Economics of the Pharmaceutical Market

Luca Maini*

April 26, 2021

See most recent version here

CONTENTS

*These notes are a work in progress. Please check back frequently for an updated version, and email lmaini@email.unc.edu for corrections.

1 PRIMER ON PHARMACEUTICAL MARKETS

Pharmaceutical products represent a large component of overall healthcare spending, but are arguably even more important in terms of their contribution to overall gains in health status over the past century. Thanks to pharmaceutical innovation we now have vaccines that protect us against once deadly infectious diseases, statins and beta blockers for controlling high cholesterol and high blood pressure, antiretroviral therapies for the treatment of HIV, and targeted oncologics to treat many types of cancers.

The key reason why pharmaceutical products have been so successful is that, by and large, they are very cheap to produce. As a result, most pharmaceuticals, when prescribed appropriately, are incredibly cost-effective. Conversely, more intensive treatments that require a lot of physician time and capital (such as surgery, therapy, etc.) tend to be a lot more expensive. Moreover, while drugs generally get cheaper to manufacture over time, medical and surgical treatments usually get more expensive, because physician salaries increase over time.

Despite the fact that drugs are cheap to produce, the debate around drugs—in the US, but also abroad—generally centers around their high price. This apparent contradiction is (partially) explained by a few peculiar features of this market, namely the fact that developing new drugs requires large investment costs, and that when new drugs come to market they generally enjoy a high degree of monopoly power because they are protected by patents from identical competitors.

Pharmaceuticals tend to be among the most regulated products on the market. Governments regulate which drugs can be sold, what prices can be charged, and who can be targeted by marketing campaigns (and these are only some of the things that governments regulate). In this initial section, we cover some basic history and facts about the pharmaceutical market.

1.1 A brief history of pharmaceutical products

The pharmaceutical industry as we understand it today originated in the second half of the 1800s. Some of the early companies that began to research and produce chemicals for the purpose of human consumption still exist today (Merck in Germany, GlaxoSmithKline in the UK, Pfizer and Eli Lilly in the US). Interestingly, a lot of these companies started as dye producers, and were able to later pivot to pharmaceutical manufacturing by leveraging their knowledge of chemical production processes.

The earliest drugs were synthesized or otherwise extracted from plants. This was the case with morphine, one of the earliest drugs, which comes from the opium poppy plant. In many cases, the active ingredient and its effect were known for a long time. What marked the transition to an actual "industry" was the focus on efficient manufacturing technologies. For example, the main component of aspirin (acetyl salicylic acid) had been referenced in written records since at least ancient Egypt. Starting around 1850, various chemists and companies came up with increasingly efficient ways of synthesizing the compound. The winner of this race was Bayer—who still produces aspirin to this day. Over the next couple of decades other breakthrough therapies were

discovered: insulin was first extracted from a cow's pancreas, and Alexander Fleming (accidentally) discovered penicillin.[1]

After that, technological progress on drugs followed three waves. In the first wave, vaccines and antibiotics all but eradicated most infectious diseases, such as syphilis, tuberculosis, dysentery, scarlet fever, whooping cough, meningococcal infections, and pneumonia. In the second wave, companies developed drugs to manage several chronic conditions: hypertension, high cholesterol and cancer. The third and final wave (which is currently ongoing) has seen the rise of what's often called personalized medicine: the development of drugs that can treat rare conditions, or target specific disease subpopulations that carry certain biomarkers. An example of this is Herceptin (trastuzumab), a monoclonal antibody that is used to treat HER2 receptor positive breast cancer.[2]

Prescription drug insurance in the US Prescription drugs were relatively affordable throughout most of the twentieth century, and, as such, were generally not included in health insurance plans (Table 1). The earliest examples of organized provision of pharmaceuticals were nonprofit union-operated pharmacies in health centers, funded by contributions of employers to union welfare funds. The first union to provide this kind of coverage was the International Ladies Garment Workers Union, in 1917 (Gondek, 1994).

The push to include prescription drugs in health insurance started with the Medicaid program. In 1967 (two years after Medicaid was established through the Social Security Amendments of 1965), 31 states offered some form of drug coverage through Medicaid. Private insurers then followed the government's example. For example, Blue Shield started its first prepaid drug insurance program in late 1967.

Medicare also included some partial drug coverage through Part B, but only for drugs that could not be self-administered, and had to be administered by a physician in an office/hospital setting. It wasn't until 2006 that Medicare introduced prescription drug coverage through Medicare Part D. Part D was originally proposed by Bill Clinton in 1999, but was only introduced under George W. Bush through the Medicare Modernization Act of 2003. Medicare Part D expanded prescription drug coverage among the elderly and made it cheaper, though by then, a large part of Medicare beneficiaries were already purchasing supplemental drug coverage on the private market.

1.2 Basic overview of drug discovery and manufacturing

The key characteristics of pharmaceutical production are the very large cost of development, and the relatively inexpensive cost of production.

[1] Oskar Minkowski and Joseph von Mering first figured out the link between insulin and the pancreas after discovering that removing the pancreas would give a dog diabetes.

[2] The culmination of this last wave is CRISPR, a gene-editing technology that attempts to treat diseases by editing the DNA molecules that carry the disease-relevant mutations. While no CRISPR-based therapy has been approved at this time, several companies have started human trials in the last year, and the scientists who pioneered this approach

Share of Prescription Drug Spending by Source of Payment, Selected Years, 1965–1998

Year	*Out-of-Pocket*	*Private Insurance*	*Medicaid*	*Other*
1965	92.6%	3.5%	0.0%	3.9
1970	82.4	8.8	7.6	1.2
1975	75.4	12.2	10.8	1.6
1980	66.0	20.1	11.7	2.2
1985	55.4	29.9	11.8	2.9
1990	48.3	34.4	13.5	3.8
1995	33.9	46.8	15.8	3.4
1996	31.6	48.8	16.1	3.5
1997	29.1	50.8	16.5	3.6
1998	26.6	52.7	17.1	3.6

Source: Report to the President: Prescription Drug Coverage, Spending, Utilization and Prices (Washington: DHHS, April 2000, Table 2-30).

Table 1: Table 2 from Berndt (2002)

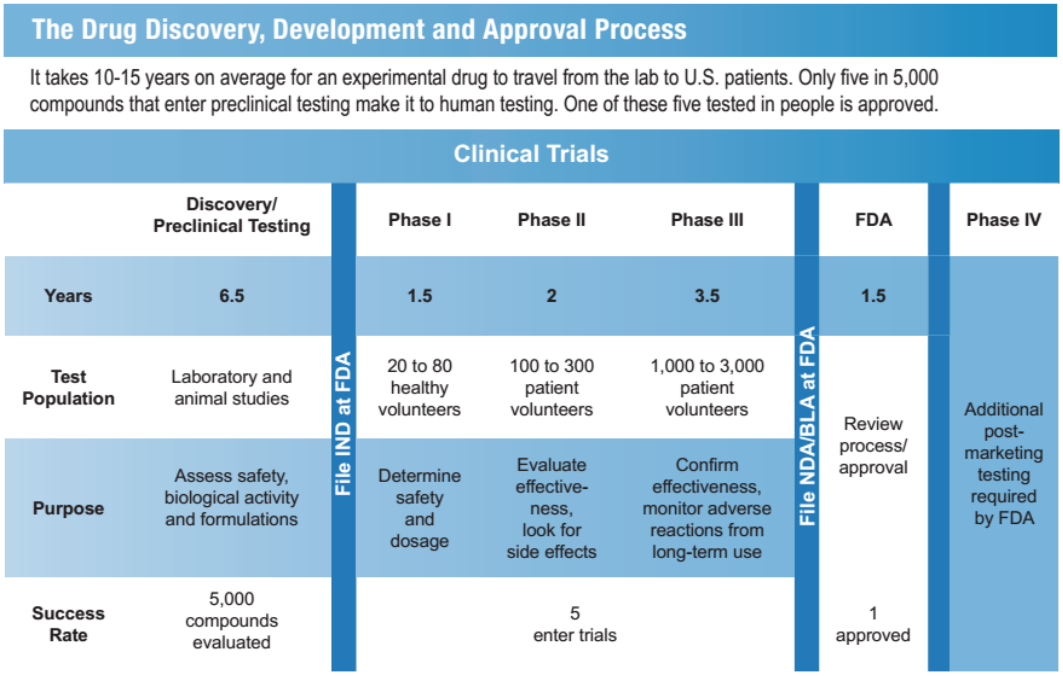
TOTAL EMPLOYMENT AND PLANNED R&D COSTS OVERALL AND IN SELECTED INDUSTRIES, 2011

Industry	NAICS code	R&D Costs paid by company ($mil)	Worldwide employees (thousands)	R&D Costs per employee ($)
Computer and electronic products	334	$77,887	2,951	$26,393
Pharmaceuticals and medicines	**3,254**	**$75,602**	**1,003**	**$75,376**
Publishing (including software)	511	$39,323	1,185	$33,184
Professional, scientific, and technical services	541	$34,407	2,799	$12,293
Transportation equipment	336	$31,639	2,596	$12,188
Machinery	333	$19,344	1,805	$10,717
All industries	31–33, 42–82	$365,211	29,327	$12,453

Source: National Science Foundation/National Center for Science and Engineering Statistics and US Census Bureau, Business R&D and Innovation Survey, 2011

Table 2: Table 1 (Lakdawalla, 2018)

Figure 1: Clinical Trial timeline



Source: *Research in Your Backyard - Developing Cures, Creating Jobs*, PhRMA.

The pharmaceutical industry is the most R&D intensive industry (Table 2). Drug development is costly both because ensuring that a pharmaceutical product is safe and effective requires many years of careful testing. Candidate compounds are first isolated and tested in the lab or on animals. Historically, the process of finding new molecular compounds happened almost entirely by trial and error. As a result, we don't actually know how most drugs work.[3] More recently, R&D has become more targeted, through the development of "rational drug design" protocols. These protocols usually involve matching of molecule characteristics to possible targets in cells and viruses. The mRNA-based COVID-19 vaccines are an example of this rational drug design. The vaccine works through a tiny snip of genetic code that is designed to stimulate a coronavirus immune response in the body (without actually exposure to the virus).

Once a candidate drug has been identified and shown to be effective in animal tests, it moves on to human testing. This occurs in three phases (Figure 1):

Phase I: small trial of (usually healthy) volunteers, meant to determine the highest dosage of the drug that can be safely administered to a patient.

Phase II: a slightly large trial of patient volunteers that evaluates effectiveness across various doses.

were awarded the Nobel prize this past year.

[3]This includes also drugs that have been used for decades, such as Tylenol and penicillin. The mechanism of action of Aspirin was also unknown for decades.

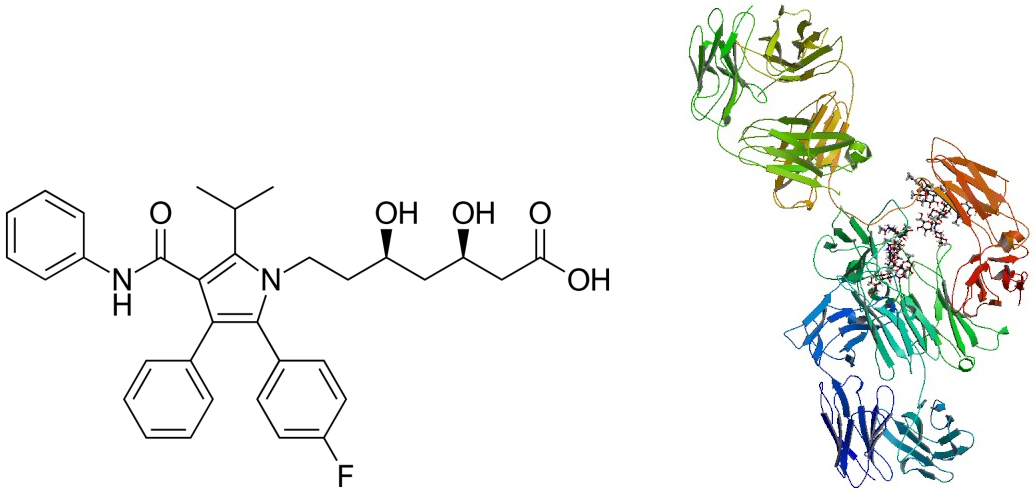


Figure 2: Small molecule (Atorvastatin, left) vs. biologic (bevacizumab, right)

Phase III: a large head-to-head trial against a placebo, or alternative therapy, meant to deliver evidence for a prospective marketing approval.

Finally, the last thing to note about the pharmaceutical R&D process is that disclosure requirements are much more stringent than in other industries. Firms are required to register and disclose the results of all clinical trials they run (in the US). In addition, a single patent usually covers the entire product (i.e. the molecule), which makes it easy for rival companies to replicate the innovation.[4] This means that the patent system is the main safeguard of intellectual property in the industry. This is quite different from what goes on in other industries, where products are usually the combination of multiple patents, and firms protect themselves through corporate secrecy.

Once a product has been approved, manufacturing tends to be a relatively straightforward and inexpensive process. As a result, the marginal cost of production of most drugs is negligible, and can often be safely approximated as zero. There are however, important exceptions and caveats to this rule.

Broadly speaking, there are two categories of drugs. The first categories is traditional, small-molecule drugs. These drugs are synthesized through established chemical processes that have been known and perfected over several decades. These drugs are cheap to make, and very easy to replicate. Within the group of small-molecule drugs, the ones that are simplest to manufacture, ship, and store are oral drugs in the form of pills and tablets. Many of these drugs are also available over the counter (that is, without a physician's prescription). Each step of the manufacturing and delivery chain is slightly more complicated for injectable drugs because these drugs tend to be more complicated to make and store.

The second category is large-molecule, so-called "biologic" drugs. Examples of biologic drugs include insulin, interferons (used to treat multiple sclerosis, some types of cancers, and hepatitis), and monoclonal antibodies (used to treat rheumatoid arthritis, psoriasis, and ulcerative colitis).

[4]Most manufacturers also take out additional patents to cover things like use, and manufacturing process.

These drugs are much more complicated, and are "grown" from organic (plant or animal) tissue (Figure 2). through genetic engineering that adapts or exploits the processes found inside living organisms. The presence of organic material in the manufacturing process means that these drugs cannot be exactly replicated without accessing the same originating cell line. Biopharmaceutical products, more commonly known as biologic drugs or biologics, are large, complex molecules derived from living sources. Some of the oldest drugs in existence are biologics. For example, most vaccines are derived from versions of the virus they are meant to inoculate against.[5] However, biologic drugs started becoming more prominent in the late 1970s, partly due to technological advancement coming from the discovery of DNA (among other things). The rise of biologics was accompanied by a switch in the industry that saw the rise of several "biotechnology" companies (biotech for short), which challenged the prominence of traditional "big pharma" companies. Some of these biotech companies (e.g. Genentech, Biogen, Genzyme) later became giants in their own rights.

Regulation of pharmaceutical products Regulatory oversight of the pharmaceutical market grew alongside the industry. In 1906, Theodore Roosevelt passed the Food and Drug Act, giving government the power to examine and seize food and drugs for reason of "adulteration" or "misbranding". The authority was originally held by the USDA Bureau of Chemistry, which was later reorganized into the Food and Drug Administration (FDA).

The FDA gained the authority to pre-screen all new drugs in 1938, on the heels of several scandals and tragedies that involved worthless cures and dangerous products such as Radithor (a radioactive solution purported as a cure-all), Lash Lure (a mascara dye that caused blindness), and Elixir Sulfanilamide (an improperly prepared antibiotic that contained a poisonous solvent). Shortly thereafter, the FDA started to designate many drugs as "prescription-only" (i.e. safe only under the supervision of a medical professional). Over the years, the agency developed several other protocols, such as the requirement that all new drugs demonstrated "substantial evidence" of efficacy (and not just safety), the obligation to use an established chemical "generic" name alongside the branded name, and restrictions in advertising to FDA-approved indications. Many of these were introduced in 1962 through the Keufaver-Harris Amendment to the original 1938 Act.

The flurry of reforms to the FDA authority also introduced some issues. The new clinical trial requirements significantly increased the time required to bring drugs to market. Since patents were acting as the main protection from competition, and since patents are taken out before starting clinical trials, the result was a much shorter period of market exclusivity. At the same time, a literal interpretation of the statute implied that the clinical trial requirement also applied to generic copies of approved drugs, which posed a significant barrier to entry.

[5]The first vaccine was the smallpox vaccine developed in 1796 by Edward Jenner, a British doctor. It consisted of a injection of a milder virus in the same family. In the late 19th century, American Home Products—a predecessor of the pharmaceutical company Wyeth—became the first company to mass-produce a version of the smallpox vaccine. Called Dryvax, their product was a live-virus preparation of another virus in the smallpox family: the vaccinia virus, from which all vaccines take their name.

The Drug Price Competition and Patent Restoration Act of 1984 (commonly known as the Hatch-Waxman Act) was passed to achieve a compromise between the need to guarantee a return to investment to developers of new drugs, and the desire to reduce the cost to bring generics to market. The Act established three important things:

1. It extended patent exclusivity of new drugs by guaranteeing seven years of exclusivity from the time of FDA approval (twelve for biologic drugs)

2. It introduced an abbreviated approval process for generic drugs that removed the requirement of clinical trials, replacing with a simple equivalency requirement.

3. It introduced some incentives for generic manufacturer to enter the market early on by providing a 6-month "generic exclusivity" period for the first generic to receive marketing approval.[6]

The generic drug industry really took off after the passage of the Hatch-Waxman Act. Generic drugs. Over the time period between 1984 and 1988, Grabowski and Vernon (1992) estimate that generic drugs could gather about a 50% market share within two years of entry, whereas today penetration reaches 90% within twelve months.

While the Hatch-Waxman proved to be very successful in shaping the traditional drug market, it did not address issues with biologic drugs, which at the time represented a vanishingly small fraction of the market. The main problem with biologic drugs is that it is impossible to create exact copies. It is still possible to create a "biosimilar" version of an existing biologic drug: one that contains the same active ingredient but may contain small differences in clinically inactive components.[7]. However, the abbreviated approval process for generics does not apply to biosimilars. As a result, prior to 2010, only one biosimilar product had been approved.

The Biologics Price Competition and Innovation Act (BPCIA) of 2009 attempted to address the gaps in the regulation of biosimilars by creating an abbreviated approval pathway for these drugs. Under the BPCIA, approval of a biosimilar product has two main criteria: (i) presentation of chemical data showing that the product is highly similar to the reference biologic; and (ii) evidence of equivalence with the reference biologic from a clinical trial. Although the BPCIA makes it easier to bring biosimilars to market, they still face three significant hurdles relative to generic drugs. First, the cost of developing a biosimilar is still much higher than developing a generic. Informal estimates place the cost of bringing a biosimilar drug to market between $100 million and $250 million, while the development of a generic small-molecule drug costs about $2–$5 million Grabowski et al. (2014).

[6]This type of generic entry, often called Paragraph IV entry in reference to the part of the Hatch-Waxman Act that describes it, applies in cases where the FDA-guaranteed seven year exclusivity period has expired, but (at least one of) the drug's patents are still valid. In these cases, a generic manufacturer can apply for FDA-approval and make an argument that the patent is not valid. In these cases, the manufacturer can challenge the entrant in court. These cases often settle with so-called "pay-for-delay" agreements (Jacobo-Rubio et al., 2020).

[7]As defined by the Food and Drug Administration (FDA), a "biosimilar" is a biologic that is highly similar to, with no clinically meaningful differences from, an existing reference or innovator biologic already in production. The phrase "meaningful differences" refers to the efficacy and safety of the biosimilars in comparison to the reference biologic.

2 Competition in Pharmaceutical Markets

There is a substantial gap between how media and conventional wisdom perceive competition in pharmaceutical markets and how economic research describes it. The most common complaint about drug prices is that they are too high, and that competition from therapeutic substitutes does nothing to limit it. Conversely, it is widely accepted in economic research that competition among pharmaceutical substitutes works just as well as in any other market, or at least as well as it could be expected given the frictions that exist in the market.

The reasons behind this discrepancy is simple: calculating the effective price of a pharmaceutical product is complicated, and oftentimes the effective price of a drug is not publicly available. The reason why calculating the price of a pharmaceutical product is complicated is because prices are usually reported in terms of physical units (such as a vial or pill, or a mass unit, like a price per gram of active ingredient). However, the value of a drug is not based on pure quantity, but rather on the effect that it has on a patient. As a result, the correct price should be expressed in terms of a treatment regimen, or—for maintenance drugs—yearly cost. Even when this adjustment is possible, most price points available to the public tend to inflate the effective price of drugs because of hidden discounts and rebates provided to large payers.

Table 3 provides an overview of several commonly used price points. With the exception of Average Sales Price (ASP) and estimated net price, all of these price points are best understood as list prices, e.g. upper bounds on effective prices that represent the starting point of negotiations between manufacturers and large institutional payers. Average Sales Price reflects some (though not all) discounts received by commercial payers, but is available only for drugs covered under Medicare Part B, and is calculated at the level of an HCPCS code, which is a categorization that often aggregates drugs producing similar molecules. Net price is an estimate that can be calculated by combining net sales reported in SEC filings with estimates of sales at list price levels. However, these estimated net prices include discounts from all types of payers (so there is no distinction between the statutory rebates to institutional payers and the rebates to commercial payers), and also include things that we would not normally consider as "rebates", such as copay assistance programs. Moreover, the data is only available for publicly traded companies (private companies don't need to disclose information to the SEC), and only for certain drugs. Data on invoice sales is also proprietary in general.

Older papers almost exclusively rely on list price data. Over time, as discounts became larger, this became more and more problematic. More recently, a few papers started using estimated net prices like the ones I described above. A good introduction to this data is Kakani et al. (2020). Their findings show that rebates grew from 32% to 48% between 2012 and 2017 (Figure 3; the numbers are probably a bit on the high side because their sample selects for relatively older drugs), with most of this growth coming from changes in within-drug rebates (and not, say a shift towards drugs with higher rebates). The findings tend to be pretty robust across classes.

Price point	Shorthand	Definition	Source	Level (relative to WAC)
Average Wholesale Price	AWP	Firm-suggested sales price for wholesalers	Firm	$\sim 20\% >$ WAC
Wholesale Acquisition Cost	WAC	Firm-reported price to wholesalers	Firm	
Direct Price	DP	Firm-reported price to individual buyers	Firm	Usually identical to WAC
National Drug Acquisition Cost	NADAC	Government survey data of pharmacy costs	CMS	Similar to WAC
Average Manufacturer Price	AMP	Acquisition cost for wholesalers	CMS	$\sim$ WAC for brand, lower for generics
Average Sales Price	ASP	Effective price	CMS	$\sim 20\% <$ WAC
(Estimated) Net price		Ratio of net sales to units sold	SEC	varies across drugs

Table 3: Pharmaceutical price points

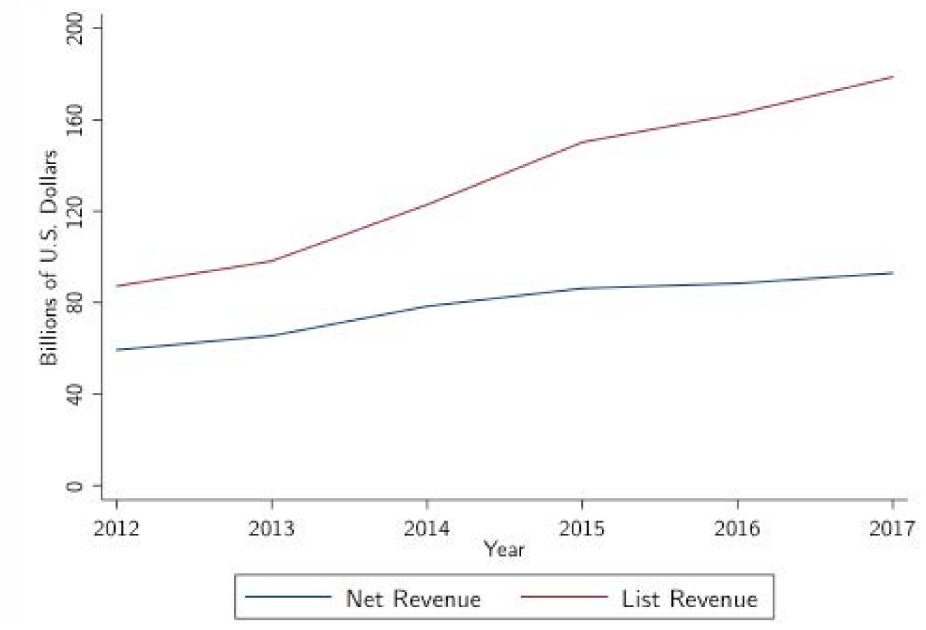


Figure 3: Annual rebate growth, 2012-2017 (Kakani et al., 2020, Fig. 1)

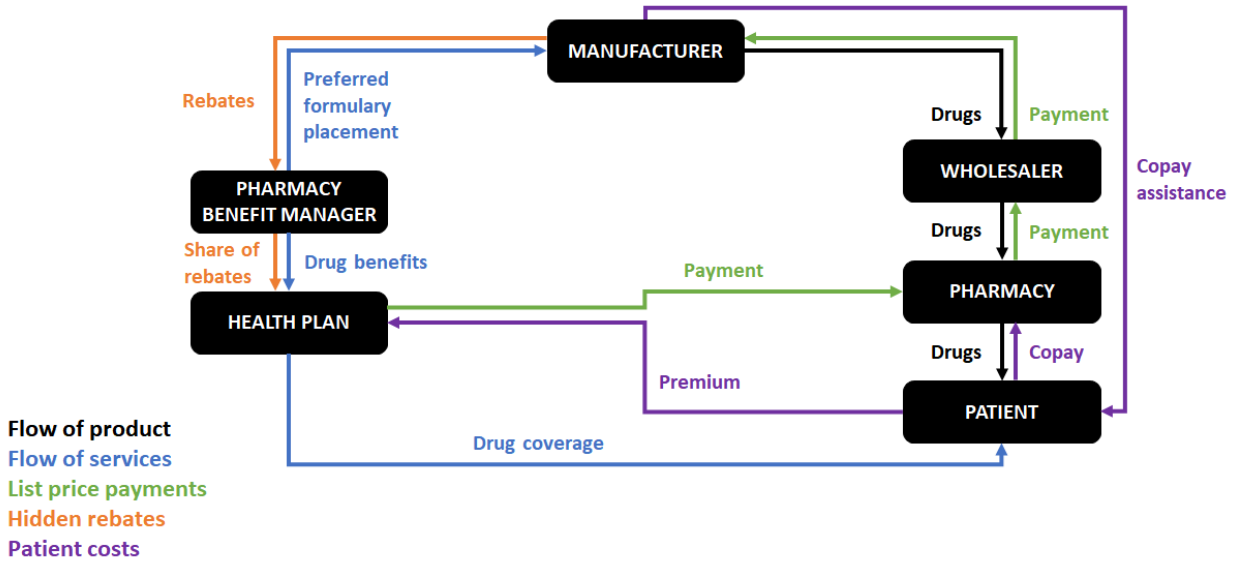


Figure 4: Simplified diagram of pharmaceutical market structure

2.1 Overview of pharmaceutical market structure

Figure 4 provides an overview of the structure of the pharmaceutical market. The diagram is simplified, but is a decent representation of the market for retail prescription drugs (i.e. drugs bought and distributed by pharmacies).

The diagram distinguishes between two separate operational chains. On the distribution side, manufacturers sell drugs to wholesalers, which then distributed it to retail pharmacies, where patients purchase them. All the transactions on this chain occur at list price levels, with both wholesalers and pharmacies earning a small distribution fee, which oftentimes is expressed as a fraction of the list price.

On the contractual side, manufacturers negotiate contracts with large payers called pharmacy benefit managers (PBMs), which in turn package menus of "contracts" (called formularies) and sell them to insurers and health plan managers. The main role of PBMs is to negotiate discounts on the list price. These discounts are generally tied to specific placement on the formulary (formulary have tiers, and drugs in lower tiers receive more favorable cost-sharing status). PBMs earn money by retaining a fraction of the rebates they negotiate, and by receiving fees tied to other services they provide such as claim processing, and managing mail-order prescriptions.

2.2 Characteristics of demand for pharmaceutical products

Before diving into competition, it's helpful to review a few key features of demand for prescription drugs. As with most other healthcare services, the choice of pharmaceutical products is probably best understood as a joint decision between the physician and the patient.

The main difference is that many pharmaceuticals are maintenance drugs for chronic conditions, meaning that patients will take them on a regular basis. A large number of drugs falls into

this category: statins for high cholesterols, beta blockers for hypertension, various categories of drugs to treat mental health conditions, treatments for asthma and other respiratory problems, insulin for diabetes, etc. Repeated use means that patients have a chance to learn which drugs work and which don't, so experimentation and learning play a role (Crawford and Shum, 2005). At the same time, patients tend to exhibit inertia in their preferences (Feng, 2020), meaning that they much less likely to switch if they are already using a drug.[8] There are many possible explanations for inertia. First, if there are a lot of similar options that are all very effective, we would not expect a lot of switching. Second, many drugs require a bit of an adjustment period at the beginning. This implies that there are fixed costs to switching. Third, patients may exhibit behavioral biases that lead them to stick with the current medication.

More broadly, it's been argued that behavioral biases may play an important role in explaining certain puzzles of drug demand, such as why many patients exhibit partial adherence to medication even when the potential long-term consequences are dire (Baicker et al., 2015; Beaulieu et al., 2006). Behavioral biases or bounded rationality may also explain another important aspect of drug demand, which is the preference for branded prescription drugs even when much cheaper (and identical) generic versions are available (Bronnenberg et al., 2015).

2.3 Competition between branded drugs

The first thing to note about competition between branded drugs is that the evidence on this type of competition faces significant limitations due to lack of available data (see the discussion of list and net price above).

We start with Lu and Comanor (1998), which examines pricing determinants of pharmaceutical products using a simple reduced-form framework. They look at 144 new molecular entities launched between 1978 and 1987. The fact that these drugs are pretty old presents obvious drawbacks (results may no longer apply), but also advantages, because the practice of hidden discounts was not as widespread at the time due to the more limited role played by insurance companies.

The empirical analysis of the paper consists of a simple linear regression design of (i) a drug's introductory price (normalized against the average price of existing competitors), and (ii) a drug's price growth at 4, 6 and 8 years, on indicators of novelty, number of competitors, presence of generics in the class, and a few more.

The findings align with what you might expect from basic economic theory. Drugs that with fewer substitutes and that represent larger technological improvements have higher launch prices. Interestingly drugs who come on the market competing against a generic product tend to set higher prices. This finding can be interpreted in two ways. One possibility is that these drugs are unobservably of higher quality. This could be because a firm that comes to market against a generic competitor must be pretty confident that their product is superior (otherwise it wouldn't justify the development cost). A second possibility is that the best strategy of these drugs is to

[8] Even in Crawford and Shum (2005), learning is incredibly limited, since around 90% of the patients in their sample stick with the first drug they are prescribed.

"split the market", and charge a lot to patients who have strong preferences for their product. This strategy avoids direct price competition with generics, which generally have much thinner profit margins.

The price growth results are a bit more mixed, though the authors try to interpret them in the light of older theories of market penetration that compare increasing and decreasing pricing strategies for high- and low-novelty drugs. You can refer to the paper for a discussion of these results.

The main selling point of the paper in my view is how carefully it constructs pricing measures, taking care of translating reported prices per package to the implied cost of a daily dose, and carefully constructing substitute classes across drugs. Without this sort of careful variable construction process it's hard to compare prices across drugs. Keep this in mind as you read the literature because many papers are not as careful.

Are drug prices too high?

A common perception for certain categories of drugs is that their prices are too high relative to their clinical value. This is the case in oncology, especially for drugs targeting late-stage cancers. Many of these drugs are approved on the basis of marginal benefits, such as increasing survival by months, or even weeks. Figure 5 shows the kind of patterns that motivate this perception. Over time, the quality-adjusted price of new pharmaceuticals has increased from an average of about \$50–100K, to almost \$200K.

The concept that a good could be priced above its value should be questioned by economists. In general, firms can never charge a price above value. If they did, they wouldn't be able to sell anything. In the pharmaceutical market however, there are caveats that apply. First, value here is subjective. The value of an extra year of life varies across individuals. The statement that the price of a drug is more than its value usually implies an assumption about how to value a clinical benefit. The standard way to do so is to use a sort of conversion rate between benefit and dollars that works in two steps. First, the benefit of a drug is converted to Quality Adjusted Life Years (QALYs). This is a measure of how much the drug improves both the quality of life of a patient and his/her survival rate.[9] Then, the number of QALYs is multiplied by a \$-per-QALY constant to reach a "value". Values of this constant vary quite a bit. The UK, who uses these assessments to price drugs, uses 30,000 BP (roughly \$50,000). Institutions that do this in the US, use numbers between \$150–250K.

A second caveat is that patients may not understand the true value of a treatment. Several papers have examined the role of physicians in treatment decisions and found that patients tend to follow physician advice (see e.g Cutler et al., 2013; Dickstein, 2017). Since physicians (for the most part) do not bear the cost of a drug, they may not take into account the price when making

[9] Some part of QALYs is also obviously subjective. How can we compare, say, the burden of depression to the burden of a physical injury? In practice, what happens is that each medical condition is attached a weight from 0 to 1 (with 0 being death and 1 being perfect health) that's calculated based on surveys. Unsurprisingly, these weights can change across countries, ages, etc.

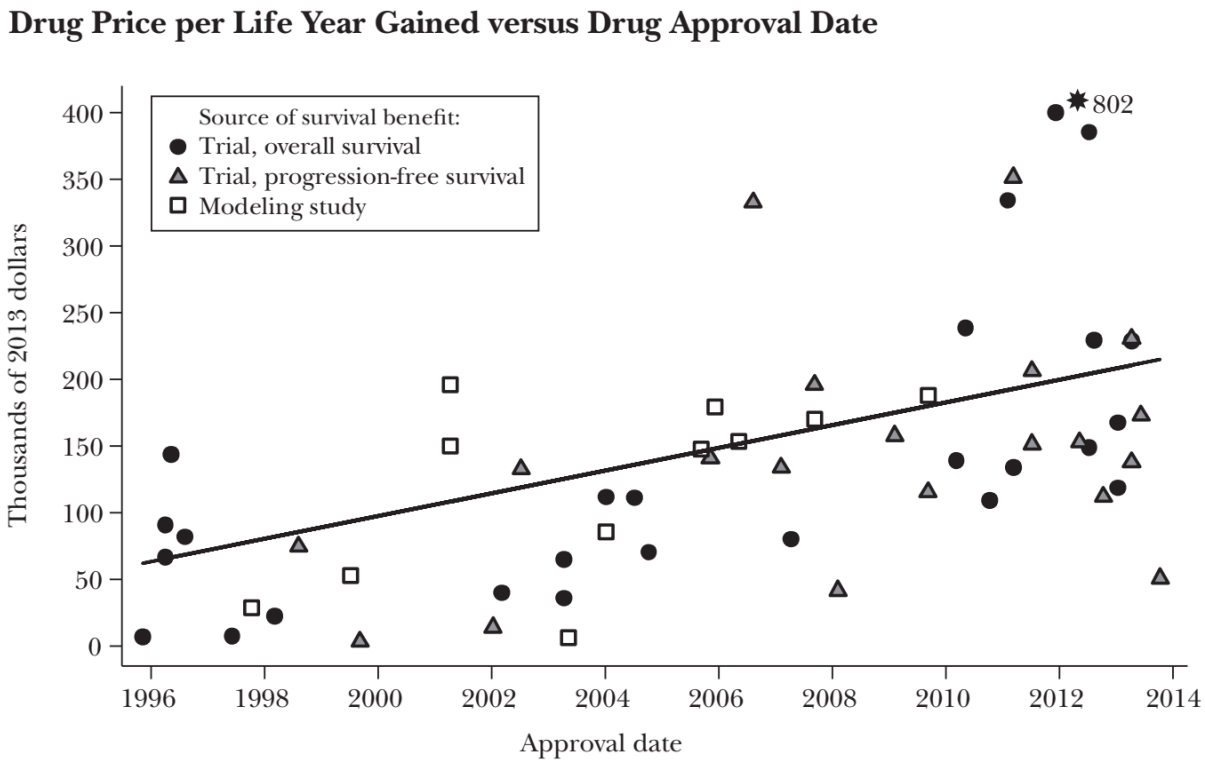


Figure 5: Figure 2 from Howard et al. (2015)

treatment decisions. This may imply that demand is very inelastic, and not necessarily tied to value.

A final caveat is that the presence of insurance allows for greater-than-value price even in a model with perfectly rational decision-makers. To see why, consider a simple model where insured patients pay a coinsurance α to purchase drugs. For simplicity, we let demand be linear and assume a price sensitivity of 1:

$$D(p) = \delta - \alpha p$$

This representation is helpful because we can interpret δ as the "value" of the drug. Then, a monopolistic price-setting firm chooses the price by solving

$$\max_p (\delta - \alpha p) \times (p - c)$$

where c is marginal cost. The first order condition implies

$$p = \frac{1}{2}\left(c + \frac{\delta}{\alpha}\right)$$

Hence, for values of $\alpha < \frac{1}{2}$, this model implies that $p > \delta$, i.e. price is greater than value. Notice that this is far from an extreme case. As Table 1 (shown previously) reports, out-of-pocket costs for

prescription drugs were below 30% in 1998, and likely have come down significantly since then.

Howard et al. (2015) look at this question and propose two alternative theories. The first is that drugs are priced according to a model of informal reference pricing. When asked about their pricing strategies, most pharmaceutical executives will point to a similarity between the price of their product and that of available alternatives, or other similar products on the market.[10] The theory is based on the idea that oncologists (or, more generally, physicians), will continue prescribe the newest and most effective medications as long as the prices are not outrageously high relative to older alternatives. There is some evidence supporting this. A famous case that appeared in the news a few years ago was Zaltrap, an oncology medication that offered basically the same effectiveness as an already existing drug, but at double the price. Several oncologists at Sloan-Kettering wrote an op-ed on the New York Times denouncing the fact, and vowing to not prescribe the product. The manufacturer soon cut the price in half.[11] At the same time, you may want to keep in mind that pricing close to your competitors isn't exactly a groundbreaking strategy, and that statements from executives should probably be taken with a grain of salt. After all, the prices used in the analysis are list prices, and in order to really understand what is going on you probably need to look at net prices.

The second theory put forth in the paper has to do with government regulation and is a bit more grounded in actual economic theory. The theory is based on firm response to incentives generated by the 340B program. This program was started in 1992 and required drug manufacturers to provide discounts (off of list price) to qualified buyers (originally, hospitals with a high-share of Medicaid patients). The 340B formula required a minimum statutory rebate of 15.1% (now 23.1%) off of the Average Manufacturer Price (a measure of list price, see Table 3 for more details). The rational response to this incentive may be to simply increase the list price. Notice that this justifies *high* prices, not necessarily *growing* prices. The growing part could arise because, over the years, more hospitals started qualifying for the program, due to a combination of Medicaid expansion and extensions of qualifying criteria (Figure 6).

Other papers

- Bhattacharya and Vogt (2003) proposes a theory of increasing prices and decreasing advertising over a drug's life-cycle that is motivated by the manufacturer's need to build public knowledge about the drug. The theory does match general patterns in the industry, though it's worth pointing out that several other models can predict the same pattern (e.g. a model of stock vs. flow of patients where low prices and high advertising are used at the beginning to build a customer base early on, or the model of history-dependent demand in Feng, 2020).

[10] Platitudes you may hear or read include saying that the price is "in line with other treatments", or "similar to the overall landscape", or "what the market can bear".

[11] See the article here: https://www.nytimes.com/2012/10/15/opinion/a-hospital-says-no-to-an-11000-a-month-cancer-drug.html

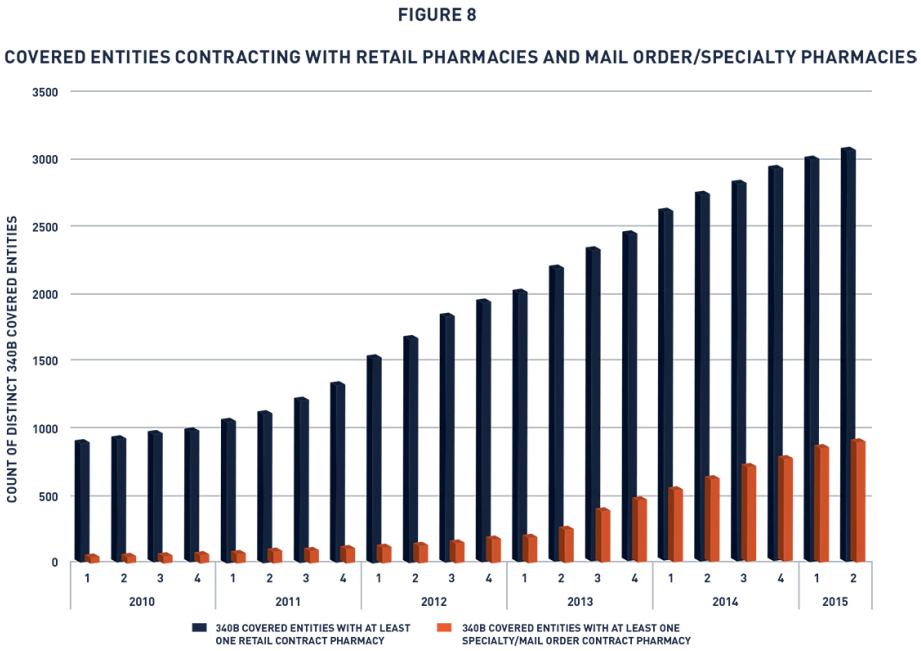


Figure 6: Number of 340B-covered entities

2.4 Competition with generics and biosimilars

What happens to prices and quantities when a drug loses exclusivity? The two main effects that have been identified in the data[12] are:

- The sales-weighted molecule price falls, often dramatically, as many cheap generic products enter the market
- The price of the branded product either remains stable or increases

The first effect is not particularly suprising: it's what simple models of Cournot or Bertrand competition (which apply because these are indeed homogeneous products) predict. The second one is a bit more surprising, because you may expect the branded product to attempt to compete with generic entrants. That's usually not what happens.

Frank and Salkever (1992) propose a theory for why branded products may increase prices following generic entry. The intuition is simple. As long as there is a fraction of consumers with strong brand preferences, the branded product will earn higher profits by trying to extract a higher per-patient surplus from this segment than by competing for the whole market at lower margins. The result is a split market where price-sensitive consumers purchase generic products, and a few price-insensitive consumers purchase the brand at much higher prices.

[12] There is other research on things like advertising, though the effects are usually more mixed. Scott Morton (2000) shows that brands generally do not use advertising to deter generic entry, and that in fact advertising may help generics (this is due to the fact that prescription for the molecule are usually automatically switched to generics by the pharmacist).

I present a simplified version of the model from Maini et al. (2021). The "loyal" market is a measure α of homogeneous consumers with valuation V_b for the brand product, and 0 for the generic product. Demand in the price sensitive market is a Hotelling model of price competition, where a measure 1 of consumers is distributed uniformly along a unit interval. The brand is exogenously located at 0, while the generic entrant is exogenously located at 1. Utility of purchasing the incumbent or the entry product for a consumer in location $x \in (0,1)$ is given by

$$U_b(x) = V_b - p_b - tx$$
$$U_g(x) = V_g - p_g - t(1-x)$$

where t is the travel cost, and $V_g \leq V_b$. The prices are assumed to be set sequentially, starting with the brand (this is to ensure a unique equilibrium).

The overall maximization problem of the incumbent is

$$\max_{p_b} p_b \left(D_L(p_b) + D_S\left(p_b, p_g^\star(p_b)\right)\right)$$

where

$$D_L(p_b) = \begin{cases} \alpha & \text{if } p \leq V_b \\ 0 & \text{otherwise} \end{cases}$$

and

$$D_S\left(p_b, p_g^\star(p_b)\right) = \min\left\{\max\left\{0, \frac{1}{2} + \frac{V_b - V_g - p_b + p_g^\star(p)}{2t}\right\}, \frac{V_b - p_b}{t}, 1\right\}$$

The segmented nature of the market implies that the incumbent has three options:

1. Capture only the loyal market and stay out of the price-sensitive market
2. Capture the loyal market and compete with the entrant over the price-sensitive market
3. Capture the entire market

The optimal choice depends on α, t and ΔV.

Equilibrium price and profits are straightforward to calculate for options 1 and 3. If the incumbent focuses on the loyal market only, it will set price $p_1 = V_b$, netting profit $\Pi_1 = \alpha V_b$. To capture the entire market instead, the incumbent must set price $p_3 = \Delta V - t$. Doing so nets profits $\Pi_3 = (1-\alpha)\, p_1$. The equilibrium outcome in the competitive case is more complicated, but one can show that the optimal price can be written as $p_2 = Kt + \frac{\Delta V}{2}$, with $K > 0$.

Figure 7 displays the optimal price as a function of ΔV for visual clarity. The key result is that, while price is generally increasing with respect to ΔV, a sufficiently low value of ΔV generates a discontinuous jump in price as the incumbent retreats to the loyal segment of the market.

You can see two patterns:

- As ΔV falls (i.e. perception of the quality gap to the brand declines), so do the returns to competing. At some point, the price required to compete will be low enough that it makes

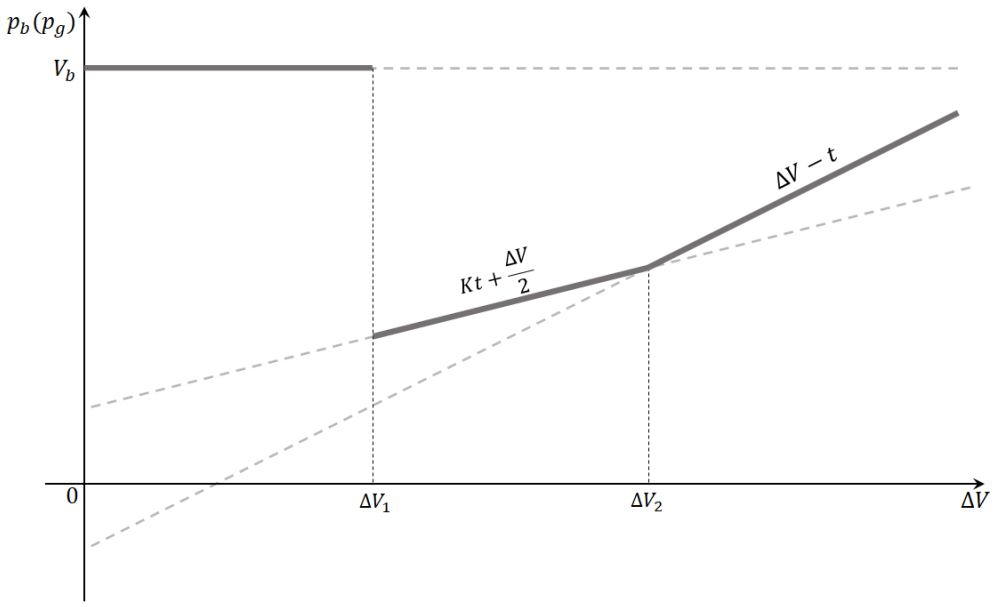
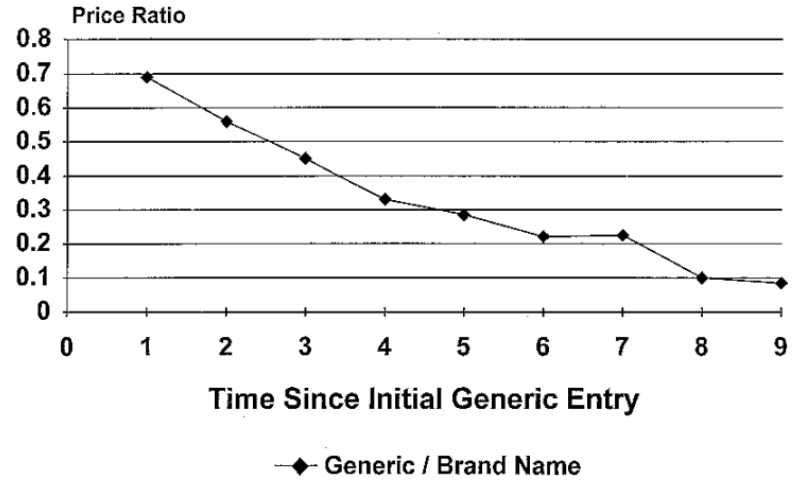


Figure 7: Optimal price of the brand as a function of ΔV

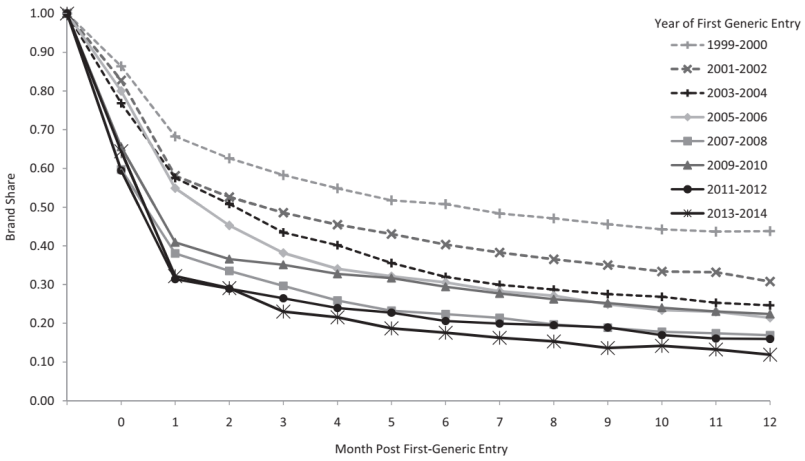
more sense to abandon the market altogether in order to be able to extract the maximum surplus from the loyal segment of the market. Conversely, for large enough ΔV, the generic stops representing credible competition in the price-sensitive market, and the brand can capture the entire market without needing to lower the price too much.

- A decline in travel cost t has two effects. On the one hand, it makes it easier for the generic to compete and decreases the brand's returns to competing. On the other hand, the decline in travel cost makes it easier for the brand to compete for consumers very close to the generic. Hence, capturing the entire market becomes a relatively more attractive proposition. The combination of these two effects implies that a fall in t makes option 2 look worse than both option 1 and option 3. At small enough t, option 2 will never be optimal, and the incumbent will either capture the entire market or leave the price-sensitive market to the entrant altogether.

The model above is used to describe competition between biologic incumbents and biosimilar (so not identical) competitors, which is a case we will tackle in the next section. When discussing brand and generic competition, we can safely assume that $t = 0$ (i.e. there are no real differences between the brand and the generic). The implication is that either the brand will be powerful enough to keep all generic entrants out, or, if a generic enters, that the brand will keep prices high and retreat to the loyal segment every time. The prediction does seem to bear in the data. Virtually no branded product competes on price with generics (Figure 8, Panel A), and loss of market share is substantial (Figure 8, Panel B).



(a) Ratio of generic-to-brand price (Figure 2 from Frank and Salkever, 1997)



(b) Brand market share after generic entry (Figure 5 from Grabowski et al., 2016)

Figure 8: Impact of generic entry on brand market share and prices

Competition between biologic brands and biosimilars

Biosimilar drugs are a relatively new phenomenon. The US introduced legislation to facilitate their approval only in 2009.[13] To date, only a handful of drugs have experienced competition from biosimilar entrants. The evidence on the impact of biosimilars is mixed. Even with a preferred path to approval, biosimilars face significant hurdles:

- There are currently no law mandating biosimilar substitution like with generics
- The accelerated pathway still requires clinical trials, and even without those the cost to develop of biosimilar is significantly higher (because biologics are much more complicated to

[13] Europe has had biosimilars for a bit longer. See Scott Morton et al. (2018) for an overview of biosimilar penetration and regulation in Europe.

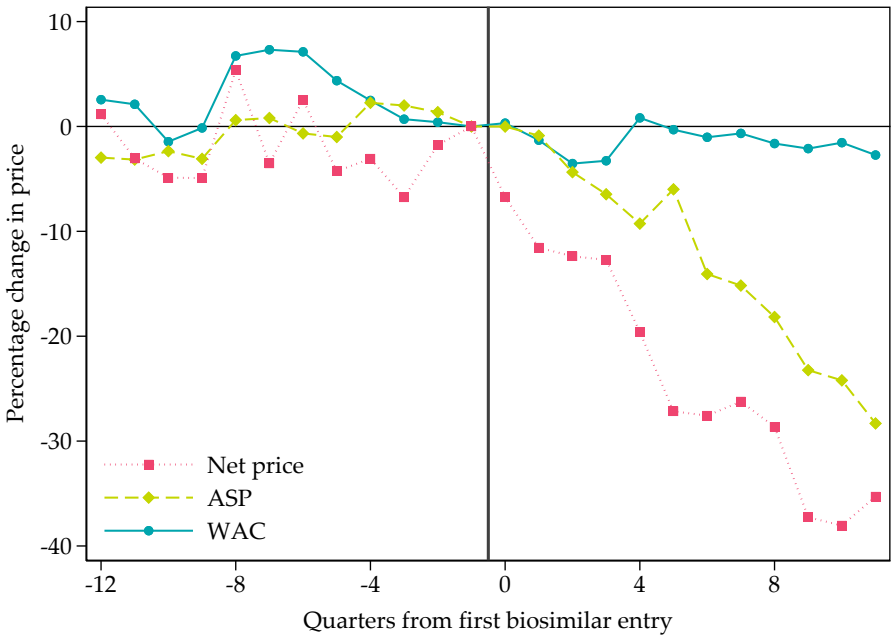


Figure 9: Change in price of originator after biosimilar entry (Figure 4 from Maini et al., 2021)

produce).[14]

- Biosimilars face much higher risk of litigation from the incumbent biologic manufacturer. Biologic drugs are protected by a much larger number of patents relative to small-molecule drugs, and infringement of any of these patents is enough to prevent the biosimilar from coming to market (Chen et al., 2017). There are currently more than a dozen approved biosimilars that have not been launched due to ongoing litigation or because of settlements with originator companies.

As a result, biosimilars have been slow to gain market share, with some medical researchers even arguing that the US government should abandon biosimilars in favor of price regulation (Atteberry et al., 2019; Trusheim et al., 2019).

Part of the difficulty in gaining market share seems to be due to the fact that—unlike with traditional generics—biologic brands have been reacting much more aggressively to biosimilar entry (Figure 9). One way to explain the effect is through the model from the previous section. If a biosimilar is perceived as horizontally different from a biologic brand, the brand will have more wiggle room to attempt to compete on the entire market. This could have further ramification if it makes it less likely that a biosimilar would enter (because an aggressive incumbent brand

[14]Informal estimates place the cost of bringing a biosimilar drug to market between $100 million and $250 million, while the development of a generic small-molecule drug costs about $2–$5 million (Grabowski et al., 2014). Also see remarks made by FDA Commissioner Scott Gottlieb to the America's Health Insurance Plans' National Health Policy Conference on March 7, 2018 (https://www.fda.gov/news-events/speeches-fda-officials/capturing-benefits-competition-patients-03072018, retrieved November 3, 2020).

means lower profits). There is some evidence that "simpler" and older biosimilars (which did not go through the abbreviated process, but had to undergo regular clinical trials, have both (i) been slightly more successful in penetrating the market, and (ii) elicited a milder reaction from incumbent brands (i.e. a reaction more in line with that of traditional brands against generics). However, these results should be taken with a grain of salt both because (i) the market is really young and firms may be learning how to compete, and (ii) the results are based on very few datapoints.

2.5 PBM–Manufacturer Negotiations

There are virtually no papers that tackle this topic. Two exceptions are Feng (2019) and Olssen and Demirer (2021). I cover them briefly here.

Olssen and Demirer (2021) This paper examines the role of rebates in Medicare Part D, with an eye towards the question of whether the government should negotiate rebates on behalf of insurers for this market. In Medicare Part D, insurers offer plans that are defined by a drug formulary, plus cost-sharing conditions (like deductible, copays, etc.). This paper focuses on the market for statins, which are cholesterol-lowering drugs used to prevent cardiac disease. In 2010 (the year their data is from), this market had three generic options and two branded options. Their focus is on the competition between these two branded drugs (Crestor and Lipitor). These two drugs compete for inclusion on two possible formulary tiers (preferred and non-preferred). Table 4, Panel A shows the composition of formularies across plans in their data, while Panel B shows market shares for each drug across all possible formulary configurations. As you may expect, better placement leads to a higher tier. Patients also seem to show an overall preference for Lipitor.

Their data includes both individual-level drug and plan choices, as well as formularies for each plan. They do not, however, observe any data on rebates.

Their model has three stages:

1. First, insurers and PBMs negotiate rebates
2. Next, insurers determine the formulary placement of statins on all the plans
3. Then, beneficiaries observe plan characteristics and formularies, and choose plans, and *simultaneously* choose which statin to purchase.

You should notice a certain similarity with models of provider-insurer negotiation we have already seen in previous chapters (e.g. Ho and Lee, 2017; Gowrisankaran et al., 2015). These papers model demand as sequential, Here, I think the simultaneity assumption is allowing for a more obvious selection channel (I'll explain why later). They do not actually solve the first stage, so the estimation stage is a slightly modified version of Ho (2006). They do however add a moment inequality estimation stage to recover bounds on the average rebate.

	Lipitor Tier		
Crestor Tier	Preferred	Non–Preferred	Off
Preferred	208 (53.2%)	72 (18.4%)	31 (7.9%)
Non–Preferred	10 (2.6%)	44 (11.3%)	1 (.3%)
Off	10 (2.6%)	0 (0%)	15 (3.8%)

Notes: Each cell shows the number of plans (percent of plans) with the indicated formulary placement for Crestor and Lipitor. The sample consists of the 391 plans with at least 1,000 enrollees. Alaska, Hawaii, New Mexico, and Nevada are excluded because of sample size restrictions. Finally, 42 plans are excluded because they use the standard benefit schedule as opposed to a tiered formulary

(a) Formulary design (Table 1 from Olssen and Demirer, 2021)

	Lipitor Tier		
Crestor Tier	Preferred	Non–Preferred	Off
Preferred	9.4%, 24.8%	14.7%, 4.8%	8.4%, 0%
Non–Preferred	7.1%, 30.9%	8.4%, 13.2%	9.6%, 0%
Off	0%, 32.4%	-, -	0%, 0%

Notes: Each cell shows the market share of Crestor (before the comma) and Lipitor (after the comma) among beneficiaries on plans with the indicated formulary placement for Crestor and Lipitor.

(b) Market shares (Table 2 from Olssen and Demirer, 2021)

Table 4: Statin formularies

Starting from the bottom, and proceeding by backward induction, utility of patient i (type t) for choosing statin k on plan j is

$$v_{ijkt} = -\alpha_t^{OOP} OOP_{ijkt}\left(f_j\right) + \alpha_t^x x_{ikt} + \xi_{kt} + \varepsilon_{ikt} \tag{1}$$

where f_j is the formulary of plan j, and x_{ikt} are individual characteristics that can affect preference for branded statins. Notice some restrictions: logit shocks are statin-specific but not plan-specific. Out-of-pocket costs are calculated on an annual basis. The calculation is not straightforward, because it depends on the formulary placement of other drugs.[15] There are some assumptions they need to make here that imply their model is at least a little bit misspecified (e.g. no moral hazard in non-statin drug choice). Notice that since they include fixed effects ξ_{kt}, they are identifying α_t^{OOP} off of individual-level variation in annual OOP costs. Therefore these assumptions on OOP costs will matter quite a bit.

[15] Why? Because if other drugs are expensive you are more likely to go past the deductible and hit catastrophic coverage.

They combine this with plan-choice utility:

$$u_{ijt} = \beta_t^{v^\star} v_{ijt}^\star - \beta_t^p p_j + \beta_t^x x_j + \zeta_{jt} + \tau_{ijt} \tag{2}$$

where $v_{ijt}^\star$ is the utility of the preferred statin, p_j is the premium, x_j are observed plan characteristics (like deductible and copays), and the rest are the usual error terms. The analogous of the term $v_{ijt}^\star$ in Ho (2006) would be the term for the expected utility of the network. Here, because of simultaneity, patients know v_{ijkt} in advance (which really means that they see ε_{ikt} in advance). That's why $v_{ijt}^\star$ enters directly. The main difference with Ho (2006) is that you need a simulation step to account for ε_{ikt} (since the econometrician cannot observe it). Details of this step are on pp. 28-29 of the paper.[16]

The model of formulary choice uses moment inequalities in estimation. The intuition is that formulary choices generate restrictions on the level of rebates. The profit function of plan j

$$\hat{\Pi}_j\left(f_j, p_j, f_{-j}, p_{-j}, r_j\right) = \sum_{i=1}^{N}\left[s_{ijt}\left(f_j, p_j, f_{-j}, p_{-j}, \hat{\theta}_t^j, \hat{\theta}_t^k\right) p_j - c_{ijt}\left(f_j, p_j, r_j, f_{-j}, p_{-j}, \hat{\theta}_t^j, \hat{\theta}_t^k\right)\right]$$

where r_j is a vector of formulary-contingent rebates received by plan j.

The form they assume for the rebate is $r_{jk}\left(f_j\right) = \gamma_k\left(f_j\right) + \nu_{2,j}$ where $\gamma_k\left(f_j\right)$ is the drug-specific average rebate under formulary f_j, and $\nu_{2,j}$ is a plan-specific shifter. Notice the shifter does not depend on the specific formulary arrangement. The identification strategy here relies on the assumption that $\nu_{2,j}$ is insurer-specific. Hence, if an insurer h has plans $j = 1, \ldots, J_h$, then, for all such j, $\nu_{2,j} = \nu_{2,h}$.

They can then build moment inequalities that constrain $\gamma_k\left(f_j\right) + \nu_{2,h}$ based on the formulary choices of various plans. The moment inequalities are based on revealed preferences, which means that you can derive restrictions on $\nu_{2,h}$ based on the plan offerings. They provide the following intuition for the identification argument:

> "Consider a rebate menu that offered 100% rebates for branded statins on the preferred tier. Under such a menu, insurers would face no cost from branded statins. Every insurer would place both Crestor and Lipitor on the preferred tier because that would increase plan demand and premium revenues without increasing costs. [...] On the other extreme, consider a case with no (0%) rebates. Then insurers would face large costs from branded statins, and our profit functions imply that most plans would remove both Crestor and Lipitor from their formularies. However, only 4% of plans exclude both Crestor and Lipitor from their formularies; as a consequence, we reject the no rebate case." (p.30)

There are a few things that could go wrong with this approach. None of them (as far as I can tell) are an easy fix, and to some extent we should cut the authors some slack, since this is a complicated

[16] Throughout, I refer to the March 2021 version of the paper. This is a working paper however, so it's possible that it will undergo more edits.

market, and they are among the first to study it. That being said, these considerations may matter.

- Two extensions that obviously break the estimation strategy are formulary-insurer specific rebates, and modeling the choice of plans as a multiple-agent game. The first one may be a bit more of a concern.

- Insurers usually do not negotiate rebates. PBMs do. This matters more in Medicare part D because PBMs offer plans *and* act as intermediaries for some insurers. Many PBMs are also vertically integrated with insurers, which adds complexity.

- It's possible that offering additional formularies may come with higher administrative costs. This is ruled out I believe, though not explicitly.

- It's a bit hard to understand how substitution patterns in plan choice affect the moment inequality. The authors appear to argue that a lot of the terms in these inequalities get crossed out. It's not entirely clear whether that is obviously the case (if I remove a plan, it matters which other plans those subscribers move to, so all plans should be affected). In theory, their model gives them all the tools to take these substitution patterns into account. In practice, it's possible that doing so is computationally infeasible. The description of the estimation on pp. 30–32 doesn't quite explain how the estimation works in practice.[17]

Ultimately, their results imply that rebates for both drugs are in the 30–50% range. They then run counterfactual analyses that show how changing rebates would affect the formulary choices of insurance companies. In doing so they have to hold a few things fixed (most importantly premiums).

Feng (2019) This paper also examines the market for statins, but uses a different modeling approach that includes a description of the rebate-setting process. The data for this project is a combination of claims from Marketscan (these are commercial claims for working adults and their dependents), and estimated (average) net prices. The definition of a formulary is almost identical to Olssen and Demirer (2021): there is a low-copay tier, a high-copay tier, and exclusion. PBMs can choose any combination of the three over the three branded drugs considered (Crestor, Lipitor, and Zocor). Generics are assigned to the lowest tier automatically.

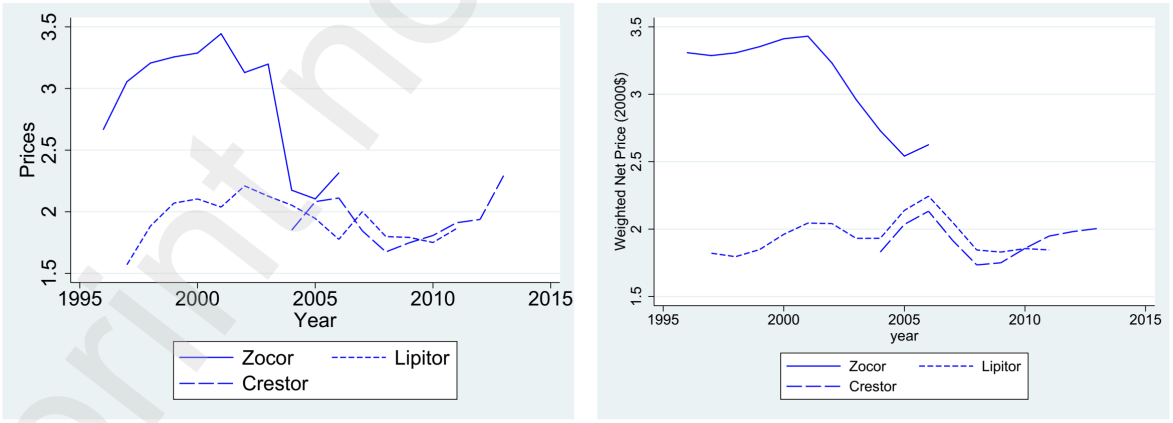
The model follows a few stages:

1. Manufacturers send rebate bids to a representative PBM

2. The PBM picks the formulary

3. Patients choose drugs

[17] It's also not entirely clear how the profit function is calculated. To some extent this is not surprising: Medicare Part D is tightly regulated, and rules imply insurers receive spending-dependent subsidies from the government. You also have to account for profits from all drugs, which is complex.



(a) Predicted Price (b) Data

Figure 10: Match of model to data in Feng (2019)

Relative to the previous paper, here there is more attention to the rebate-setting process, but the aspects of plan choice and competition are significantly downplayed. This is a reflection of differences in data sources. Olssen and Demirer (2021) can match formularies to plan enrollment and drug choice data, but do not have data on rebates, and only use a single year. Feng (2019) has some data on average rebates, but cannot match formularies to claims, and does not observe plan choices. Hence, the chocie of using a single representative PBM seems appropriate. An additional wrinkle of this paper is that it uses dynamic demand (this comes from an empirical result in a companion paper—Feng, 2020—which finds significant inertia in drug choices). As a result, the choices of rebate bids depend on a value function that carries past demand as a state variable.

Also notice the similarities and differences with the insurer-provider bargaining literature. The formulary here is similar to the provider network, but with more degrees of freedom since it's not only in-or-out, and rebates represent transacted prices. The model is simpler because there is no data on payer-specific rebates. The closest paper is Gowrisankaran et al. (2015), which also does not model plan choice.

Manufacturer j bids in period t are determined as the maximizing argument to the value function

$$V_{jt}\left(x_t; N_t\right) = \max_{P_{jt}} P_{jt} N_t \sum_{k=0}^{n} x_{kt} \mathcal{P}\left(m_t = j \middle| m_{t-1} = k, f\left(\mathbf{P}\right)\right) + \beta V_{j,t+1}\left(x_{t+1}\left(\mathbf{P}\right); N_{t+1}\right)$$

where $\mathbf{P}$ is a vector of premium bids, x_t represents the state variable, and N_t is an exogenous measure of market size. The probability inside the sum is essentially a conditional market share (probability that a patient who chose k in the previous period chooses j today). That's where dynamics come in.

Modeling the PBM's objective function is quite complicated, and the solution is to use a flexible control function that includes terms for consumer surplus, expected spending on drugs, aversion to outright exclusion from formularies, and a residual term for ulterior profit motives that PBMs

might have, such as a preference for generics (because PBMs earn some money from running their own mail-order prescription drug on the side):

$$K_t(x_t; N_t) = \max_{f_t} N_t \left[W(f_t | x_t) - \theta \sum_j P_{jt} s_{jt}(f_t | x_t) - \chi \sum_j \mathbb{I}(p_j = \infty) x_{jt} + M(s_t(f_t | x_t)) + \sigma_\omega \omega_{ft} \right] + \beta K_{t+1}(x_{t+1}(f_t); N_{t+1})$$

The PBM also has a value function because formulary decisions today will influence market shares, which is the state variable. Hence, formulary decisions today may have dynamic implications for the future. The point here is that PBM objectives are complicated. There is also a more "structured" approach to modeling PBMs in the paper, where many identical PBMs compete with each other, but it requires a lot of assumptions in order to make everything work.

The estimation searches for values of the various parameters (the weights in the PBM objective function, plus

> θ, the weight on expected cost, is identified by the average price levels in the data. The more PBM customers care access versus costs, the more likely drug companies will set higher prices, because they know PBMs will still put them in a favorable position. χ captures the exclusion penalty, which is identified by how much higher previous-period market share affect prices in the current period. [The parameters in $M(s_t(f_t | x_t))$ are] identified off any abnormal gap between Lipitor and Zocor prices in the pre-2003 period, [and] off lower brand drug price levels after generic entry, as they would face PBMs looking to switch people to generics. (pp. 23–24)

The model fits pricing patterns well (Figure 10), and it seems that dynamic incentives matter to some extent. The intuition here is that history-dependent demand generates incentives to start at lower prices and then slowly increasing them. This can (partially) justify why Zocor (an older drug) has much higher prices than Lipitor, even though Lipitor is a more popular drug.

3 PRICE REGULATION

Drug prices are among the most tightly regulated both in the US (where regulation generally applies only to drugs purchased by the federal government), and abroad. In this section I cover the most important examples of this regulation, and discuss how they affect the strategic choices of pharmaceutical companies.

3.1 Drug price regulation in the US

The two main pieces of price regulation in the US are the statutory formulas that determine how much Medicaid and Medicare pay for drugs that are administered to beneficiaries of these two

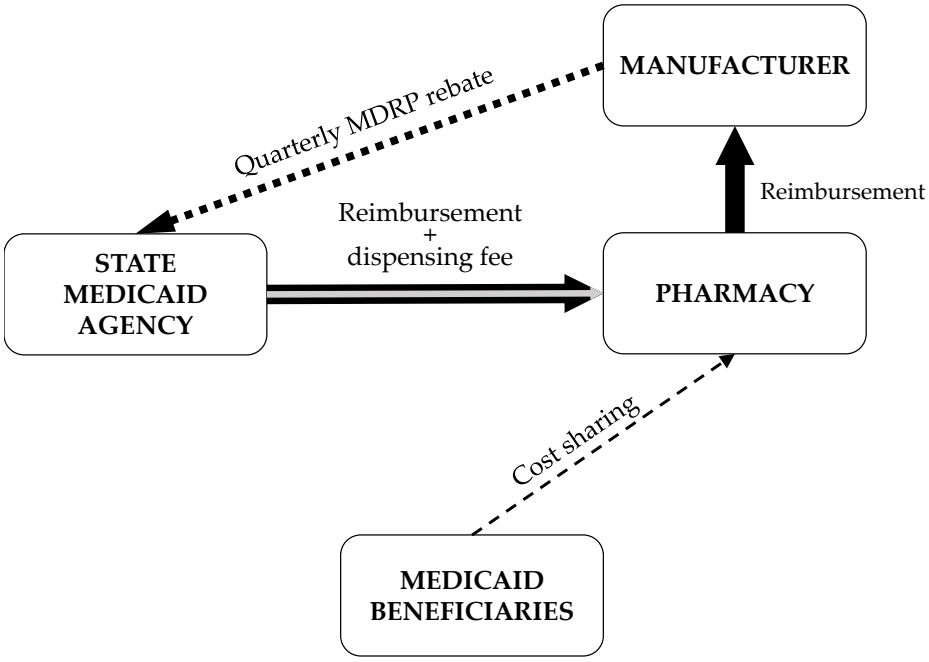


Figure 11: Stylized Flow of Medicaid Drug Payments

programs. In both cases, formulas are necessary because CMS is prohibited from negotiating prices directly with manufacturers. This is a provision in the Medicare Modernization Act of 2003, which established Medicare Part D. At the time, the perception in the pharmaceutical market was that having the government involved in the prescription drug market would ultimately lead to lower prices, and they wanted to have a stopgap to prevent further government involvement.[18]

Medicaid

All states offer drug coverage as part of Medicaid, and Medicaid is required to cover virtually all drugs approved by the Food and Drug Administration. Medicaid payments for most prescription drugs can be broken down in two steps (see Figure 11 for simplified schematic). First, state governments reimburse pharmacies that dispense prescriptions to Medicaid beneficiaries. Reimbursement rates vary across states, but are generally based on acquisition costs and include a dispensing fee. Medicaid beneficiaries usually pay minimal or no out-of-pocket costs. Then, manufacturers remit a quarterly rebate to states under a program called the Medicaid Drug Rebate Program (MDRP). The size of the rebate depends on a statutory formula that varies for innovator (i.e., branded) and non-innovator drugs (i.e., generics).

[18]The industry was right, in a way, since the introduction of Part D did give rise to the use of formularies, which in turn gave PBMs a powerful negotiating tool that lead to lower prices (Duggan and Scott Morton, 2010). However, Medicare Part D also expanded the market significantly by making prescription drug insurance more easily available to a large chunk of Medicare beneficiaries. The overall effect was a clear increase in revenue (which later papers exploited to study the relationship between market size and innovation).

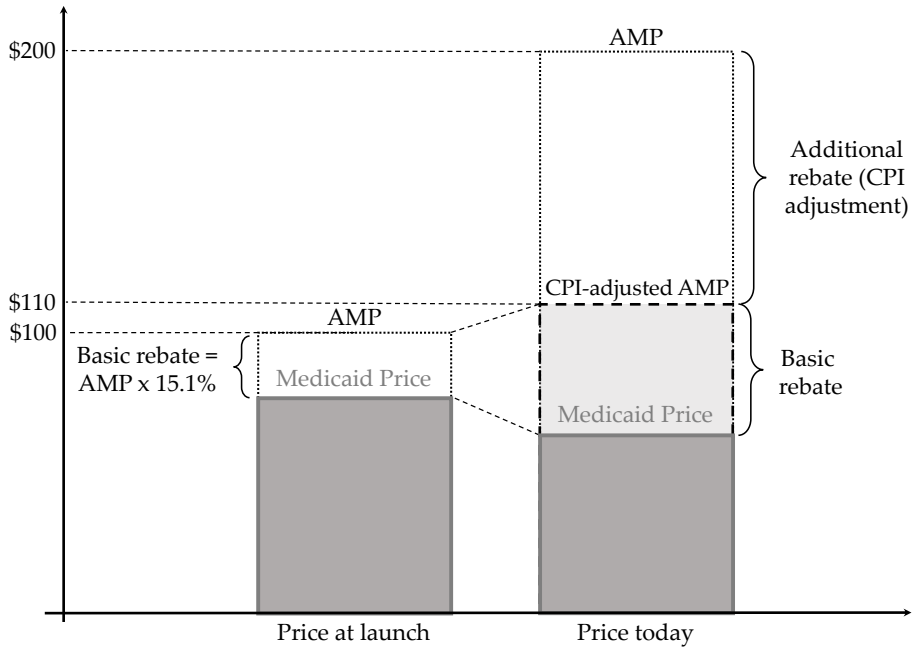


Figure 12: Cross-sectional Analysis of a Rebate Change

The basic formula for an innovator drug's rebate is

$$R = N \times p \times r$$

where N is the total number of Medicaid prescriptions in a given quarter, p is a measure of price, and r is the rebate fraction. CMS measures p from the Average Manufacturer Price (AMP). Crucially, AMP is an invoice price; it does not incorporate discounts provided to commercial and public payers.

The rebate r has two components. The first (known as the "basic rebate") is equal to the greater of a statutory minimum rebate and the largest discount offered to any commercial payer. Discounts to Medicare Part D plans, the VA, and certain other institutional and public payers are exempt from the calculation of this rebate. The second component, called the *inflation penalty*, is the difference between the current AMP and the inflation-adjusted AMP at launch. Many drugs experience price growth above the rate of inflation, so the inflation penalty often makes up a significant fraction of the overall rebate.

As Figure 12 illustrates, this formula entails some counter-intuitive implications. The statutory minimum rebate is expressed as a fraction of the *current* AMP, but the inflation penalty anchors the price to the AMP *at launch* (adjusted for inflation). As a result, increasing the AMP above the rate of inflation *decreases* the effective per-unit price that Medicaid pays.

The Medicaid rebate formula affects the pricing strategies of firms that produce drugs with high exposure to the Medicaid market in three ways.

1. To maximize the per-unit reimbursement, firms may try to **set a high initial list price** (Duggan and Scott Morton, 2006);

2. Firms are generally incentivized to **avoid giving out discounts above the statutory minimum rebate** to private payers (Scott Morton, 1997; Feng et al., 2020);

3. Firms should **minimize list price growth** over time (Feng et al., 2020).

Each of these three effects has been explored in the literature. Scott Morton (1997) was the first to point out that the Medicaid best-price clause would incentivize lower discounts. However, without data on rebates, she was only able to show this empirically for generic drugs (who generally do not give out rebates, and whose list price data tends to be accurate). Duggan and Scott Morton (2006) showed that drugs with a high Medicaid market share tend to have higher list prices.

Finally, Feng et al. (2020) exploit data on estimated net prices, and a 2010 change to the rebate formula to show that high Medicaid market share drugs give out lower discounts and tend to experience lower list price growth over time. The change in the rebate formula is a key component of the estimation strategy here, because it creates a natural experiment that can be exploited to draw causal inference. The policy change raised the minimum statutory rebate from 15.1% to 23.1%. Intuitively, this has the effect of relaxing the best-price constraint. For example, a firm that was giving out a best-price equivalent to a 20% discount would have triggered the best-price clause before 2010, but not afterwards. As a result, the drug manufacturer's bargaining position with commercial payers becomes worse. Intuitively, the manufacturer can no longer argue that increasing the discount would lower its Medicaid profits, so it may be forced to surrender a lower price.

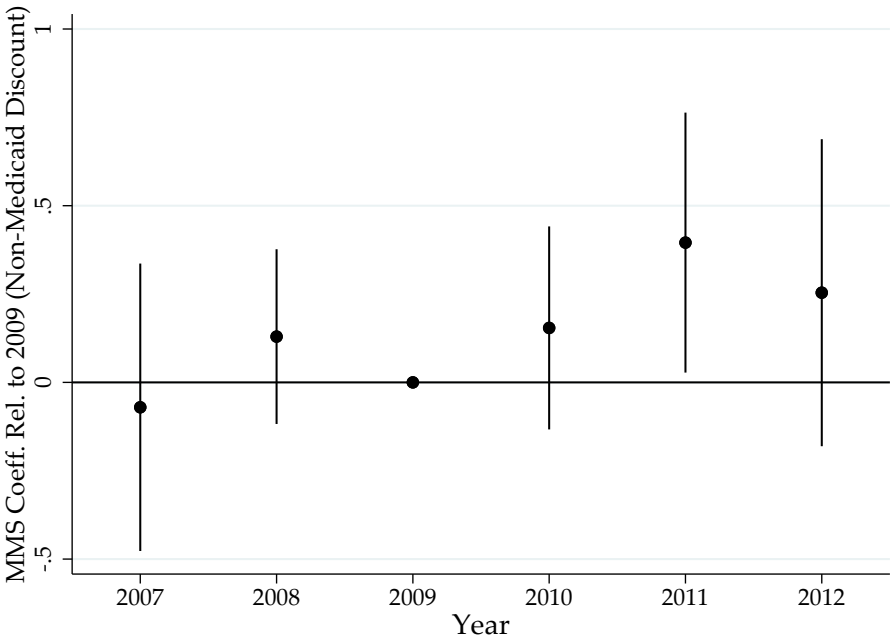
The paper uses a diff-in-diff strategy based on ex-ante exposure to the Medicaid market that relies on the idea that drugs with a high Medicaid market share will be more affected by the reform. The main estimating equation is

$$Y_{it} = \gamma_i + \beta_1 \times MMS_{i,2009} + \beta_2 \times MMS_{i,2009} \times \text{Post } 2010_t + \varepsilon_{it}$$

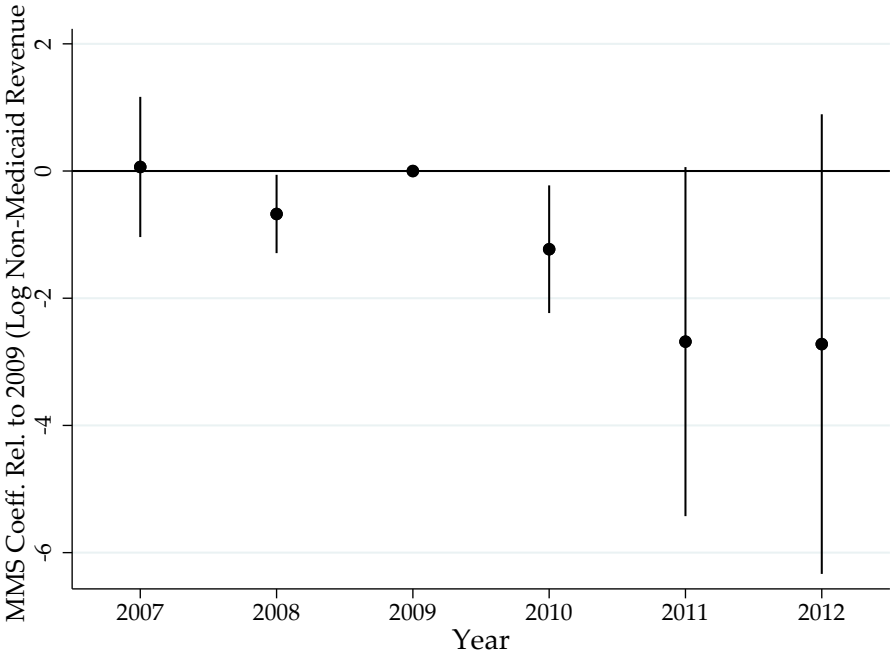
where γ_i are drug fixed effects, and $MMS_{i,2009}$ is Medicaid market share in 2009 (right before the reform). shows that the change lead to both lower prices and lower revenue (Figure 13, Panels A and B respectively). The overall implied difference amounts to about 2.5% lower net prices and spending on drugs in the non-Medicaid market, which translates to ~6–7 billion USD.

Medicare Part B

Medicare Part B operates in a slightly different way than Medicaid, partly because Part B covers only drugs administered directly by physicians, whereas most of what Medicaid reimburses is prescription drugs bought in pharmacies. The key difference is that Medicare reimburses providers, and not pharmacies. This matters because pharmacies buy drugs at list prices, but providers often don't, because they negotiate discounts with pharmaceutical companies. The formula is also



(a) Non-Medicaid Discounts



(b) Log net Non-Medicaid Revenue

Figure 13: Impact of Medicaid reform (Figure 5 from Feng et al., 2020)

different, as Medicare Part B does not have a best-price clause, or any inflation adjustment.

Before 2005, Medicare would reimburse physicians based on Average Wholesale Price (AWP). This is a measure of list price that does not reflect what providers actually pay. As a result, providers could make a healthy profit margin on most drugs administered under Part B. Under this policy, firms had an incentive to keep AWP high, but not because it would directly increase their profits (higher AWP does not necessarily mean providers are paying more). Rather, a high AWP meant a high profit margin for physicians, which could in turn boost demand.[19]

After 2005, CMS switched to a system based on reimbursing the Average Sales Price (ASP) plus a 6% markup. The ASP is a measure of the actual acquisition price of physicians, which CMS started collecting directly from pharmaceutical companies (technically, it is defined as the average price of the previous six months). The policy change implied two things. First, the average reimbursement price fell (sometimes by quite a bit). Second, the formula for reimbursement started being tied to actual transaction prices, generating an incentive for manufacturers to keep prices higher.

Ridley and Lee (2020) examine the implications of this policy change and find that it lead to higher prices for drugs covered by Part B. They use monthly prices for drugs sold in th US between January 1999 and December 2010. Their measures of price are WAC (throughout the entire period) and ASP (only after 2005).[20]

Their main estimating equation is

$$p_i = \beta_0 + \beta_1 T_i + \beta_2 Post_i + \beta_3 \left(T_i - T_{2005m1}\right) \times Post_i + \\ \delta New\ Molecule_i + \gamma Priority\ Review_i + \zeta_{c(i)} + h_i'\eta + \varepsilon_i$$

where T_i is the time in years between January 1999 and the launch date of drug i, and $Post_i$ is an indicator for months after the 2005 reform. They also include an additional interaction term between $Post_i$ and T_i to account for potential changes in trends. They expect β_2 and β_3 to be positive and significant as drugs increase their launch price after the reform. Since they use drug fixed effects here, the identification is coming from drugs who launch new formulations before and after the reform.

They report a very large effect: prices double after the reform (Table 5). One concern here is that drug prices are trending up over time, and the main analysis is basically using a regression-discontinuity design that doesn't have a clear control group. They do have a diff-in-diff specification, which delivers very similar results (Figure 14). Part of this may be the fact that they use list prices. Another concern is that they report null results for existing drugs, which contradicts the theory. Once again, this points to rebates playing a role. They are very open about limitations of the study (a whole section is dedicated to it), which is really good practice. Keep in mind that

[19] This is tied to the idea that physicians respond to financial incentives, which we discussed in the "Pay for Performance" chapter.

[20] They claim that WAC is close to the price paid by providers, though that's not quite true because WAC is a list price, and providers can take advantage of many discounts.

TABLE 5.—ESTIMATES OF THE EFFECT OF A REIMBURSEMENT CHANGE ON DRUG LAUNCH PRICES USING INTERRUPTED TIME SERIES REGRESSION

	log (price per ml in 2010 dollars)					
Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)
Post	1.902*	2.115**	1.803	1.937**	1.941*	1.983*
	(1.114)	(1.021)	(1.096)	(0.958)	(1.129)	(1.105)
Year 2006					0.0459	0.0398
					(0.858)	(0.861)
Year 2007					−0.243	−0.278
				(1.149)	(1.083)	
Year 2008					−0.438	−0.648
				(1.205)	(1.241)	
Year 2009					−0.414	−0.724
					(1.461)	(1.280)
Year 2010					−0.0610	−0.437
					(1.704)	(1.612)
Years since 1999	0.0747	−0.144	0.154	−0.00179	0.0949	−0.0685
	(0.104)	(0.191)	(0.234)	(0.228)	(0.243)	(0.231)
Years since 2005 × Post			−0.120	−0.231		
			(0.304)	(0.281)		
New molecule	1.186*	0.982	1.175*	0.967	1.186*	0.990
	(0.621)	(0.687)	(0.634)	(0.711)	(0.629)	(0.732)
Priority review	−0.772	−0.804	−0.778	−0.838	−0.797	−0.854
	(1.056)	(1.016)	(1.059)	(1.008)	(1.046)	(1.045)
Constant	1.896***	−1.124	1.647**	−1.719	1.825**	−1.344
	(0.429)	(2.499)	(0.819)	(3.067)	(0.801)	(3.140)
Competition controls	No	Yes	No	Yes	No	Yes
Class fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	233	233	233	233	233	233
R^2	0.82	0.84	0.82	0.84	0.82	0.84
Adjusted R^2	0.76	0.78	0.76	0.78	0.75	0.77

Price is wholesale acquisition cost. *Post* = 1 if the drug launch was 2005 or later. Standard errors (clustered by therapeutic class) are in parentheses. *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.

Table 5: Effect of Medicare reimbursement change on drug prices (Table 5 from Ridley and Lee, 2020)

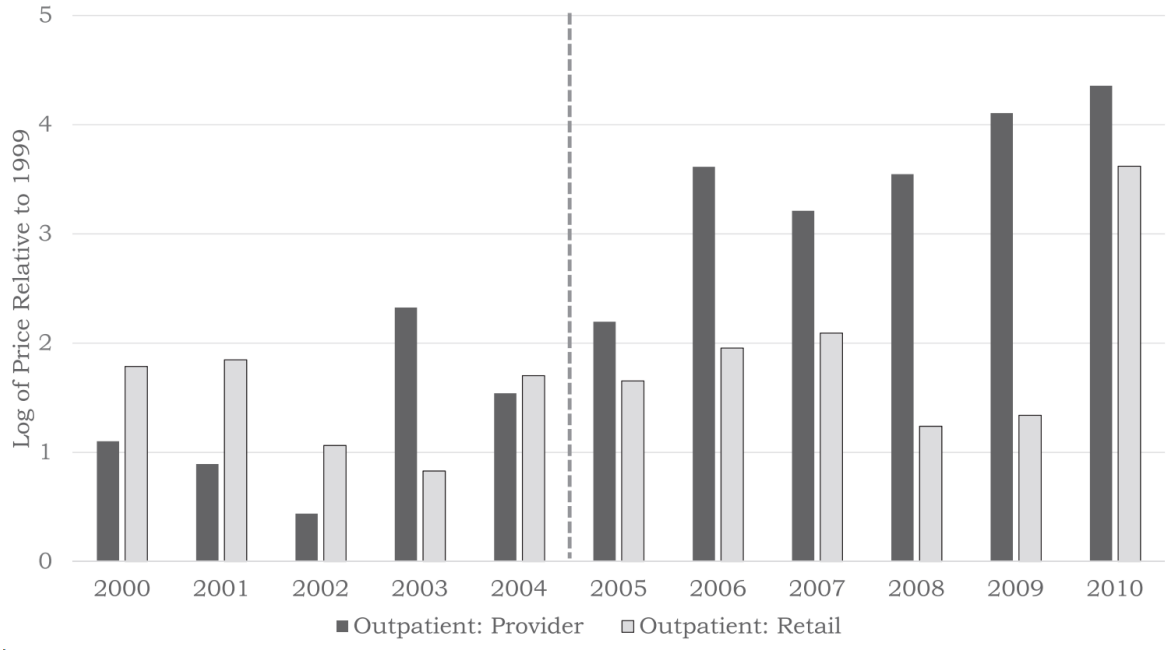


Figure 14: Price changes of outpatient and retail drugs (Figure 4 from Ridley and Lee, 2020)

no paper is meant to be the ultimate word on a topic, and consensus is built over many pieces of evidence. In this specific case, their work adds to what is a pretty substantial body of work showing that these types of government formulas have a big impact on the commercial market.

Other literature on US regulation

- Duggan and Scott Morton (2010) examine the impact of the introduction of Medicare Part D on drug prices. The paper documents a surprising phenomenon: the introduction of Medicare Part D lead to lower drug prices. The result is surprising because Medicare Part D expanded insurance access, and insured patients are less price sensitive (because they bear a fraction of the drug's cost). They point to the parallel introduction of formularies as the answer to the puzzle. Two things to keep in mind are that (i) even without formularies, a bargaining model may have delivered the same implication (because each payer now represents more patients), and (ii) the results are based on list price data, which shouldn't actually capture this bargaining/formulary channel.

- Yurukoglu et al. (2017) use the same policy change as Ridley and Lee (2020) and argue that it lead to drug shortages. The idea is that cutting reimbursements lead to reduced incentive to invest in capacity, which in turn decreased reliability of manufacturing plants. They use a dataset on drug shortages to compare shortages of Part B covered drugs to shortages of other drugs.

3.2 Drug price regulation abroad

In developing countries outside the US, governments usually negotiate the price of drugs directly. The two main tools used to do so are Health Technology Assessments (HTAs, also sometimes referred to as Cost-Effectiveness Analysis, or CEA), and reference pricing. The economic implications of HTAs are pretty straightforward. HTAs impose caps that apply to an individual country, and pharmaceutical companies generally accept those caps because marginal cost of production are very low (meaning there are always gains from trade).[21] The economic implications of reference pricing are more interesting, so I focus on those here.

Reference pricing is the practice of benchmarking the price of a drug using the price of either other similar drugs (internal reference pricing), or the price of the same drug in other markets (external or international reference pricing, ERP for short). ERP is the most common of the two.

ERP limits the ability to optimally price discriminate across countries with different income levels. Firms can react in one of two ways (or both):

- they can keep prices artificially high in lower-income countries whose prices are referenced by high-income countries, and

[21] The only interesting economic implications of HTAs have to do with innovation. You can argue that HTAs could depress incentives to innovate if they are used to keep prices lower, but there is also reason to like the "directional" incentives: prices tend to be higher for drugs that perform better in clinical trials, so firms would be incentivized to try and develop more valuable drugs.

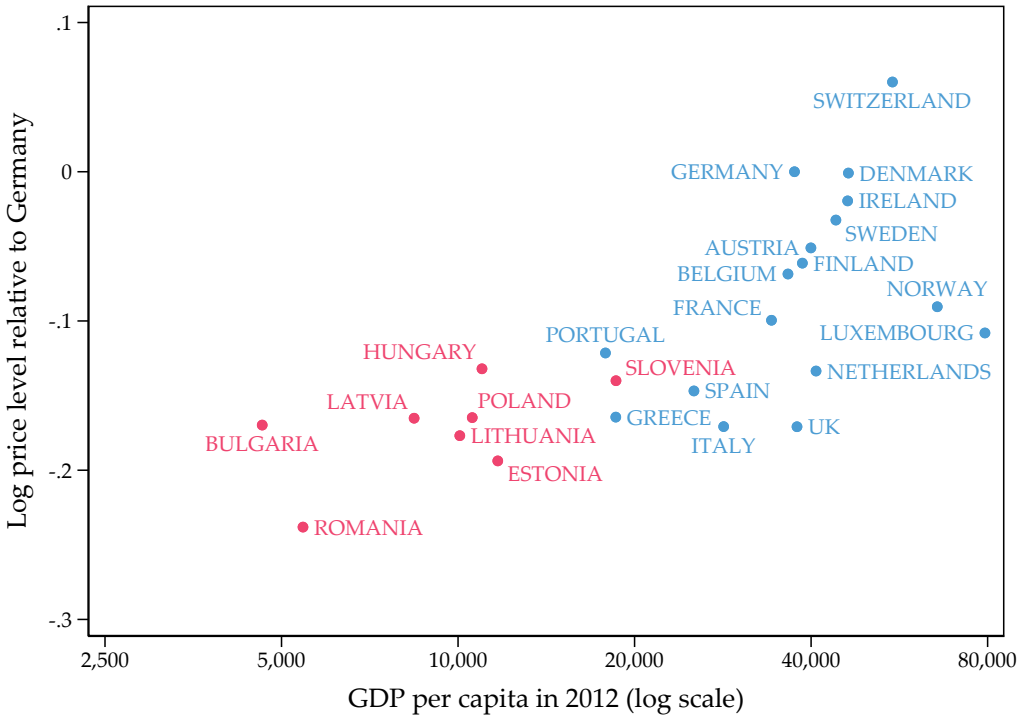


Figure 15: Correlation between drug prices and GDP per capita

- they can delay the launch of new drugs (in those same countries)

Both effects are fairly hard to detect in the data. In the first case, it's hard to quantify what "high" means. A simple plot of price levels against income per capita across European countries (most of which use reference pricing) reveals a non-linear relationship with prices flattening out at around €20,000 (Figure 15). This is suggestive of an active effort of pharmaceutical companies to avoid prices that are too low, but there is no way to easily assess the actual impact of the policy using a simple reduced-form approach.

In the second case, the difficulty arises because delays have many possible causes. (You should keep in mind here that delays in access to drugs are measured in the order of years, and sometimes decades, when comparing high-income countries and low- and middle-income countries). Reduced-form analysis generally finds that using price controls (including ERP) and adoption of weaker patent protection standards are associated with slower access (see e.g. Kyle, 2007; Cockburn et al., 2016). However, these papers model entry using hazard models that limit the ability of the authors to (i) isolate the effect of specific policies, and (ii) conduct counterfactual analysis.[22]

Here I cover two structural papers that examine these policies.

[22] You should also note that while strict price controls like HTAs could reasonably explain delays, the application of ERP does not, because ERP generates delays in *referenced* countries, not in the country that is doing the referencing.

The impact of reference pricing on prices and revenue: Dubois et al. (2019)

The motivation behind this paper is to try to figure out what would be the impact of introducing ERP to the United States. The most recent working version of this paper that I could find is centered around a policy that would cap US prices using prices in Canada. Since this policy has never actually been used in the US, the main counterfactual is a simulation based on demand estimates in both countries.

They estimate a static demand and supply model where demand is based on BLP, and prices are set through bargaining with the government in Canada, and through unrestricted Nash-Bertrand pricing in the US. Through this exercise, they recover the bargaining parameter of the Canadian government, and marginal cost estimates for all drugs.

Using the empirical estimates from the model they then simulate a counterfactual where US prices are capped at Canadian levels. To get some initial intuition for what we may expect to happen we can think back to the effect of the Medicaid best-price clause from Feng et al. (2020). In that paper, the presence of the best-price clause implies that prices will be higher in the commercial market. The same thing should intuitively happen here: firms will bargain more aggresively with Canada because they know that any price concessions will need to be matched in the US. The Nash Bargaining game in the counterfactual looks like this:

$$\max_p \left[\Pi_j^{US}\left(p_j^{US}\left(p, p_{-j}^{CA}\right), p_{-j}^{US}\right) + \Pi_j^{CA}\left(p, p_{-j}^{CA}\right) - \Pi_j^{US}\left(p_j^{US}\left(\infty, p_{-j}^{CA}\right), p_{-j}^{US}\right)\right]^{\rho_j} \times \left[\Delta_j W_{CA}\left(p, p_{-j}^{CA}\right)\right]^{1-\rho_j}$$

Here, the key is that $p_j^{US}\left(p, p_{-j}^{CA}\right) < p_j^{US}\left(\infty, p_{-j}^{CA}\right)$, so the manufacturer can credibly threaten to walk away if p gets too low. For comparison, the bargaining without US reference pricing would simply be

$$\max_p \left[\Pi_j^{CA}\left(p, p_{-j}^{CA}\right)\right]^{\rho_j} \times \left[\Delta_j W_{CA}\left(p, p_{-j}^{CA}\right)\right]^{1-\rho_j}$$

because the US market is completely separate.

Their results show that in the vast majority of cases, prices go up by a lot in Canada, and only fall marginally in the US (Figure 16). This is not particularly surprising, because the US is much larger than Canada (about 10 times larger in terms of pure market size), and so even a small price drop in Canada is incredibly costly for the US. The perhaps more surprising takeaway is that overall profits for the pharmaceutical market go up, while consumer welfare goes down. This is a bit strange because the pharmaceutical industry has been quite vocal in their opposition to reference pricing. However, keep in mind that these results approximate a market in steady state. If ERP was implemented suddenly, firms would probably be unable to adjust prices of drugs they have already launched, and would lose a lot of money on those. The effect on new drugs however, could be quite different, and may turn out to favor manufacturers.

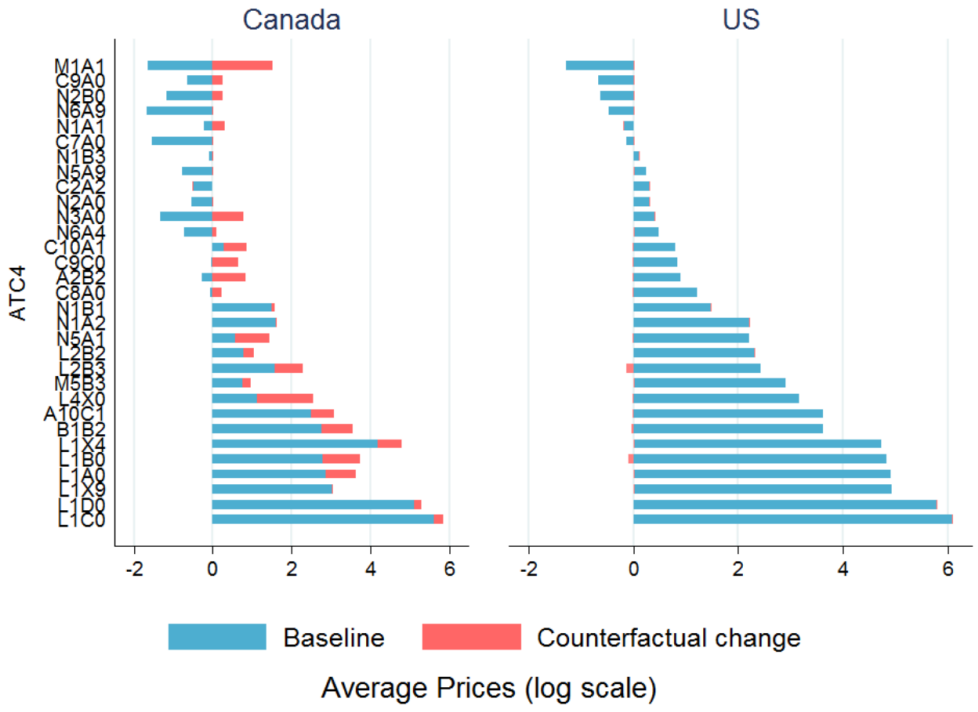


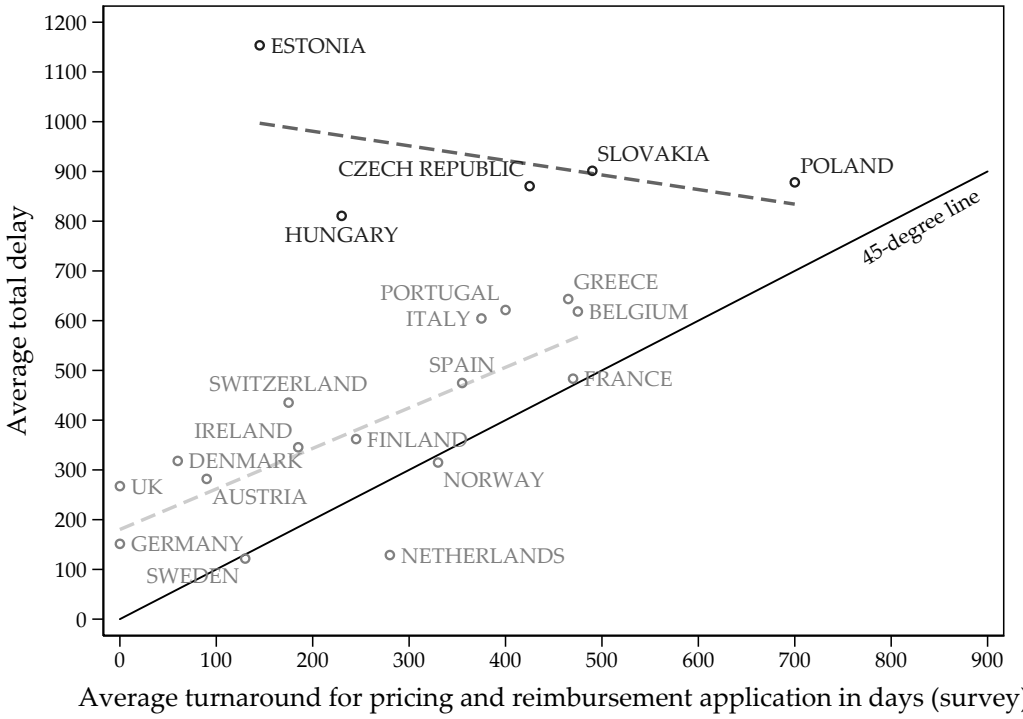
Figure 16: Counterfactual change in US & Canada prices from ERP (Figure 5.1 from Dubois et al., 2019)

The impact of reference pricing on delays: Maini and Pammolli (2019) This paper studies the impact of external reference pricing policies in Europe. The focus of the authors is on disentangling ERP-motivated delays from delays that arise for other idiosyncratic reasons. In Europe, drugs received marketing approval simultaneously in all Member States, but usually encounter additional entry delays because individual governments retain the ability to regulate prices independently from one another and drugs cannot be sold before the firm and the government agree on pricing terms.[23] Survey evidence suggests that these delays can be substantial (Figure 17: up to two years on average, depending on the country), but are not observed separately from strategic delays in the data, because the econometrician only observes the approval date and the launch date).

To isolate the impact of ERP the paper develops a single-agent model of entry that explicitly incorporates the price externality generated by ERP across markets. Intuitively, a manufacturer considering an additional launch faces a tradeoff between the additional revenue from the new market, and the possible fall in revenue in other countries through the reference pricing channel. The model then interprets delays in small countries with low price levels as strategic if these countries are referenced by larger countries with higher price levels. Conversely, delays in large markets with high price levels are usually interpreted as idiosyncratic. Variation across drugs and within country variation also matters because some countries have specific preferences for certain

[23] Technically, drugs *can* be sold, but if the government does not reimburse them, no one will buy them.

Figure 17: Average application turnaround versus Average delay



drugs.

Setup

To keep the estimation feasible, demand and price are estimated separately. Demand is a standard nested logit, but without a price coefficient. Price is expressed as a two-part, flexible control function that is a weighted average of an internal government benchmark p_{ijt}^{gov} and the reference price p_{ijt}^{ref} (which is based on previous period prices, with a functional form that is assumed to be known). The weights μ_j are country-specific, and estimated as part of the model.

$$p_{ijt} = \begin{cases} p_{ijt}^{\text{gov}} & \text{if } p_{ijt}^{\text{ref}} \geq p_{ijt}^{\text{gov}} \\ (1-\mu_j)\, p_{ijt}^{\text{gov}} + \mu_j p_{ijt}^{\text{ref}} & \text{if } p_{ijt}^{\text{ref}} < p_{ijt}^{\text{gov}} \end{cases} \tag{3}$$

Firms know demand and prices in advance, and choose when to enter in each market. However, rather than entering right away, firms have to apply in each country, and then their application is accepted with probability $(1-\psi_j)$. If an application is not accepted, it is delayed until the next period, when a new shock is drawn. The value function of the problem is

$$V_t\left(S_{t-1}\right) = \max_{\mathcal{A}_t=\{A_\tau(S_{\tau-1})\}_{\tau=t}^{T}} \sum_{S} \left(\sum_{\tau=t}^{T} \beta^{\tau-t} \Pi_\tau\left(S\right) \right) \cdot \mathcal{P}\left(S \middle| S_{t-1}, \mathcal{A}_t\right)$$

where the state space variable $S_t$is the entry sequence as of time t, S is the overall entry sequence, and

$$\Pi_\tau(S) = \mathbb{E}\left[\sum_j D_{ij\tau} \times p_{ij\tau}(S)\right]$$

The transition probabilities $\mathcal{P}(S|S_{t-1}, \mathcal{A}_t)$ depend on the actions of the firm, which are given by the set of state-specific vector $\mathcal{A}_t$, and on the random delay shocks.

The estimation has to get around a series of hurdles that I skip over in this overview. Basically, because the model is too complicated to solve, they use revealed preference moment inequalities. These are based on the idea that the observed entry strategy should maximize revenue (Olssen and Demirer, 2021 use an analogous intuition in their setting). There is an additional issue here, which is that the observed entry strategy may not be the optimal one (because the firm may have gotten unexpectedly delayed by the application process), but the basic structure of the moment inequality extends with some additional work.

The final estimates suggest that reference pricing is responsible for up to 1 year of delays in a set of lower-income Eastern European countries (a little over half of all observed delays). The model can also reject the hypothesis that ERP is responsible for delays in the remaining Western European countries. Curiously, the overall amount that firms earn by engaging in strategic delays is fairly small (around €20 million for the average drug), which suggests that paying firms with lump sums to give up delays may be a feasible strategy.

Other literature on international regulation

- Dubois et al. (2021) study drug procurement in low- and middle-income countries, and find that centralized procurement of drugs by the public sector leads to lower prices but that the induced price reduction is smaller when the supply side is more concentrated.

- Chaudhuri et al. (2006) study the welfare impact of the Trade-Related Intellectual Property Rights (TRIPS) Agreement, which increased patent protection standards for members of the World Trade Organization (WTO), in India. As a result of the agreement, local drug manufacturers were forced to withdraw from the market (to avoid infringing on existing patents). They find that withdrawal was linked to substantial consumer welfare loss from higher prices. (TRIPS was passed to incentivize *diffusion* of drugs, so the tradeoff is between faster entry of novel drugs and potentially lower access to drugs conditional on entry. The potential benefits of TRIPS are not quantified in the paper).

4 INNOVATION

The literature on innovation in the pharmaceutical market follows a few different strands.

- One part of the literature in this field studies the relationship between market size and innovation (which could be more broadly interpreted as the relationship between financial

reward and innovation). By and large, papers on this topic have focused on estimating how market size or revenue affect the number of new products on the market. Some studies also try to quantify the quality of marginal innovations. More recently, there has been some work on how consolidation affects R&D as well.

- Another part of the field focuses on questions that are almost closer to the operations research literature. These include questions examining what affects a company's choice to pursue a certain type of product/therapeutic area, how public funding and basic scientific knowledge contribute to private R&D, and how firms learn from each other.

- Lastly, some papers also think more broadly about the optimal way to set up a reward system for innovators. This is part of a broader literature that goes beyond pharma. While I do not cover much of this literature in this notes, a good starting point is Kremer (1998). Heidi Williams and coauthors (2013; 2017; 2019) have done some specific work on this looking at how patents on specific genes affect follow-on innovation.

4.1 Financial incentives and innovation

The literature on financial incentives and innovation is generally focused on estimating the elasticity of innovation with respect to profits (or revenue). Usually, this takes the form of the number of additional drugs that can be expected if profits increase by 10% or so. Empirically, there are usually two distinct approaches to this question. One tries to find instrumental variables that can explain market size, but are exogenous to the innovation process. The second relies on natural experiments based on health policy reforms and interventions.

The IV approach

The first paper in this space was Acemoglu and Linn (2004), which uses demographic change in the US as an instrument for market size, and found a very large elasticity (about 4). Here I focus on a similar, more recent paper: Dubois et al. (2015).

The paper uses a similar IV strategy, but uses global sales, and a more recent dataset. The focus of the analysis is empirical, but they do have a short model (most of the derivation is in the Appendix), that links entry and market size and incorporates a few empirically relevant aspects of the pharmaceutical market:

1. Endogenous entry

2. Varying levels of investment (also endogenous)

3. Horizontal differentiation among competing products

4. Lower margins in more crowded markets

The empirical analysis has two steps. First, they build a measure of expected revenue that is based on observed sales data by therapeutic category and country. This is not a simple accounting exercise because (i) they have only eleven years of data (which is not enough to capture the entire life-cycle of more than a handful of drugs), and (ii) expectations require predictions about what competitors will enter in the future. They use some imputation techniques to get around these limitations.

Second, they use an instrument for their variable of expected revenue. The instrument is motivated both by the fact that there is probably some reverse causality between innovation and revenue (e.g. if there are no drugs in a class, there will be no observed revenue), and because their measure of revenue is probably very noisy. The instrument they use is a measure of potential demand that they obtain by combining population data by age and gender, GDP, and mortality data (by country). The main threat to identification here is that innovation may affect some of these variables as well, especially mortality. The strategy will remain valid as long as these effects are not large in the short run.

The first model they estimate is

$$\ln N_{ct} = \alpha \ln W_{ct} + \gamma_c + \delta_t + \varepsilon_{ct}$$

where c is a therapeutic class, and t is a period. α is the parameter of interest and measures the elasticity of N_{ct} (number of new drugs) to W_{ct} (expected revenue). They estimate this equation using both OLS and 2SLS (using population and mortality by disease and GDP as instruments). Since N_{ct} is "truncated" at 0, and W_{ct} is unobserved when $N_{ct} = 0$, the linear regression may be misspecified. They estimate a zero-truncated Poisson maximum-likelihood estimation to get around that.

The estimates from the regressions vary between 3% and 30%, with an average of 23% and the estimate from the authors' preferred specfication around 14%. The implied revenue required to undertake the investment needed to bring one extra drug to market is around $2.5 billion, which is roughly in line with what one might expect given the approximate cost of R&D (~1 billion per new molecular entity), and the profit margins suggested by the industry (~50%).

The natural experiment approach

The second approach is to exploit a change in the incentives generated by policy reform. Examples include Finkelstein (2004) on vaccines, and two papers on the recent Medicare Part D expansion. I focus on these two papers here.

The first paper (chronologically), is Blume-Kohout and Sood (2013). Unlike Dubois et al. (2015), the main outcome variable that this paper is interested in is the number of preclinical and clinical trials undertaken by pharmaceutical companies. This measure has both advantages and drawbacks relative to a simple count of new drugs. On the one hand, it is a more precise estimate of the strategic choices of manufacturers (because failure to introduce new products may simply

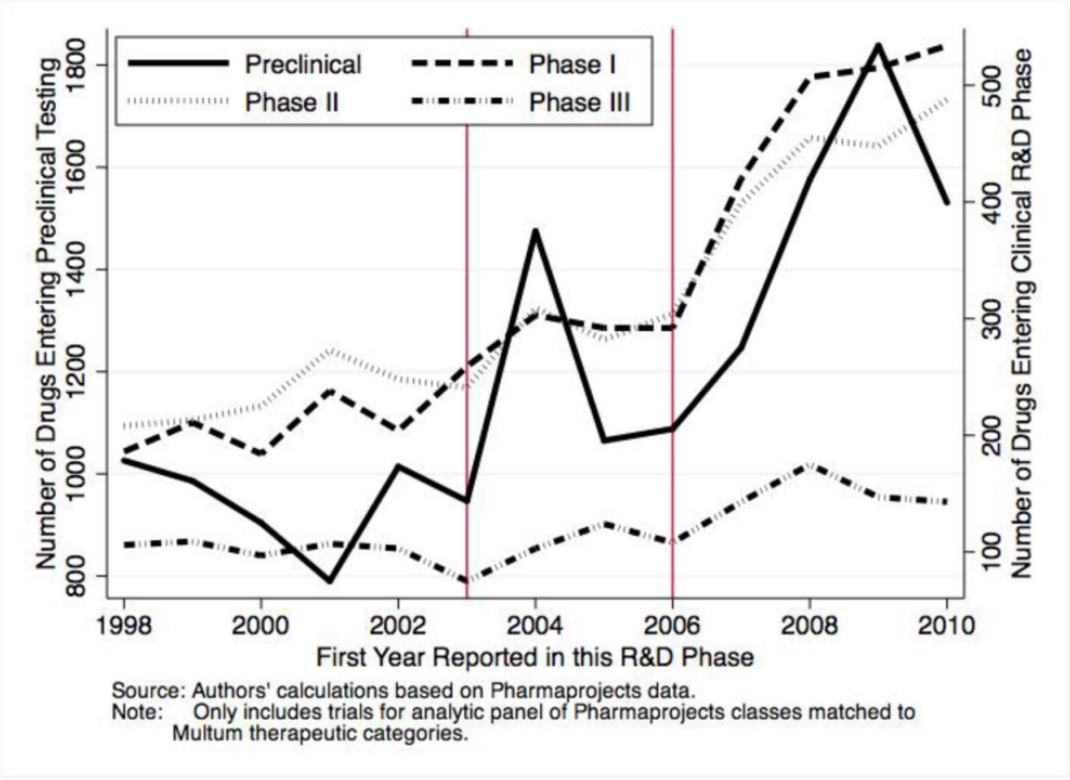


Figure 18: Number of Drugs in the R&D Pipeline (Figure 1 from Blume-Kohout and Sood, 2013)

be due to bad luck). On the other hand, it is a measure that is less relevant to consumer welfare.

The paper exploits the introduction of Medicare Part D as a plausibly exogenous shock to profitability that affected some therapeutic classes more than others. In particular, drugs that treat diseases with high relative prevalence among patients aged 65 and above were much more affected by the reform. The overall impact combines two channels. The first, and most obvious channel is the direct change in projected revenue for a new drug in one of these classes. The second is a liquidity channel: companies marketing drugs with high exposure to the Medicare market experienced an increase in their cash flow, which in turn might have had an impact on their overall R&D efforts.

Figure 18 gives an idea of the type of reduced-form patterns in the data that motivate the analysis. Medicare Part D was introduced in 2003 through the Medicare Modernization Act, and the market itself started in 2006. There seems to be an immediate reaction in Preclinical testing of new compounds, which later translated into a higher number of drugs in Phase I and II, including a small impact on the number of compounds in Phase III trials.

In the empirical analysis, they match clinical trials for new compounds to specific drug classes, and then calculate a Medicare Market Share for each class (basically a measure of the exposure of the class to the Medicare population. They then run a simple regression-discontinuity design to

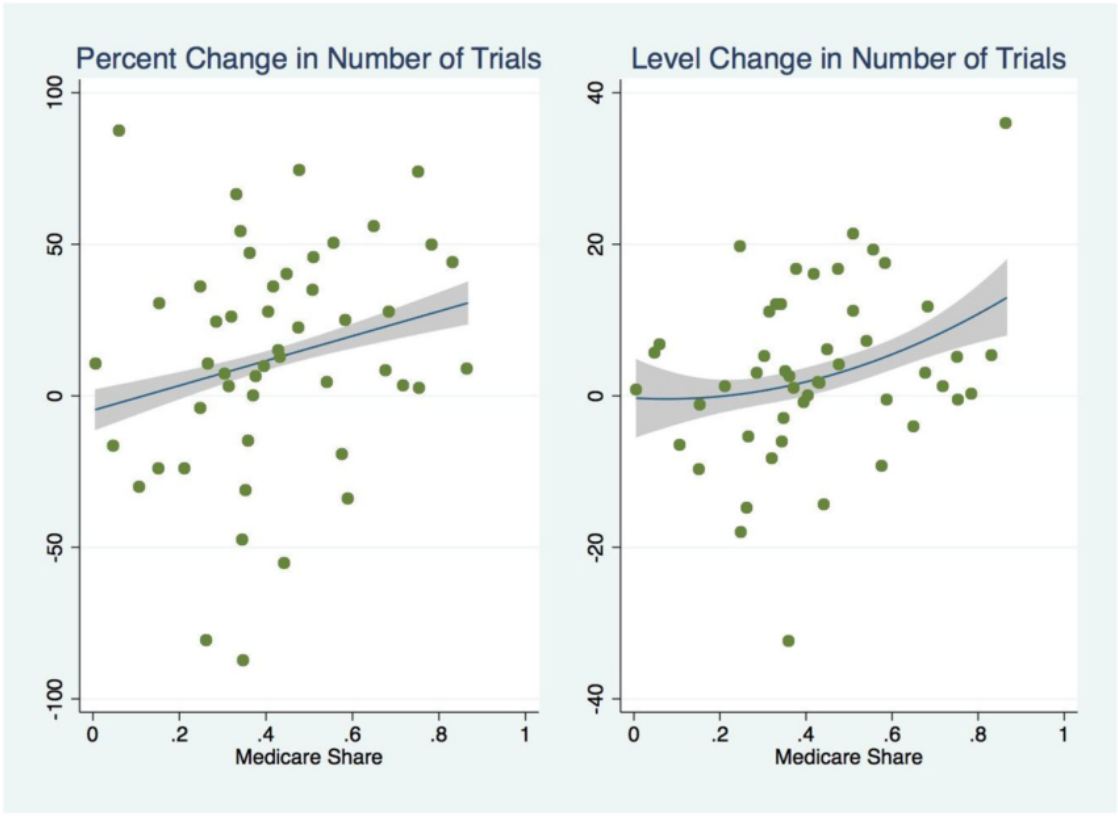


Figure 19: Relationship between Medicare share and impact of the reform on R&D (Figure 3 from Blume-Kohout and Sood, 2013)

estimate the impact of the introduction of Medicare in each class:

$$N_{ctp} = \alpha + \delta_t + \beta \times \mathbb{I}\{t > 2005\} + \varepsilon_{ctp}$$

where c is a drug class, t is a year, and p is a clinical trial phase. Figure 19 plots their results (aggregated across all levels of clinical trials), which imply an increase of about 3 trials for a drug with average Medicare market share.

The second paper is Dranove et al. (2020).[24] This paper uses a very similar framework to Blume-Kohout and Sood (2013), but uses an additional variable to measure the quality of the marginal innovations generated by the introduction of Medicare Part D.

The additional measure of quality they introduce is based on target-based actions (TBAs) associated to a drug. TBAs essentially represent the mechanisms through which a drug operates in the body. The idea is that drugs that use TBAs that have already been tested in previous drugs are less innovative than drugs that try TBAs that have not been tested yet. Their metric is a bit more sophisticated, and takes into account also the stage at which the test has occurred (e.g. Stage I-III).

[24] A similar paper had circulated as early as 2014 as an NBER working paper with the title "Pharmaceutical Profits and the Social Value of Innovation".

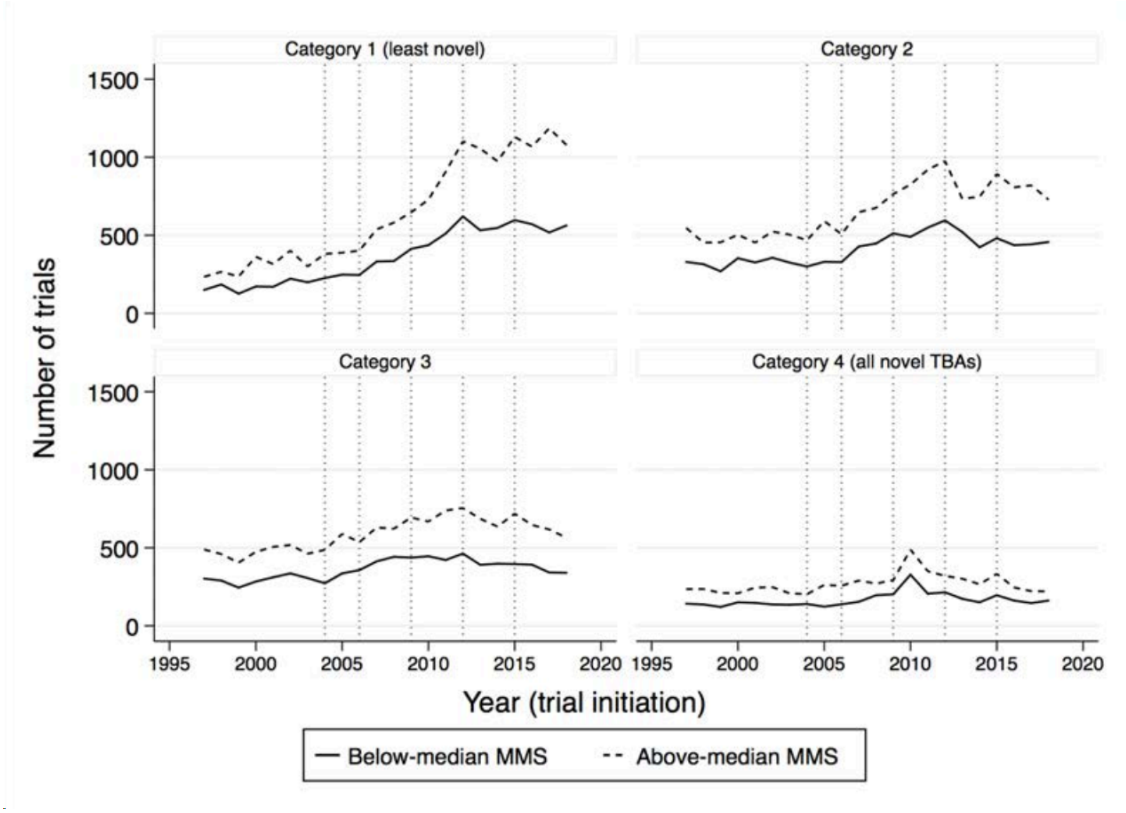


Figure 20: Number of initiated trials by Medicare market share and novelty (Figure 5 from Dranove et al., 2020)

It's calculated as

$$\text{NOVELTY}_i = \frac{1}{1 + D_{k(i)}}$$

where $D_{k(i)}$ is the number of other products in Phase k or higher that have tested the TBA(s) of drug i, and $k\,(i)$ is the current phase of drug i.

Figure 5 shows the key patterns behind their result. While the reform seems to have had an impact on the number of new clinical trials (confirming the results of Blume-Kohout and Sood, 2013), these new trials seem to be concentrated among drugs rated as least innovative according to their measure. These results are consistent with the idea that market size can incentivize firms to start testing a few additional drugs, but probably does not spur a lot of innovative research, or at least not in the short run.

Overall, the implication of this literature is that there is almost certainly a positive link between market size and innovation. This is not particularly suprising, but it's reassuring, in a way. It's still somewhat of an open question whether marginal changes in expected revenue can have a very large impact on the probability of developing very valuable/innovative products. As you explore the literature in this space you should keep in mind the following:

- R&D takes time, so it's hard to come up with an event study design that cleanly identifies

an effect (we do not expect R&D output or even inputs to adjust immediately)

- Firms may not behave according to fully rational models. R&D carries a lot of uncertainty, and the executives making R&D decisions may not operate according to a standard cost-benefit analysis model. They may worry about losing too much money on a product that does not deliver any revenue, they may suffer from sunk cost fallacy, and they may face unobserved budget constraints.
- Pharma R&D (even more so than R&D in other fields) faces sharp technological constraints that can prevent progress even in the most profitable of therapeutic areas. There are still many diseases that are uncurable, even though a cure would earn very high revenue. Oncology and Alzheimer are the two most obvious examples here.
- (Perhaps most importantly) this is a young field, and evidence beyond greater market size $\implies$ more innovation is scarce. There is ample room to review and subvert the current conventional wisdom.

Issues in Antitrust: Cunningham et al. (2020)

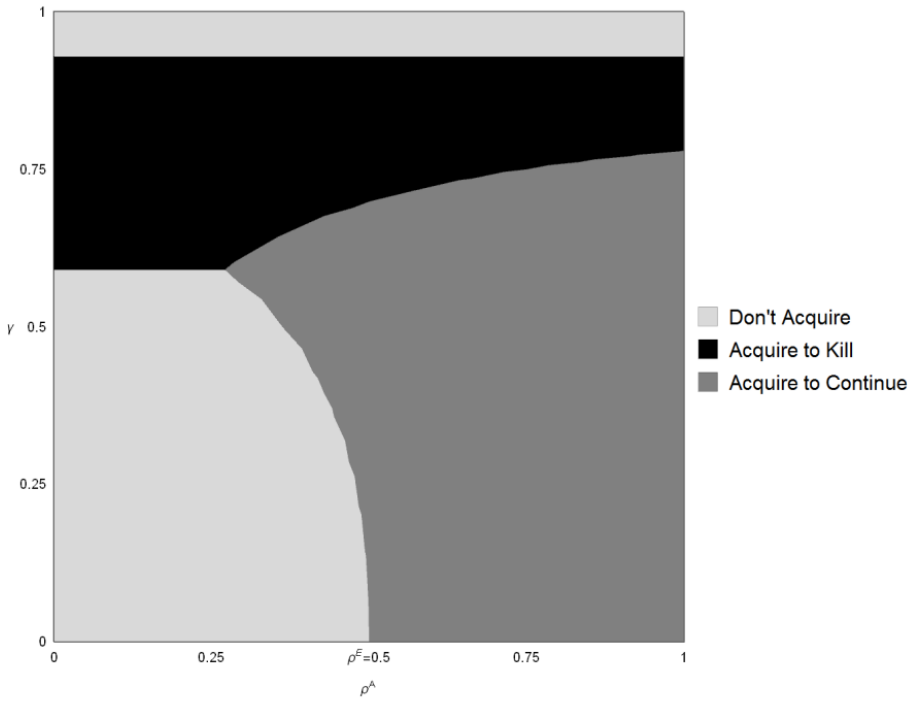
The last paper I cover in this space is only tangentially related to the issue of market size and innovation, but it does highlight the presence of some counterintuive incentives in the R&D space. The basic argument of the paper is that under certain conditions, incumbent firms will have a lower incentives to innovate, and, sometimes, will even have an incentive to acquire and shut down potential competitors before they make it to market. This sort of result matters not only in pharma, but also in tech, and the concept of "killer acquisitions" (the title of the paper) has been incorporated in antitrust terminology.

The paper proposes a model over three periods. At time $t = 0$ an entrepreneur E with a single project is born. The project represents a potential new product for a market where $n \geq 1$ firms operate. One of these firms, a potential acquirer A, decides (also in period 0) whether to acquire the entrepreneur at an endogenously determined price P.

At time $t = 1$ the project is either terminated or completed (by whoever owns it at that time, either A or E). Finally, at $t = 2$, all firms engage in competition and earn profits.

If we forget about the acquisition stage for a second, it should be obvious that an incumbent firm has a weaker incentive to develop a new product than a new entrant. That's because the new product will cannibalize some of the profits that the incumbent is earning from their existing products. The entrant does not consider this business-stealing effect and therefore has a stronger incentive to continue development. (Obviously this holds all else equal, and there could be differences in the ability of each party to carry the project to completion).

The model also derives two simple comparative statics. First, the difference in incentives will disappear as n increases. That's because the cannibalization effect is strongest with a monopolist, and weakest when the market gets close to perfect competition. Second, the difference in incentives will increase with the remaining patent life of the incumbent's product. Again, this makes



This graph plots the optimal acquisition decisions—"Don't Acquire" (light gray), "Acquire to Kill" (black), and "Acquire to Continue" (dark gray)—as a function of the acquirer's development capability ρ^A and the degree of substitutability γ. The other parameter values are $\alpha^A = \alpha^E = 100$, $\rho^E = 0.5$, $L = 20$, $k = 80$, and $n = 2$.

Figure 21: Optimal acquisition strategies (Figure 4 from Cunningham et al., 2020)

sense because it the incumbent's product is about to go generic, then by the time the new product is developed the incumbent may not actually be an incumbent anymore.

Finally, they consider the acquisition choice. For a killer acquisition to occur, three conditions must hold:

1. The entrepreneur must have an incentive to continue developing the project
2. The incumbent would rather terminate development
3. The marginal expected gain from terminating development of the incumbent is greater than the marginal expected gain from development of the entrepreneur.

The last condition is what allows the acquisition to go through. Figure 21 provides some intuition for what determines whether a killer acquisition will occur. The two variables on the y- and x-axis are the degree of substitutability between the two products, and the incumbent's development capability (i.e. the probability that, if developed, the product will successfully reach the market). Intuitively what makes killer acquisitions most likely is that the two products are very close substitutes. In these cases, the entrepreneur stands to gain the least from entering (because she will

have to compete closely on price with the incumbent), and the incumbent has the most to lose (for the same reason).[25]

Their empirical strategy is relatively simple (though it involves a battery of robustness checks and tests of alternative explanations that I do not discuss here). They measure development of a project as progressing to a new phase of clinical trials. They regress this outcome variable on an indicator for post-acquisition, and include an interaction for overlap of indication between the acquired drug and any drug owned by the acquiring company.

Their results suggest that acquisition in an overlapping indication is associated with a 3.7% lower likelihood of development after acquisition (Table 6). Acquired products are also generally less likely to be developed (by 2%), so this is an overall effet of 5.7%. The results are robust to various specifications. They also run several other tests of their model's hypothesis and find broad confirmation for all of them: development is even less likely in less competitive markets, and if the relevant product owned by the acquiring company is not near patent expiration.

4.2 Research on other determinants of R&D

Here I provide a brief list of other papers studying various determinants of R&D. Note that this is not meant to be comprehensive. Rather, it's a starting point to dig into other questions that are floating around in the literature.

- Costinot et al. (2019) use data on global burden of disease by country to show that firms headquartered in certain countries are more likely to develop drugs treating diseases that are prominent in those countries. The main implication is as a test of the home-market effect in trade, but it has implications for R&D as well (one could think to match these results to Dubois et al. (2015) to get more nuanced predictions on direction of R&D)

- Azoulay et al. (2019) examines the role of public funds in pharmaceutical R&D. This is an area where there has been a lot of interest recently, because some have argued that the pharmaceutical market is essentially extracting rents from primary research that is sponsored by the government and various non-profit organzations (like universities).

- Kao (2019) looks at how information from cancer genome mapping studies has affected the R&D decision-making of pharmaceutical companies.

- Yin (2008) studies how the passage of the Orphan Drug Act (which provided additional years of exclusivity and R&D tax breaks to firms that developed drugs whose primary indication treats a disease with low prevalence in the US) affected R&D choice of drug manufacturers.

[25] The grey bar at the top where no acquisitions can occur is because the model assumes that the new project also has higher vertical quality relative to the incumbent product, so for a high-enough degree of substitutability the old product is kicked out of the market.

	Development Event = 1					
	(1)	(2)	(3)	(4)	(5)	(6)
I(Acquired) × I(Post) × Overlap	−0.037***	−0.033**	−0.029*	−0.041**	−0.043**	−0.054**
	(0.013)	(0.014)	(0.015)	(0.019)	(0.021)	(0.024)
I(Acquired) × I(Post)	−0.020***	−0.016**	−0.017**	−0.024**	−0.018	−0.018
	(0.006)	(0.007)	(0.009)	(0.010)	(0.011)	(0.013)
I(Acquired) × Overlap	0.004	0.009	0.026**			
	(0.008)	(0.009)	(0.011)			
I(Acquired)	−0.002	−0.004	−0.011			
	(0.004)	(0.005)	(0.012)			
Before(−3) × Overlap						−0.031
						(0.032)
Before(−2) × Overlap						0.012
						(0.032)
Before(−1) × Overlap						−0.040
						(0.030)
Before(−3)						0.015
						(0.017)
Before(−2)						0.020
						(0.017)
Before(−1)						−0.003
						(0.016)
Observations	143,569	143,569	143,569	143,569	134,662	143,569
R-squared	0.038	0.252	0.289	0.366	0.662	0.370
Vintage FE	Y	Y	Y			
Age FE	Y					
Age × TC × MOA		Y	Y	Y	Y	Y
Originator [Target company] FE			Y			
Project FE				Y	Y	Y
Propensity score reweighted					Y	

Table 6: Overlapping acquisitions and project development (Table 2 from Cunningham et al., 2020)

- Krieger et al. (2018) looks at how liquidity affects the choice to develop more novel drugs. They use an interesting measure of novelty based on the chemical similarity between two drugs.
- Hamilton et al. (2018) uses a very detailed demand-side model to explore how consumption affects innovation. This is one of the few papers that attempts to say anything about the direction of innovation (as opposed to simply quantifying the intensity of R&D).

5 Advertising

Pharmaceutical companies advertise in two ways: through sales representatives that target physicians (i.e. detailing), and through direct-to-consumer advertising (DTCA). DTCA is only legal in the US and New Zealand, and is often regarded as malicious because of fears that it may drive prescriptions to inappropriate patients. However, DTCA only represents a very small fraction of the overall advertising budget of pharmaceutical companies—estimates vary, but most figures suggest roughly a 6-to-1 ratio beteween detailing and DTCA spending.

Despite this disparity, most economic research has tended to focus on DTCA rather than detailing, for two reasons. First, data on detailing was historically hard to come by. Second, detailing is very targeted not only to specific geographic areas but to specific physicians and hospitals. As a result, issues of selection and endogeneity are hard to solve (this is a common issue in advertising in general, but detailing is particularly "targeted", which makes things hard). Since the Sunshine Act of 2010, data on detailing has become more prevalent, and CMS publishes yearly payments made from pharmceutical companies to US physicians. A number of papers have used this new data to look at how detailing affects prescription patterns.

5.1 Welfare impact of direct-to-consumer advertising

Evidence from most economic research that looks at direct-to-consumer advertising tends to challenge the commonly held view that DTCA is harmful and leads to inappropriate prescriptions. The alternative view (held by pharmaceutical manufacturers), is that advertising provides information to patients about treatments they would otherwise be unaware of.

What makes these two alternative hypotheses hard to parse is that they both suggest that DTCA will result in a higher number of prescriptions, which means that additional data is needed to figure out which of the two is dominant.

Shapiro (2020) examines these two alternative views in the context of local TV ads for antidepressants. To differentiate between the two explanations, the paper does not only show that advertising leads to an increase in prescriptions of antidepressants, but also how this increase in prescriptions leads to a fall in workplace absenteism. Clinically speaking, antidepressants have been shown to reduce workplace absences in individuals diagnosed with depression.

To establish a causal link between advertising and the outcome variables of interest, one must address the fact that TV ads are targeted to areas and time periods where the firm expects to have a large impact ex-ante. The identification strategy of the paper—which is introduced in an earlier paper, Shapiro (2018)—relies on geographical variation in advertising in counties that are located alongside the border of designated market areas (DMAs). DMAs are advertising markets that are made up of multiple counties. Since ads run on local channels, the entire DMA sees the same exact ads. The author exploits the fact that counties that run alongside the DMA's border are similar to counties just outside the border. Moreover, since ads are targeted to the entire DMA, the characteristics of these border counties only have a small impact on the quantity and type of ads in the DMA. This is especially true when the DMA is centered around a large city (e.g. Cleveland and Columbus, Figure 22), whose consumers are going to drive ad spend.

In order to look at how advertising affects workplace absenteism, the paper exploits a dataset that links prescription claims data to data on workplace performance. It then combines this data with data on local TV ad spending.

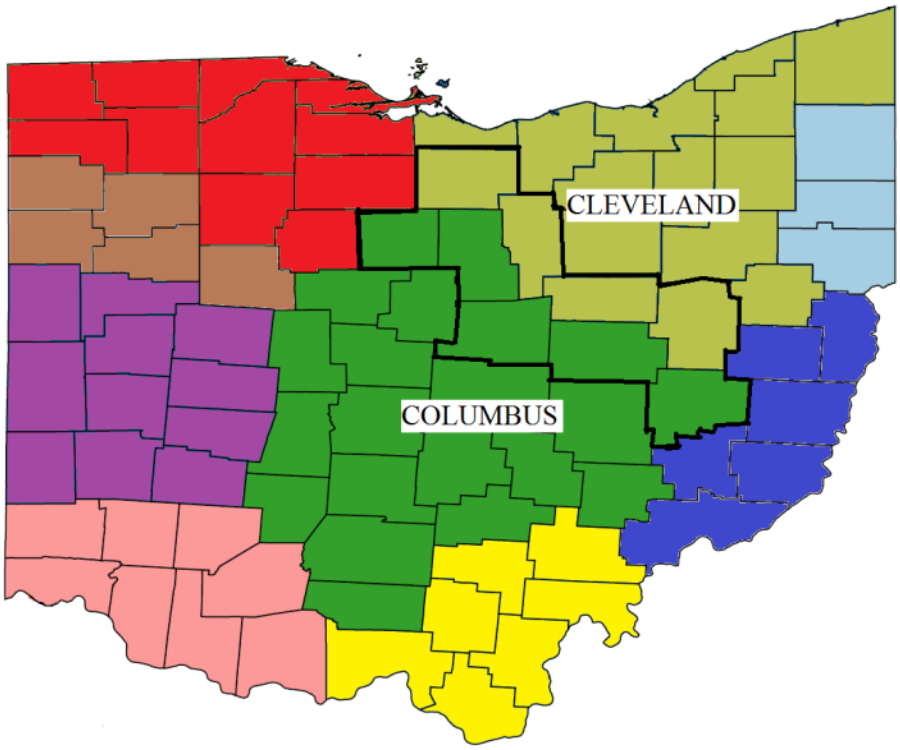


Figure 22: Advertising market border lines in Ohio (Figure 2 from Shapiro, 2020)

The estimating equation is

$$Y_{ibdt} = \beta_1 f_1 \underbrace{(GRP_{dt})}_{\substack{\text{current period} \\ \text{ad spending}}} + \beta_2 f_2 \underbrace{\left(\sum_{\tau=t-t_0}^{t-1} GRP_{d\tau} \right)}_{\text{past ad spending}} + \alpha_i + \alpha_{bt} + \varepsilon_{ibdt}$$

where i indexes an individual, b is a border area (e.g. the area with the bold black border in Figure 22), d is a DMA, and t is a period. The main coefficient of interest is either β_1 or β_2 (it varies depending on the outcome variable Y_{ibdt}). In the main analysis, both $f_1(x)$ and $f_2(x)$ are $\log(1+x)$, and the sum of past ad spending is taken over the previous 6 months.

The first set of results (in Table 7) shows the impact of advertising on prescriptions. The main coefficient of interest in this case is β_1 (concurrent advertising). The paper finds a small (0.31%), but positive and significant effect of advertising on new prescriptions (Column 1). All other effects on prescription-related outcomes are not statistically distinguishable from zero. These results establish that advertising affects new prescriptions, but also that it does not have a large impact on other outcomes such as switching to brand prescriptions, price, and probability of initiating therapy.

The second set of results (in Table 8) shows the impact of advertising on work absences. The coefficients of this regression need to be interpreted carefully because of the possibility for spu-

	(1) NewRX	(2) Switch	(3) Copay	(4) Price	(5) GenericRate	(6) Therapy
$Log(1+GRP)$	0.00031	0.00286	0.1499	-0.61605	-0.00335	-0.00011
	(0.00015)	(0.0028)	(0.16599)	(0.38631)	(0.00246)	(0.0006)
$Log(1+GRP_{past})$	-0.00019	0.00666				-0.00002
	(0.00027)	(0.0047)				(0.0002)
Individual FEs	x	x				x
Border-Side FEs			x	x	x	
Border-Month FEs	x	x	x	x	x	x
Mean DV	0.0099	0.3615	$10.84	$61.96	0.59867	0.0212
R-squared	0.0745	0.19234	0.0686	0.02373	0.03573	0.44033
Observations	12,064,669	906,012	1,053,769	1,053,769	1,054,561	12,438,333

Table 7: Impact of advertising on prescriptions (Table 3 from Shapiro, 2020)

rious correlations. For example, the current level of advertising may be targeted to areas with high levels of undiagnosed depression, which would correlate wth absenteism as well. Because of mean reversion, we would then expect a negative correlation with lagged advertising. Another The change in results from column 1 to column 4 shows this pattern. Columns 2 and 3 introduce FEs, which help soak up some this unwanted variation. Column 4 ultimately uses a more aggressive strategy, that focuses only on border counties (throwing out the remaining observations) and uses border-month FEs. These fixes effects soak up period-specific variation that affects the area around the border, potentially leaving advertising as the main source of variation for absenteism. The coefficient implies that the average individual in the sample gains about 0.11 of monthly time at work from a 10% increase in ad spending in the previous 6 months. Overall, the benefits of this reduction in time missed outweigh the cost of new prescriptions by an order of magnitude.

The final results of the paper confirm the positive effect of ads. The author shows that the absenteism gains are concentrated among workers that started prescriptions for antidepressants, and that those who start new prescriptions do not look more likely to have adverse events, or to discontinue treatment down the line.

Other works:

The first papers in this field focused on showing the direct impact of DTCA on demand.

- Shapiro (2018) exploits the border strategy to show that there are significant spillover effects from advertising (meaning that competitors in the same therapeutic area benefit from each other's advertising campaigns). This in turn leads to an equilibrium where firms underinvest in advertising because they do not capture the spillover benefits.

- Sinkinson and Starc (2018) uses shocks to advertising caused by the political advertising cycle to show that ads lead to significant business stealing among branded drugs (but not

	(1)	(2)	(3)	(4)
$Log(1+GRP)$	0.2100	0.2518	-0.0650	-0.0339
	(0.0990)	(0.0768)	(0.0904)	(0.0288)
$Log(1+GRP_{past})$	-0.5356	-0.3710	-0.2757	-0.1382
	(0.2886)	(0.1477)	(0.1830)	(0.0602)
Individual FEs		x	x	x
Month FEs			x	
Border-Month FEs				x
Mean DV	2.432	2.432	2.432	2.876
R-squared	0.00639	0.27876	0.32319	0.35549
Observations	16,310,368	16,303,199	16,303,199	3,363,046

Table 8: Impact of advertising on workplace absenteism (Table 4 from Shapiro, 2020)

generics). They also confirm the spillover effects in Shapiro (2018). Their view of DTCA is also relatively benign.

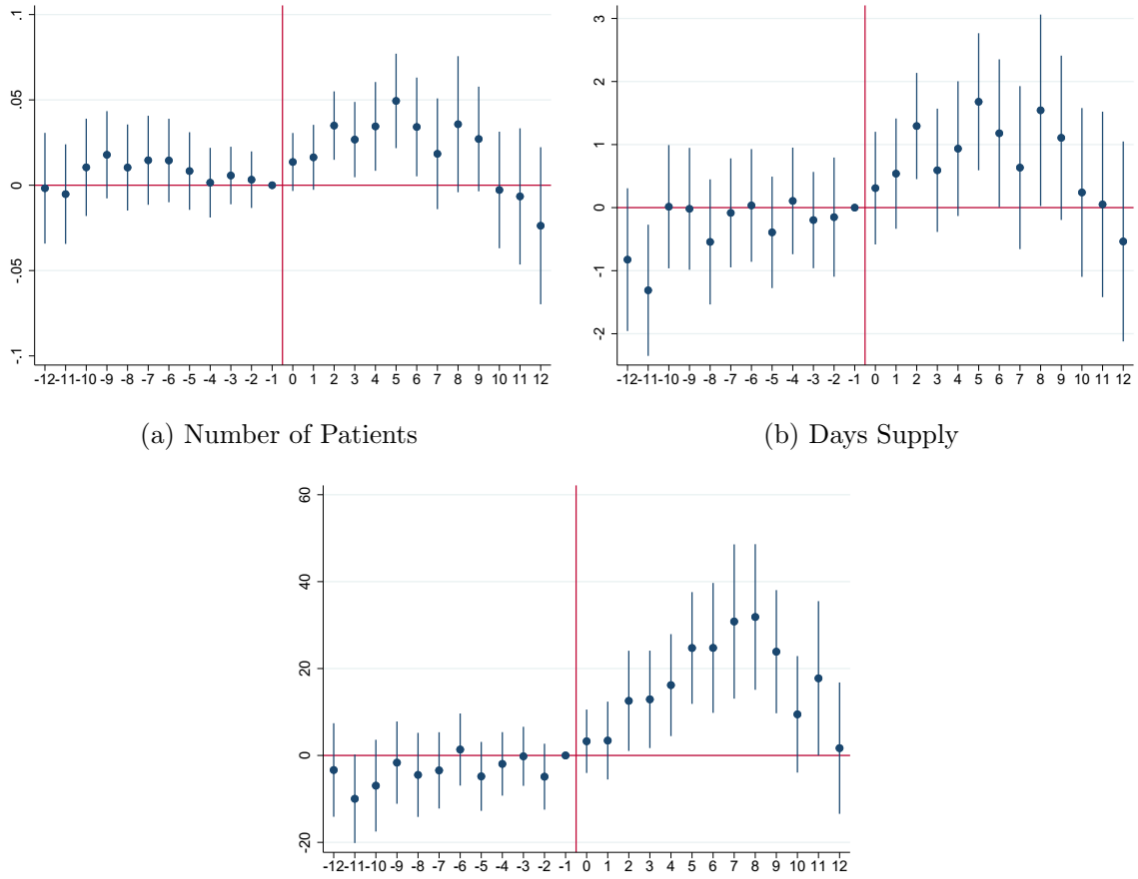
5.2 Detailing and physician behavior

With the release of the Open Payment dataset, several researchers have started looking at the effect of detailing. The main obstacle to extracting causal estimates from this data is the same as with the DTCA literature: detailing is targeted, and therefore variation in the sums received by physicians is likely linked to unobserved characteristics. The problem is much more subtle than with DTCA however. While DTCA is targeted "broadly" to geographic areas, detailing is much more precise. As a result, identifying sources of exogenous variation is exceedingly difficult. Here I briefly present a couple of recent papers and discuss their results and estimation strategies.

Carey et al. (2021)

This paper combines the Open Payment data with prescription claims from the 20% random sample of Medicare enrollees. They cover the span from 2013 to 2015. Their estimation strategy is basically to use a lot of fixed effects to try and isolate the residual impact of payments to specific physicians. In particular, their regressions control for physician-drug fixed effects δ_{pd} and drug-period fixed effects δ_{dt}:

$$y_{pdt} = \sum_{r \neq -1} PresPaid_{pd}\beta_r + X_{pdt}\Gamma + \delta_{pd} + \delta_{dt} + \varepsilon_{pdt}$$



(a) Number of Patients

(b) Days Supply

(c) Expenditure

Figure 23: Impact of detailing payments (Figure 3 from Carey et al., 2021)

The outcomes variable y are the number of patients, days supply, and expenditure for a given physician p, drug d and month t. $PresPaid_{pd}$ is an indicator for whether the physician is ever paid by the manufacturer of drug d (at any point in the sample), while r denotes the first time that a physician receives a payment. Their results show a positive impact of detailing payments on all three outcome variables (Figure 23).

The main concern here is that this design essentially assumes that the payment event is a natural experiment, even though it isn't (to see why compare the design to studies of hospital or insurer mergers, which also use fixed effects, but generally have to resort to additional adjustments to claim a causal effect). Another potential issue is that some of the fixed effects may soak up important variation. For example, δ_{dt} will capture variation generated by broad detailing campaigns (say, if the manufacturer decides to push for a given product by increasing payments to all physicians). This may partly explain why their estimates are somewhat smaller than those of other studies.

In a second set of results, the paper finds that these payments do not appear to lead to any obvious negative welfare effects on patients, as targeted physicians are not slower to switch to

generics, and quality of prescribed drugs does not seem to suffer.

Agha and Zeltzer (2019)

This paper uses a similar estimation strategy, but it's more focused on peer effects. The analysis is motivated by the observation that a large part of spending on detailing is concentrated among a few selected physicians, consistent with a theory where key opinion leaders can influence prescribing patterns of a large network of peers. These larger payments are generally given for speaking engagements. In contrast, payments to most physicians are given largely for meals.

The paper finds evidence that after these large payments, peer physicians (who did not receive direct payment) increase their prescriptions by 2%. This effect represents about a quarter of the overall effect of detailing (which they also estimate, finding an overall impact similar to Carey et al., 2021).

Grennan et al. (2018)

This paper is somewhat of an outlier, in that it does not use Open Payment data. Instead, it relies on proprietary data payments for branded statins in 2010-2011. It also differentiates itself because it uses an instrumental variable strategy to obtain causal identification.

The empirical strategy relies on variation at the hospital and geographical level on conflict-of-interest policies. The construction of the IV is a bit convoluted, but the basic intuition is that drug manufacturers will generally devote fewer resources to areas where a high number of hospitals have restrictive conflict-of-interest policies. Hence, physicians whose hospitals do *not* have such policies, but whose neighboring hospitals do, will see fewer sales reps and receive fewer payments.

Interestingly, they find much larger effects of these payments (73%. compared to about 6% in the previous two papers). This can be because of a few reasons. First, they argue that a simple OLS estimator would underestimate the relationship because reps are likely to target physicians who would normally prescribe fewer drugs (becayse they provide the largest marginal gains). Second, they use data from the pre-Open Payment period and the fact that payments now need to be disclosed may have lead to a decline in the payments and their effect, for example because physicians do not want to give the appearance that they are favoring specific manufacturers.

6 BIBLIOGRAPHY

Acemoglu, Daron and Joshua Linn, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *The Quarterly Journal of Economics*, aug 2004, *119* (3), 1049–1090.

Agha, Leila and Dan Zeltzer, "Drug Diffusion Through Peer Networks: The Influence of Industry Payments," *NBER Working Paper*, 2019.

Atteberry, Preston, Peter B. Bach, Jennifer A. Ohn, and Mark Trusheim, "Biologics are natural monopolies (part 1): Why biosimilars do not create effective competition," *Health Affairs Blog*, 2019, (Part 1), 1–9.

Azoulay, Pierre, Joshua S. Graff Zivin, Danielle Li, and Bhaven N. Sampat, "Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules," *Review of Economic Studies*, 2019, *86* (1), 117–152.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein, "Behavioral Hazard in Health Insurance," *The Quarterly Journal of Economics*, 2015, *130* (4), 1623–1667.

Beaulieu, Nancy, David M. Cutler, Katherine Ho, George Isham, Tammie Lindquist, Andrew Nelson, and Patrick O'Connor, "The business case for diabetes disease management for managed care organizations," *Forum for Health Economics and Policy*, 2006.

Berndt, Ernst R., "Pharmaceuticals in U.S. health care: Determinants of quantity and price," *Journal of Economic Perspectives*, 2002, *16* (4), 45–66.

Bhattacharya, Jayanta and William B. Vogt, "A simple model of pharmaceutical price dynamics," *Journal of Law and Economics*, 2003, *46* (2), 599–626.

Blume-Kohout, Margaret E. and Neeraj Sood, "Market Size and Innovation: Effects of Medicare Part D on Pharmaceutical Research and Development," *Journal of Public Economics*, 2013, *97* (January), 327–336.

Bronnenberg, Bart J., Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro, "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium," *The Quarterly Journal of Economics*, 2015, *130* (4), 1669–1726.

Carey, Colleen, Ethan M.J. Lieber, and Sarah Miller, "Drug firms' payments and physicians' prescribing behavior in Medicare Part D," *Journal of Public Economics*, 2021, *197*, 1–45.

Chaudhuri, Shubham, Pinelopi K. Goldberg, and Panle Jia, "The Effects of Global Patent Protection in Estimating Pharmaceuticals : A Case Study of Quinolones in India," *American Economic Review*, 2006, *96* (5), 1477–1514.

Chen, Brian, Tony Y. Yang, and Charles L. Bennett, "Challenges to Biosimilar Substitution," *JAMA*, 2017, *318* (12), 1186.

Cockburn, Iain M., Jean O. Lanjouw, and Mark Schankerman, "Patents and the global diffusion of new drugs," *American Economic Review*, 2016, *106* (1), 136–164.

Costinot, Arnaud, Dave Donaldson, Margaret Kyle, and Heidi Williams, "The more we die, the more we sell? A simple test of the home-market effect," *Quarterly Journal of Economics*, 2019, *134* (2), 843–894.

Crawford, Gregory S. and Matthew Shum, "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 2005, *73* (4), 1137–1173.

Cunningham, Colleen, Florian Ederer, and Song Ma, "Killer Acquisitions," 2020.

Cutler, David, Jonathan Skinner, Ariel Dora Stern, and David Wennberg, "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," 2013.

Dickstein, Michael J., "Physician vs. Patient Incentives in Prescription Drug Choice," 2017.

Dranove, David, Craig Garthwaite, and Manuel I. Hermosilla, "Expected Profits and the Scientific Novely of Innovation," 2020.

Dubois, Pierre, Ashvin Gandhi, and Shoshana Vasserman, "Bargaining and International Reference Pricing in the Pharmaceutical Industry," 2019, (May), 1–75.

_ , Olivier de Mouzon, Fiona Scott-Morton, and Paul Seabright, "Market size and pharmaceutical innovation," *The RAND Journal of Economics*, 2015, *46* (4), 844–871.

_ , Yassine Lefouili, and Stéphane Straub, "Pooled procurement of drugs in low and middle income countries," *European Economic Review*, 2021, *132*, 1–42.

Duggan, Mark and Fiona M. Scott Morton, "The Distortionary Effects of Government Procurement: Evidence From Medicaid Prescription Drug Purchasing," *Quarterly Journal of Economics*, 2006, *121* (1), 1–30.

_ and _ , "The effect of medicare part D on pharmaceutical prices and utilization," *American Economic Review*, 2010, *100* (1), 590–607.

Feng, Josh, "Pricing Intermediaries in Prescription Drug Markets: Impact and Policy Implications," 2019.

_ , "History-Dependence in Drug Demand: Implications for Entry Timing," 2020.

_ , Thomas Hwang, and Luca Maini, "Profiting from Most-Favored Customer Procurement Rules: Evidence from Medicaid," 2020, pp. 1–30.

Finkelstein, Amy, "Static and dynamic effects of health policy: Evidence from the vaccine industry," *Quarterly Journal of Economics*, 2004, *119* (2), 527–564.

Frank, Richard G. and David S. Salkever, "Pricing, Patent Loss and the Market for Pharmaceuticals," *Southern Economic Journal*, 1992, *59* (2), 165.

_ and _, "Generic entry and the pricing of pharmaceuticals," *Journal of Economics and Management Strategy*, 1997, *6* (1), 75–90.

Gondek, K., "Prescription drug payment policy: Past, present, and future," *Health Care Financing Review*, 1994, *15* (3), 1–7.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town, "Mergers when prices are negotiated: Evidence from the hospital industry," *American Economic Review*, 2015, *105* (1), 172–203.

Grabowski, Henry G. and John M. Vernon, "Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act," *The Journal of Law and Economics*, oct 1992, *35* (2), 331–350.

_ , Rahul Guha, and Maria Salgado, "Regulatory and cost barriers are likely to limit biosimilar development and expected savings in the near future," *Health Affairs*, 2014, *33* (6), 1048–1057.

Grabowski, Henry, Genia Long, Richard Mortimer, and Ani Boyo, "Updated trends in US brand-name and generic drug competition," *Journal of Medical Economics*, 2016, *19* (9), 836–844.

Grennan, Matthew, Kyle Myers, Ashley Swanson, and Aaron Chatterji, "Physician-Industry Interactions: Persuasion and Welfare," *SSRN Electronic Journal*, 2018, pp. 1–65.

Hamilton, Barton H., Andrés Hincapié, Robert A. Miller, and Nicholas W. Papageorge, "Innovation and Diffusion of Medical Treatment," *NBER Working Paper Series*, 2018, (24577), 1–59.

Ho, Kate and Robin S. Lee, "Insurer Competition in Health Care Markets," *Econometrica*, 2017, *85* (2), 379–417.

Ho, Katherine, "The welfare effects of restricted hospital choice in the US medical care market," *Journal of Applied Econometrics*, nov 2006, *21* (7), 1039–1079.

Howard, David H., Peter B. Bach, Ernst R. Berndt, and Rena M. Conti, "Pricing in the Market for Anticancer Drugs," *Journal of Economic Perspectives*, feb 2015, *53* (1), 139–162.

Jacobo-Rubio, Ruben, John L. Turner, and Jonathan W. Williams, "The distribution of surplus in the us pharmaceutical industry: Evidence from paragraph iv patent-litigation decisions," *Journal of Law and Economics*, 2020, *63* (2), 203–238.

Kakani, Pragya, Michael Chernew, and Amitabh Chandra, "Rebates in the Pharmaceutical Industry: Evidence from Medicines Sold in Retail Pharmacies in the U.S.," 2020.

Kao, Jennifer, "Charted Territory: Evidence from Mapping the Cancer Genome and R&D Decisions in the Pharmaceutical Industry," *Jmp*, 2019, pp. 1–61.

Kremer, Michael, "Patent buyouts: A mechanism for encouraging innovation," *Quarterly Journal of Economics*, 1998, *113* (4), 1137–1167.

Krieger, Joshua, Danielle Li, and Dimitris Papanikolaou, "Developing Novel Drugs," *SSRN Electronic Journal*, 2018.

Kyle, Margaret K., "Pharmaceutical Price Controls and Entry Strategies," *Review of Economics and Statistics*, 2007, *89* (February), 88–99.

Lakdawalla, Darius N., "Economics of the pharmaceutical industry," *Journal of Economic Literature*, 2018, *56* (2), 397–449.

Lu, Z. John and William S. Comanor, "Strategic pricing of new pharmaceuticals," *Review of Economics and Statistics*, 1998, *80* (1), 108–118.

Maini, Luca and Fabio Pammolli, "Reference Pricing as a Deterrent to Entry: Evidence from the European Pharmaceutical Market," 2019.

_ , Josh Feng, Thomas Hwang, and Jacob Klimek, "Biosimilar Entry and the Pricing of Biologic Drugs," 2021.

Olssen, Alexander L. and Mert Demirer, "Drug Rebates and Formulary Design: Evidence from Statins in Medicare Part D," 2021.

Ridley, David B. and Chung-Ying Lee, "Does Medicare Reimbursement Drive Up Drug Launch Prices?," *The Review of Economics and Statistics*, 2020.

Sampat, Bhaven and Heidi L. Williams, "How do patents affect follow-on innovation? Evidence from the human genome," *American Economic Review*, 2019, *109* (1), 203–236.

Scott Morton, Fiona, "The Strategic Response by Pharmaceutical Firms to the Medicaid Most-Favored-Customer Rules," *RAND Journal of Economics*, 1997, *28* (2), 269–290.

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

_ , Ariel Dora Stern, and Scott Stern, "The Impact of the Entry of Biosimilars: Evidence from Europe," *Review of Industrial Organization*, aug 2018, *53* (1), 173–210.

Shapiro, Bradley T., "Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants," *Journal of Political Economy*, 2018, *126* (1), 381–437.

_ , "Promoting Wellness or Waste? Evidence from Antidepressant Advertising," 2020.

Sinkinson, Michael and Amanda Starc, "Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals," *Review of Economic Studies*, 2018, *86* (2), 1–46.

Trusheim, Mark, Preston Atteberry, Jennifer A. Ohn, and Peter B. Bach, "Biologics are natural monopolies (part 2): A proposal for post- exclusivity price regulation of biologics," *Health Affairs Blog*, 2019, (Part 2), 10–13.

Williams, Heidi L., "Intellectual property rights and innovation: Evidence from the human genome," *Journal of Political Economy*, 2013, *121* (1), 1–27.

_ , "How do patents affect research investments?," *Annual Review of Economics*, 2017, *9*, 441–469.

Yin, Wesley, "Market incentives and pharmaceutical innovation," *Journal of Health Economics*, 2008, *27* (4), 1060–1077.

Yurukoglu, Ali, Eli Liebman, and David B. Ridley, "The Role of Government Reimbursement in Drug Shortages," *American Economic Journal: Economic Policy*, may 2017, *9* (2), 348–382.