny with price guarantees for manufacturers who meet those high standards—but at the cost of increased prices.
Our trilemma framing is not, to be sure, an axiom of futility: some reforms will improve outcomes in two corners much more than they cause harm in the third, and in some generic drug markets the tradeoffs will be much less stark than in others. What the trilemma framing accomplishes is to highlight the compromises that characterize off-patent drug policy. It thus helps to explain how some of the puzzling outcomes in off-patent drug markets—sudden price spikes, persistent shortages, and safety lapses—trace back to choices by policymakers to prioritize (rightly or wrongly) other corners. And it inspires reflection on approaches that might alleviate some of the tension among the price, quantity, and quality objectives—for example, government efforts to accelerate manufacturing innovation.

In Section 2, we describe the regulatory and competitive landscape for off-patent drugs in the United States. Section 3 introduces the trilemma. Section 4 considers policies that seek to transcend the trilemma through antitrust enforcement, government manufacturing, and public investments in supply chain resilience. Section 5 concludes.

2 The U.S. Pharmaceutical Patent Afterlife

The United States, like most other countries, provides 20 years of patent protection for inventions that satisfy certain statutory requirements. Congress has provided several additional protections for drugs and vaccines that sometimes extend the period of market exclusivity beyond the 20-year patent term. For example, the Orphan Drug Act of 1983 provides seven years of exclusivity for drugs that treat rare diseases or conditions—generally, diseases or conditions affecting fewer than 200,000 people within the United States. The seven-year term starts from the date of FDA approval, so if a qualifying drug is approved late in the 20-year patent term, Orphan Drug Act exclusivity may last beyond the expiry of patent protection. Likewise, the Hatch-Waxman Act of 1984 allows a patent term restoration of up to five years to compensate for some of the time lost to clinical testing and FDA review of a new drug application (though the patent term cannot be stretched beyond 14 years after FDA approval).

Once patent protection and any additional periods of market exclusivity have ended, a pharmaceutical product enters (what we call) the “patent afterlife.” Even then, though, competitors must obtain FDA approval before they can begin to sell a pharmaceutical product. This section provides an overview of the regulatory regime that governs the U.S. pharmaceutical patent afterlife.

2.1 Small-Molecule Drugs vs. Biologics

Regulation in the U.S. pharmaceutical patent life differs depending on whether the relevant product is a small-molecule drug or a biologic. Small-molecule drugs have simple chemical structures that can be described relatively easily by scientists. As a result, it is generally straightforward to determine whether two small-molecule drugs are chemically equivalent.
Most small-molecule drugs are administered orally (e.g., as tablets, capsules, or liquids), but some are administered via injection (including many small-molecule chemotherapy drugs and anesthetics). Sometimes, the term “generic” is used to describe exclusively small-molecule generics.

Biologics have more complex structures and are typically derived from living material. Examples include vaccines, monoclonal antibodies, and insulin. Because of their complexity, the chemical structures of biologics are much harder (and sometimes impossible) for scientists to describe, making comparisons between biologics much more difficult. Biologics are usually—though not always—administered by injection or infusion.

The 20th century has been termed “the era of the small molecule” (Economist 2014). Most pharmaceutical products in the United States still are small-molecule drugs, including over-the-counter products such as aspirin and ibuprofen as well as familiar prescription drugs such as fluoxetine (brand name Prozac) and atorvastatin (Lipitor). But many of the blockbuster drugs of recent decades have been biologics, including the arthritis drug adalimumab (Humira) and the autoimmune disease treatment infliximab (Remicaid). By 2018, eight of the ten best-selling prescription drugs around the world were biologics (Yip 2020).

2.2 Small-Molecule Drugs in the Patent Afterlife

Any new drug—whether it is a small-molecule drug or a biologic—must be approved by the FDA before it can be sold in the United States. Small-molecule drugs in the patent afterlife can be grouped into three categories based on their pathway to FDA approval: (1) brand-name drugs approved via a new drug application (NDA); (2) generics approved via an abbreviated new drug application (ANDA); and (3) drugs approved under section 505(b)(2) of the Federal Food, Drug, and Cosmetic Act, which offers a hybrid between an NDA and an ANDA.

2.2.1 Brand-Name Small-Molecule Drugs

A new drug application for a small-molecule drug must show the product is safe and effective for its intended use, that its benefits outweigh its risks, and that it will be manufactured in a controlled way that will preserve its quality. To generate evidence in support of these applications, the sponsor must conduct costly clinical trials. According to one recent estimate, the average cost of clinical trials across phases I, II, and III is $114 million in 2013 dollars (DiMasi, Grabowski, and Hansen 2016). The average cost per approved compound is much higher because most clinical trials are not successful.

---

2 The $114 million estimate is an average for small-molecule drugs and biologics. For phase I and phase III clinical trials, the authors find no statistically significant cost differences between small-molecule drugs and biologics. Phase II clinical trials appear to be more expensive for biologics than for small-molecule drugs (DiMasi, Grabowski, and Hansen 2016, online appendix B).
A brand-name drug enters the patent afterlife when patent protection (and any other period of market exclusivity) expires. Patent expiration may occur less than 20 years after the drug is first marketed because some of the patent term is generally consumed by the time spent on clinical trials and regulatory approval (Budish, Roin, and Williams 2015). Firms often prolong patent life through “evergreening,” or filing later patents on secondary innovations related to their drugs, though these patents tend to be lower quality (Frakes and Wasserman 2020) and are more likely to be held invalid or not infringed during litigation (Hemphill and Sampat 2013). Looking specifically at drugs for which the patent term was extended under the Hatch-Waxman Act, Lietzan and Liebecker (2020) estimate that the average effective patent life for these drugs is 11.58 years.

Once a brand-name drug enters the patent afterlife, it does not necessarily face immediate competition from generics. For a brand-name drug based on a “new chemical entity” (roughly speaking, a drug whose active ingredient has never before been approved by the FDA), the Hatch-Waxman Act grants five years of “data exclusivity.” Data exclusivity prevents a generic firm from relying on the brand-name drugmaker’s original clinical trial data. Technically, data exclusivity does not prevent market entry if the generic firm conducts its own clinical trials, whereas “market exclusivity” (as under, e.g., the Orphan Drug Act) completely blocks market entry. But as a practical matter, data exclusivity has a similar effect to market exclusivity because of the cost of conducting new clinical trials (Thomas 2017). If a brand-name drug based on a new chemical entity is approved with less than five years of patent protection remaining, data exclusivity is likely to extend the brand-name drugmaker’s monopoly beyond the patent term.

2.2.2 Small-Molecule Generics

Once the patent term and any non-patent market and data exclusivity periods are over, a would-be generic manufacturer still needs FDA approval before it can sell a small-molecule drug. The Hatch-Waxman Act created the standard pathway to FDA approval for small-molecule generics: an abbreviated new drug application, or ANDA. Because small-molecule drugs generally comprise only 20 to 100 atoms, laboratory measurements can be used to show that a generic is chemically equivalent to the innovator (i.e., brand-name) drug. The regulatory standard is that an ANDA must contain evidence of “bioequivalence,” meaning that there is no “significant difference in the rate and extent to which the active ingredient . . . becomes available at the site of drug action.” A manufacturer of a small-molecule generic that satisfies this bioequivalence standard does not need to conduct its own clinical trials to establish safety and efficacy. Instead, it can rely on the studies that supported the brand-name drug’s application (provided that the brand-name drugmaker no longer enjoys data exclusivity).

FDA approval of an ANDA for a small-molecule generic has potential state-law consequences. As of September 2019, 19 states had enacted laws that require pharmacists to substitute a lower-cost generic for a brand-name small-molecule drug when the physician’s prescription references the brand-name product. The other 31 states and the
District of Columbia had laws that allow—but do not require—pharmacists to substitute lower-cost generics for brand-name small-molecule drugs. State laws also vary in whether they require pharmacists to notify patients or obtain patient consent when they substitute a generic for a brand-name drug. In all states, physicians can prevent generic substitution by prescribing the brand-name drug and indicating “dispense as written” on the prescription (Sacks et al. 2021).

Historically, most ANDAs targeted drugs that were no longer covered by patents (FTC 2002). However, Hatch-Waxman allows—and encourages—generic developers to seek approval before a brand-name manufacturer’s patents have expired. An ANDA seeking market entry before patent expiration must include a “paragraph IV” certification stating that the relevant patents are invalid or that the brand-name manufacturer’s patents are not infringed by the generic product. Hatch-Waxman awards the first paragraph IV filer with 180 days of generic market exclusivity once its ANDA is approved, meaning that the FDA will not allow a second generic on the market during this time. The 180-day exclusivity period is intended to incentivize generic developers to challenge patents that were improperly granted or that are asserted to cover drugs beyond their proper scope.

### 2.2.3 Hybrid Drugs Approved Under Section 505(b)(2)

Hatch-Waxman also created another pathway—known as a “505(b)(2) application”—for hybrid small-molecule drug products that depend on both an existing drug’s clinical trial data plus new clinical trial data. These hybrids typically have the same active ingredient as an already approved drug but a different formulation or delivery mechanism. Drugs approved under section 505(b)(2) are eligible for three years of data exclusivity before other firms can rely on the new clinical trials. One well-known example of a drug approved under section 505(b)(2) is the Narcan nasal spray, used to stop or reverse the effects of opioid and heroin overdoses. The FDA had previously approved a drug using the same active ingredient (naloxone hydrochloride) in injectable form. Section 505(b)(2) allowed the pharmaceutical product ZIMHI to be approved based on safety and efficacy studies from the earlier naloxone hydrochloride NDA as well as new evidence regarding the safety and efficacy of the nasal spray pathway (FDA 2021b).

Approvals under section 505(b)(2) have increased significantly since the 1990s. In 2011, the FDA for the first time approved more section 505(b)(2) applications than NDAs (Gaffney 2015). However, the ANDA route for small-molecule generics remains much more common than the 505(b)(2) pathway, with the FDA granting more than ten times as many ANDA approvals each year as 505(b)(2) approvals (Darrow, He, and Stefanini 2019).

---

3 The 505(b)(2) pathway has also been used for some biologic drugs, but as of March 23, 2020, new drugs that rely on an existing biologic drug’s clinical trial data must be approved under the “biosimilar” pathway described below.
2.3 Biologics in the Patent Afterlife

Because of the relative recency of the biologics boom (Mullard 2022), comparatively few biologics have reached the patent afterlife. When they ultimately do, biologics fall into one of two categories: (1) brand-name biologics that have lost patent protection and market exclusivity, and (2) “biosimilars.”

2.3.1 Brand-Name Biologics

The analog to an NDA in the biologics context is a biologics license application (BLA). As with NDAs for small-molecule drugs, BLAs require costly clinical trials to demonstrate safety and efficacy. One salient difference between small-molecule pharmaceuticals and biologics is the length of data exclusivity for innovative drugs. Recall that under Hatch-Waxman, brand-name small-molecule drugs based on new chemical entities enjoy five years of data exclusivity. By contrast, the Biologics Price Competition and Innovation Act of 2010 (BPCIA)—enacted as part of the Affordable Care Act—provides twelve years of data exclusivity starting from the date that a new biologic is licensed by the FDA.

2.3.2 Biosimilars

In addition to its data exclusivity provisions for brand-name biologics, the BPCIA created an analog to an ANDA for biologic drugs: an abbreviated biologics license application (aBLA). The greater complexity of biologics makes it more difficult to demonstrate that a second product will have the same clinical effect, which is why generic biologics are termed “biosimilars” (rather than “bioequivalents”). To show biosimilarity, an aBLA must provide evidence that the product’s characteristics are “highly similar” to the original biologic and that it has “no clinically meaningful differences” in clinical response studies.

If the aBLA also demonstrates that there is little risk in switching between the original biologic and the new biosimilar, the FDA may deem the biosimilar to be “interchangeable” with the brand-name biologic. When a biosimilar is designated as interchangeable with a brand-name biologic, no other manufacturer can receive an interchangeability designation for 12 to 42 months (with the length of the exclusivity period depending on the state of patent litigation). Thirteen states require pharmacists to substitute interchangeable biosimilars when a prescription references the brand-name biologic; the remaining states allow but do not require substitution (Sacks et al. 2021). As of June 2022, the FDA had approved approximately three dozen biosimilars, but it had designated only two of those biosimilars as interchangeable with a brand-name biologic.

The scientific differences between small-molecule and biologic drugs are reflected in different development costs. For small-molecule generics, the ANDA process typically takes one to three years and $1 million to $5 million, with no new clinical trials. By contrast, biosimilar development typically takes eight to ten years and around $100 million—including tens of millions in clinical response trials to confirm biosimilarity (Atteberry et al. 2019).
Table 1 summarizes the main categories of drugs that we will consider in this chapter.

Table 1. Types of Drugs in the Patent Afterlife

<table>
<thead>
<tr>
<th>Small-Molecule Drugs</th>
<th>Biologics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Brand-name small-molecule drug:</strong> Small-molecule drug approved under a new drug application (NDA) for which patent protection and market exclusivity have expired</td>
<td><strong>Brand-name biologic:</strong> Biological product approved under a biologics license application (BLA) for which patent protection and market exclusivity have expired</td>
</tr>
<tr>
<td><strong>Small-molecule generic:</strong> Small-molecule drug approved under an abbreviated new drug application (ANDA)</td>
<td><strong>Biosimilar:</strong> Biological product approved under an abbreviated biologics license application (aBLA)</td>
</tr>
<tr>
<td><strong>505(b)(2) hybrid:</strong> Small-molecule drug approved under hybrid review process based on existing drug’s clinical trial data plus new data</td>
<td>- <strong>Interchangeable biosimilar:</strong> Biosimilar deemed interchangeable with brand-name biologic by FDA</td>
</tr>
<tr>
<td></td>
<td>- <strong>Non-interchangeable biosimilar:</strong> Biosimilar without interchangeability determination</td>
</tr>
</tbody>
</table>

2.4 Competition in the Pharmaceutical Patent Afterlife

The Hatch-Waxman Act is generally viewed as a success in increasing generic entry for small-molecule pharmaceutical drugs. By September 1994, ten years after Hatch-Waxman’s enactment, the FDA had approved 1.91 ANDAs for every approved brand-name small-molecule drug (Heled 2021). Out of 206 new small-molecule drugs approved in tablet or capsule form between 1995 and 2010, the vast majority (167, or 81.1 percent) had at least one generic version approved by the FDA as of the end of 2021 (Gupta et al. 2022). Yet despite these successes, there are still more than 360 small-molecule drugs that are no longer protected by patents or other forms of market exclusivity but that have no generic competitors (FDA 2022b), and many more with limited competition.

In contrast to Hatch-Waxman’s qualified success, the BPCIA is generally viewed as failing at its goal of increasing access to biologics in the United States. Ten years after enactment of the BPCIA, the FDA still had approved only 0.1 follow-on products for every approved brand-name biologic (Heled 2021). And for some biologics like vaccines, the FDA has not even issued regulations for how a firm would show that its vaccine is biosimilar to an existing vaccine. In other words, there is no such thing as a generic (or biosimilar) vaccine as a regulatory matter. Competition in vaccine markets thus requires an independent BLA based on new clinical trials.
Overall, competition in the pharmaceutical patent afterlife is far from robust. Conti and Berndt (2020) examine unique markets for off-patent small-molecule and biologic drugs from 2004 to 2016 and conclude that after a drug loses patent protection or other exclusivity, the median number of manufacturers in each market is typically only two or three. Over half of generics only have one supplier (NASEM 2018). As a number of scholars have noted, the generic markets attracting the least competition—including small-molecule drugs with limited markets (Scott Morton and Boller 2017) and most biologics (Atteberry et al. 2019)—often have a particularly high ratio of entry costs to available profits, giving them characteristics of natural monopolies.

From a social welfare perspective, competition is not an end in itself but a means to other welfare-relevant ends: lower prices, sufficient quantities, and higher quality. Yet as explained in the next section, lack of competition in the pharmaceutical patent afterlife has important—and negative—implications for patient welfare.

3 Three Goals of Generic Drug Policy

Here, we introduce three key goals of policy in the patent afterlife—low prices, adequate quantities, and high quality—and explain why they form a trilemma: policy reforms that pursue one or two of these goals generally sacrifice on the third. The price-quantity-quality trilemma is conceptually distinct from the innovation-access dilemma that dominates discussions of the optimal patent term. In the standard economic model of patents (Nordhaus 1969), the optimal term is finite—at some point, the extra innovation incentive from a longer patent term no longer justifies the deadweight loss of monopoly. For present purposes, we will remain agnostic as to when that point is reached. Our focus here is on what happens afterwards.

3.1 Price

The most common complaint regarding drugs in the patent afterlife in the United States is their high price (e.g., Rosenthal 2014). Two dimensions of drug pricing are welfare-relevant: (1) the price paid to manufacturers by purchasers (including individual and group health plans, Medicare, and Medicaid), and (2) the out-of-pocket price paid by patients. The two dimensions are closely related, but they are also importantly distinct. Indeed, some policies that reduce out-of-pocket prices for patients may raise prices paid to manufacturers (and vice versa).

In recent years, large and abrupt increases in the list prices for generic drugs have become front-page news (e.g., Pollack and Tavernise 2015). From the first quarter of 2010 to the first quarter of 2015, more than a fifth of established generic small-molecule drugs (315 out of 1441) experienced a sudden price increase of 100 percent or more (GAO 2016). In one of

---

4 We thus bracket the question of whether patent terms or market exclusivity should be extended for some types of drugs and/or shortened for others (Buccafusco and Masur 2021).
the most notorious examples, Turing Pharmaceuticals—then run by the since-convicted “pharma bro” Martin Shkreli—hiked the price of the six-decade-old anti-parasite drug Daraprim by 5000 percent (Lupkin 2019).

Yet cross-country comparisons complicate the narrative of high U.S. generic drug prices. A RAND Corporation study commissioned by the U.S. Department of Health and Human Services found that as of 2018, U.S. unbranded generic drug prices were 16 percent lower than prices in other countries within the OECD (Mulcahy et al. 2021). The same study found that unbranded generic drugs accounted for a much larger share of drug volume in the United States than elsewhere in the OECD (84 percent versus 35 percent). In other words, relative to other high-income countries, U.S. purchasers are paying less for unbranded generics and buying more of them.

The RAND study’s findings do not imply that high prices are a spurious concern for off-patent drug policy. First, the cross-country comparison excludes biologics, which now constitute more than two-fifths of total U.S. pharmaceutical spending (IQVIA 2020). (As noted above, biologic drugs do not technically have “generics”—a drug designed to have the same clinical effect as a branded biologic is a “biosimilar.”) Second, the comparison also excludes non-originator drugs sold under brand names, such as drugs approved by the FDA through the hybrid section 505(b)(2) pathway. Once all small-molecule non-originator drugs are included in the cross-country comparison, U.S. purchasers pay—on average—21 percent more than OECD peers. Third, the category of unbranded generics does not include off-patent drugs that do not (yet) face a generic competitor, for which prices may remain high while effective market exclusivity persists. Fourth, for some small-molecule drugs (like the estrogen derivative estradiol), prices remain high even after generics enter the market (Thomas 2018). And even insofar as U.S. generic drug prices are in line with other OECD countries, one still might conclude that OECD countries are overpaying for generic drugs across the board.

Most importantly, the RAND comparison focuses on prices paid to manufacturers, not out-of-pocket prices paid by patients. In countries with universal health coverage, patient copays for prescription drugs are generally subject to low caps. For example, under the United Kingdom’s National Health Service program, copays for outpatient drugs are capped at £9.35 (NHS 2022), or approximately $11.40 in U.S. dollars as of June 2022. Outpatient drugs are free for many patients—including all children under 16, all adults over 60, and anyone who is pregnant or has had a child within the last 12 months—and inpatient drugs are free for everyone.

For patients with health insurance in the United States, by contrast, copays for prescription drugs can be substantial even after those drugs enter the patent afterlife. For example, average out-of-pocket spending on insulin for patients covered by Medicare Part D was $49 per prescription, or $520 per year, in 2019 (Cubanski and Damico 2022). And out-of-pocket costs can be significantly higher for the roughly 10 percent of Americans who have no health insurance (Cohen et al. 2021). Out-of-pocket patient costs are important because
they can impede access to medicines. In one recent survey, three in ten U.S. adults reported not taking medicines as prescribed at some point in the past year “due to cost” (Kearney et al. 2021).

For purposes of the analysis below, our primary focus is on prices paid to manufacturers, not out-of-pocket prices paid by patients. We focus on prices paid to manufacturers for three reasons. First, most people in the United States (more than 60 percent) are covered by private health insurance plans. Higher costs incurred by those plans are likely to be passed through partly or fully to patients—for example, through higher premiums or, in the case of employer-sponsored plans, potentially through lower wages (Kolstad and Kowalski 2016). Thus, prices paid to manufacturers by insurers or by self-funded employer plans remain relevant to patients even when those prices are not reflected directly in copays. And high premiums potentially push more individuals into the ranks of the uninsured. Second, in many cases, higher prices paid to manufacturers are reflected directly in higher out-of-pocket prices for patients. For example, Medicare Part B mandates a 20 percent copay for prescription drugs administered in an outpatient setting. Third, more than 45 percent of prescription drug costs in the United States are paid by federal, state, and local governments (CMS 2021). Prices paid to manufacturers by government purchasers affect fiscal capacity, including the capacity of federal, state, and local governments to provide greater relief to patients.

But while the twin goals of reducing prices paid to manufacturers and reducing out-of-pocket prices paid by patients will often align, they will sometimes come into conflict. Higher copays potentially encourage patients to economize on health care spending (Zeckhauser 1970). For example, a patient facing a high copay for a brand-name drug may be more likely to choose a lower-priced generic (insofar as the patient faces a choice under the relevant state’s substitution law) or more likely to ask her physician to prescribe a biosimilar in place of a brand-name biologic. Capping or eliminating copays may therefore result in higher health system costs, some of which may be passed back to the broader patient pool (e.g., through higher premiums).

### 3.2 Quantity

A second area of concern regarding generic drugs in the United States is quantity. As of June 2022, the FDA classified 120 drugs as “currently in shortage” (FDA 2022a). Drug shortages are not a Covid-19-specific phenomenon: in 2011, the FDA recorded 251 new shortages (FDA 2021a). Although shortages can arise with respect to on-patent and off-patent drugs, the FDA reports that two-thirds of drugs that went into shortage during a five-year period from 2013 to 2017 were drugs with a generic version on the market (FDA 2020). Shortages are especially common with respect to sterile injectables, such as anesthetics and chemotherapy treatments (Yurukoglu, Liebman, and Ridley 2017).

From an economic perspective, the very idea of a “shortage” is somewhat puzzling: unless the supply and demand curves are both vertical lines, there must be some price at which
they intersect. Within a Marshallian supply-demand framework, there are at least three ways to make sense of widespread reports of drug “shortages.” First, demand for the relevant drug may be inelastic. For example, insulin is often cited in microeconomics textbooks as an example of a product with inelastic demand (e.g., Brown 1995; Cowen and Tabarrok 2009). If demand is entirely unresponsive to price, then a temporary supply shock may result in disequilibrium. Second, purchasing practices may introduce rigidities that prevent prices from rising to market-clearing levels. For example, most medical centers acquire drugs through group purchasing organizations (GPOs), which act as intermediaries between purchasers and manufacturers (Bruhn, Fracica, and Makary 2018). Long-term contracts between GPOs and manufacturers—including contracts that lock a GPO into using a particular manufacturer as its sole source for a drug—may prevent prices from responding rapidly to changes in demand and supply. Third and finally, the term “shortage” might be understood to refer broadly to supply shocks that cause quantity to fall and price to rise, even if there is no extended period of disequilibrium between supply and demand. Although “shortage” is arguably a misnomer (i.e., the market clears at the new, higher price), the result is still that patients cannot obtain drugs at prices that prevailed prior to the shock.

Whether or not drug shortages are “[r]eal shortages” (Stomberg 2018) in the economic sense, the health costs of generic drug supply shocks are potentially significant. In 2018, the American Medical Association adopted a policy declaring drug shortages to be an “urgent public health crisis” (American Medical Association 2018). In several surveys, physicians and pharmacists report higher rates of adverse drug outcomes and even patient deaths due to drug shortages (for a literature review, see Phuong et al. 2019). Retrospective cohort studies document negative clinical outcomes associated with specific shortages. For example, Vail et al. (2017) find that patients with septic shock in hospitals affected by a 2011 norepinephrine shortage experienced higher rates of in-hospital mortality. Gross et al. (2017) find higher rates of *C. difficile* infection among patients at hospitals that switched to other antibiotics during a nationwide shortage of piperacillin/tazobactam in 2014.

Not all shortages are associated with negative clinical outcomes. For example, Trifilio et al. (2013) find no reduction in remission rates among patients with acute myeloid leukemia who switched from the chemotherapy drug daunorubcin to an alternative (idarubicin) during a daunorubicin shortage. Indeed, there is some evidence that older patients benefited from the shortage-induced switch. Moreover, the number of new shortages appears to be on the decline. According to the FDA, new drug shortages fell from a high of

---

5 Notwithstanding insulin’s utility as a textbook example of inelastic demand, many patients actually appear to be price-sensitive: a study of patients at the Yale Diabetes Center found that more than a quarter had cut back on insulin use due to cost (Herkert et al. 2019).

### 3.3 Quality

A final area of concern regarding generic drugs in the United States is quality, including both safety and efficacy. From a theoretical perspective, there are several reasons to be concerned about the quality of generic drugs. While these concerns may loom larger in theory than in practice, the theoretical perspective reminds us why the quality of generics should not be taken for granted.

One source of concern is extreme quality uncertainty in the prescription drug market. Patients have no real way to detect manufacturing defects even after taking a drug. In the absence of any manufacturing defect, prescription drugs typically are less than 100 percent effective and produce adverse reactions in a subset of patients. A patient’s own experience is thus minimally informative about potential manufacturing flaws. Quality uncertainty potentially leads to a “market for lemons,” in which manufacturers of high-quality products cannot monetize their quality investments (Akerlof 1970). Those manufacturers may respond by diluting their quality standards or by exiting the market.

Second, reputational mechanisms—which address quality uncertainty in other markets—are particularly weak in the market for generic drugs. As noted above, all states require or permit pharmacists to substitute a cheaper available generic when a doctor’s prescription specifies a brand-name small-molecule drug (Sacks et al. 2021). Patients do not necessarily choose—and may not even know—the identity of the generic manufacturer. The low salience of reputation reduces incentives for generic manufacturers to invest in quality.

Third, no single generic firm internalizes the full costs of manufacturing defects. For example, the infamous Cutter incident—in which live polio virus contaminated 120,000 doses of the Salk polio vaccine in 1955—shook confidence in (and reduced uptake of) other vaccine manufacturers’ products as well (Oshinsky 2006). Acting on its own, a generic manufacturer may choose a level of investment in quality assurance that is lower than the optimal level from the entire industry’s (or society’s) perspective.

Fourth and finally, ex-post liability for manufacturing defects cannot resolve all the problems of quality uncertainty in the market for generic drugs. Some defects will likely go undetected, since patients and physicians will not be able to distinguish the consequences of manufacturing defects from the normal operation of a non-defective drug. Moreover, some of the social costs of manufacturing defects—for example, reduced uptake of other manufacturers’ products—likely won’t be recoverable in tort. And all this is on top of

---

6 For a critique of the FDA’s methodology for identifying shortages, see Lutter (2022). The University of Utah Drug Information System reports a higher number of shortages but a similar trend: from a high of 267 new shortages in 2011 to 129 in 2020 and 114 in 2021 (ASHP 2022).
familiar shortcomings of tort law—most relevantly, the fact that limited liability for
corporate shareholders shields actors from the full costs of their torts (Leebron 1991).

Anecdotal evidence partly bears out these theoretical concerns about generic drug quality. In 2013, generic manufacturer Ranbaxy accepted a $500 million fine for safety violations at two of its Indian factories, including delays in reporting “unknown impurities” detected in an epilepsy drug that eventually led to a 73-million-pill recall (Thomas 2013). Discoveries of the likely carcinogen NDMA have led to major FDA-initiated recalls of off-patent drugs to treat heartburn (Johnson 2019), high blood pressure (Edney, Berfield, and Yu 2019), and diabetes (Blankenship 2020a). Investigative journalists have also uncovered anecdotes of safety violations at generic manufacturing plants, such as fabrication of data for regulators (e.g., Eban 2019; Stockman 2021).

Evaluating the actual health costs of these quality failures is more challenging. The possible presence of contaminants of unknown risk at various points along the generic drug supply chain will not always translate to patient harm, and these problems may be episodic rather than systemic. It is also unclear whether quality problems are substantially higher for generics than for brand-name drugs—quality concerns also can strike before pharmaceuticals enter the patent afterlife. Although critics of the generics industry point to the outsourcing of most generic manufacturing to India and China and problems with the FDA’s foreign inspection program, many brand-name drugs are also sourced from overseas, sometimes from the same factories as generics. Moreover, the FDA does not publicly disclose where a drug is made because this information is considered a trade secret, which makes it hard to determine whether manufacturing origin affects patient outcomes.

Randomized controlled trials generally have not found significant differences in the clinical efficacy of generic and brand-name products (Kesselheim et al. 2008; Kesselheim et al. 2010). But these studies do not resolve the quality debate, for two reasons. First, randomized trials compare generic drugs to brand-name drugs, but if brand-name drugs also have quality problems, researchers may observe no statistically significant differences between generic and brand-name treatment groups even though quality problems plague both. Understanding the health consequences of quality issues in the drug supply would require a comparison of real-world marketed drugs with the idealized, defect-free versions. Second, blinded randomized controlled trials cannot capture the consequences that arise when patients know they are switching from brand-name drugs to generics. Some studies have found that switching from brand-name to generic drugs was associated with worse clinical outcomes and more adverse events (for a systematic review, see Straka, Keohane and Liu 2017), possibly because of negative perceptions of generics (Colgan et al. 2015) and related psychosomatic effects (Goldszmidt et al. 2019). FDA efforts to bolster confidence in generic drug quality may have positive health consequences even if those consequences flow through psychosomatic channels.
3.4 The Generic Drug Trilemma

Each of these three goals—low prices, sufficient quantity, and high quality—theoretically lies within reach. But efforts to advance one or two of these goals predictably conflict with the third. We first consider policies that would satisfy one corner of the trilemma (i.e., would achieve one of the three goals), and then consider policy combinations that would satisfy two corners.

Price. The most straightforward way to reduce prices of prescription drugs is to impose price caps. For example, in 1998, the government of Ontario—Canada’s most populous province—introduced the so-called “70/90” regulations under which the first generic entrant was prohibited from charging more than 70 percent of the brand-name product price and subsequent generic entrants were initially capped at 90 percent of the first generic price (Zhang et al. 2016). Since 2018, several U.S. states have created prescription drug affordability boards authorized to establish price ceilings for prescription drugs (Williamson 2021). The Build Back Better Act, which passed the House in late 2021, would effectively cap the price of insulin and a limited number of single-source drugs, including some off-patent drugs without generic competitors.

Price caps, to be sure, are not certain to achieve their immediate goal of reducing prices. For example, Anis, Guh, and Woolcott (2003) find that price caps set by the Ontario 70/90 regulations served as focal points for generic manufacturers: firms offered the maximum price allowable under the regulation even when they might have set lower prices in an unregulated market. But even well-designed price caps cannot escape from the law of unintended consequences. Price caps reduce incentives for generic entry and increase incentives for exit, as evidenced by increased generic exit in Ontario (Zhang et al. 2016). Policies laser-targeted at reducing prices are therefore likely to have negative effects on quantity (Scott Morton 2001; Rye 2012).

Policymakers also can push down prices by reducing regulatory barriers to market entry. Conrad and Lutter (2019) find a strong negative correlation between the median generic price of a particular drug (relative to the brand price before generic entry) and the number of generic producers of the drug: the median price with one generic producer is 61 percent of the brand price, falling to 21 percent of the brand-name price with four producers. The number of generic producers per drug would likely increase—and prices would fall—if the FDA’s ANDA review process were less demanding. But that review process exists to ensure quality, and while it may be possible for the FDA to make some cost-cutting changes to the ANDA process without any serious impacts on safety or efficacy, agency efforts to reduce regulatory barriers to entry will ultimately come into conflict with the quality objective.

Quantity. For the same reason that price caps negatively affect quantity, the most straightforward way to address drug shortages is to guarantee higher prices. When Congress asked the FDA to analyze the “drug shortage crisis” in 2018, the agency concluded that the root cause of shortages was economic forces that lead to a “race to the bottom” in
generic drug pricing (FDA 2020). But by construction, price guarantees would run counter
to the goal of lowering drug prices. Quality issues, meanwhile, are the immediate reason for
most drug shortages in the United States: according to the FDA, 62 percent of new drug
shortages from 2013 to 2017 were triggered by quality issues (FDA 2020). Relaxing quality
standards would thus reduce the number of shortages—but of course, safety and efficacy
would suffer as a result.

**Quality.** To address concerns about generic drug quality, regulators could impose more
rigorous quality requirements on new entrants, coupled with strict liability regimes
targeting firms whose generics have manufacturing defects. Heightened ex ante scrutiny or
ex post liability, however, raise the costs of supplying the market and thus are likely to
undermine the other two corners of the trilemma by decreasing quantity and raising prices.
For example, Atal, Cuesta, and Sæthre (2019) examine Chile’s introduction of
bioequivalence requirements for generic drugs and found that this “stronger quality
regulation decreased the number of drugs in the market by 25% [and] increased average
paid prices by 10%.”

**Price Plus Quantity.** Just as a rigorous quality standard will increase both prices and the
risk of shortages, policymakers could achieve both lower prices and higher quantities by
reducing ex ante regulatory barriers and ex post liability risks. For example, Sachs (2019)
proposes that after a price spike exceeding a certain threshold for a particular drug, the
FDA could “preclear” generic manufacturers to make and sell that drug before completing
the full ANDA or aBLA approval process. (One could imagine a similar “preclearance”
mechanism triggered by drug shortages.) Yet as Sachs acknowledges, shortcutting the
generic drug approval process “creates serious problems” for safety and efficacy.

Similarly, Cohen et al. (2019) note that the FDA already has statutory authority to
authorize importation from Canada, and they suggest that the agency use this authority to
allow generic drug imports when prices rise suddenly in the United States. (They also argue
for expanding the FDA’s importation authority to include a select group of other countries
beyond Canada.) Yet as Bruser and McLean (2014) note, the FDA’s safety standards are
often more stringent than those of Canada’s pharmaceutical regulator. In other words,
importation can lower prices and reduce the risks of shortages because it effectively reduces
the quality threshold. While U.S. patients may be better off on balance if the FDA were to
allow generic imports in response to price spikes, the choice would not be tradeoff-free.

**Price Plus Quality.** Like price and quantity, price and quality are mutually realizable
goals. For example, policymakers could combine binding price caps with rigorous safety and
efficacy regulations. But the combination of price caps and stringent quality standards
would heighten the risk of shortages. Exacting quality standards would raise manufacturer
costs, spurring exit and deterring entry without high prices as an inducement.

Arguably, the U.S. childhood vaccine experience in recent years reflects the consequences of
policies that prioritize price and quality over quantity. The 1993 Vaccines for Children
program provides eligible children with covered vaccines at no out-of-pocket cost to their families. Eligible children are those who are uninsured or eligible for Medicaid, as well as children who receive vaccines at Federally Qualified Health Centers and all children who are members of an Indian tribe. More than half of young children in the United States are eligible for vaccines through the program (HHS 2020). Covered vaccines are all pediatric vaccines included on a list maintained by the CDC Advisory Committee on Immunization Practices. One provision in the 1993 law imposes price caps on preexisting pediatric vaccines: the U.S. Department of Health and Human Services is prohibited from paying more through the program for any vaccine than the May 1993 price adjusted for subsequent increases in the Consumer Price Index.⁷

Congress’s choice to cap prices for a significant portion of the childhood vaccine market without making any downward adjustments to quality standards puts the trilemma framework to the test. Sure enough, the United States subsequently saw a rash of childhood vaccine shortages. By 2004, eight of eleven routine childhood vaccines had gone into shortage, with the probability of shortages correlated to low prices for the relevant vaccine (Ridley, Bei, and Liebman 2016). Thirty-five states had temporarily suspended or reduced immunization requirements for daycare and/or school programs—evidently in response to supply constraints (GAO 2002).

The childhood vaccine experience illustrates the perils of transplanting the framework of an innovation-access tradeoff into the generic drug space, where a different policy tradeoff dominates. As Ridley, Bei, and Leibman (2016) summarize, one of the reasons why Congress imposed price caps through the Vaccines for Children program was a belief that it is “no longer necessary to worry about incentives for innovation for these vaccines because the costs of innovation were paid long ago by manufacturers.” But in the patent afterlife, prices affect dimensions other than innovation as well. Specifically, when prices are constrained to be low but quality standards raise production costs, manufacturers are less likely to invest in maintaining and expanding supply. Whether or not Congress made the “right” tradeoff among price, quality, and quantity as part of the 1993 Act, the choice to cap prices without conceding on quality clearly was a tradeoff—though one not immediately acknowledged at the time.

**Quantity Plus Quality.** Finally, policymakers could ensure sufficient quantities and maintain rigorous quality standards if they were willing to guarantee high prices for generic drugs. In this respect, “money answereth all things” (Ecclesiastes 10:19). Manufacturers would have strong incentives to invest in spare capacity—and in measures to protect safety and efficacy—if profit margins on drugs in the patent afterlife were wide enough. But of course, the one objective that cannot be satisfied by throwing money at a problem is economizing on price.

---

Figure 1 illustrates the three corners of the trilemma and some of the policy reforms that satisfy one or two corners.

![Image of the Generic Drug Trilemma]

**Figure 1. The Generic Drug Trilemma**

### 4 Partial Solutions to the Trilemma

Although policymakers cannot escape the price-quantity-quality trilemma, certain solutions will manage the trilemma better than others. In some cases, policy reforms can achieve gains for one or two goals that far outweigh the losses for the remaining corner. The trilemma also helps illustrate that no single goal is worth pursuing at any cost. Rather, ideally, society should seek gains in any one corner of the trilemma only up to the point that marginal benefit of those gains for social welfare equals the marginal cost of welfare losses at the other corners. Reaching this optimum, however, is easier said than done.

In some circumstances, competitive markets may be the best mechanism for balancing these tradeoffs, and greater antitrust scrutiny may help facilitate such competition. But the government also can generate competition itself by entering the market as a producer (or by contracting with other entities to manufacture off-patent drugs). In addition, the trilemma highlights the need for policies to reduce price-quality-quantity tradeoffs, such as public investments in supply chain resilience. Here, we describe generic drug policies grounded in each of these three approaches—antitrust enforcement, government production, and promotion of manufacturing innovation. Although none of these approaches offers a complete solution, each set of policies offers promise for improving outcomes in the patent afterlife.
4.1 Antitrust Enforcement

One consequence of the United States’ choice to allow unregulated pricing with restricted entry into the off-patent drug market is to leave the market vulnerable to anticompetitive conduct. Robust enforcement of antitrust laws can address some anticompetitive practices, but it does not offer a total escape from the price-quantity-quality trilemma.

Much of the literature on generic drug entry focuses on the period before patent protection expires—the patent life rather than the patent afterlife. A significant policy concern during a drug’s patent life is that the relevant patents may have been erroneously granted by the U.S. Patent and Trademark Office (Frakes and Wasserman 2019), or that they may not actually cover the drug in question. One of the innovations of the Hatch-Waxman Act—the paragraph IV process described in Section 2.2.2—was to create a route for generic drug manufacturers to challenge a patent as invalid or not infringed. In effect, Congress has sought to enlist generic drugmakers as “patent police.”

Patent policing by generic manufacturers does implicate the innovation-access tradeoff. The paragraph IV process is designed to reveal instances in which brand-name drugs do not merit the innovation incentive that comes with patent protection. However, paragraph IV has not fully lived up to its designers’ high hopes. One threat to the integrity of the process is the practice of “pay for delay,” in which a brand-name manufacturer pays or otherwise compensates a generic firm in exchange for the generic firm dropping a patent challenge (Hemphill 2006). A further threat is common ownership: when generic drug manufacturers and brand-name counterparties have the same institutional shareholders, they are more likely to settle paragraph IV disputes (Xie and Gerakos 2020). Finally, when generic drug manufacturers also have patented products in their portfolios, they are less likely to pursue paragraph IV challenges to judgment—possibly for fear of establishing precedents that will undermine their own patents (Carrier, Lemley, and Miller 2020).

Antitrust enforcement could potentially help invigorate patent policing by generic manufacturers. Stricter scrutiny of pay-for-delay deals might deter generic and brand-name manufacturers from entering anticompetitive agreements. Policy interventions to reduce common ownership (Elhague 2016; Posner, Scott Morgan, and Weyl 2017) might encourage more aggressive patent challenges under Hatch-Waxman and the BPCIA. Merger review might play a role in preventing brand-name and generic manufacturers from combining.

Our focus here, though, is not on the ways in which generic manufacturers fulfill their patent policing function, but on outcomes after the relevant patents have expired or been invalidated. Anticompetitive practices continue into the patent afterlife, but they differ from the practices associated with on-patent drugs. Often, anticompetitive practices in the patent afterlife exploit elements of the regulatory infrastructure designed to ensure the safety and efficacy of generic drugs. And the market for off-patent drugs is more vulnerable

---

8 The BPCIA sets forth an elaborate process—colloquially known as the “patent dance”—through which makers of brand-name biologics and biosimilars may resolve patent disputes.
to anticompetitive conduct as a result of entry barriers erected with safety and efficacy in mind.

Historically, one mainspring of anticompetitive conduct in the patent afterlife has been the requirement—borne out of concerns regarding generic drug quality—that generic drug manufacturers demonstrate bioequivalence. In order to satisfy the requirement, a potential generic drug manufacturer must test its own product against samples of the corresponding brand-name drug. Some brand-name manufacturers have sought to prevent potential generic competitors from performing those tests by refusing to provide samples of the brand-name drug (Pear 2018)—what one former Federal Trade Commission official has described as a “sample blockade” (Kades 2021). Turing Pharmaceuticals—the company once led by Martin Shkreli, who was later convicted of securities fraud—used this strategy to maintain its monopoly over the anti-parasite drug Daraprim long after all the relevant patents expired.9

“Safety protocol filibusters” (Kades 2021) are another anticompetitive practice that brand-name manufacturers have used to block generic entry during the patent afterlife. These “filibusters” involve drugs that are part of the FDA’s risk evaluation and mitigation strategies (REMS) program. For drugs that carry a risk of serious adverse effects, the FDA requires manufacturers to develop safety protocols to reduce the risk of adverse outcomes. For example, a REMS protocol might require healthcare professionals to complete a training module and obtain certification before prescribing or dispensing a particular drug. Historically, the FDA has required brand-name and generic manufacturers of the same drug to develop a single, shared REMS system. The generic manufacturer could not begin to sell the drug until the shared REMS was in place. That requirement encouraged brand-name manufacturers to engage in foot-dragging—or “filibustering”—during negotiations to establish the shared REMS, thus delaying generic entry. The FDA had the authority to waive the single, shared REMS requirement, but the agency rarely used that power (Dabrowska 2018).

The CREATEAct of 2019 seeks to address both the sample blockade and safety protocol filibuster issues. The statute—signed into law as part of a larger budget bill in December 2019—allows a potential generic entrant to sue to obtain samples of the brand-name drug. The statute also gives the FDA greater flexibility to allow generic manufacturers to establish their own REMS. Anecdotal evidence suggests that these reforms have been successful (Kades 2021). However, sample blockades and safety protocol filibusters are not the only anticompetitive practices in the patent afterlife. For example, brand-name manufacturers still may—and do—use “citizen petitions” to challenge generic drugmakers’ assertions that the brand-name and generic products are bioequivalent (Carrier 2018). These citizen petitions can delay generic approval for months or more even when they are ultimately denied.

---

Sample blockades, safety protocol filibusters, and sham citizen petitions all are distinct anticompetitive practices, but they are connected by a common thread: in each case, brand-name manufacturers exploit the fact that prices of off-patent drugs are unregulated while entry remains restricted. Those entry restrictions, in turn, are designed with safety and efficacy in mind. In other words, efforts to protect quality enable brand-name manufacturers to take actions that inflate price.

Even when anticompetitive conduct does not involve direct exploitation of quality protections, the entry barriers erected for quality-related reasons leave the market for off-patent drugs more vulnerable to collusion. For example, two executives at the generic manufacturer Heritage Pharmaceuticals pled guilty in 2017 to charges that they conspired with competitors to fix prices (DOJ 2017). A third executive at Sandoz, the generic unit of Novartis, pled guilty to related charges in 2020 (Blankenship 2020b). An ongoing set of lawsuits filed by forty-eight states alleges “rampant” collusion among generic manufacturers of prescription topical products (Stuart 2020). Quality controls—such as rigorous requirements to establish bioequivalence—facilitate this type of collusion by limiting the number of players in the market for any given drug and making it harder for new entrants to undermine existing cartels. In largely denying a motion to dismiss the states’ antitrust claims, a federal district court emphasized that the states “plausibly outline a regulatory regime” that would provide a motive for price fixing because “[h]igh barriers to entry . . . make an industry more conducive to collusion.”10

Robust enforcement of antitrust laws—as well as legislative changes such as the CREATES Act—may help to alleviate some of the costs of a market with unregulated prices plus significant barriers to entry. Ultimately, though, antitrust enforcement is treating a symptom of the underlying choice to prioritize quantity and quality over price. As long as prices are unregulated and barriers to entry are high, antitrust enforcement will continue to play an important role in the off-patent drug market—but even the most rigorous antitrust scrutiny will not escape the trilemma. For example, in an empirical study of high off-patent drug prices in the United States compared with other countries, Ganapati and McKibbin (2021) note the greater market power of U.S. drug suppliers as a factor limiting price decreases in U.S. generic markets. But they also conclude that given the high fixed costs of entry, price controls are likely needed to substantially lower generic prices. Of course, capping prices without adjusting quality requirements will implicate concerns about quantity.

4.2 Government Manufacturing

The primary strategy for managing the price-quantity-quality trilemma in the United States has been to promote competition through antitrust enforcement and to ensure quality through FDA oversight. One open question is whether these targeted interventions are superior to a more direct government role in supplying generic and biosimilar drugs. In

other words, can the federal government manage the price-quantity-quality tradeoff more effectively by acting as a police officer and seeking to root out anticompetitive behavior and unsafe manufacturing practices or—alternatively—by manufacturing safe and effective generics and biosimilars itself?

The proposed Affordable Drug Manufacturing Act—first introduced in 2018 by Representative Jan Schakowsky and Senator Elizabeth Warren—would pursue the latter pathway. The bill would create an Office of Drug Manufacturing, located within the U.S. Department of Health and Human Services (HHS), which would be authorized to manufacture—or enter into contracts with other entities to manufacture—certain drugs. The authorization would apply only to drugs for which patent protection and any period of market or regulatory exclusivity under Hatch-Waxman and the BPCIA has expired. Moreover, the authorization would be limited to drugs marketed in the United States by fewer than three manufacturers—and, even then, only if the price has increased faster than the rate of inflation within the past five years, the drug is included on the FDA’s drug shortage list, or the drug is listed by the World Health Organization as an essential medicine and the HHS secretary determines that the current price is a barrier to patient access (Warren 2018; Warren 2020). The federal government then would sell the manufactured drugs at a “fair price” determined by the new HHS office.

The government manufacturing approach merits strong consideration. As Warren (2018) notes, the idea has some precedent in the Strategic National Stockpile, an HHS-managed cache of vaccines and antidotes accumulated in preparation for a possible bioterrorist attack or other public health emergency. And the federal government is already a large direct purchaser of pharmaceuticals through the Department of Defense (DoD), the Department of Veterans Affairs (VA), the Indian Health Service, the Federal Bureau of Prisons, and the Department of State. The two largest purchasers, the DoD and the VA, spent nearly $15 billion on pharmaceutical procurement in 2018, which accounted for nearly 5 percent of total U.S. drug expenditures (CBO 2021). Federal pharmaceutical contracting has directly impacted an even larger number of Americans during the Covid-19 pandemic, when all Americans could freely access vaccines that the government procured from Pfizer, Moderna, and Johnson & Johnson.

One key difference between the Schakowsky-Warren proposal and the Covid-19 vaccine experience is that under the Schakowsky-Warren proposal, the federal government would not necessarily control the entire U.S. supply of the drugs that it manufactures or procures. The federal government would simply be an additional source from which pharmacies, hospitals, group purchasing organizations, and other prescription drug buyers could procure the drugs they need. Thus, the federal government would not have primary responsibility for allocating and distributing drugs across the country. Moreover, by overshooting on quantity and contracting with a diverse set of suppliers, the government could minimize the risk that a quality-control issue at one plant will trigger a national shortage.
To be sure, government manufacturing is not a panacea to the trilemma. Manufacturing (or arranging for the manufacture of) millions of prescription drug doses lies far outside HHS’s core competency—it remains to be seen whether the agency could amass the requisite expertise and logistical capacity to implement a large-scale manufacturing effort. And the risks of waste and mismanagement—familiar from other government procurement programs (e.g., Liebman and Mahoney 2017; Decarolis et al. 2020)—would remain. At the same time, it bears emphasis that the status-quo approach already imposes significant burdens on federal authorities, who bear responsibility for enforcing antitrust laws amid legal and economic uncertainty. Rooting out anticompetitive conduct in the patent afterlife is not necessarily easier for the federal government than introducing competition itself.

### 4.3 Promoting Innovation in Manufacturing

At the heart of the price-quantity-quality trilemma lies the challenge of scaling up production of high-quality drugs in response to a price spike or shortage. The substantial cost and time required to bring a generic drug or biosimilar to market means that an incumbent manufacturer can raise prices and reap large profits for an extended period before new entrants can undercut it, and drug shortages can persist for months or years before anyone else fills the void. A final set of solutions strikes at the heart of the trilemma by enabling faster, more flexible manufacturing processes that do not compromise on safety or efficacy. And unlike for patented drugs, where efforts to encourage innovation often come at the expense of access, innovation and access are aligned in the patent afterlife. That is, manufacturing innovations have the potential to increase access to a cheap and plentiful supply of high-quality off-patent drugs.

One of the most promising manufacturing innovations is a switch from batch processing to continuous production. Batch processing of pharmaceuticals can be analogized to baking cookies in a home kitchen. The various steps (assembling ingredients, mixing the cookie dough, dropping the dough onto a cookie sheet, baking in the oven, and cooling on a rack) occur sequentially. The entire batch moves onto the next step only after all the members of the batch have passed through the preceding step. By contrast, continuous manufacturing looks something like an assembly line. New ingredients may be fed into the assembly line even as fully finished products are rolling off the end. The switch from batch processing to continuous manufacturing has the potential to cut production times dramatically. As an FDA publication notes, “manufacturing that takes a month with batch technology might take only a day with continuous manufacturing” (FDA 2019a). The fast ramp-up means that continuous manufacturing methods could—in theory—respond rapidly to price spikes or shortages.

Notwithstanding the benefits of switching from batch to continuous processing, the FDA—as of early 2022—had approved only six drugs that utilize continuous manufacturing (FDA 2022c). A recent report by the National Academies of Sciences, Engineering, and Medicine attributes the slow adoption of continuous manufacturing partly to regulatory
As the National Academies report notes, the FDA evaluates manufacturing practices in the context of individual drug approvals. A manufacturer that sought to switch from batch to continuous production would need to obtain FDA approval for each drug affected by the switch. While the FDA has expressed support for continuous manufacturing (FDA 2019b; FDA 2022c), the costs—including regulatory costs—of changing to continuous manufacturing have dissuaded firms from making the switch for existing drugs.

Policymakers seeking to accelerate the switch to continuous manufacturing could pursue several strategies (Price 2014). For example, an expedited FDA review pathway for new drug applications based on continuous manufacturing could incentivize more firms to make the switch for more drugs. Federal funding for the development and implementation of continuous manufacturing could add further impetus. To that end, a bill passed by the House in October 2021 would authorize $100 million in funding over a five-year period for new “National Centers of Excellence in Advanced and Continuous Pharmaceutical Manufacturing” at institutions of higher education. (As of mid-2022, the legislation had stalled in the Senate.) If the federal government assumed a more active role in drug manufacturing—as under the Schakowsky-Warren proposal—it also could encourage or require the firms with which it contracts to implement continuous manufacturing themselves.

5 Conclusion

This chapter introduces the generic drug trilemma as a conceptual framework for evaluating policy reforms in the pharmaceutical patent afterlife. By highlighting the tradeoffs among three key goals of generic drug policy—low prices, adequate quantities, and high quality—the trilemma can help policymakers and analysts assess whether interventions that improve outcomes in one corner justify sacrifices elsewhere. The trilemma framework also inspires a search for policies that can loosen the tension among price, quantity, and quality—such as government manufacturing of generic drugs and government promotion of manufacturing innovations that enable rapid responses to price spikes and shortages. Finally, our analysis has focused on the U.S. context, but the price-quantity-quality trilemma is global in scope. One reason to begin with the domestic context is that the problems of price, quantity, and quality are more tractable in the United States, where resource constraints are less binding and a strong quality regulation regime already exists. Successful navigation of the policy trilemma in the U.S. context is likely to yield lessons for abroad. In the global context, access-to-medicines advocates have primarily emphasized one element of the trilemma—price—and have focused on removing barriers to generic entry. Yet serious quantity and quality concerns apply in less developed countries as well (WHO 2016; Chokshi, Mongia, and Wattal 2015). Our analysis of the U.S. context serves as a reminder that price is only one element of the global access-to-medicine
challenge and that a holistic approach to access-to-medicine problems will need to be three-dimensional.

References


https://www.ashp.org/drug-shortages/shortage-resources/drug-shortages-statistics

https://accessiblemeds.org/resources/reports/2021-savings-report


https://www.fiercepharma.com/manufacturing/fda-recommends-metformin-recalls-for-5-drugmakers-after-carcinogen-contamination


https://www.economist.com/business/2014/12/30/going-large


https://www.fda.gov/regulatory-information/search-fda-guidance-documents/quality-considerations-continuous-manufacturing
https://www.fda.gov/media/131130/download

https://www.fda.gov/media/150409/download

———. 2021b. “FDA approves naloxone injection to counteract opioid overdoses.”
October 18, 2021.

https://www.accessdata.fda.gov/scripts/drugshortages/default.cfm

Last updated June 16, 2022.
https://www.fda.gov/drugs/abbreviated-new-drug-application-anda/list-patent-exclusivity-drugs-without-approved-generic

June 27, 2022.


https://www.nber.org/papers/w27579

*Regulatory Focus.* April 8, 2015.

https://www.nber.org/papers/w29206


*Clinical Infectious Diseases* 65 (4): 613–18.


