How Much Are Medical Innovations Worth? A Detailed Analysis Using Thousands of Cost-Effectiveness Studies

Abe Dunn, Lasanthi Fernando, and Eli Liebman*†

July 21, 2023

Abstract

Medical innovation is thought to be a key driver of medical cost growth and improved life expectancy, but linking innovations to their impacts on costs and quality of care is challenging. In this paper we extract and organize quality information from thousands of cost-effectiveness studies to measure the quality improvements of over one hundred distinct treatments for 13 major conditions. We combine these quality measures with private insurance claims data from millions of individuals to account for how treatments diffuse and their impact on the cost and quality of treatment. Across different conditions, we find significant heterogeneity in the diffusion of new technologies and the changes in the cost, quality, and welfare from treatments. Similar to markets outside of health care, we find innovations can improve consumer welfare substantially, leading quality-adjusted prices to decline for 7 of the 13 conditions we study. However, we also observe a phenomena arguably unique to healthcare, cases where the quality improvements from new and diffusing innovations do not keep pace with their higher costs, reducing consumer welfare. This implies that suppliers may be capturing more than 100% of total surplus for these drugs.

JEL: E31,O30, I10

Keywords: Innovation, Price index, Quality adjustment, Medical spending

*Abe Dunn and Lasanthi Fernando: Bureau of Economic Analysis. Eli Liebman: Department of Economics, Terry College of Business, University of Georgia (email: eli.liebman@uga.edu).

†We thank Aakash Arora, Maeve Breathnach, Jack Cenatempo, Justin Erwin, Vishal Parmar, Abdur Rahman Bin Irfan, and Cole Wittbrodt for their excellent research assistance. We thank Michael Chernew, David Cutler, Qing Gong, Scott Grosse, Joe Newhouse, David Ridley, Louise Sheiner, Marcia Weaver, and participants in IHEA, SEA and ASSA conference sessions for helpful comments. The views expressed in this paper are those of the authors and not necessarily the views of the Bureau of Economic Analysis.

1 Introduction

U.S. health care spending has risen from 5% of GDP in 1960 to 17.7% in 2019 (Hartman et al., 2021). Over this same period, life expectancy has increased by 9 years (Arias et al., 2019). A number of prominent papers argue that if gains in life expectancy are due to technological improvements in medical care, then the increase in health care costs may reflect welfare improvements (Murphy and Topel, 2006; Hall and Jones, 2007), even if they are simultaneously a major contributing factor to overall cost growth (Chernew and Newhouse, 2011). While there are numerous papers that suggest cost growth in health care is "worth it," there is also substantial literature arguing that the price of innovation is excessive in the U.S.[1] At the heart of the debate is a measurement issue, as it is difficult to determine precise costs and benefits of distinct technological improvements.

The goal of this paper is to provide evidence for how innovation shapes the quality and cost of treatment in health care markets. Measuring the value of medical innovation is notoriously difficult. In most non-medical care markets, a common approach to measuring value (and quality) is to apply methods that rely on revealed preferences or hedonics. However, there are numerous market distortions in health care which complicate the use of these methods.[2] Because of these distortions, a common approach to measuring the value of medical care is to use outcome measures (e.g., mortality) (Sheiner and Malinovskaya, 2016; Cutler et al., 1998, 2022). However, the outcomes-based approach does not identify which technologies are driving the associated changes in outcomes and cost, which is the primary

[1]For example, Hall and Jones (2007); Murphy and Topel (2006); Cutler and McClellan (2001); Cutler et al. (2022) all view innovation as improving welfare. Cutler (2018); Shrank et al. (2019); Kesselheim et al. (2016) all argue that the price of medical care is excessive.

[2]Examples of papers which use revealed preference approaches to value innovation in non-health care markets include Feenstra (1994); Trajtenberg (1989); Petrin (2002); Gowrisankaran and Rysman (2012), among many others. The distortions in health care markets include: insurance coverage insulating patients from risk (moral hazard), insurers distorting demand through formulary design, asymmetric information between patients and providers (principal agent problems), and imperfect information which all complicate revealed preference and hedonics approaches (Dauda et al., 2022).

focus of this paper.

Our paper takes a unique approach to measure innovation, quality, and cost in the health care sector by leveraging the knowledge accumulated in the medical literature. In particular, we use the Tufts Cost-Effectiveness Analysis Registry (CEAR) database of over 8,000 cost effectiveness studies to assign measurable characteristics of quality to specific treatments and medical conditions. The purpose of cost-effectiveness studies is to determine the net benefit of innovative treatments, so the literature clearly has the potential to shed light on the value of medical innovations, but in practice this requires synthesizing the results of thousands of cost-effectiveness studies. This paper handles this issue by: (1) classifying the treatments in each study so they can be matched across studies; (2) developing a methodology to compute average quality for each treatment, which takes account of measurable differences across studies; and (3) matching these treatments with the medical claims data of millions of commercially-insured individuals. This newly combined data set contributes a novel and rich source of information for understanding and measuring cost growth and innovation in the health care sector.

To link our quality measures to theoretical objects of interest, we use the framework for consumer welfare derived in Cutler et al. (1998). In their framework, consumer welfare is a function of (1) the costs of care (which we measure using observed prices in insurance data and a separate dataset which we use to adjust for rebates),[3] (2) the health produced from medical care (which we measure by combining the Tufts quality measures with insurance claims that captures how treatments diffuse), and (3) the value of a statistical life year (we present results for a range of assumptions). As in Cutler et al. (1998), our consumer welfare measure directly connects to our quality-adjusted price index, which is a measure of the expenditure change necessary to maintain the same level of utility across periods.

[3]Throughout the draft we use the term cost and spending interchangeably, unless we refer specifically to "marginal costs." The cost we measure refers to the amount insurers and patients pay for a treatment.

We focus on 13 conditions (asthma, atrial fibrillation, colon cancer, cystic fibrosis, hypertension, hepatitis C, HIV, lung cancer, multiple sclerosis, osteoporosis, rheumatoid arthritis, schizophrenia, and venous thromboembolism) where we feel our methodology most accurately captures the innovations present and how the quality of treatment is changing. In particular, we focus on conditions: (1) where most of the treatment (or innovation) for the condition is through pharmaceuticals – where the mapping from the Tufts CEAR database to the insurance claims database is feasible and the mapping from treatment to outcomes is less complex; and (2) where the set of drugs we observe in the Tufts data accounts for nearly all pharmaceutical spending for that condition. The 13 conditions we study are important in their own right as they account for $210 billion, or 8%, of total medical expenditure and 14% of pharmaceutical expenditures in 2018.[4]

We find a lot of heterogeneity in trends across conditions, but conceptually conditions fall into two categories, which we refer to as innovative or non-innovative markets. In innovative markets, as new treatments enter and take market share, welfare changes are affected by both price and quality changes. In contrast, in non-innovative markets, welfare changes are determined mostly by price, as the quality of available products changes little. While the focus of our paper is on how innovation shapes markets, non-innovative markets demonstrate how these markets mature, including eventual patent expiration, providing a more long-run and complete view of how innovation impacts markets.

These two types of markets are exemplified by treatments for hepatitis C and rheumatoid arthritis over our study period. The main distinction is that hepatitis C is an innovative market with several new entrants over our sample period, while the major innovations for rheumatoid arthritis occurred in the late 1990s and early 2000s, and would be categorized as a non-innovative market over our sample period. For hepatitis C, there is considerable cost

[4]These results are based on the BEA Health Care Satellite Account (HCSA), which reports spending by condition and are weighted to BEA's Personal Consumption Expenditure (PCE) estimates of medical care spending (Dunn et al., 2015).

growth, which is driven by the diffusion of expensive new treatments. This leads to large increases in the cost of treating hepatitis C, but once quality is accounted for, consumer welfare increases and quality adjusted prices fall quickly. In contrast, rheumatoid arthritis has large cost increases because of within-drug price increases, with prices of key drugs more than doubling over our sample period. Meanwhile, the market shares for each treatment are remarkably stable over time. Hence for rheumatoid arthritis, we see relatively little quality improvement and rapidly rising quality-adjusted prices.

Overall, looking across all the conditions we study, we find considerable evidence of innovation. For all the conditions we consider (except colon cancer) there is at least some quality improvement and seven of our thirteen conditions have consumer welfare improving when assuming that the value of a statistical life year (VSLY) is \$100k, implying declining quality-adjusted price indexes . If we assume a VSLY of \$500k, then we find nine of our thirteen conditions have declining quality-adjusted prices. While we caution against extrapolating outside the sample, we aggregate across these conditions, weighting by spending. We find that without adjusting for quality, on average prices for these conditions increase by 70% from 2007 to 2018. Meanwhile, our quality adjusted price index, assuming a VSLY of \$100k, rises by 45% – a reduction of 1.5 percentage points in the compound annual growth rate. If we assume a \$500k VSLY, the quality-adjusted price index falls by 55%. This suggests that price indexes which do not account for quality improvements may be overstating price growth.

Looking more specifically at individual conditions, we find a number of novel and interesting results. Surprisingly, several of the markets where our measure of consumer welfare *declined* had a lot of innovation. These markets had large quality improvements, but those quality improvements were small relative to the cost increase. This is surprising because without distortions, a standard model of demand would suggest that an innovation would not diffuse if its price was so high that it lowered consumer welfare. To understand how

this occurs, we take the example of Orkambi, a breakthrough therapy for cystic fibrosis. In our data, insurers and patients combine to pay more than $150k per year for Orkambi, or well over $1 million in lifetime costs. However, patients pay on average just $1.5k per year out-of-pocket. We find a sizeable quality improvement of taking Orkambi of 0.85 quality-adjusted-life years (QALYs), where 1 QALY is a year of life in perfect health. Orkambi contributes to major health improvements for patients with cystic fibrosis. However, when accounting for the total cost (and the fact that other consumers may be bearing it through their insurance premium) this innovation reduces overall consumer welfare as the cost growth overwhelms the quality improvements in our framework.

This is not an isolated case. We find that for five of our conditions, consumer welfare in 2018 is lower than it would have been in a counterfactual we compute that removes new treatments from the market. This is consistent with the findings in Kyle and Williams (2017) who find high-cost, low-quality drugs diffuse faster in the U.S. than other countries.

However, we emphasize that while we find numerous examples of new innovations that lower consumer welfare, these innovations increase total welfare as the high price of these drugs increases producer profits. One interesting implication of consumer surplus falling due to innovation is that it means that producers are receiving more than 100% of the surplus from their innovations in 2018. A famous result in innovation economics is that patents provide insufficient incentives for innovation, as the innovator is not able to capture all the consumer surplus (and monopoly pricing creates deadweight loss)(Nelson, 1959; Arrow, 1962). Our results show that may not be the case if distortions lead products with negative consumer surplus to diffuse.

Another surprising result emerges in non-innovative markets. Non-innovative markets, like colon cancer and hypertension have falling price indexes and rising consumer welfare because of patent expiration. This highlights how the effects we discussed in the previous paragraphs change over the long-run. The high prices we see and reductions in consumer

welfare are potentially a short-run phenomenon. Interestingly, these two conditions are the only cases where we observe quality (and total welfare) decline. This occurs when older, lower quality treatments lose patent protection. Out-of-pocket prices fall for these relatively lower quality drugs and they gain market share, reducing the average quality of care – though at a lower price. That is, surprisingly, in our framework the only time we see total welfare declining is after generic entry. However, this effect can be short-lived. For hypertension, another higher quality drug comes off patent a few years later and average quality rises again. These novel observations about non-innovative markets and patent expiration highlight the importance of looking at the long-run impacts of innovation.

As noted by Bryan and Williams (2021), one of the fundamental challenges in measuring the value of innovation is taking measures of innovation, such as patents or clinical trial investments, and connecting them to "changes in welfare, which depend on how new innovations impact prices and health outcomes, but opportunities to construct such direct linkages to welfare-relevant outcomes are quite rare." In this paper, we construct these linkages and find a number of unconventional (but intuitive) results about how innovation and patent expiration shape welfare in health care markets. These novel results highlight the importance of empirical evidence in health care markets where multiple distortions can cloud theoretical predictions and conventional wisdom about innovation.

2 Literature Review

Our paper relates to multiple literatures on innovation in health care markets. Outside of healthcare, innovation generally leads to welfare improvements. However, the reverse is possible in markets where distortions may lead to inefficient pricing and the diffusion of products where costs exceed the benefits (Chandra and Skinner, 2012). Motivated by the distortions in health care markets, Chandra et al. (2016) explore whether health care is an

exception. Like us, they find the diffusion of higher quality care. This suggests that health care markets are responsive to quality, which they refer to as a "signpost of competition." However, as discussed in the previous paragraph, we find several cases of new treatments that diffuse which are higher quality, but not cost-effective.

For the purposes of our paper, the main advantage of assigning quality and cost measures at the treatment level is it allows us to be specific about which innovations are driving quality improvements and changes in costs. Given the difficulty of measuring individual innovations, one common approach to measure innovation is to control measurable drivers of spending (e.g., age, insurance, price, and incomes), then, following the logic of Solow (1957), the residual is attributed to innovation.[5] Our granular approach both allows us decompose the share of spending due to innovation, by observing these innovations directly, but also weigh the cost growth against measures of quality.[6]

Having such rich data on innovation and cost growth allows us to look at many conditions and innovations in a systematic, yet granular fashion. Because we apply a systematic methodology, our results are also more comparable across conditions, leading to general insights about how innovations shape cost and quality. We view this as an important contribution. While prior case studies have led to advancements in the literature, they vary in their assumptions and methodologies, as they adapt to the unique institutional details and features of each condition and innovation, making it difficult to generalize results, or gauge the relative magnitudes across studies.[7]

Our paper also relates to recent work by Cutler et al. (2022) and Weaver et al. (2022) who

[5]See Schwartz (1987), Newhouse (1992), Cutler (1995), Smith et al. (2009), and Smith et al. (2022).

[6]As discussed in the previous paragraph, innovation may not improve consumer welfare if there are market distortions. This has implications for interpreting the estimated share of cost growth due to innovation. As we have direct measures of quality, we can separate the cost growth due to innovation from the quality growth due to innovation.

[7]Papers which focus on individual or a few cases include: Almond et al. (2010); Cutler et al. (1998); Cutler and McClellan (2001); Romley et al. (2020); Shapiro et al. (2001); Berndt et al. (2002); Frank et al. (2004); Lucarelli et al. (2022); Eggleston et al. (2019); Dauda et al. (2022). See Sheiner and Malinovskaya (2016) for a more complete review.

use population-level measures of spending and health to derive measures of productivity and quality-adjusted medical-care price indexes across a comprehensive set of medical conditions. Outcome-focused measures better capture the economic object of interest (improved health) and abstract from the often non-linear process by which treatment impacts health (Chernew and Newhouse, 2011). On the other hand, the outcome-based approach requires observable outcomes and strong assumptions regarding whether the observed change in health is attributable to improvements in medical care. It also does not link specific innovations to health outcomes, which is essential for our paper.[8] In addition, the outcome-based approach may better capture quality improvements for conditions where outcomes (e.g., mortality or disability) are easier to measure, whereas our approach may better capture conditions where treatments may improve, rather than lengthening life. Hence, we view our paper as complementary to these outcomes-based papers. The methods answer similar questions, but rely on different assumptions.

Finally, our paper relates to Hult et al. (2018) and Dunn et al. (2022) who also use the Tufts CEAR data to construct quality-adjusted indexes. Dunn et al. (2022) show that medical innovations typically lead to quality-adjusted prices declining, but both of these papers use the CEAR data at a very aggregate level and Dunn et al. (2022) imposes strong assumptions regarding how technologies diffuse. In other words, these papers take a top-down approach, while this paper takes a bottom-up approach by matching specific treatments in the Tufts CEAR database to the diffusion of treatments in medical claims data. While the bottom-up approach requires a substantial amount of additional data work, it also provides much more detailed and rich information about which technologies are used in practice.[9]

[8]Notably, Cutler et al. (2022) acknowledge the potential challenge of attributing changes in population health to the medical care sector, so they also apply a disease model to cardiovascular conditions, which, similar to our paper, relies on the medical literature to measure the quality of treatment, rather than the observed health outcomes. For cardiovascular conditions, they find the two approaches yield similar results.

[9]Properly weighting innovations based on their usage is critical for accurately measuring welfare. For example, a major breakthrough innovation for a rare disease may have less of a welfare impact than a marginal innovation which diffuses broadly.

3 Background

3.1 The Case of Hepatitis C and Rheumatoid Arthritis

While we construct our index for 13 conditions, we begin our exposition with a focus on hepatitis C and rheumatoid arthritis as both observe costly new treatments, but the dynamics in each market are different in ways that help demonstrate how our methodology works and some of the main takeaways.

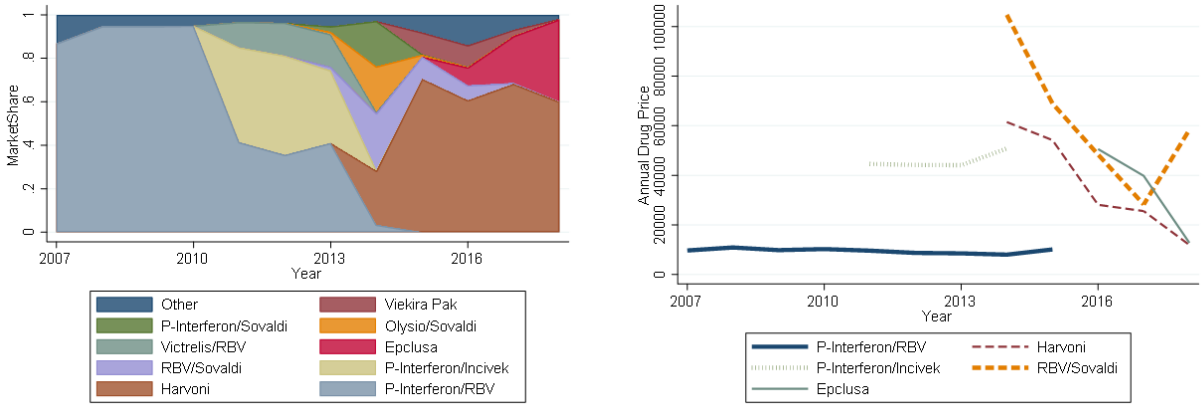
Hepatitis C is a viral infection that can cause inflammation of the liver. The condition is serious, but it may take months or years for symptoms to develop and for the disease to progress. If left untreated the disease can cause liver cancer, liver disease, liver failure, and potentially death. It has been estimated that over the 2013 to 2016 period, around 2.4 million individuals in the U.S. had hepatitis C (Hofmeister et al., 2019). Hepatitis C has a 10-15 year prognosis and is not typically fatal with proper treatment.

Hepatitis C drugs were in the national spotlight in 2014 after Gilead priced its breakthrough treatment, Sovaldi, at $84,000 per treatment regimen, a controversial decision at the time, but the drug was seen as curative with fewer side effects than the alternatives. While the cost of Sovaldi made headlines, in the context of our paper, hepatitis C is interesting because there was actually a sequence of important new innovations.

Figure 1 shows how the prices (adjusted for rebates using the SSR Health's rebate estimates) and market shares for the top hepatitis C treatments evolved during our sample period. At the beginning of our sample in 2007, the original standard treatment for hepatitis C was Pegylated Interferon (P-Interferon) and Ribavirin (RBV), which had low cure rates and severe side effects. In 2011, Incivek and Victrelis entered the market. These drugs were more expensive, but also higher quality than Interferon. However, these drugs were soon followed by Sovaldi (launched in December 2013), which was both a much higher cost and more effective than all previous alternatives. In fact, despite the relatively high price, it was

perceived by many as cost effective.[10] Finally, Harvoni, Epclusa, and Viekira Pak entered starting in late 2014. These drugs are even more effective than Sovaldi and also much less expensive, likely due to the greater competition among highly effective treatments upon entry.

Figure 1: Market share and prices for the top hepatitis treatments over time



Notes: The left panel of this figure presents the market shares by year for the 9 highest volume drugs for hepatitis C across our entire sample period in the MarketScan data. The right panel presents the average price per year of the 5 highest volume drugs in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

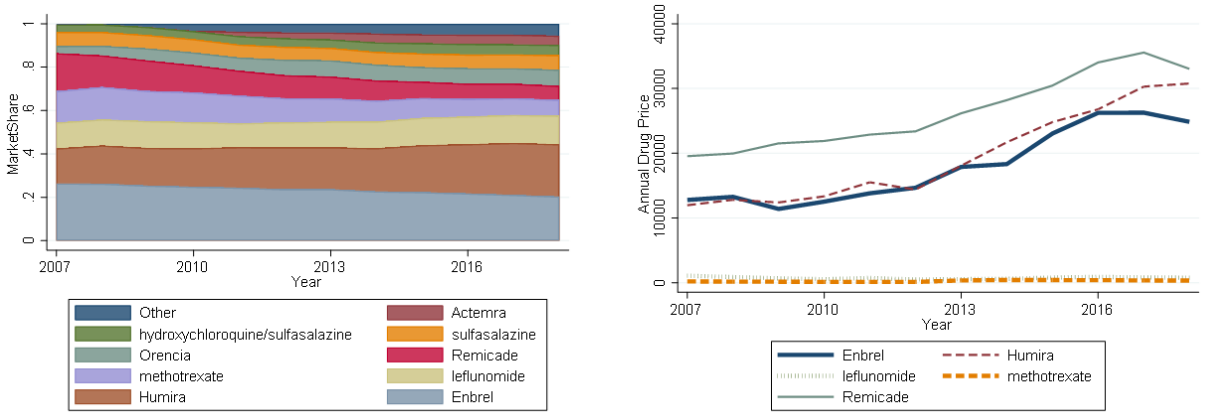
Rheumatoid arthritis provides a nice contrasting case to hepatitis C, where the market shares and prices for rheumatoid arthritis are shown in Figure 2. Rheumatoid arthritis is a chronic autoimmune condition associated with inflammation, severe joint pain, and, if untreated, joint deterioration. There are about 1.5 million people in the U.S. with rheumatoid arthritis. Rheumatoid arthritis can afflict those of any age, but the likelihood of onset increases with age and is the highest for those in their 60s. As with hepatitis C, rheumatoid

[10] "UK Says Sovaldi Is Worth It. We Should Listen." Forbes, August 2014.

arthritis is typically not fatal with proper treatment.

The baseline treatment for rheumatoid arthritis is methotrexate. Typically, patients begin treatment with methotrexate, which entered the market in 1947 to treat cancer and shown to be useful in treating rheumatoid arthritis in the 1980s. For some patients, methotrexate is less effective and over time the effectiveness of methotrexate may wane. When this occurs, there are a number of higher-cost disease-modifying antirheumatic drugs (DMARDs), the most popular of which are etanercept (Enbrel) and adalimumab (Humira), which entered the market in 1998 and 2002, respectively. This new generation of drugs are seen as highly effective at preventing significant joint deterioration and can reduce joint pain. However, these new drugs were already in the market prior to our sample period. Hence, we see almost no change in market shares (Figure 2), which implies limited changes in quality from new treatments for rheumatoid arthritis patients. The average patient in 2007 is getting a similar basket of treatments to a patient in 2018.

Figure 2: Market share and prices for the top rheumatoid arthritis treatments over time



Notes: The left panel of this figure presents the market shares by year for the 9 highest volume drugs for rheumatoid arthritis across our entire sample period in the MarketScan data. The right panel presents the average price per year of the 5 highest volume drugs in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

At the same time, rheumatoid arthritis treatments have gained notoriety for cost increases (Hopkins, 2021). As we show in Figure 2, the price for Enbrel doubled, while the price for Humira has nearly tripled (after adjusting for rebates and economy-wide inflation). Within-molecule price increases mean that costs are rising quickly, even if there are no major quality improvements for this condition.

In summary, hepatitis C and rheumatoid arthritis are two conditions which have highly effective new treatments and been noted for rising costs in recent years. However, for hepatitis C, these cost increases coincide with the diffusion of new innovative drugs, while for rheumatoid arthritis the market has matured and so there is not a similar shift towards potentially higher quality treatments.

The goal of this paper is to better understand how these changes in the market translate into quality improvements and cost increases. With each new innovation there is a potential

change in the value of medical care spending. Our methodology estimates average quality measures for these treatments and matches them with their respective market shares to better understand how the quality of treatment for the average person with each condition changes over time. We summarize the total welfare change using a quality-adjusted price index, which we describe in more detail below.

3.2 Cost-effectiveness Studies

The goal of this subsection is to highlight how cost-effectiveness studies, which are comparisons of two treatments, provide information which can be aggregated to compute consumer welfare at the disease level, as in Cutler et al. (1998).

Cost-effectiveness analysis is one of the most widely applied tools to guide policy surrounding the allocation of medical care resources (Meltzer and Smith, 2011). A standard cost-effectiveness analysis compares the cost and effectiveness of a medical intervention (I), such as a new treatment, with a "comparator" or standard of care (SOC) treatment (i.e., a commonly used treatment for a particular condition, or no treatment). The costs included in these studies include all costs, including insurer costs and out-of-pocket costs to patients and commonly cover the lifetime differences in costs. The effectiveness is typically measured in healthy years of life or quality-adjusted-life years (QALYs), where QALYs account for both the length of life and the quality-of-life. One QALY represents one year of life in perfect health.

To communicate the main methodological points, we start by focusing on just two conceptual technologies, I and SOC. Let S_I and S_{SOC} be the lifetime costs for the innovation, I, and standard of care treatment, SOC, respectively; and also let H_I and H_{SOC} be the effectiveness of treatment, as measured in QALYs, for the innovation and SOC treatment, respectively.

An important feature of cost-effectiveness studies is that the costs and health outcomes

for both I and SOC are measured identically across the two treatments, covering the same population and applying identical study features. This allows for the precise measurement of the relative cost and effectiveness of I compared to SOC, holding other variables fixed. The outcome of interest in cost-effectiveness studies is often the incremental cost-effectiveness ratio (ICER):

$$\text{Cost-effectiveness Ratio} = \frac{S_I - S_{SOC}}{H_I - H_{SOC}}. \tag{1}$$

The meaning of the ratio changes depending on whether the numerator and denominator are positive or negative values, but for the typical case where both the numerator and the denominator are positive, the ratio measures the cost increase per unit of healthy life years gained from the innovation. If a dollar value can be placed on healthy life years gained, then researchers can determine whether the innovation provides a net benefit. The dollar value placed on a QALY is often measured as a value of a statistical life year (VSLY), which conceptually captures an individual's value of living an additional year in good health. The innovation offers a net benefit relative to the SOC treatment if $VSLY > \frac{S_I - S_{SOC}}{H_I - H_{SOC}}$. That is, the innovation is beneficial if the benefits per healthy year of life (the left-hand side) exceed the cost per healthy year of life (the right-hand side).

Alternatively, the elements of this ratio can be re-arranged to express the net benefit or consumer welfare from the innovation in a dollar amount.

$$\Delta \text{ Consumer Welfare}_{I,SOC} = VSLY \cdot (H_I - H_{SOC}) - (S_I - S_{SOC}) \tag{2}$$

The first term, $VSLY \cdot (H_I - H_{SOC})$, captures the incremental dollar value in health benefits from innovation, relative to the SOC treatment, and the second term captures the change in cost, relative to the SOC treatment. One important observation from this equation is that a cost-effective treatment will increase consumer welfare if it replaces its comparator

(assuming a VSLY), while a treatment that is not cost-effective will lower consumer welfare if it replaces its comparator.

As an example, Chhatwal et al. (2015) calculate the net benefit from the introduction of Sovaldi for patients with hepatitis C.[11] From this study, Sovaldi improved health by 1.12 QALYs, while increasing costs by $24k. This implies a ICER of $21,000, an additional life year from using Sovaldi costs $21,000. Typically, policymakers have a threshold VSLY, where an ICER higher than that threshold would be considered to be cost-effective. The net welfare calculation (Equation 2) assumes a VSLY. Researchers in the literature have commonly selected VSLY ranging from $50,000 to around $400,000 (Neumann et al., 2014; Cutler et al., 2022; Viscusi, 2020). Assuming the value of a statistical life year is $100,000, then Sovaldi provides $112,000 in benefits due to extended life. Its incremental cost is $24,000. The net welfare calculation says that Sovaldi provides $88,000 in additional welfare for those who take it. Even though it is more expensive than comparators, the incremental health benefits are larger than the cost.

4 Consumer Welfare and Quality-Adjusted Price Indexes

In this section we describe the utility-based price index. The theory used to construct the index for the treatment of a condition has been outlined and discussed in other papers including Cutler et al. (1998), Sheiner and Malinovskaya (2016), Dauda et al. (2022), and Dunn et al. (2022). For most of our paper, we focus on a price index for a specific disease, indexed by d. Condition-based inflation measures are recommended in National Research Council (2002) and National Research Council (2011) and are useful when measuring quality

[11] This example is for expositional purposes and is the same example as that used in Dunn et al. (2022). In this example, Sovaldi is compared to pegylated interferon and ribavirin. Estimates from the study are converted to 2018 dollars.

changes, which typically affect the treatment of specific conditions.

As discussed in Fisher and Shell (1972), a utility-based cost-of-living price index measures the relative expenditures needed to maintain the same level of utility across periods, given changes in prices, and in our case, quality. This idea connects directly to the cost-effectiveness discussion in the previous section, but instead of calculating the consumer welfare from switching away from the standard of care treatment, SOC, to the innovative treatment, I, we calculate the consumer welfare of receiving a typical treatment at a point in time, $t-1$, relative to treatments received at time, t. Changing the subscripts in equation (2) accordingly, we obtain the following equation for the consumer welfare change over time:

$$\Delta \text{ Consumer Welfare}_{d,t,t-1} = VSLY \cdot (H_{d,t} - H_{d,t-1}) - (S_{d,t} - S_{d,t-1}). \tag{3}$$

As derived in Cutler et al. (1998), Consumer Welfare$_{t,t-1}$ is the compensating variation, accounting for the change in the price and quality of treatment.[12]

The associated price index measures the percent change in treatment expenditures needed to purchase a fixed level of utility across the two periods. This can be formed as a ratio where the denominator is the base-period average treatment cost and the numerator is calculated by subtracting the consumer welfare change from the base-period cost of treatment. Specifically, the price index for disease, d, is:

$$\begin{aligned}
\text{Price Index}_{d,t,t-1} =& \frac{S_{d,t-1} - \Delta \text{ Consumer Welfare}_{d,t,t-1}}{S_{d,t-1}} \\
=& \frac{S_{d,t-1} - [VSLY \cdot (H_{d,t} - H_{d,t-1}) - (S_{d,t} - S_{d,t-1})]}{S_{d,t-1}} \\
=& \frac{S_{d,t}}{S_{d,t-1}} - \frac{VSLY \cdot (H_{d,t} - H_{d,t-1})}{S_{d,t-1}}.
\end{aligned} \tag{4}$$

[12]This equation for consumer welfare is derived by taking a first-order Taylor series expansion of the utility function in Cutler et al. (1998). One important implication of this is it assumes away the risk premia of insurance and wealth effects. This simplification means that the marginal utility of a dollar is constant.

The first line of equation (4) shows that the price index falling (being less than 1) means that consumer welfare is rising, and vice versa. The middle line of equation (4) shows the complete formula. Suppose that for a condition, the average treatment cost in period $t-1$ is \$50,000. Similar to the discussion above, suppose that the diffusion of a new treatment leads to a 0.2 increase in QALYs for the average patient, but adds \$10,000 in average treatment costs. Assuming a VSLY of \$100,000, the change in consumer welfare is \$10,000: \$20,000 in improved quality of life, minus \$10,000 in net treatment costs. The cost of purchasing a bundle, which keeps utility constant, declined by 20% once quality changes are accounted for.

The last line of equation (4) separates the quality improvement from the price change. The first component, $\frac{S_{d,t}}{S_{d,t-1}}$, is the change in the cost of treating the disease across the two periods, ignoring changes in the quality of treatment over time. The second term is an adjustment term that accounts for the dollar value of the change in the quality of the treatment over time, $\frac{VSLY \cdot (H_{d,t} - H_{d,t-1})}{S_{d,t-1}}$, so that the index declines as the quality of treatment improves.[13]

In contrast to the consumer welfare calculated from cost-effectiveness studies, which involve just two treatments, the consumer welfare calculation used in the price index includes the full basket of treatments across two points in time, t and $t-1$. Typically there are a variety of treatments for each condition and different individuals are receiving different treatments at a single point in time, and the mix of treatments changes over time. Our measure of health care quality or cost in a given period is the average quality or cost of each treatment. Let $\mathcal{R}_{d,t}$ be the set of treatments available to a patient at time t, and $w_{r,d,t}$ be the share of the population with a condition that adopts treatment r at time t. Then, the average QALY at time t is:

[13]The formula we apply in this paper is a Laysperes formula that uses the initial period as the base period. Alternatively a Pacche formula could be applied that uses the last period as the base, as discussed by Dunn et al. (2022).

$$H_{d,t} = \sum_{r \in \mathcal{R}_{d,t}} w_{r,d,t} H_{r,d} \tag{5}$$

This can be interpreted as the average health benefit received by the population in time period t. The average cost of treatment is calculated similarly.[14]

5 Data

In order to construct the quality-adjusted indexes, we need to estimate the cost and QALY of each treatment, as well as determine the share of patients receiving each type of treatment. To do this, we use two main datasets: (1) Tufts CEAR data; and (2) the Merative™ MarketScan® Research Databases.

Tufts CEAR data

The Tufts CEAR data is compiled by the Center for Evaluation of Value and Risk in Health at Tufts University. The data consist of reviews of more than 8,000 original cost-effectiveness analyses which have been published in English and are indexed by Medline, from the years 1976-2018, though the bulk of studies start after 1990. Each study includes at least one comparison, which is a comparison between an intervention, often a new treatment, and a comparator, which is often a standard of care treatment. For example, in the case of hepatitis C, a study may include a comparison between Sovaldi versus P-Interferon and another comparison between Harvoni versus P-Interferon. The unit of observation in the raw data is a comparison and there are a total of more than 22,039 comparisons.

Each study is independently read by two reviewers to ensure accuracy. Reviewers record more than 40 variables, the important ones for our study includes the QALY and cost for each treatment in a comparison, strings which describe both treatments in each comparison,

[14] Specifically, the costs are calculated as: $S_{d,t} = \sum_{r \in \mathcal{R}_{d,t}} w_{r,d,t} S_{r,t,d}$. As we discuss in greater detail later, claims data are used to estimate treatment costs.

a detailed disease classification (i.e. asthma, hepatitis C) and 3-digit ICD-10 code classifications, though we focus on the detailed classification. It also includes information on the study: the journal, author, author affiliation, funding, year published, country the study was performed in, among other characteristics.

The Tufts CEAR data contain detailed information about each study, but it is not in a form that is readily combined with other data sources or across studies within the Tufts CEAR as the treatment information is given in sentence form. For example, a treatment may be "Sofosbuvir, 12 weeks + pegylated interferon-alpha-2a and ribavirin, 12 weeks" or "Pan-Genotypic direct-acting antiviral agent regimen." We had at least two research assistants review and independently classify each treatment into specific pharmaceutical interventions, i.e. molecules. Accuracy was then verified by an additional review of the independent classifications.[15] This classification contrasts our data with Hult et al. (2018) and Dunn et al. (2022) that use the raw CEAR data, but do not connect treatments across studies or link specific technologies to medical care claims.

To focus our analysis, we concentrate on 13 conditions where most of the treatment (or innovation) for the condition is through pharmaceuticals as these are much easier to classify in the Tufts and merge to claims data.[16] Limiting the data to these conditions leaves us with 5,414 comparisons, out of the 22,039 initial comparisons. That is, these 13 conditions account for about 25% of the Tufts data. Appendix Section OA.D.2 explores how well the Tufts data captures spending in the MEPS data. Drugs we classify in the Tufts data account for at least 79% of MEPS drug spending on their respective conditions, for all conditions, except atrial fibrillation (63%).[17]

[15]Focusing on the molecule level abstracts from some information, such as dose, form, or length of treatment, but this information is not consistently included in the Tufts CEAR data.

[16]While we tried classifying procedures, the terminology in Tufts did not always map easily to procedure codes in claims data.

[17]We do this exercise with the 2007-2017 MEPS data because the MEPS data includes diagnosis codes on drug claims. For this statistic, we limit our analysis to drugs that account for at least 5% of market share, but present results without that restriction in the appendix. This calculation also does not include cystic

For the 13 conditions we focus on, we keep all comparisons where two classified drugs are compared to each other,[18] and both drugs have a non-outlier QALY and cost estimates associated with them.[19] In particular, we drop 2,232 comparisons where either cost or QALY information is missing, 197 observations with outlier costs or QALYs, and 1,549 comparisons where one or both of the treatments is not classified. In addition, there are 349 comparisons where we could classify at least one of the treatments as a placebo (or "no treatment") or "standard of care." While these categories do not map to specific drugs, they provide information which may be useful when comparing to drugs indirectly. In our main specification we drop these categories, leaving 1,087 comparisons, but results are robust to including them. In our main sample we have 151 treatments across the 13 conditions.

MarketScan Data

After classifying each treatment, we link the CEAR data to insurance claims by molecule. We use the Merative™ MarketScan® Research Databases from 2007-2018. The MarketScan database contains retrospective insurance claims for a sample of commercially-insured patients who are under-65. We limit our sample to those who are not in capitated plans, enrolled for 360 days, and have drug benefits. This accounts for 220,658,074 member-years. Individuals in our data have unique identifiers which we can link to claims files, so we can match diagnoses and treatments to individuals.

There are two types of claims files, medical claims and pharmacy claims. Each observation

fibrosis, as cystic fibrosis is masked in the MEPS due to confidentiality thresholds. It also does not include colon cancer and lung cancer where chemotherapy drugs are mostly given in an office setting and MEPS does not report procedure codes needed to identify specific treatments.

[18]Many of the observations in the Tufts data include non-drug interventions which are vague or difficult to match to procedure codes (e.g. surgery), difficult to observe in claims data (e.g. diet and exercise booklets provided). Sometimes there are vague drug references like "statin therapy" which cannot easily be matched to a particular molecule.

[19]One common situation is the study will report the difference in QALYs or cost, but not the level of the two which leaves missing values. Outliers are QALY estimates greater than 100, cost estimates are greater than $10,000,000. Because our estimates are based on proportional effects, we also classify observations where the cost or QALY of one treatment is 5 times as large an another as an outlier. This is typically the case with very small QALY estimates, for example if one treatment provides 0.05 QALYs and another provides 0.3.

in the data corresponds to a line item in an "explanation of benefits" form; therefore each claim can consist of many records and each encounter can consist of many claims. Medical claims have information on the diagnosis (characterized by ICD-9 and ICD-10 codes), the procedures performed, and the price (this is the actual amount paid by the insurer and the member, combined). Pharmacy claims data have information on the price paid at the pharmacy and the specific drug prescribed, by NDC code (which incorporates a molecule-manufacturer-dose-form). Pharmacy claims do not have diagnosis or procedure codes.

To account for manufacturer rebates, we supplement the MarketScan data with SSR Health Data, which has also been used by Kakani et al. (2020) in the economics literature. SSR Health, LLC collects data from drug manufacturers' U.S. Securities and Exchange Commission (SEC) filings on revenue net of rebates and merge that with measures of revenue gross of rebates collected by Symphony Health to estimate the share of revenue that is rebated. See our data appendix, Section OA.D, for more details about how we clean the Tufts data and merge it to the MarketScan data, as well as how we incorporate rebates into estimates.

6 Methods

Section 6.1 discusses how we estimate QALYs using the Tufts data. Section 6.2 discusses how we use the MarketScan data to estimate costs.

6.1 Estimating QALYs from Tufts Data

Treatment level QALYs can be taken directly from the raw Tufts CEAR data for specific studies, but this is not the preferred approach for obtaining precise estimates for several reasons: (1) the Tufts data often have multiple observations for each treatment, necessitating some averaging; (2) there is variation in study design, populations, and assumptions which

will affect each treatment in a comparison; and (3) there is variation in the drugs that treatments are compared to.[20] We use a regression to address all three of these issues (i.e., (1) compute an average; (2) account for heterogeneity in study design; and (3) difference out comparison-specific factors).

Cost-effectiveness studies are centered around comparisons between an intervention and a comparison treatment. While the intervention and comparison treatments are different, all other aspects of the comparison are identical, including the identical populations, settings, and study parameters. In the Tufts data, the unit of observation is a comparison, which we subscript with u. We reshape the Tufts data so the comparator and intervention treatments, subscripted by $c \in \{\text{intervention, comparator}\}$, are separate observations that are part of the same comparison, u. Therefore, in our regressions the unit of observation is a specific comparison and treatment with the unique subscript, u, c.

Each observation also corresponds to a given treatment r and disease d. Denote the set of treatments used for disease d as $r \in \mathcal{R}_d$. Many different studies may contain a common treatment (e.g., Sovaldi appears in multiple observations), and there are many studies for a given disease (e.g., there are many comparisons and treatments for hepatitis C).

To average across quality measures for specific treatments, we use a linear regression model, that allows us to control for the different features of each study. The specific regression is:

$$log(H_{u,c,d}) = \gamma_{r,d} + \gamma_{u,d} + \epsilon_{u,c,d} \tag{6}$$

where the dependent variable is the log of the QALY. The $\gamma_{r,d}$ and $\gamma_{u,d}$ are treatment and comparison-specific fixed effects, respectively. The $\epsilon_{u,c,d}$ is the error term. We use logs

[20]For example, Harvoni is compared to P-Interferon and ribavirin twice, P-Interferon, ribavirin, and Sovaldi seven times, and ribavirin and Sovaldi five times. Finally, in some cases the intervention and comparator drugs are switched from study to study, i.e. we observe both Harvoni as the comparator and Sovaldi as the intervention, and Sovaldi as the intervention and Harvoni as the comparator.

because it places less weight on outlier observations and we also think it is likely that differences across treatments and comparison groups lead to proportional effects on health (e.g.,treatment A is 20% more effective than treatment B). However, as a robustness check we also repeat the analysis in levels and obtain similar results.

The treatment specific fixed effect, $\gamma_{r,d}$, provides a measure of the log difference treatments, relative to the left out alternative. This is the main coefficient of interest, as it provides an average relative value of each treatment which will form the basis of our estimates of treatment QALYs. The comparison specific effect, $\gamma_{u,d}$, is intended to difference out observed and unobserved heterogeneity across studies that are present in both the intervention and the comparator. As mentioned previously, it might be that a particular comparison has a different target population, different assumptions on the discount factor, or other study or comparison-specific factors, which will be captured with the $\gamma_{u,d}$ fixed effect.

While the estimate of $\gamma_{r,d}$ is key to our analysis, $\gamma_{r,d}$ are estimates of proportional effects and need to be converted into levels. To do this, we need an estimate of what they are proportional to. One option would be to choose a value of $\gamma_{u,d}$ from a particular study or choose an average of $\gamma_{u,d}$. Rather than take these approaches, we account for observable differences across studies. Specifically, to standardize the value of $\gamma_{u,d}$, we run a regression of the value of $\gamma_{u,d}$ on the characteristics of each study to create a standardized value of $\gamma_{u,d}$ that accounts for the different characteristics of the cost-effectiveness studies (e.g., age, sex or time horizon of the study). See appendix section OA.D.3 for additional details and robustness checks where we apply alternative methods to standardize $\gamma_{u,d}$. These differences do not change results much.

After obtaining regression coefficients, we retransform our estimates into levels using a smearing factor as proposed in Duan (1983). This gives us a prediction for the level of QALYs for a given treatment, relative to an omitted treatment. For disease level estimates, we can calculate the average QALY by taking a quantity weighted average across all treatments in

a given year, as in Equation 5.

The same steps can be taken to extract cost information from cost-effectiveness studies by replacing QALYs with costs on the left-hand side of Equation (6). We report results in Appendix Table OA34, however we strongly prefer using the MarketScan data to estimate costs, as cost-effectiveness studies only reflect costs at a single point in time.[21]

6.2 Estimating Costs from Claims Data

To calculate costs, denoted $C_{d,t}$, for a condition d, we begin by summing over all the expenditures a person has for condition d in a given year.[22] We deflate all expenditures to 2018 dollars using the aggregate Personal Consumption Expenditure (PCE) deflator. We include inpatient and outpatient claims in our annual spending measure, but because we do not map this spending to the CEAR data, this spending is assumed to have no innovation. Therefore, our results are likely understating true quality changes.

While we compute average annual costs from the claims data, the theoretical exercise underlying a QALY is to measure the lifetime improvement in health. Hence, we rescale all costs by a "lifetime scaling factor." To compute this scaling factor, we take into account how spending evolves over time for an individual and the expected length of life for someone with condition d. While computing lifetime costs is challenging, we find that our main points are fairly robust to the methodological details.

First, we calculate the evolution of expenses for someone with disease d. For example,

[21] A central result of this paper is the importance of within-molecule price increases or decreases (e.g., due to patent expiration). Because the Tufts costs are based on a single point in time, this price trend is absent for conditions where most of the price change is within-molecule. In addition, they only reflect costs for a particular setting (which may be in a very different health system as many studies are not based in the United States).

[22] Inpatients and outpatient claims include diagnosis codes, so for those claims we sum allowed amounts for any claims where condition d is the first listed diagnosis. Drug claims do not include diagnosis codes, which complicates knowing whether a prescription was for a given condition. In our preferred specification, we include all drugs that we classify using Tufts CEAR data, which covers most drug expenditures for our selected conditions. In Appendix Table OA35 we allocate all drug claims to medical conditions to pick up drugs that are not in the Tufts. Our results do not change much.

someone with hepatitis C typically has one expensive year of treatment, then relatively few expenses thereafter (some monitoring), whereas someone with rheumatoid arthritis typically takes expensive DMARDs every year for a lifetime. To measure this cost progression, we look at spending patterns over time for a large panel of individuals with disease d, which we use to extrapolate spending in the future.

We use the cost trend from the panel of individuals to capture how costs evolve after the initial diagnosis. To construct the net present value of lifetime cost we combine our estimate of how cost evolve with information on the probability of dying at a given age using life tables and the age distribution of individuals with disease d in the MarketScan data, assuming a discount rate of 0.97. This information is used to construct a scaling factor that we can use to multiply the cost of a typical year of treatment, $C_{d,t}$, into a lifetime cost estimate.[23] See Appendix Section OA.D.4 for a more thorough discussion of this calculation.

7 Results

We begin by describing detailed results for hepatitis C and rheumatoid arthritis. Then, we show summary results for all 13 conditions.

7.1 Detailed Results for Hepatitis C and Rheumatoid Arthritis

The left panel of Table 1 presents results for the nine highest revenue hepatitis C drugs. Column 1 indicates the baseline drug which all other drug's QALYs are compared to. Column 2 presents estimates of the average QALY for each drug, relative to the baseline drug. The second generation of drugs, Victrelis and Incivek, are more effective than first generation P-Interferon/RBV: they respectively provide 0.8 and 1.4 additional QALYs relative

[23] We implicitly assume that for each condition, how annual costs are scaled to lifetime costs does not change over time.

to P-Interferon/RBV based treatment. However, these drugs were not as effective as P-Interferon/Sovaldi, which provides 2.3 QALYs relative to P-Interferon/RBV based therapies. The second generation of drugs exited the market in 2015 because they were less effective than Sovaldi and anticipated falling market shares.[24] Finally, the newest generation of drugs Harvoni, Viekira Pak, and Epclusa are the most effective. Each provides more than 2.7 QALYs compared to P-Interferon/RBV.

Table 1: QALY Estimates for Hepatitis C and Rheumatoid Arthritis Drugs

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
Epclusa	0	2.763
Harvoni	0	2.834
Olysio/Sovaldi	0	2.853
P-Interferon/Incivek	0	1.376
P-Interferon/RBV	1	0.000
P-Interferon/Sovaldi	0	2.287
RBV/Sovaldi	0	1.186
Victrelis/RBV	0	0.812
Viekira_Pak	0	2.816

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
Actemra	0	3.164
Enbrel	0	2.153
Humira	0	2.295
Orencia	0	2.641
Remicade	0	1.907
hydroxychloroquine/sulfasalaz	0	2.165
leflunomide	0	0.048
methotrexate	1	0.000
sulfasalazine	0	-0.259

Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. The left panel presents results for hepatitis C, the right panel for rheumatoid arthritis. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

The right panel of Table 1 presents results for the nine highest revenue rheumatoid arthritis drugs, where methotrexate is the baseline treatment. Enbrel has 2.2 QALYs relative to methotrexate, suggesting that Enbrel was a large innovation at the time of its introduction. The newer generation of DMARDs, such as Humira, Orencia, and Actemra all have higher estimated QALYs to Enbrel. These estimates are picking up the generational difference in drug quality. Furthermore, all of these newer drugs appear to be highly effective providing at

[24] "From Riches to Rags: Vertex Discontinues Incivek as Sales Evaporate." Wall Street Journal, August 2014. "Merck stops production of HCV drug due to low demand." Drug Topics, January 2015.

least 2.2 additional QALYs relative to the baseline. In fact, the QALY improvements from the newer generation of rheumatoid arthritis drugs appears to be similar in magnitude to the QALY improvements we see from hepatitis C drugs, relative to the baseline treatment.

While each of the new innovations improve quality, the overall welfare change in the market depends on how much these treatments are used and how costly they are. A highly effective treatment which few people use may provide less welfare than a slight improvement which diffuses broadly. We combine information in Figures 1 and 2 and Table 1 to estimate how average quality and costs are changing.

Quality and price index trends for hepatitis C are shown in Table 2. Column 1 of Table 2 calculates how quantity weighted QALYs are changing over time for the treatment of hepatitis C, relative to 2007. In 2011, when Incivek and Victrelis enter, the average treated hepatitis C patient receives 0.79 more QALYs than they would have in 2007. In 2014, with the emergence of Sovaldi, that number jumps to 2.2 QALYs and in 2018 it is 2.9 QALYs after the entry of Harvoni.[25]

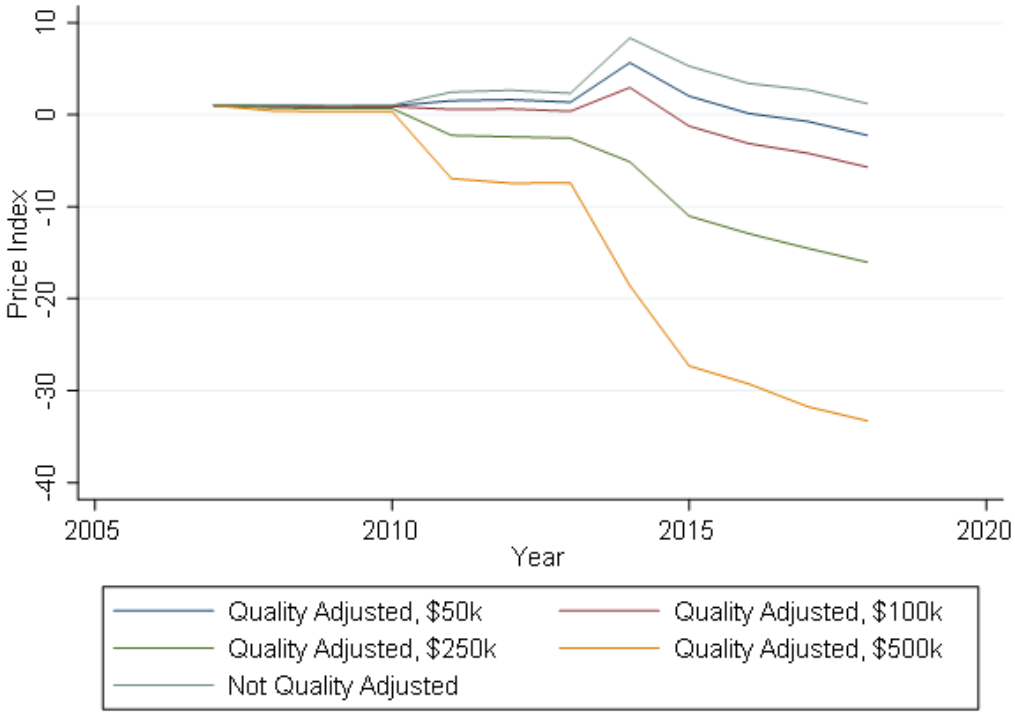
[25]These numbers are somewhat larger than the differences in Table 1. This is because there are other treatments like interferon (rather than P-Interferon) and P-Interferon without ribavirin that have fewer QALYs than the baseline treatment and positive market share in 2007.

Table 2: Price Indexes and Changes in Welfare by Year for Hepatitis C

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	42	1.000	1.000	0	1.000	0
2008	0.056	45	1.078	0.945	2	0.413	25
2009	0.056	42	0.990	0.858	6	0.331	28
2010	0.058	43	1.028	0.890	5	0.338	28
2011	0.792	104	2.460	0.578	18	-6.948	335
2012	0.850	112	2.650	0.631	16	-7.447	356
2013	0.821	98	2.330	0.382	26	-7.413	354
2014	2.269	351	8.346	2.959	-82	-18.589	825
2015	2.746	222	5.280	-1.240	94	-27.322	1,193
2016	2.752	143	3.405	-3.130	174	-29.267	1,275
2017	2.903	114	2.708	-4.185	218	-31.757	1,380
2018	2.905	51	1.205	-5.693	282	-33.285	1,444

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure 3.

Figure 3: Price Indexes for Hepatitis C



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. A subset of these indexes are also shown in Table 2. These results are constructed using data from Tufts, MarketScan, and SSR Health.

The average lifetime cost is shown in column 2, which is the average annual cost of hepatitis C in a given year multiplied by our lifetime multiplier. The average person who received hepatitis C treatment in 2007 has an estimated lifetime cost of $42k. Column 3 presents the price index without quality adjustment, which is simply the average lifetime cost of treatment in that year divided by the average lifetime cost in 2007. We can see the change in drug generation reflected in the costs. Costs are roughly $42k until 2011, then they rise to roughly $104k due to the entry of Incivek and Victrelis. Then in 2014, following the launch of Sovaldi, the average cost jumps to $351k, an 835% increase from 2007. However, Sovaldi's market dominance was short lived. Prices dropped sharply as competitors entered

at lower price points. By 2018, prices had fallen to \$50k, only 20% higher than 2007.

Columns 4 and 5 show the quality adjusted price index and change in consumer welfare assuming a \$100k VSLY. In addition, price indexes for \$50k, \$250k, and \$500k are shown graphically in Figure 3. Given columns 1 and 2, one can construct all the other estimates in this table or using any other assumed VSLY using equations 3 and 4. For example, in 2015, the average QALY was 2.74 QALYs higher than in 2007 when most patients were receiving interferon based treatments. At a \$100k VSLY, this represents \$274k of welfare. Given the \$180k difference in average costs, this represents a \$94k gain in consumer welfare. Likewise, the index is $\frac{222-100 \cdot 2.74}{42} = -1.24$. If consumers value life more, the quality adjustment gets larger. If one assumes the VSLY is instead \$500k, the index becomes $\frac{222-500 \cdot 2.74}{42} = -27.3$. The price index for hepatitis C is negative in the last few years of the sample. This indicates that the gain in health is so large that individuals would actually need to be paid more than the price of the new technology to use the older technology, in order to maintain the same level of utility across periods.

Once one accounts for QALY differences, the prices appear to be declining with each subsequent generation of new drug. The second generation, in 2011-2013, is roughly 2.5 times as expensive as the first generation of drugs, but at roughly 0.8 additional QALYs means that quality adjusted prices are lower than the original generation. In 2014, the introduction of Sovaldi meant a large unadjusted price increase, and even reduction of consumer welfare at \$100k VSLY, but at larger assumed VSLY, this meant prices falling further. The most recent generation of drugs both reduced costs and had higher quality leading to very large quality adjusted price declines.

In summary, hepatitis C is a condition which has been an innovative market in the last decade. While the treatments have been controversial due to their high costs, the treatments appear cost effective so the quality-adjusted indexes are well below 1, while the high prices lead to unadjusted price indexes above 1.

Trends for rheumatoid arthritis, shown in Table 3, contrast starkly with the trends for hepatitis C, as rheumatoid arthritis has arguably been a non-innovative market over the entire sample period. Column 1 calculates how quantity weighted QALYs are changing over time for rheumatoid arthritis. In contrast to hepatitis C, average QALYs are not rising by as much, a 0.2 increase between 2007 and 2018. Recall that the major innovations for rheumatoid arthritis took place in the late 1990s and early 2000s, with the introduction of Enbrel and Humira, but we observe relatively few innovations over our study period, as reflected in the lack of market shares shifting in our data (Figure 2). Consequently, there is little change in our estimated average quality of the treatments.[26] As the previous results make clear, this is driven by the lack of diffusion of these newer drugs, rather than the lack of efficacy of these treatments; the newer generation of rheumatoid arthritis drugs have similar relative QALYs as the newest generation of hepatitis C drugs. Column 2 shows that the average lifetime cost in 2007 for a patients with rheumatoid arthritis is \$198k. Lifetime costs almost double during our sample period to \$394k. As discussed in subsection 3.1, this is due to large within-molecule price increases. Drugs like Enbrel and Humira more than doubled their price during our sample period.

Changes in consumer welfare and quality-adjusted price indexes for rheumatoid arthritis are shown in columns 3-7. The price indexes are also presented graphically in Figure 4. In this case, while quality increased some, the high cost of the condition in the base period and the large price increases mean that even after adjusting for quality, the price indexes are increasing. If one assumes a \$100k VSLY, prices grew by 89% over our sample period.

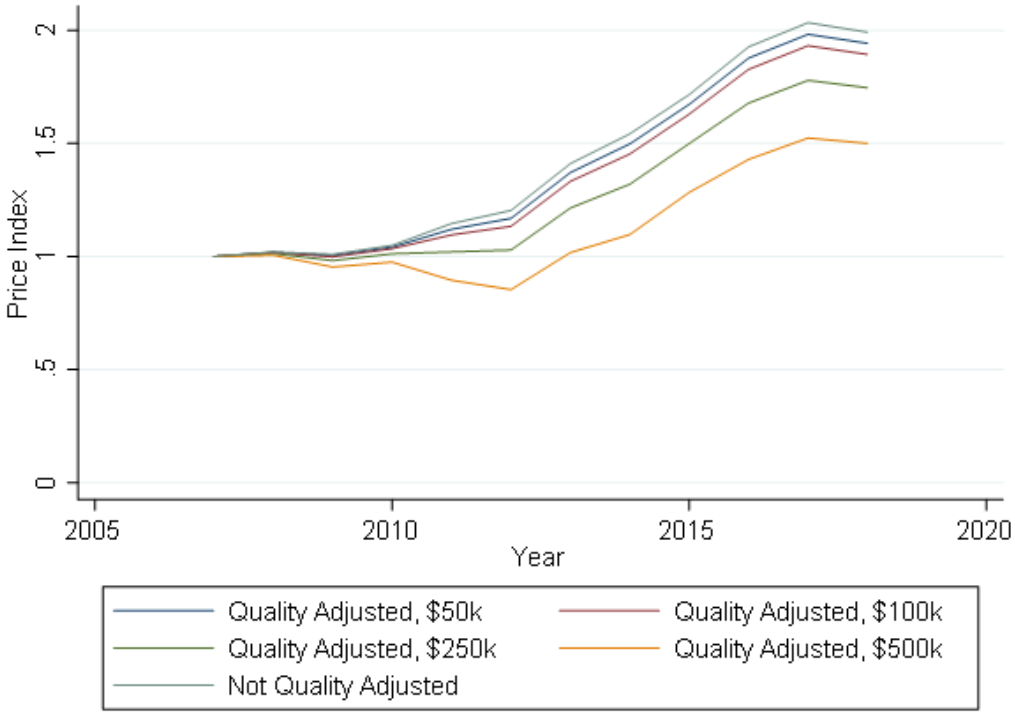
[26]The small change in QALYs that we observe is largely driven by the entry of Actemra, which was approved in 2010 and coincides with a distinct jump in QALYs in 2010.

Table 3: Price Indexes and Changes in Welfare by Year for Rheumatoid Arthritis

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	198	1.000	1.000	0	1.000	0
2008	0.006	202	1.020	1.017	-3	1.006	-1
2009	0.022	200	1.010	0.999	0	0.954	9
2010	0.030	208	1.049	1.034	-7	0.975	5
2011	0.099	227	1.146	1.096	-19	0.895	21
2012	0.139	239	1.204	1.134	-27	0.854	29
2013	0.156	280	1.411	1.332	-66	1.018	-3
2014	0.177	306	1.543	1.454	-90	1.097	-19
2015	0.171	340	1.716	1.630	-125	1.284	-56
2016	0.197	382	1.927	1.828	-164	1.429	-85
2017	0.202	403	2.034	1.932	-185	1.524	-104
2018	0.195	395	1.992	1.894	-177	1.500	-99

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3and 4. The price indexes are also graphed in Figure 4.

Figure 4: Price Indexes for Rheumatoid Arthritis



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. A subset of these indexes are also shown in Table 3. These results are constructed using data from Tufts, MarketScan, and SSR Health.

7.2 Results for Other Conditions

We run these same analyses for all 13 conditions. Detailed tables and figures like those presented for hepatitis C and rheumatoid arthritis are presented for the remaining 11 conditions in Appendix Section OA.A. Table 4 summarizes the results by showing the 2018 price indexes for all 13 of the conditions. This table presents the last row of tables 2 and 3, except it presents the 2007 cost rather than the 2018 cost and we present total welfare in column (7), which we discuss in section 7.3. Column (1) shows the total change in QALYs over the period from 2007-2018 and column (2) shows the base period expenditures for treatment in

2007. Columns (3), (4), and (6) show the price index values for 2018 with various assumptions about the VSLY. Column (5) shows the change in consumer welfare assuming a VSLY of $100k.

Table 4: Price Indexes and Changes in Welfare for Each Condition Between 2007 and 2018

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.002	16	1.013	0.999	0	0.941	0
AtrialFibrillation	0.461	16	2.977	0.136	14	-11.228	46
ColonCancer	-0.060	403	0.543	0.558	178	0.618	-6
CysticFibrosis	0.233	718	4.349	4.316	-2,380	4.187	23
HIV	0.188	348	1.505	1.451	-157	1.235	19
HepatitisC	2.905	42	1.205	-5.693	282	-33.285	291
Hypertension	0.040	9	0.684	0.250	7	-1.487	4
LungCancer	0.654	281	1.899	1.666	-187	0.734	65
MultipleSclerosis	0.440	507	3.025	2.938	-982	2.590	44
Osteoporosis	0.035	7	1.680	1.185	-1	-0.799	3
RheumatoidArthritis	0.195	198	1.992	1.894	-177	1.500	20
Schizophrenia	0.120	39	0.826	0.514	19	-0.732	12
VenousThromboembolism	0.105	6	1.292	-0.364	9	-6.989	11

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time.

Starting with Column (1) of Table 4, we see a lot of heterogeneity in the change in mean QALYs. All conditions we measure had quality improvements during this time period, except colon cancer (which we explain in Section 7.3). However, these changes vary by

condition. To highlight where changes in QALYs (and costs) are coming from, Figure 5 shows market shares for 6 selected conditions. Hypertension and colon cancer along with asthma and schizophrenia, (see Appendix Figures OA1 and OA31), have relatively small changes in market shares, and few new entrants over our sample period. Other conditions in Figure 5 may be categorized as innovative. Osteoporosis has the entry of denusumab and cystic fibrosis has the entry of Orkambi. Atrial fibrillation and multiple sclerosis have multiple new entrants that take considerable market share.

One important takeaway from Column (1) is that two things need to happen for significant quality improvement: (a) the condition needs to have new treatments which make large improvements in quality; (b) these treatments need to diffuse. Rheumatoid arthritis has highly effective new treatments, but they were mostly introduced prior to our sample period and the market shares for treatments are fairly constant, so quality improvements are relatively small. On the other hand, osteoporosis had the entry and diffusion of denusumab, but we estimate that denusumab has only an incremental improvement in quality, so quality gains are modest. It is worth noting that capturing these quality changes requires both sources of data, to measure both the quality improvement (i.e, cost-effectiveness studies) and diffusion (i.e., claims data).[27]

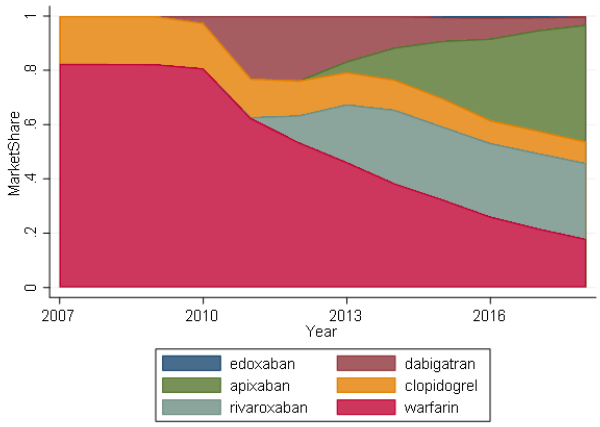
Columns (3)-(6) of Table 4 show the changes in consumer welfare and price index values for each of the conditions in 2018. Price trends differ considerably for each condition. To help us summarize the results, Figure 6 shows prices for the top 5 treatments for selected conditions.

Within conditions categorized as "non-innovative" markets, there are conditions where costs are rising and those where costs are falling. Rheumatoid arthritis, which we showed

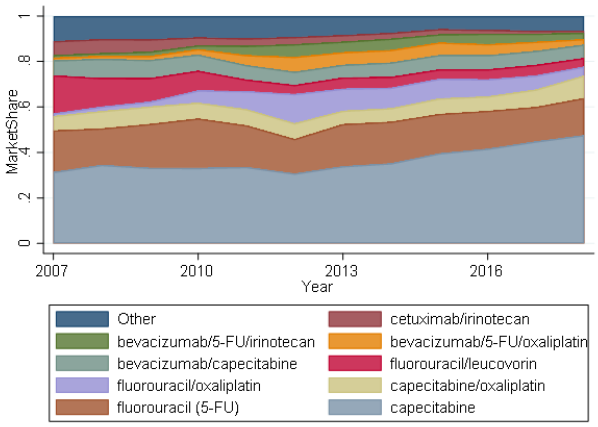
[27]Without merging cost-effectiveness studies to claims data, as in the top-down approach of Dunn et al. (2022), the Tufts data would have suggested that rheumatoid arthritis was highly innovative as the Tufts data suggests that the new DMARDs have large quality improvements over older generations of drugs. Likewise, the claims data show large changes in market share for osteoporosis, but, on their own, the claims do not provide proper context to assign quality.

in Figure 2, is the clearest example of costs rising. There are rapid within-molecule price increases and almost no change in the treatment mix. Hypertension, colon cancer, and schizophrenia (shown in Appendix Figure OA32) have multiple drugs come off patent during our sample period and little entry. These conditions have declining unadjusted price indexes. For this latter group, consumer welfare increases mostly because prices are falling, rather than quality improving.

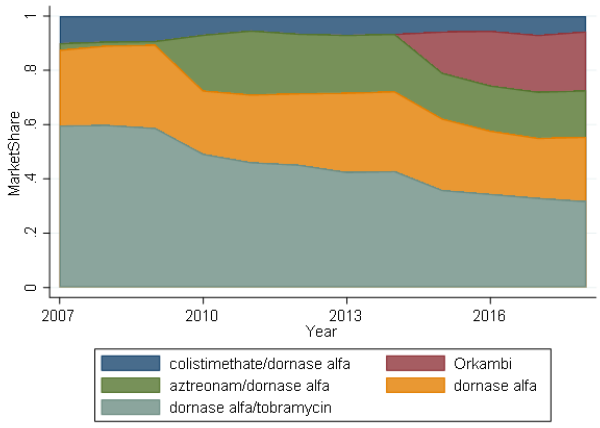
Figure 5: Market share for top treatments over time



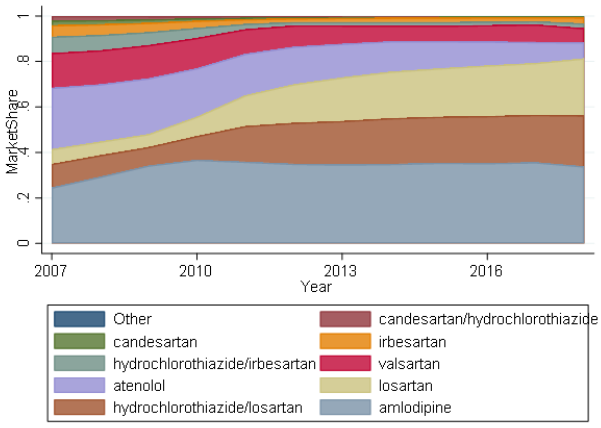
(a) Atrial Fibrillation



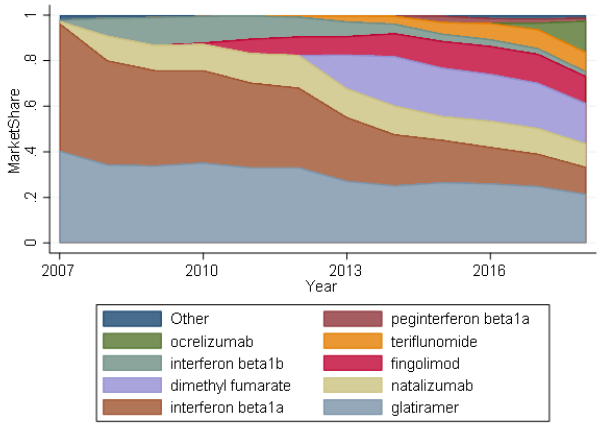
(b) Colon Cancer



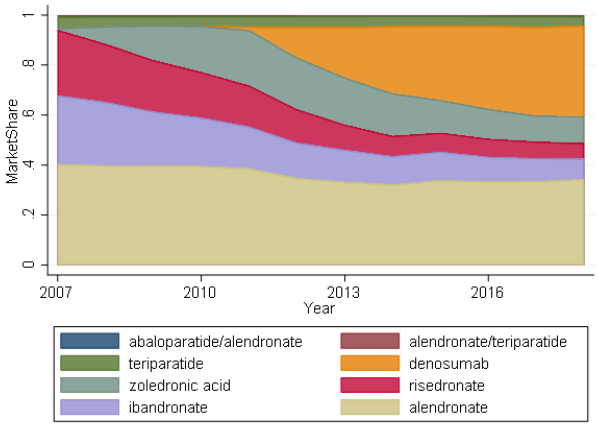
(c) Cystic Fibrosis



(d) Hypertension



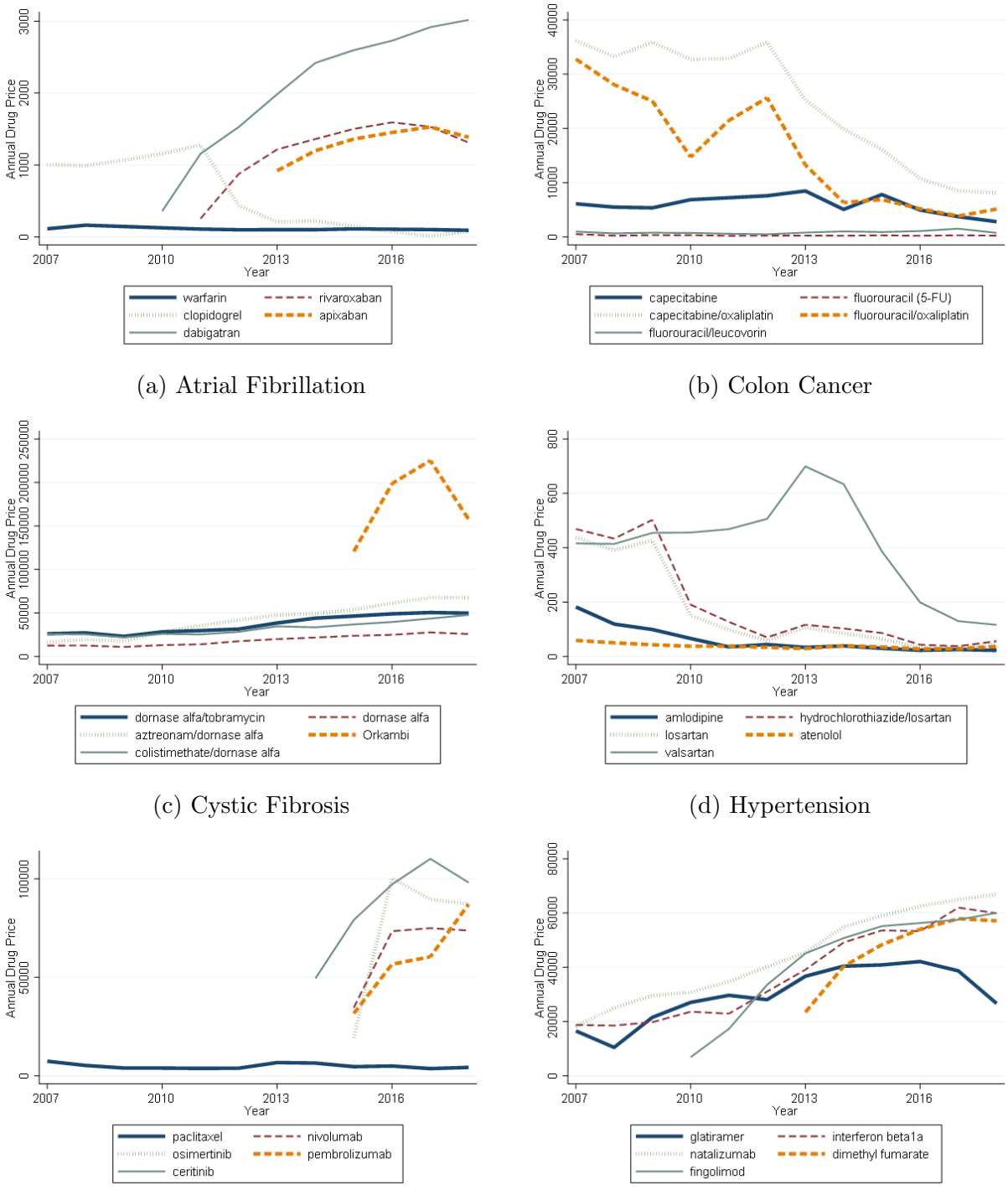
(e) Multiple Sclerosis



(f) Osteoporosis

Notes: This figure presents the market shares by year for the highest volume drugs for selected condition across our entire sample period in the MarketScan data.

Figure 6: Prices for the top 5 treatments for selected conditions



(a) Atrial Fibrillation

(b) Colon Cancer

(c) Cystic Fibrosis

(d) Hypertension

(e) Lung Cancer

(f) Multiple Sclerosis

Notes: This figure presents the average price per year of the 5 highest volume drugs in our sample for various conditions (except lung cancer where we focus on newer entrants). Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Atrial fibrillation, cystic fibrosis, and lung cancer, along with hepatitis C are "innovative" markets, all four of these conditions have new entrants who enter at price points considerably above other treatments in the market (Figure 6). For atrial fibrillation, anticoagulants such as rivaroxaban and apixaban, entered the market in 2011 and 2012, respectively and replaced the much cheaper warfarin. We estimate these drugs have a 0.1 to 0.2 QALY improvement over warfarin. Indeed, in 2019, the American College of Cardiology (ACC) and the American Heart Association (AHA) recommended these newer anticoagulants as the preferred drug class over warfarin (January et al., 2019), which can also be seen in their large increases in market share (Figure 5). Because these treatments cost $1-3k, which is considerably more than warfarin, the unadjusted price index for atrial fibrillation nearly triples during our sample period (Column 3 of Table 4). However, because the quality improvements would be worth $10k-$20k (assuming $100k VSLY), atrial fibrillation has quickly declining quality-adjusted price indexes.

Cystic fibrosis is an especially interesting case of an innovative condition. Cystic fibrosis costs are partly driven by a very high cost entrant, Orkambi. Orkambi was controversially priced at least $150k per year and has taken over about 20% market share by 2018.[28] Because of this, costs for cystic fibrosis quadruple in our sample period, where we estimate lifetime costs are well over $1 million by 2018.[29] [30] However, Orkambi was viewed as a breakthrough therapy and indeed we estimate that it adds 0.856 QALYs compared to tobramycin, a sizeable improvement. While cystic fibrosis costs are rising due to high quality innovations, our

[28]For example, see "A Drug Costs $272,000 a Year. Not So Fast, Says New York State." New York Times, June 2018. We find in the MarketScan data the average cost of Orkambi was closer to $150k per year.

[29]Prices for other cystic fibrosis drugs doubled or tripled in price during this time, which also factors in as Orkambi only accounts for 20% market share.

[30]One important difference between high cost drugs for cystic fibrosis and rheumatoid arthritis, versus hepatitis C is that for cystic fibrosis or rheumatoid arthritis the drugs are taken for multiple years. The lifetime scaling factors are on the order of 25- 30 for rheumatoid arthritis and cystic fibrosis, which means even if a year of treatment would be similar in price, we would view these drugs as being about 7.5-10 times more expensive than hepatitis C which has a lifetime scaling factor of 3.8.

framework still finds rapidly increasing quality adjusted price indexes (Table 4). The VSLY assumptions we make suggest that the large improvement in quality are not worth the cost growth, consumer surplus falls sharply.[31]

Lung cancer, multiple sclerosis, and HIV follow a similar pattern to cystic fibrosis.[32] They are clearly innovative with multiple new entrants yielding large quality improvements. However, these new entrants are very expensive. For these cases, despite the quality improvements (rising QALYs), the costs are rising rapidly enough that quality-adjusted prices are still rising and consumer welfare is falling.

While we caution against extrapolating outside the sample, in Appendix Section OA.E, we show results which aggregate across conditions, weighting by spending. The unadjusted price index rises by 70%, while the $100k VSLY index rises by 45% – a reduction of 1.5 percentage points from the compound annual growth rate. At $500k VSLY, the index falls by 55%.

In summary, we have examined conditions where it would be difficult to measure the quality of treatment using other methods. Our methodology finds a lot of heterogeneity in trends across conditions, but fairly large quality adjustments for nearly all conditions, suggesting that quality adjusted prices are growing more slowly than indexes that do not account for quality.

7.3 Total Welfare and Producer Surplus

While we find consumer welfare is falling for many innovative conditions, we should caution that these innovations may not be reducing total welfare if drug manufacturers are profiting off of the high prices. In this subsection, we do a back-of-the-envelope calculation for per-

[31] As mentioned previously, if one only considers consumers out-of-pocket payments, then this drug would have generated large consumer surplus gains.

[32] We drop Pre-Exposure Prophylaxis (PrEP) treatments, such as Truvada, from our analysis of HIV because they are preventative innovations rather than treatment innovations. However, we note that PrEP are important innovations for HIV during our sample period.

patient total welfare and producer surplus to make this point. In our framework, we can define average producer surplus as the difference between the revenue for the basket of treatments $S_{d,t}$ and the marginal cost of producing those treatments, $mc_{d,t}$:

$$\Delta\text{Producer Surplus}_{d,t,t-1} = (S_{d,t} - S_{d,t-1}) - (mc_{d,t} - mc_{d,t-1}), \tag{7}$$

where $mc_{d,t}$ is the marginal cost of production for the average bundle of treatments for disease d, at time t. This standard definition of producer surplus ignores the fixed cost of research and development of treatments. Therefore, our paper focuses on the value these new treatments provide, while others, such as DiMasi et al. (2003), estimate the cost of developing new treatments.

Adding together consumer welfare from Equation 3 and producer surplus from Equation 7 provides a measure of per-patient welfare:

$$\Delta\text{Total Welfare}_{d,t,t-1} = VSLY \cdot (H_{d,t} - H_{d,t-1}) - (mc_{d,t} - mc_{d,t-1}). \tag{8}$$

In our baseline case, we assume that the marginal cost of production is constant over time. This simplifies the change in total welfare from Equation 8 to be $VSLY \cdot (H_{d,t} - H_{d,t-1})$, which is just the health benefit of the treatments. We think this assumption is a lower bound on marginal costs, as newer drugs are likely more expensive to produce (especially biologics). This means we will likely be overstating the total welfare gains. However, we think this bias is small as the marginal cost of drug production is generally low. In Appendix Section OA.B, we present results where we assume that marginal costs are 20% of the price we observe in the data (which we think is extreme) to show how this assumption impacts our findings.

Column (7) of Table 4 presents results for total welfare. In our framework, total welfare is simply the health benefit multiplied by the VSLY, so it is \$100k multiplied by column (1). For rheumatoid arthritis, consumer welfare is falling because prices are rising, but those

high prices are profits for drug companies, so per-patient total welfare is rising during our sample period.

Interestingly, the one situation in our framework where we see total welfare falling in our model is after generic entry (of a relatively lower quality treatment). This is exemplified by colon cancer where consumers substitute from higher quality bevacizumab to lower quality older generation drugs (capecitabine and oxaliplatin) once their patents expire (see panel (b) of Figure 5 and panel (b) of Figure 6). While this is exemplified by colon cancer, it is not the only case where this pattern emerges. Indeed, hypertension had lower average QALYs in 2008-2011 relative to 2007 because relatively lower quality amlodipine's patent expired in 2007. Its price fell and it gained market share (panel (d) of Figure 5 and panel (d) of Figure 6).[33] However, for hypertension this decline was short lived as slightly newer and higher quality drug, losartan, had its patent expire in late 2009 and it gains market share, raising average QALYs.

Unconventional (but intuitive) results like this or consumer welfare falling in innovative markets highlights the importance of empirical evidence for how innovation (and patent expiration) impact welfare. It is well understood that market distortions play an important role in health care markets. However, understanding the relative importance of these distortions is ultimately an empirical question. In the latter case, our results suggest that insurance insulating consumers from the full price of extremely high cost drugs may be reducing consumer welfare in the aggregate. Likewise, insurance using their formulary design to steer patients to generics may be reducing total welfare if these drugs are less effective.[34]

As the results for colon cancer, hypertension, and schizophrenia demonstrate, in the long run, patents will expire and prices will be closer to marginal costs. Once that happens,

[33]Note that we are showing the price that is paid by the consumer and the insurer, not solely the consumers out-of-pocket price. However, out-of-pocket prices fall for these drugs after patent expiry.

[34]We think the insurance example is the most intuitive explanation for these dynamics, but physician incentives, lack of information among consumers, among other distortions likely also play an important role in our findings.

the total welfare gains in Table 4 may eventually accrue to consumers, leading to consumer welfare improving and quality-adjusted prices falling.

To better demonstrate the long-run impacts of these innovations Table OA25 in the appendix presents results from a counterfactual where we reduce the prices of all on-patent drugs by 85% (from their 2018 prices) while holding market shares constant at 2018 levels. This assumes that there would be no further innovation or diffusion after 2018, but would demonstrate how the current set of innovations impact consumer welfare once all those innovations go off patent. In this counterfactual consumer welfare is higher than in 2007 for all conditions, except cystic fibrosis. That is, in the long run these innovations improve consumer welfare.

7.4 Robustness checks

For our main robustness check, we test the sensitivity of our estimates to the QALY measure. There are a number of reasons why we may be overstating or understating true quality changes.[35] We think the most sensible approach is to see how our results change if we assume we are off by a factor of 2. Specifically, we multiply our estimated QALYs by two or by one half. Table OA26 shows results with QALYs doubled. Table OA27 shows results with QALYs cut in half. For the price index calculation, re-scaling QALYs is isomorphic to assuming different VSLYs. Hence, the impact of this robustness check is similar to what we find when we change the VSLY assumption. The main qualitative results are similar when we change the VSLY assumptions in our main tables and the same is true when we change QALYs.

[35]For example, we could be overstating the value of innovation if there are publication biases leading to more QALYs for new treatments (for example p-hacking or conflicts of interest, though we try to control for conflicts of interest below). On the other hand, we are only capturing the quality changes for a discrete set of treatments, while ignoring the potential quality improvements of other spending (e.g., physician or hospital spending). This would lead to us to understate results as we are ignoring quality improvements from new treatments, tests or imaging, but capturing cost growth for those services.

One important robustness check we do is to add weight to studies which the Tufts reviewer deem to be higher quality and studies with authors with academic affiliations in the Tufts CEAR regressions. We also lower the weight on studies authored or sponsored by industry.[36] Table OA29 presents the results, which are very similar to the equal weighting results, and are not sensitive to changes in the weighting scheme we use or varying which variables we include.

There is also some potential sensitivity to the lifetime cost estimates, especially related to how we scale annual costs. While we check some differences in specification on the scaling factor in Appendix Section OA.D.4, we discuss the widest range of estimates here, which we think represent upper and lower bounds for cost estimates. Our lower bound uses annual costs. We think is an unreasonable assumption as it assumes all treatment costs are completed in one year and then stop. With annual costs, all conditions except cystic fibrosis have falling prices when we assume the VSLY is \$100k. At the other extreme, we only use the age distribution for a condition and the life tables to calculate a lifetime scaling factor. That is, we assume treatment costs are constant over time. For many conditions, this assumption deviates from what we observe in the data as treatments may last for one year (hepatitis C) or costs tend to be higher in the first year (perhaps related to diagnostics or surgery) and then fall. Hence, we view this as an upper bound and it amounts to multiplying annual costs by a 23-28. For these results, four conditions have falling price indexes at \$100k VSLY. Three of those conditions are colon cancer, schizophrenia, and hypertension which have falling costs, and hepatitis C where quality improvements lead to falling price indexes. Eight conditions have falling price indexes with \$500k VSLY. As expected, with such a wide

[36]In particular, the Tufts data has a 1-7 measure of study quality, as judged by their classifiers. Quality depends on whether methods and results were communicated clearly, assumptions were reasonable, and whether sensitivity and subgroup analyses were included. In addition, the Tufts will list if authors have academic or industry affiliations, and whether the study was sponsored by industry. In the table we present, we set the weight of each study to its quality score. A study rated as a "7" is weighted seven times as much as study rated as a "1." We also add two points for studies with an author with an academic affiliation and subtract two points if the study had an author with industry affiliation or was sponsored by industry.

range of assumed values, results are different and should be viewed as very wide bounds on our central estimates.

We do a number of other robustness checks, though we leave the details to Appendix Section OA.B. Our main estimates focus on the primary treatment class, but we also include multiple classes of treatments for conditions that have multiple treatment classes. For example, rheumatoid arthritis has some comparisons between nonsteroidal anti-inflammatory drugs (NSAIDs) which are not directly or indirectly compared to DMARDs, as they are often used as a complement to DMARDs. This increases the number of treatments from 151 to 194. This adds some complications to the methodology, but does not change results much as the market share for these drugs is typically small relative to the main class of drugs.[37]

As Lucarelli et al. (2022) note, cost-effectiveness studies may understate the value of new treatments if there is heterogeneity in preferences. To address this, in a robustness check we approximate adding idiosyncratic noise to the QALY, then take an expected maximum. Our approach builds on the intuition from Ackerberg and Rysman (2005), and essentially amounts to adding a term which scales with $log(n_{d,t}) \times \gamma_d$, where $n_{d,t}$ is the number of treatments in the market in year t and γ_d imposes that QALYs increase by 25% of the range of QALYs in the market when a second treatment is added. In both cases, we think these overstate the value of heterogeneity and they increase QALYs considerably. Still, we find that many innovative conditions have falling consumer welfare.

As an additional robustness check, we bring in additional drug claims. For our main estimates we ignore all non-Tufts drugs as the MarketScan data does not have diagnosis codes on drug claims. We use the MEPS data (which has diagnosis codes on drug claims) to classify drugs in MarketScan to conditions and incorporate this spending in our estimates.

We also do a number of robustness checks on the Tufts quality regressions (Equation 6)

[37]Drugs which are not in the main class of treatments are still used in the cost calculation in the MarketScan data, so the cost calculation is unaffected by this change. For the QALY calculation, we calculate the change in QALYs by drug class, then take the average QALY change across classes, weighted by spending.

where we run the QALY regression a number of different ways. We also run QALY specifications that pull in hundreds of additional cost-effectiveness studies by including broader treatment categories like "no treatment," "placebo," "standard of care," and "usual care" among other terms. This potentially adds some noise, but also adds 349 additional comparisons to the regressions. For each of these robustness checks, results are very similar.

For our total welfare estimates, we explore the robustness of our results to a different assumption about marginal costs, which we think is extreme, but illustrate the robustness of our total welfare result. Table OA28 presents results where we assume that marginal costs are 20% of the negotiated price we observe in the claims data, so the change in producer surplus $0.8 \times (S_{d,t} - S_{d,t-1})$. This means that the change in total welfare becomes:

$$\Delta \text{ Total Welfare}_{d,t,t-1} = VSLY \cdot (H_{d,t} - H_{d,t-1}) - 0.2 \times (S_{d,t} - S_{d,t-1}). \tag{9}$$

This robustness check does not impact any of the price indexes or consumer welfare, but does reduce total welfare. In this case, we see that there are four conditions where total welfare declines during our sample period.

7.5 Which Treatments Are Driving Our Results?

In this section, we explore which treatments are driving our results. We consider all treatments, including both new innovations, but also the welfare changes from treatments that are in the sample throughout the entire period. To do this, we separate out the two mechanisms that drive our earlier results, (1) within-molecule price changes and (2) diffusion of new drugs replacing older generation drugs.

To fix ideas let $M_{r,d,t}$ represent drug r's market share in year t for disease d. For pre-existing market share, $(M_{r,d,2007})$ we replace that drug's 2018 price with its 2007 price. This captures within-molecule price changes. For newly obtained market share, $M_{r,d,2018} -$

$M_{r,d,2007}$, we replace that drug's price and QALY in 2018, with the basket average price and QALY in 2007. This accounts for the QALY and cost differences when a drug diffuses.[38] Once we make these substitutions for drug r, we recalculate the 2018 consumer welfare measure, then difference that from the main result. We then rank treatments based on which ones have the largest differences.

Table 5 shows the drugs which contributed to the biggest consumer welfare increases assuming a VSLY of \$100k. The QALY column shows the QALY difference between that drug and the 2007 basket average. The price in 2007 and 2018 are the prices of an annual course of treatment for only that drug (i.e. not including inpatient and outpatient spending on the condition). If the price in 2007 is missing, then the drug was not in the market in 2007.

[38]For those taking the drug in 2007 and 2018, the relevant welfare counterfactual for them is how the price has changed, they shouldn't expect a quality change. For newly obtained market share, the counterfactual isn't that same drug in 2007 (in which case we would capture no quality change), but rather what they would have taken in 2007, which we assume is the average bundle in 2007.

Table 5: Drugs which account for biggest consumer welfare gains - $100k VSLY

Rank	Condition	Drug	QALYs	Price in 2007	Price in 2018
1	HepatitisC	Harvoni	2.9		11885
2	HepatitisC	Epclusa	2.9		12741
3	AtrialFibrillation	apixaban	.8		1386
4	ColonCancer	capecitabine	-.2	6101	2805
5	ColonCancer	capecitabine/oxaliplatin	-.2	36151	8118
6	AtrialFibrillation	rivaroxaban	.6		1312
7	Schizophrenia	aripiprazole	.5	3080	1355
8	ColonCancer	fluorouracil/oxaliplatin	-.2	32778	5132
9	VenousThromboembolism	rivaroxaban	.1		1008
10	Schizophrenia	olanzapine	.6	3779	241
13	Hypertension	losartan	.1	438	44

Notes: This table presents the 10 drugs (plus losartan which is 13th) which contribute the most to consumer welfare increases across all the drugs and conditions in our sample. To calculate this, we calculate the difference in consumer welfare between the 2018 number and our counterfactual without that one drug. The QALYs column is the difference in QALYs between the drug and the 2007 basket average for that condition. The prices in 2007 and 2018 are the annual average price we observe for that drug only. These prices are not adjusted for lifetime costs.

There are two types of drugs which account for the biggest increases in consumer welfare during our sample period. First, new entrants which are highly effective, not too costly, and take considerable market share. Harvoni and Epclusa are the top two drugs in terms of increasing consumer welfare. While they are more expensive than P-Interferon, they are not significantly more expensive and they provide substantial quality improvements. Apixaban and rivaroxaban for atrial fibrillation, which are discussed above, are also in the top 10. These drugs are 0.6 to 0.8 QALYs higher quality than the 2007 average bundle, corresponding to $60k and $80k in consumer welfare.[39] These drugs are more expensive than warfarin, but we can see their price in 2018 is small relative to those QALY gains (roughly $1.3k in a given year). In addition, we can see in Figure 5 that these drugs took over

[39]The QALYs for this comparison are larger than the discussion above because in the discussion in Section 7.2 we are comparing these drugs to warfarin, whereas in this section we are comparing them to the 2007 average bundle.

considerable market share from warfarin, which is a necessary condition in our framework for creating quality improvements. These are examples of highly effective (but higher price) new treatments diffusing, suggesting that higher quality is valued by the health care sector, which is consistent with the findings in Chandra et al. (2016).

The other type of drug which drives consumer welfare improvements in our sample are widely used drugs which go off-patent reducing costs considerably. These include capecitabine and oxaliplatin for colon cancer, aripiprazole for schizophrenia, and losartan for hypertension.[40]

Table 6 shows the drugs which reduce consumer welfare the most. Like the drivers for increasing consumer welfare, there are two types of drugs which reduce consumer welfare: high cost drugs which raise their prices considerably, and new innovations whose high cost exceeds the quality improvement. Examples of drugs that raise their price considerably include aztreonam, Humira, and interferon beta1a.

The other set of drugs which lower consumer welfare are from costly new innovations. Orkambi costs $157k per year, while providing nearly $90k in total welfare due to the substantial increase in QALY improvement (relative to the 2007 basket of treatments for cystic fibrosis). The quality improvement is large, which helps explain why insured individuals would use it, but consumer welfare is highly negative. Ocrelizumab, dimethyl fumarate, Stribild are additional drugs which we estimate as being high quality, but their costs are large enough that our methodology suggests the costs are not worth the benefits for consumers. That these drugs which reduce consumer welfare are diffusing provides some evidence of "inefficiencies" in U.S. drug markets, similar to the findings in Kyle and Williams (2017). To explore this more, we isolate the impacts of new entrants impacts on these markets in the next section.

[40]Note that capecitabine and oxaliplatin are higher quality than their comparator, fluorouracil, in our regressions. This table shows them as having negative QALYs because in this table they are compared to the 2007 average QALY which includes the higher quality bevacizumab.

Table 6: Drugs which account for biggest consumer welfare reductions - $100k VSLY

Rank	Condition	Drug	QALYs	Price in 2007	Price in 2018
1	CysticFibrosis	Orkambi	.9		157501
2	CysticFibrosis	aztreonam/dornase alfa	.2	16550	67506
3	CysticFibrosis	dornase alfa/tobramycin	0	26155	49807
4	MultipleSclerosis	ocrelizumab	.8		75100
5	MultipleSclerosis	dimethyl fumarate	.4		57115
6	RheumatoidArthritis	Humira	.9	11970	30770
7	HIV	Stribild	.1		21258
8	MultipleSclerosis	interferon beta1a	.1	18759	59915
9	MultipleSclerosis	fingolimod	.8		60057
10	MultipleSclerosis	natalizumab	1.3	18243	66920

Notes: This table presents the 10 drugs which contribute the most to consumer welfare reductions across all the drugs and conditions in our sample. To calculate this, we calculate the difference in consumer welfare between the 2018 number and our counterfactual without that one drug. The QALYs column is the difference in QALYs between the drug and the 2007 basket average for that condition. The prices in 2007 and 2018 are the annual average price we observe for that drug only. These prices are not adjusted for lifetime costs.

8 How Does Innovation Effect Markets?

In this section we focus on how innovation, specifically the entry of new treatments, impacts markets. To do this, we compute a counterfactual where we remove new entrants from the data. This counterfactual isolates the effects of innovation, by stripping out the influence of within-drug price changes which are large in many cases. We begin by asking how much of the growth in costs is due to new entrants? Then, we look at how new entrants shape consumer, producer, and total surplus.

8.1 What Share of Spending Growth is Due to Innovation?

A number of papers in the health literature attempt to measure the contribution of health care innovation on spending growth (Chernew and Newhouse, 2011). As mentioned previ-

ously, given the difficulty of the problem researchers often follow Solow (1957) and control for measurable drivers of spending (e.g., age, insurance, price, and incomes), and the residual is attributed to innovation. Instead, for our subset of conditions, we are able to identify and track new treatments directly. In this section, we ask how much of spending growth is driven by new innovations, where we define innovations as new molecules or combinations that were not in the market at the start of our sample period.

One challenge with this question is determining a counterfactual outcome for what would have happened in the absence of innovation. For this counterfactual, we categorize treatments as "new" or "old" based on whether they were in the market at the start of our sample period (2007). The thought experiment we have in mind is removing all "new" drugs from the market. We then need to make assumptions about which drugs those patients would have taken in the absence of the "new" drugs and counterfactual prices for "old" drugs. We reallocate all the market share from "new" drugs to "old" drugs in proportion to the "old" drug market share in 2018.[41] We keep the "old" drugs at their 2018 prices.[42]

Table 7 presents results. Column 1 presents how much costs grew in our data without differentiating between innovation and within-drug price growth.[43] The average cost of rheumatoid arthritis grew by $197k in our data. Column 2 tells us how much of the new cost is due to innovation. Only $12k of that cost increase was due to innovation (Actemra). Hence, innovation only accounts for 6% of the total rheumatoid arthritis cost growth in our data (column 3). As shown above, most of the cost growth for rheumatoid arthritis is within-drug, rather than new drugs entering at higher price points.

There is a lot of heterogeneity across conditions. Hypertension had no new drugs in the

[41] This proportional substitution assumption is consistent with a type 1 extreme value error assumption widely used in discrete-choice models.

[42] Using 2018 prices (rather than 2007 prices) better captures changes in the market that would likely have occurred even in the absence of innovation, like the "old" drugs coming off patent and general inflation. If prices in the absence of "new" drugs would have been higher in the counterfactual, due to reduced competition, then our estimates will be biased downward.

[43] These numbers can be computed by multiplying Column 2 with [Column (3) minus 1] in Table 4.

Tufts CEAR data, so none of the cost growth was driven by innovation. Hepatitis C only has new drugs (no old treatment is used in 2018), so innovation accounts for 100% of its cost growth. Costs for lung cancer and venous thromboembolism drugs would have fallen in the absence of innovation, as some treatments went off patent, so innovation accounts for more than 100% of the cost growth we observe.

Table 7: Counterfactual: Removing All New Drugs

	(1) Baseline Cost Growth 2018 - 2007 ($1,000s)	(2) Cost Growth due to Innovation ($1,000s)	(3) Share of Cost Growth due to Innovation	(4) Change in Consumer Welfare due to Innovation $100k VSLY ($1,000s)	(5) Change in Producer Surplus due to Innovation ($1,000s)	(6) Change in Total Welfare due to Innovation $100k VSLY ($1,000s)
Asthma	0	0	0.316	0	0	0
AtrialFibrillation	32	7	0.209	73	7	80
ColonCancer	-184	0	-0.000	0	0	0
CysticFibrosis	2,403	667	0.277	-648	667	19
HIV	176	89	0.505	-81	89	8
HepatitisC	9	9	1.000	282	9	291
Hypertension	-3	0	0.000	0	0	0
LungCancer	252	295	1.169	-230	295	65
MultipleSclerosis	1,026	204	0.199	-186	204	18
Osteoporosis	5	2	0.517	0	2	3
RheumatoidArthritis	197	12	0.061	-5	12	7
Schizophrenia	-7	3	-0.421	1	3	4
VenousThromboembolism	2	2	1.083	5	2	7
Aggregate	276	66	0.239	-54	66	12

Notes: Column 1 presents the cost growth we see without the counterfactual. This can be calculated as Column 2 multiplied by [Column (3) minus 1] in Table 4. Column 2 tells us the amount of cost growth due to innovation. This is calculated by determining the counterfactual where we replace all "new" drugs with "old" drugs in proportion to "old" drug market share in 2018. We then calculate the cost growth between 2007 and the 2018 counterfactual. Column 2 presents the difference between the cost growth we observe and this counterfactual. Column 3 then computes the share of cost growth that is due to innovation (column 2 divided by column 1). Column 4 presents the change in consumer welfare due to innovation. Column 5 presents producer surplus which is the same as column 2 as we assume marginal costs are constant. Column 6 presents the change in total welfare due to innovation, which is just $100k multiplied by the change in QALYs due to innovation (not shown). These numbers are similar to Table 4 because most of the quality improvements are due to innovation.

While we make no claim that these conditions are representative, we take a revenue weighted average across conditions to compute aggregate measures. Our average cost growth across all conditions is $276k during our sample period. The counterfactual (with no innovation) has cost growth of $66k. Therefore, we find that about 24% of cost growth during this sample period is due to new innovation. This estimate is within the range, but on the lower end of the literature measuring innovation as the residual of cost growth that cannot be explained by other factors (e.g. aging, insurance expansion, income growth, etc.).[44] This is likely a lower bound for these conditions, as we only focus on innovations that we can measure and do not account for innovations which occurred slightly before our sample period (e.g., rheumatoid arthritis drugs).[45]

8.2 What Share of Surplus Goes to Consumers and Producers?

One important question in the innovation literature is what share of surplus captured by the innovator (Nordhaus, 2004). As noted by Nordhaus (2004), most of this literature is theoretical as measuring the welfare effects of innovation is notoriously difficult. To do this calculation, we use the same counterfactual as the previous section of removing all new drugs in the market in 2018, and assuming users purchase "old" drugs in proportion to their 2018 market share of "old" drugs (and at 2018 prices). Rather than focus on cost growth, the idea here is to calculate how welfare has changed with the introduction of the new goods, accounting for both the cost and quality changes relative to the counterfactual of no new innovations. This is calculated by looking at the difference in observed welfare in 2018 with our counterfactual welfare estimate of no innovation. We assume $100k VSLY for each

[44] For example see Newhouse (1992) and Smith et al. (2022). It is also closely in line to Dunn et al. (2023) who calculates the correlation between Tufts CEAR studies and cost growth by condition to determine the share of growth due to innovation.

[45] In Appendix Section OA.C.1, we explore the share of cost growth that is due to within-molecule inflation. We find within-molecule price growth is more important than innovation in driving higher costs.

calculation, though we present results using other VSLYs (and assumptions about marginal costs) in the appendix.

Columns 4, 5, and 6 show the change in consumer, producer, and total welfare due to innovation. Given our constant marginal cost assumption, cost growth due to innovation (Column 2) is also the change in producer surplus due to innovation.

For venous thromboembolism, we estimate that consumer welfare increased by $8k (Table 4, column 5). Of that, $5k of consumer welfare is due to the entry of new drugs. Producer surplus rose by $2k due to innovation so total surplus due to innovation is $7k higher. Therefore, we estimate that producers captured about 29% of the surplus from innovations in 2018.

For cystic fibrosis, consumer welfare falls by $648k due to the entry of Orkambi.[46] Total surplus increases by $19k, but producer surplus grows by $667k due to Orkambi. Hence, producers received 3,560% of the surplus in the cystic fibrosis market. As this table, which strips out within-drug price growth, demonstrates, this result is also not unique to cystic fibrosis. Five of our ten conditions that have meaningful new entrants are estimated to have lower consumer welfare in 2018 because of those new entrants in 2018 (and we see this with Sovaldi in 2014, in Table 2 as well). In addition, averaging across all conditions in our sample also finds average consumer welfare falls due to innovation.[47]

We think that this finding where consumer surplus is falling due to innovation is likely a feature that is unique to health care markets. Without distortions, a standard model of demand would suggest that an innovation would not diffuse if its price was so high that it lowered consumer welfare. However, distortions such as insurance, uninformed consumers, or provider incentives (or physician detailing), may lead to consumers to purchase a drug

[46] In Column 5 of Table 4 we find that consumer welfare for cystic fibrosis falls by $2.38m. The remainder of the reduction in consumer surplus is because other cystic fibrosis drugs raise their prices considerably during our sample period.

[47] It is also important to note that total welfare is rising in all of these cases. These drugs greatly improve the quality of care for patients and are quite profitable for manufacturers.

which is not cost effective. Indeed, the diffusion of Orkambi is not surprising as the average consumer in our data only pays $1.5k per year out-of-pocket for Orkambi. Hence, welfare from the point of view of an Orkambi user rises significantly, while the costs of Orkambi are spread across other enrollees in that insurance plan.

While we view this result as unconventional, it is also consistent with the cost effectiveness literature. Recall from Section 4 that when a treatment is not cost effective, that means that it would lower consumer surplus if it replaced its comparator. Indeed, both the cost effectiveness studies for Orkambi in the Tufts data find that it is not cost effective at any conventional VSLY, yet Figure 5 shows that Orkambi diffuses broadly.[48] That is, even outside of the context of our model (and assumptions), one should expect to find falling consumer surplus simply by taking these cost effectiveness studies at face value. Indeed, many of the innovative treatments which we think lower consumer welfare have studies that show that they are not cost-effective.[49]

To test the sensitivity of these results to various assumptions, we replicate Table 7 in the appendix using $50k VSLY (Table OA31) and $500k VSLY (Table OA32). Even with a $500k VSLY, three of thirteen conditions have declining welfare due to innovation and the average across conditions remains negative.[50] We also test the sensitivity of our constant marginal cost assumption by assuming that marginal costs are 20% of the prices we observe in the claims data as in Equation 9. Table OA33 presents the results. This assumption does not impact our the share of cost growth due to innovation or our consumer surplus results,

[48]Both Dilokthornsakul et al. (2017) and Sharma et al. (2018) find large improvements in health, but at extremely high costs.

[49]It is worth noting that because we are using their quality data, our quality information should be similar in order and magnitude to Tufts studies, on average. However, our cost data and assumptions are very different and independent from the cost data and assumptions in these studies.

[50]We note that in our framework, multiplying the VSLY by a number has the same effect as multiplying the change in QALYs by that number, so if our QALY estimates are off by a factor of 5 (but the VSLY is $100k) Table OA32 also shows results for that case. If you want to test the sensitivity to multiplying QALYs or VSLY by a factor X, you can do so by multiplying total surplus by X (as total surplus is simply a factor of QALYs and VSLY). To get the change in consumer welfare, simply subtract off producer surplus (column 2) from that total surplus number.

but dampens producer surplus and total welfare. For this calculation, three conditions have total welfare falling due to innovation, though we think this assumption on marginal costs is fairly extreme.

There are a number of interesting implications from this result. First, consumer surplus falling due to innovation means that producers are receiving more than 100% of the surplus from their innovations in 2018. A famous result in innovation economics is that that patents provide insufficient incentives for innovation, as the innovator is not able to capture all the consumer surplus (and monopoly pricing creates deadweight loss) (Arrow, 1962; Nelson, 1959). Our results show that may not be the case if distortions lead products with negative consumer surplus to diffuse.

Second, our results have important implications for pricing policy. Some countries in the OECD and notably the United Kingdom's National Institute for Health and Clinical Excellence (NICE) restricts medicines that do not meet a specific cost effectiveness threshold. In the context of our framework, this is loosely equivalent to imposing that new innovations increase consumer (not total) welfare. Our results show that this restriction is likely binding, which is consistent with evidence that these drugs are blocked or diffuse more slowly in other countries (Kyle and Williams, 2017).[51]

Finally, we can see how Schumpeterian profits can be fleeting. For example, Sovaldi's entry in the hepatitis C market lowered consumer welfare (Table 2), so producers were receiving more than 100% of the surplus in 2014. However, with the entry of Harvoni, Epclusa and Viekira Pak, prices fell rapidly and producers only receive 3% of the change in surplus in this market by 2018. Colon cancer is another case where the fleeting nature of profits due to innovation are on display in our results. Lucarelli et al. (2022) document

[51]This point also makes clear an important caveat that that pharmaceutical prices for the commercially insured in the United States are much higher than in other countries or public payers within the United States (Anderson et al., 2003). When viewed through the lens of incentives for innovation, our results are likely an upper bound on those incentives.

how prices and quality change for colon cancer treatments from 1993-2005 when there are numerous new entrants and a rapid rise in the cost of treating colon cancer. Our study complements theirs by documenting more recent trends for colon cancer treatments from 2007-2018, with the major change being the earlier innovations going off patent and prices for these treatments falling considerably.

9 Conclusion

If spending increases are due to technological advances that are improving or extending life, then they may be "worth it." However, determining how specific innovations are driving spending growth and changes in quality presents difficult measurement challenges. We use thousands of cost-effectiveness studies combined with information on millions of individuals to take a granular look at the causes of quality improvement and spending growth for 13 conditions.

Our granular look at each condition provides some lessons for trying to understand factors that influence welfare changes in the health sector, as captured by quality-adjusted price indexes. First, we find a lot of heterogeneity in spending growth trends, causes of spending growth, and the amount of quality improvements across the 13 conditions. This speaks to the importance of having a scalable framework that can be applied consistently across conditions. Overall, we find quality improvements for 12 of the 13 conditions and for many conditions the quality improvements are large in magnitude. This suggests that price indexes which do not account for quality improvements may overstate price growth.

Overall, the results raise important questions about how health care markets implicitly value quality improvements, amid numerous market distortions. Similar to other sectors of the economy, we provide evidence that innovation has led to sizeable total welfare gains. However, in contrast to other sectors of the economy, we find diffusion of higher quality new

innovations where the costs appear to exceed the benefit from a consumer's perspective. In other words, innovation has arguably led to a reduction in consumer welfare.

In the long run, we argue that the patents for these innovations will expire, likely leading to lower costs, consumer health improving, and higher consumer welfare. On the other hand, as Keynes famously said: "In the long run, we are all dead."

References

Ackerberg, D. and M. Rysman (2005). Unobserved product differentiation in discrete choice models: Estimating price elasticities and welfare effects.

Almond, D., J. J. Doyle Jr, A. E. Kowalski, and H. Williams (2010). Estimating marginal returns to medical care: Evidence from at-risk newborns. *The quarterly journal of economics 125*(2), 591–634.

Anderson, G. F., U. E. Reinhardt, P. S. Hussey, and V. Petrosyan (2003). It's the prices, stupid: why the united states is so different from other countries. *Health Affairs 22*(3), 89–105.

Arias, E., J. Xu, and K. D. Kochanek (2019). United states life tables, 2016.

Arrow, K. (1962). Economic welfare and the allocation of resources to invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pp. 609–26. Princeton University Press.

Berndt, E. R., A. Bir, S. H. Busch, R. G. Frank, and S.-L. T. Normand (2002). The medical treatment of depression, 1991–1996: productive inefficiency, expected outcome variations, and price indexes. *Journal of Health Economics 21*(3), 373–396.

Bryan, K. A. and H. L. Williams (2021). Innovation: market failures and public policies. In *Handbook of industrial organization*, Volume 5, pp. 281–388. Elsevier.

Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2016). Health care exceptionalism? performance and allocation in the us health care sector. *American Economic Review 106*(8), 2110–2144.

Chandra, A. and J. Skinner (2012). Technology growth and expenditure growth in health care. *Journal of Economic Literature 50*(3), 645–680.

Chernew, M. E. and J. P. Newhouse (2011). Health care spending growth. In *Handbook of health economics*, Volume 2, pp. 1–43. Elsevier.

Chhatwal, J., F. Kanwal, M. S. Roberts, and M. A. Dunn (2015). Cost-effectiveness and budget impact of hepatitis c virus treatment with sofosbuvir and ledipasvir in the united states. *Annals of internal medicine 162*(6), 397–406.

Cutler, D. M. (1995). Technology, health costs, and the nih. *Unpublished paper prepared for NIH Roundtable on Economics*.

Cutler, D. M. (2018). What is the us health spending problem? *Health Affairs 37*(3), 493–497.

Cutler, D. M., K. Ghosh, K. L. Messer, T. Raghunathan, A. B. Rosen, and S. T. Stewart (2022). A satellite account for health in the united states. *American Economic Review 112*(2), 494–533.

Cutler, D. M. and M. McClellan (2001). Is technological change in medicine worth it? *Health affairs 20*(5), 11–29.

Cutler, D. M., M. McClellan, J. P. Newhouse, and D. Remler (1998). Are medical prices declining? Evidence from heart attack treatments. *Quarterly Journal of Economics 113*(4), 991–1024.

Dauda, S., A. Dunn, and A. Hall (2022). A systematic examination of quality-adjusted price index alternatives for medical care using claims data. *Journal of Health Economics 85*, 102662.

Dilokthornsakul, P., M. Patidar, and J. D. Campbell (2017). Forecasting the long-term clinical and economic outcomes of lumacaftor/ivacaftor in cystic fibrosis patients with homozygous phe508del mutation. *Value in Health 20*(10), 1329–1335.

DiMasi, J. A., R. W. Hansen, and H. G. Grabowski (2003). The price of innovation: new estimates of drug development costs. *Journal of health economics 22*(2), 151–185.

Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association 78*(383), 605–610.

Dunn, A., L. Fernando, and E. Liebman (2023). A direct measure of medical innovation on health care spending: A condition-specific approach.

Dunn, A., A. Hall, and S. Dauda (2022). Are medical care prices still declining? a re-examination based on cost-effectiveness studies. *Econometrica 90*(2), 859–886.

Dunn, A., L. Rittmueller, and B. Whitmire (2015). Introducing the new BEA health care satellite account. *Survey of Current Business 95*(1), 1–21.

Eggleston, K., B. K. Chen, C.-H. Chen, Y. I. Chen, T. Feenstra, T. Iizuka, J. T. Lam, G. M. Leung, J.-f. R. Lu, B. Rodriguez-Sanchez, et al. (2019). Are quality-adjusted medical prices declining for chronic disease? evidence from diabetes care in four health systems. Technical report, National Bureau of Economic Research.

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *The American Economic Review*, 157–177.

Fisher, F. and K. Shell (1972). *The Economic Theory of Price Indices, New York and London*. Academic Press.

Frank, R. G., E. R. Berndt, A. B. Busch, and A. F. Lehman (2004). Quality-constant "prices" for the ongoing treatment of schizophrenia: an exploratory study. *The Quarterly Review of Economics and Finance 44*(3), 390–409.

Gowrisankaran, G. and M. Rysman (2012). Dynamics of consumer demand for new durable goods. *Journal of political Economy 120*(6), 1173–1219.

Hall, R. E. and C. I. Jones (2007). The value of life and the rise in health spending. *The Quarterly Journal of Economics 122*(1), 39–72.

Hartman, M., A. B. Martin, B. Washington, A. Catlin, and N. H. E. A. Team (2021). National health care spending in 2020: Growth driven by federal spending in response to the covid-19 pandemic. *Health Affairs 41*(9).

Hofmeister, M. G., E. M. Rosenthal, L. K. Barker, E. S. Rosenberg, M. A. Barranco, E. W. Hall, B. R. Edlin, J. Mermin, J. W. Ward, and A. B. Ryerson (2019). Estimating prevalence of hepatitis c virus infection in the united states, 2013-2016. *Hepatology 69*(3), 1020–1031.

Hopkins, J. (2021). House committee takes on pricing, patents for top-selling drug humira. *Wall Street Journal*.

Hult, K. J., S. Jaffe, and T. J. Philipson (2018). How does technological change affect quality-adjusted prices in health care? systematic evidence from thousands of innovations. *American Journal of Health Economics 4*(4), 433–453.

January, C. T., S. L. Wann, H. Calkins, L. Y. Chen, J. E. Cigarroa, J. C. Cleveland, P. T. Ellinor, M. D. Ezekowitz, M. E. Field, H. P. A. Furie, Karen L., K. T. Murray, J. B. Shea, C. M. Tracy, and C. W. Yancy (2019). 2019 aha/acc/hrs focused update of the 2014 aha/acc/hrs guideline for the management of patients with atrial fibrillation: A report of the american college of cardiology/american heart association task force on clinical practice guidelines and the heart rhythm society. *Circulation 140*(2).

Kakani, P., M. Chernew, and A. Chandra (2020, March). Rebates in the pharmaceutical industry: Evidence from medicines sold in retail pharmacies in the u.s. Working Paper 26846, National Bureau of Economic Research.

Kesselheim, A. S., J. Avorn, and A. Sarpatwari (2016). The high cost of prescription drugs in the united states: origins and prospects for reform. *Jama 316*(8), 858–871.

Kyle, M. and H. Williams (2017). Is american health care uniquely inefficient? evidence from prescription drugs. *American Economic Review 107*(5), 486–490.

Lucarelli, C., S. Nicholson, and N. Tilipman (2022). Price indices and the value of innovation with heterogenous patients. *Journal of Health Economics 84*, 102625.

Meltzer, D. O. and P. C. Smith (2011). Theoretical issues relevant to the economic evaluation of health technologies. In *Handbook of health economics*, Volume 2, pp. 433–469. Elsevier.

Murphy, K. M. and R. H. Topel (2006). The value of health and longevity. *Journal of political Economy 114*(5), 871–904.

National Research Council (2002). *At what price?: conceptualizing and measuring cost-of-living and price indexes*. National Academies Press.

National Research Council (2011). *Accounting for health and health care: approaches to measuring the sources and costs of their improvement*. National Academies Press.

Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of political economy 67*(3), 297–306.

Neumann, P. J., J. T. Cohen, M. C. Weinstein, et al. (2014). Updating cost-effectiveness—the curious resilience of the $50,000-per-qaly threshold. *NEnglJMed* 371(9), 796 – −797.

Newhouse, J. P. (1992). Medical care costs: how much welfare loss? *Journal of Economic perspectives 6*(3), 3–21.

Nordhaus, W. D. (2004). Schumpeterian profits in the american economy: Theory and measurement.

Olek, M. J. and D. Mowry (2022). Initial disease-modifying therapy for relapsing-remitting multiple sclerosis in adults. *Up to Date*.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy 110*(4), 705–729.

Romley, J. A., A. Dunn, D. Goldman, and N. Sood (2020, January). *Quantifying Productivity Growth in the Delivery of Important Episodes of Care Within the Medicare Program Using Insurance Claims and Administrative Data*. University of Chicago Press.

Schwartz, W. B. (1987). The inevitable failure of current cost-containment strategies: why they can provide only temporary relief. *JAMA 257*(2), 220–224.

Shapiro, I., M. D. Shapiro, and D. W. Wilcox (2001). A price index for cataract surgery. *Medical Output and Productivity, Chicago: University of Chicago Press for the National Bureau of Economic Research*, 411–438.

Sharma, D., S. Xing, Y.-T. Hung, R. N. Caskey, M. L. Dowell, and D. R. Touchette (2018). Cost-effectiveness analysis of lumacaftor and ivacaftor combination for the treatment of patients with cystic fibrosis in the united states. *Orphanet journal of rare diseases 13*, 1–12.

Sheiner, L. and A. Malinovskaya (2016). Measuring productivity in healthcare: an analysis of the literature. *Hutchins center on fiscal and monetary policy at Brookings*.

Shrank, W. H., T. L. Rogstad, and N. Parekh (2019). Waste in the us health care system: estimated costs and potential for savings. *Jama 322*(15), 1501–1509.

Small, K. A. and H. S. Rosen (1981). Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, 105–130.

Smith, S., J. P. Newhouse, and M. S. Freeland (2009). Income, insurance, and technology: why does health spending outpace economic growth? *Health affairs 28*(5), 1276–1284.

Smith, S. D., J. P. Newhouse, and G. A. Cuckler (2022, December). Health care spending growth has slowed: Will the bend in the curve continue? Working Paper 30782, National Bureau of Economic Research.

Solow, R. M. (1957). Technical change and the aggregate production function. *The review of Economics and Statistics*, 312–320.

Trajtenberg, M. (1989). The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of political Economy 97*(2), 444–479.

Viscusi, W. K. (2020). Failing to think properly about the value of a statistical life. *The Regulatory Review*.

Weaver, M. R., J. Joffe, M. Ciarametaro, R. W. Dubois, A. Dunn, A. Singh, G. W. Sparks, L. Stafford, C. J. Murray, and J. L. Dieleman (2022). Health care spending effectiveness: Estimates suggest that spending improved us health from 1996 to 2016. *Health Affairs 41*(7), 994–1004.

Online Appendix OA.A Detailed Results for Other Conditions

OA.A.1 Asthma

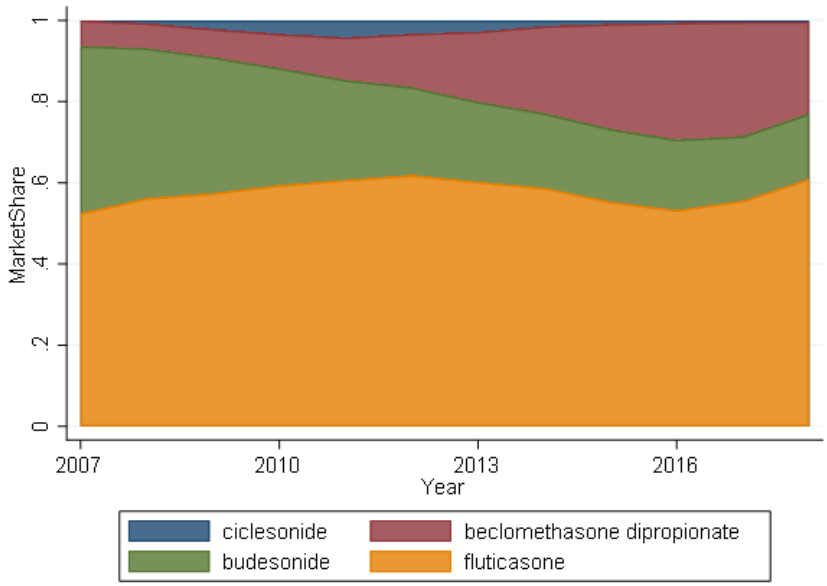
Asthma has no new entrants, relatively stable market shares of drugs, and modest within-molecule cost growth. It has very modest growth in its unadjusted price indexes and quality growth.

Table OA1: QALY Estimates for Asthma

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
beclomethasone_dipropionate	0	0.005
budesonide	0	-0.007
ciclesonide	0	-0.091
fluticasone	1	0.000

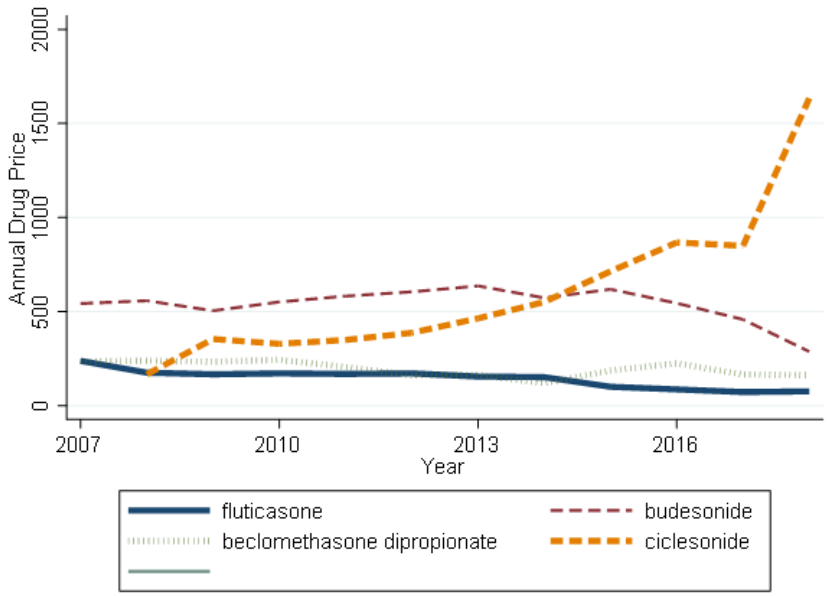
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA1: Market shares for the top Asthma treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Asthma across our entire sample period in the MarketScan data.

Figure OA2: Prices for the top Asthma treatments over time



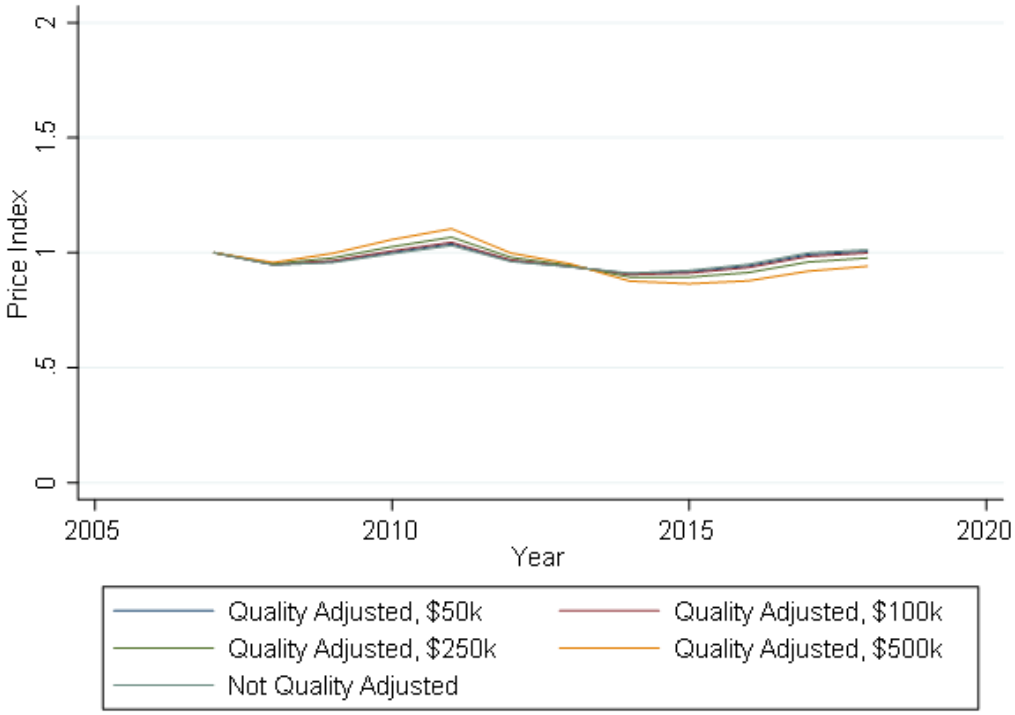
Notes: This figure presents the average price per year of the 5 highest volume drugs for Asthma in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA2: Price Indexes and Changes in Welfare by Year for Asthma

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	16	1.000	1.000	0	1.000	0
2008	-0.000	16	0.947	0.949	1	0.957	1
2009	-0.001	16	0.957	0.965	1	0.997	0
2010	-0.002	16	0.995	1.008	0	1.057	-1
2011	-0.002	17	1.031	1.045	-1	1.104	-2
2012	-0.001	16	0.961	0.968	1	0.998	0
2013	-0.000	15	0.938	0.941	1	0.951	1
2014	0.001	15	0.912	0.904	2	0.876	2
2015	0.002	15	0.923	0.911	1	0.865	2
2016	0.002	16	0.950	0.935	1	0.878	2
2017	0.003	16	0.999	0.983	0	0.919	1
2018	0.002	17	1.013	0.999	0	0.941	1

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA3.

Figure OA3: Price Indexes for Asthma



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

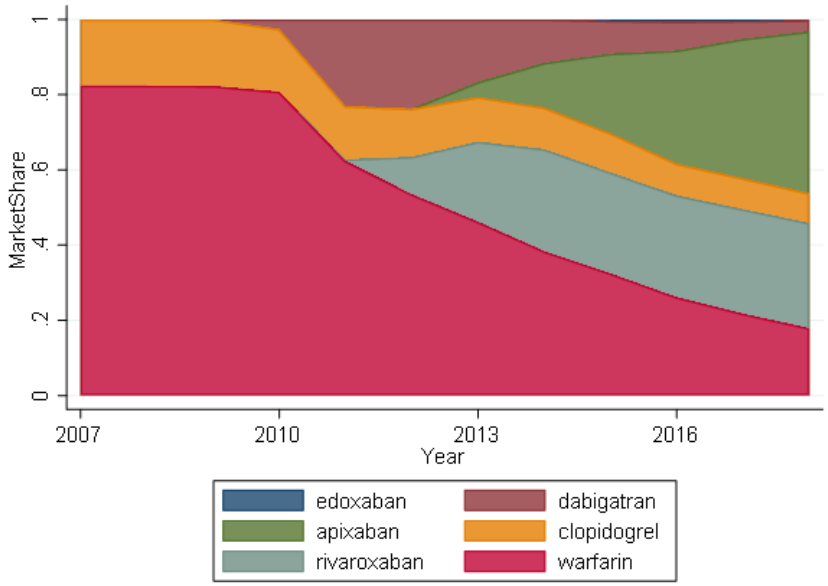
OA.A.2 Atrial Fibrillation

Table OA3: QALY Estimates for Atrial Fibrillation

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
apixaban	0	0.364
clopidogrel	0	-2.580
dabigatran	0	0.273
edoxaban	0	0.332
rivaroxaban	0	0.166
warfarin	1	0.000

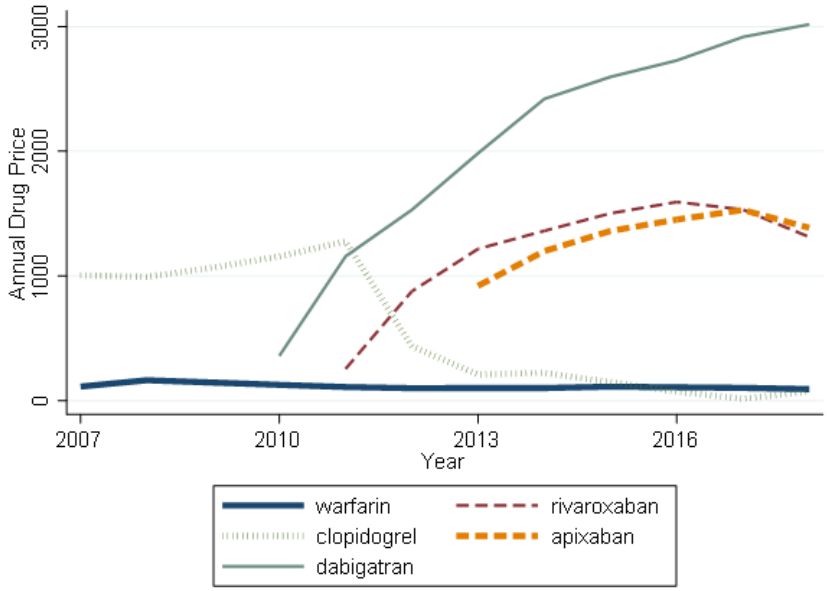
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA4: Market shares for the top Atrial Fibrillation treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Atrial Fibrillation across our entire sample period in the MarketScan data.

Figure OA5: Prices for the top Atrial Fibrillation treatments over time



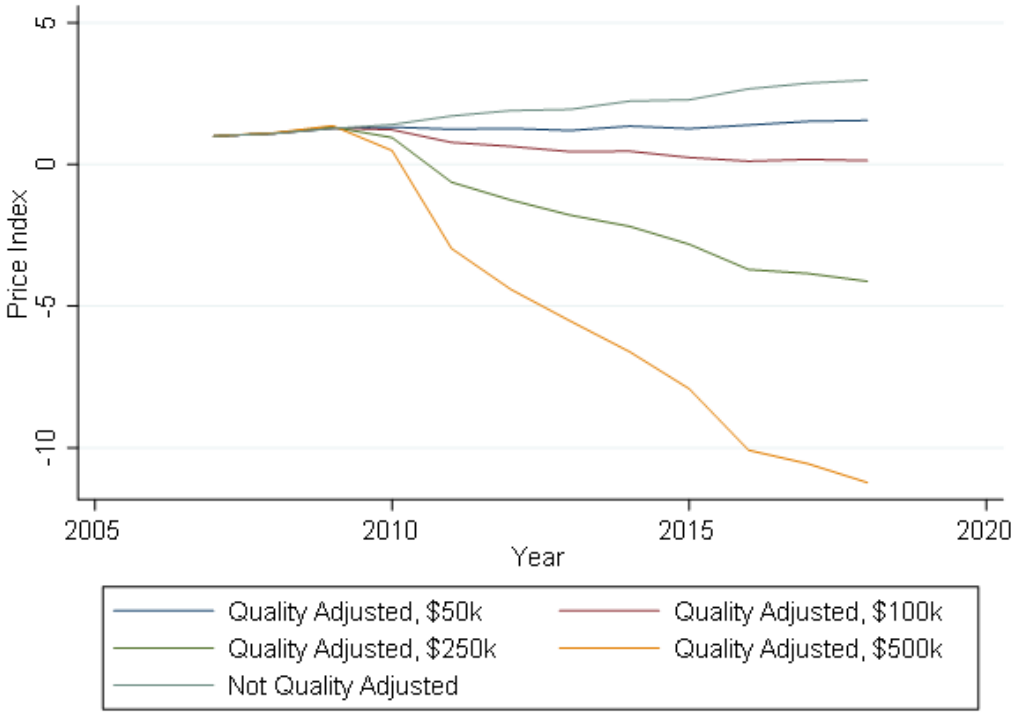
Notes: This figure presents the average price per year of the 5 highest volume drugs for Atrial Fibrillation in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA4: Price Indexes and Changes in Welfare by Year for Atrial Fibrillation

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	16	1.000	1.000	0	1.000	0
2008	-0.001	18	1.089	1.093	-2	1.111	-2
2009	-0.003	21	1.266	1.286	-5	1.364	-6
2010	0.030	23	1.407	1.224	-4	0.490	8
2011	0.152	28	1.711	0.775	4	-2.967	64
2012	0.205	31	1.901	0.638	6	-4.414	88
2013	0.242	32	1.943	0.449	9	-5.527	106
2014	0.287	36	2.237	0.467	9	-6.615	124
2015	0.331	37	2.282	0.242	12	-7.916	145
2016	0.414	43	2.667	0.117	14	-10.085	180
2017	0.436	46	2.864	0.179	13	-10.562	188
2018	0.461	48	2.977	0.136	14	-11.228	198

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA6.

Figure OA6: Price Indexes for Atrial Fibrillation



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

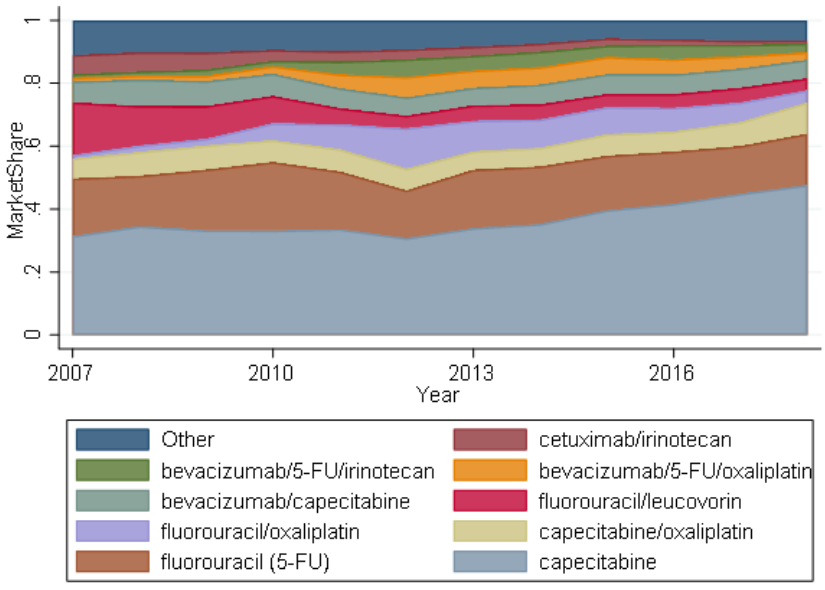
OA.A.3 Colon Cancer

Table OA5: QALY Estimates for Colon Cancer

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
bevacizumab/5-FU/irinotecan	0	0.617
bevacizumab/5-FU/oxaliplatin	0	0.429
bevacizumab/capecitabine	0	1.368
capecitabine	0	0.178
capecitabine/oxaliplatin	0	0.114
cetuximab/irinotecan	0	0.974
fluorouracil/leucovorin	0	-0.079
fluorouracil/oxaliplatin	0	0.167
fluorouracil_(5-FU)	1	0.000

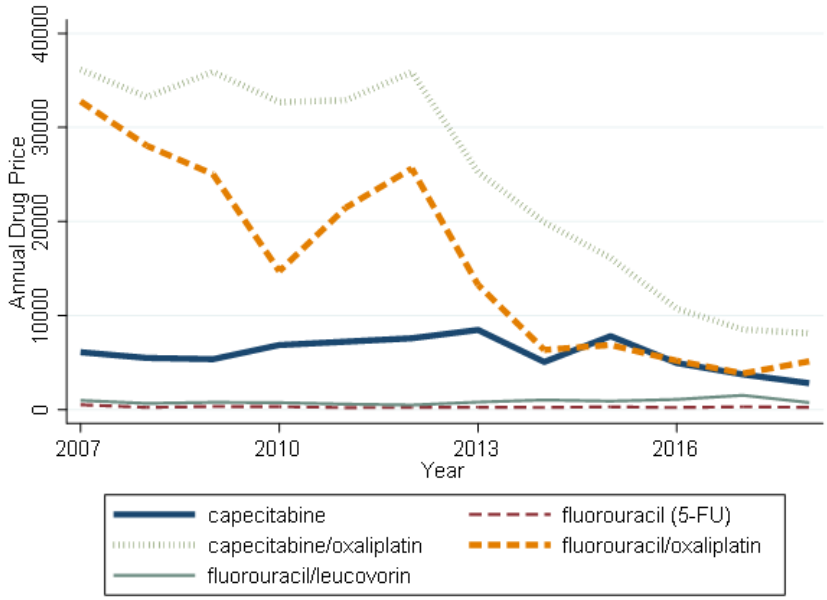
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA7: Market shares for the top Colon Cancer treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Colon Cancer across our entire sample period in the MarketScan data.

Figure OA8: Prices for the top Colon Cancer treatments over time



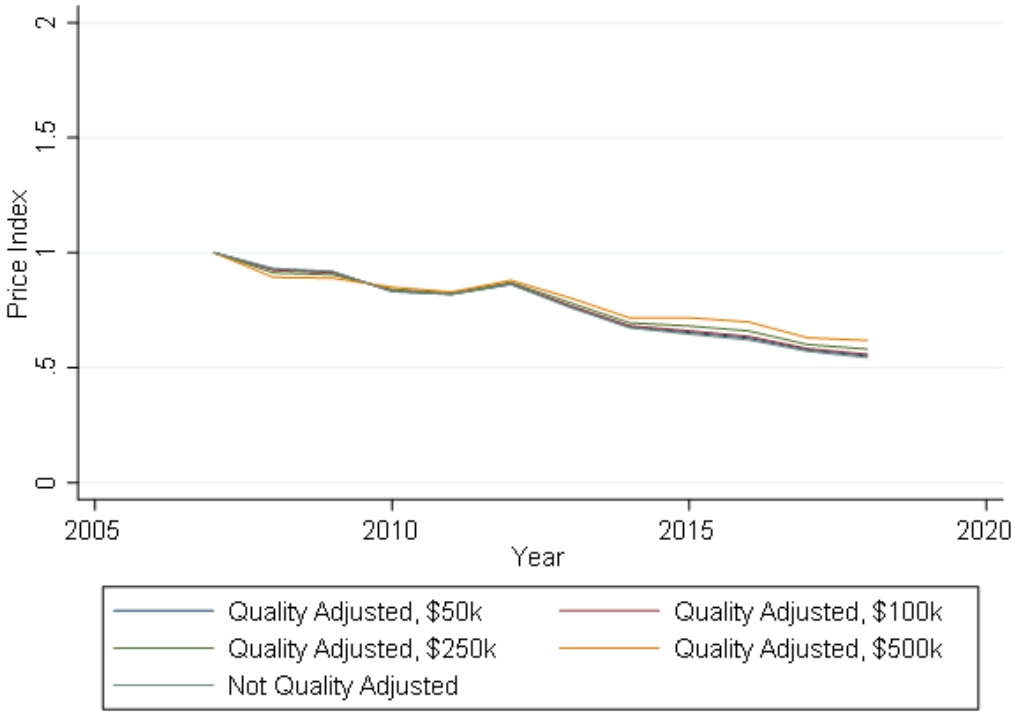
Notes: This figure presents the average price per year of the 5 highest volume drugs for Colon Cancer in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA6: Price Indexes and Changes in Welfare by Year for Colon Cancer

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	403	1.000	1.000	0	1.000	0
2008	0.031	376	0.932	0.924	31	0.893	43
2009	0.024	371	0.919	0.913	35	0.889	45
2010	-0.016	335	0.831	0.834	67	0.850	61
2011	-0.008	330	0.819	0.821	72	0.829	69
2012	-0.015	348	0.862	0.865	54	0.880	48
2013	-0.034	307	0.761	0.769	93	0.803	79
2014	-0.035	271	0.673	0.682	128	0.716	114
2015	-0.057	260	0.646	0.660	137	0.717	114
2016	-0.063	250	0.621	0.636	147	0.699	122
2017	-0.048	230	0.571	0.582	168	0.630	149
2018	-0.060	219	0.543	0.558	178	0.618	154

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA9.

Figure OA9: Price Indexes for Colon Cancer



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

OA.A.4 Cystic Fibrosis

Cystic fibrosis is a genetic disease that affects the respiratory, digestive, and reproductive systems. It causes thick, sticky mucus to build up in the lungs, pancreas, and other organs, leading to infections, inflammation, and progressive damage over time. In 2015, Orkambi was approved by the FDA. It was viewed as a breakthrough therapy and indeed we estimate that it adds 0.856 QALYs compared to tobramycin, a sizeable improvement. However, it was controversially priced at least $100k per year.[52] As shown in Figure 5 below, Orkambi has taken over about 20% market share since entry. While 0.856 is a sizeable QALY improve-

[52] "A Drug Costs $272,000 a Year. Not So Fast, Says New York State." New York Times, June 2018.

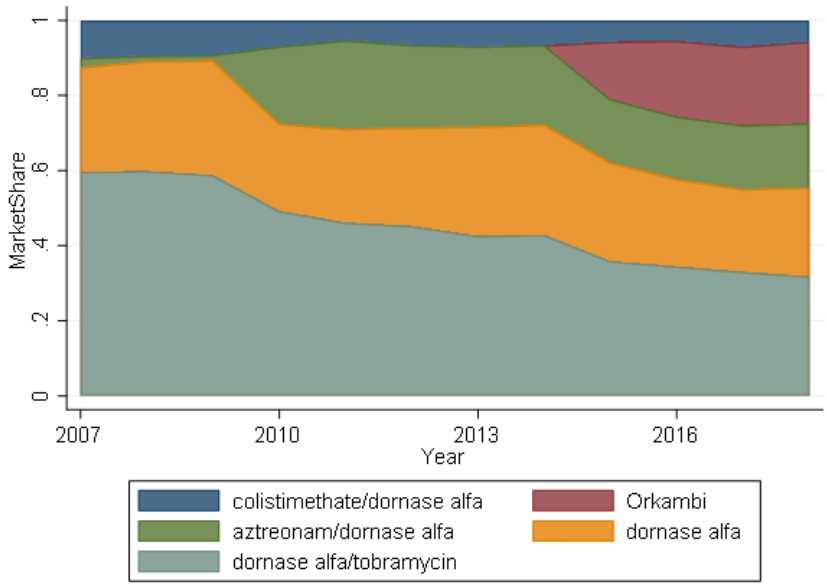
ment, so is a $100k price tag on an annual basis – or over $1m in net present value terms. Our methodology has to evaluate the trade-off, and we find given most assumptions on the VSLY we find quickly increasing quality adjusted price indexes (Table 4). It is also worth noting that all other treatments for cystic fibrosis raised their prices considerably during this time period, though the scaling (due to Orkambi's high price) masks that trend.

Table OA7: QALY Estimates for Cystic Fibrosis

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
Orkambi	0	0.894
aztreonam/dornase_alfa	0	0.209
colistimethate/dornase_alfa	0	-0.184
dornase_alfa	0	-0.002
dornase_alfa/tobramycin	1	0.000

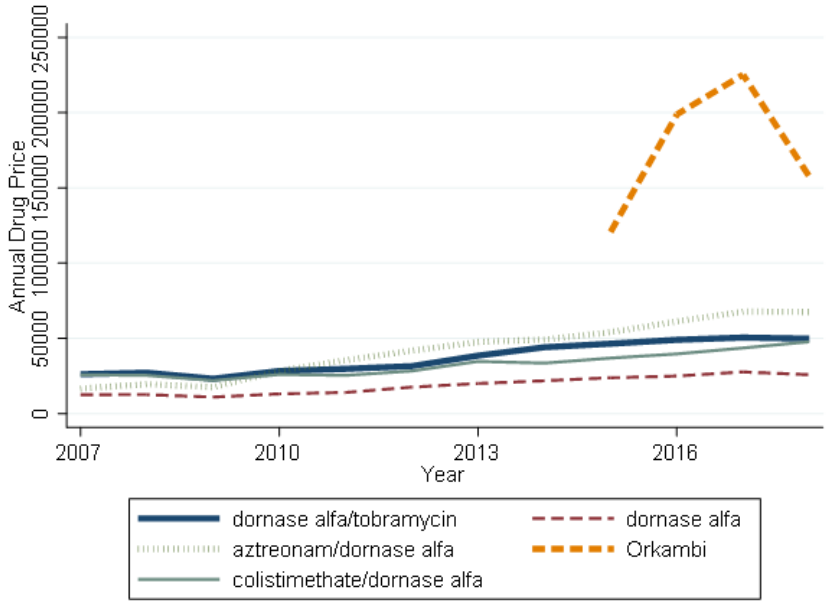
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA10: Market shares for the top Cystic Fibrosis treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Cystic Fibrosis across our entire sample period in the MarketScan data.

Figure OA11: Prices for the top Cystic Fibrosis treatments over time



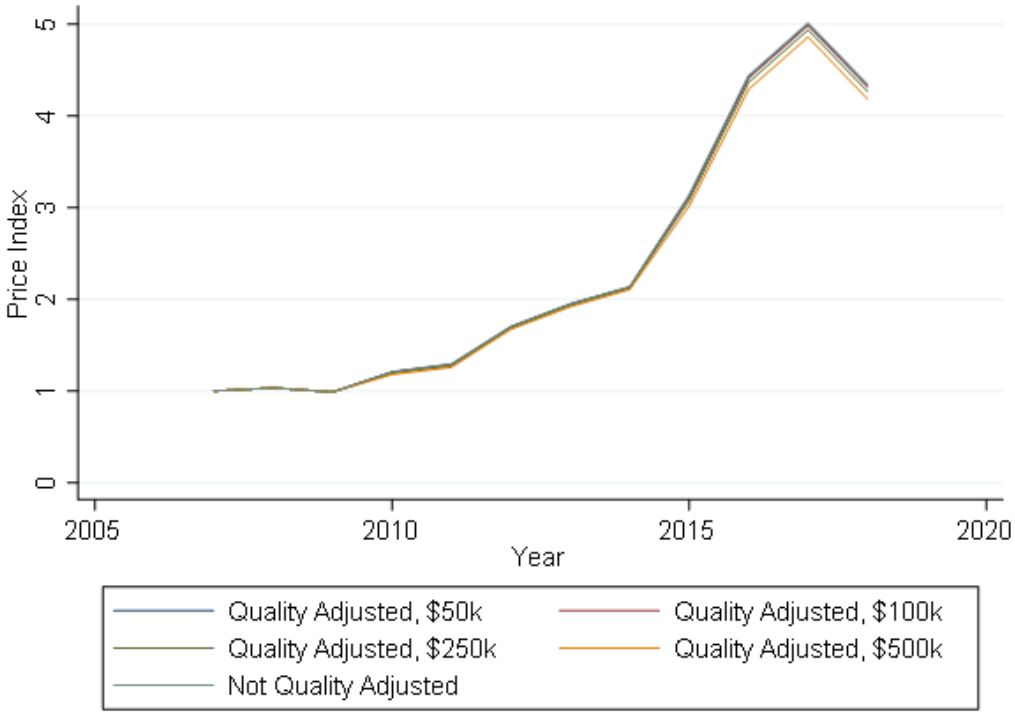
Notes: This figure presents the average price per year of the 5 highest volume drugs for Cystic Fibrosis in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA8: Price Indexes and Changes in Welfare by Year for Cystic Fibrosis

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	718	1.000	1.000	0	1.000	0
2008	-0.001	742	1.033	1.034	-24	1.034	-25
2009	-0.001	709	0.988	0.989	8	0.989	8
2010	0.044	869	1.211	1.205	-147	1.181	-130
2011	0.053	929	1.295	1.288	-206	1.258	-185
2012	0.047	1,224	1.706	1.699	-502	1.673	-483
2013	0.045	1,401	1.952	1.946	-678	1.920	-660
2014	0.045	1,535	2.138	2.132	-812	2.107	-794
2015	0.174	2,253	3.140	3.116	-1,518	3.019	-1,449
2016	0.218	3,189	4.444	4.414	-2,450	4.292	-2,362
2017	0.223	3,601	5.018	4.987	-2,861	4.863	-2,772
2018	0.233	3,121	4.349	4.316	-2,380	4.187	-2,287

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA12.

Figure OA12: Price Indexes for Cystic Fibrosis



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

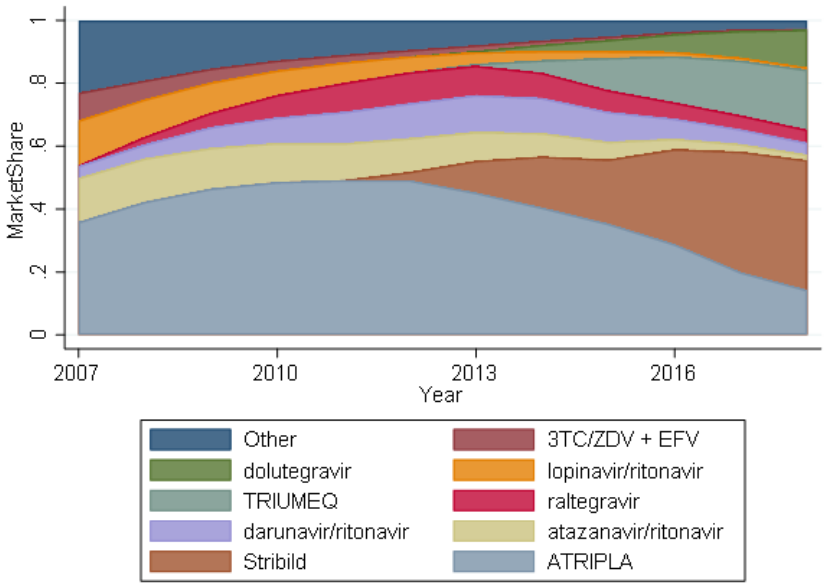
OA.A.5 HIV

Table OA9: QALY Estimates for HIV

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
3TC/ZDV_+_EFV	0	-0.325
ATRIPLA	1	0.000
Stribild	0	0.046
TRIUMEQ	0	0.190
atazanavir/ritonavir	0	0.087
darunavir/ritonavir	0	0.277
dolutegravir	0	0.347
lopinavir/ritonavir	0	-0.434
raltegravir	0	0.001

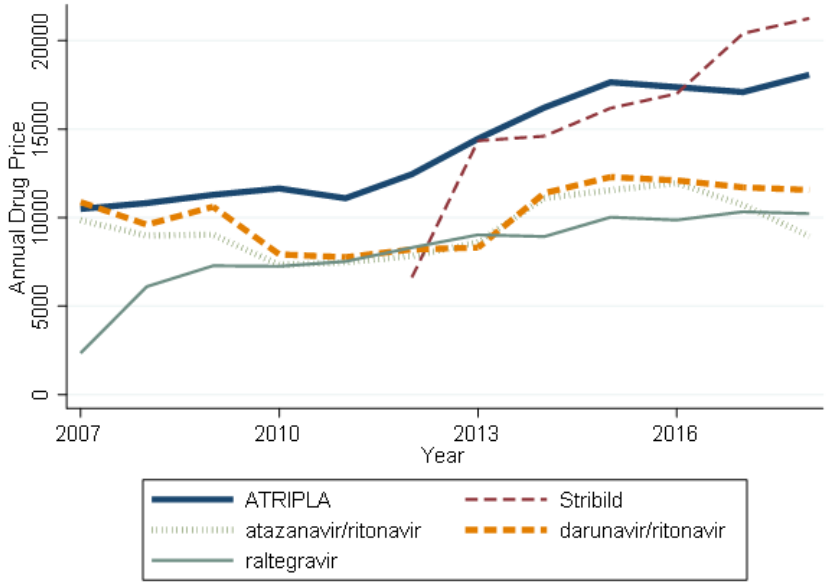
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA13: Market shares for the top HIV treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for HIV across our entire sample period in the MarketScan data.

Figure OA14: Prices for the top HIV treatments over time



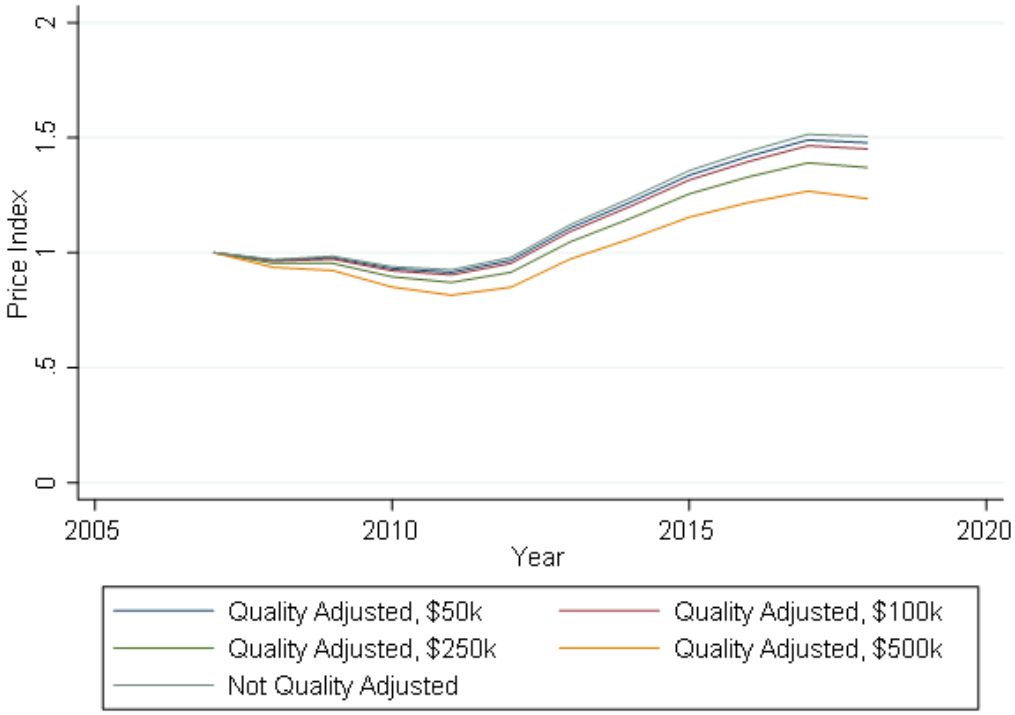
Notes: This figure presents the average price per year of the 5 highest volume drugs for HIV in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA10: Price Indexes and Changes in Welfare by Year for HIV

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	348	1.000	1.000	0	1.000	0
2008	0.024	338	0.970	0.964	13	0.936	22
2009	0.043	343	0.985	0.972	10	0.922	27
2010	0.062	327	0.939	0.922	27	0.851	52
2011	0.077	322	0.926	0.904	34	0.815	64
2012	0.090	341	0.980	0.954	16	0.850	52
2013	0.104	390	1.122	1.092	-32	0.972	10
2014	0.123	430	1.236	1.201	-70	1.059	-21
2015	0.141	472	1.357	1.316	-110	1.154	-54
2016	0.155	501	1.441	1.397	-138	1.218	-76
2017	0.172	527	1.514	1.465	-162	1.267	-93
2018	0.188	524	1.505	1.451	-157	1.235	-82

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA15.

Figure OA15: Price Indexes for HIV



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

OA.A.6 Hypertension

Hypertension's average price of treatment declines by 32% during our sample period and there is a modest quality improvement. As shown in Figure OA16, amlopidine and losartan came off patent in 2007 and 2009, leading both to gain market share. We estimate that amlopidine is slightly lower quality than average, and people substitute towards it as it comes off patent. This leads to declining quality for hypertension in the initial years of our sample period. Losartan is somewhat higher quality than the average hypertension treatment, so its increase in market share, also due to its expiring patent status, led to a modest increase in mean QALYs by the end of the sample period. Overall, price declines due to patent

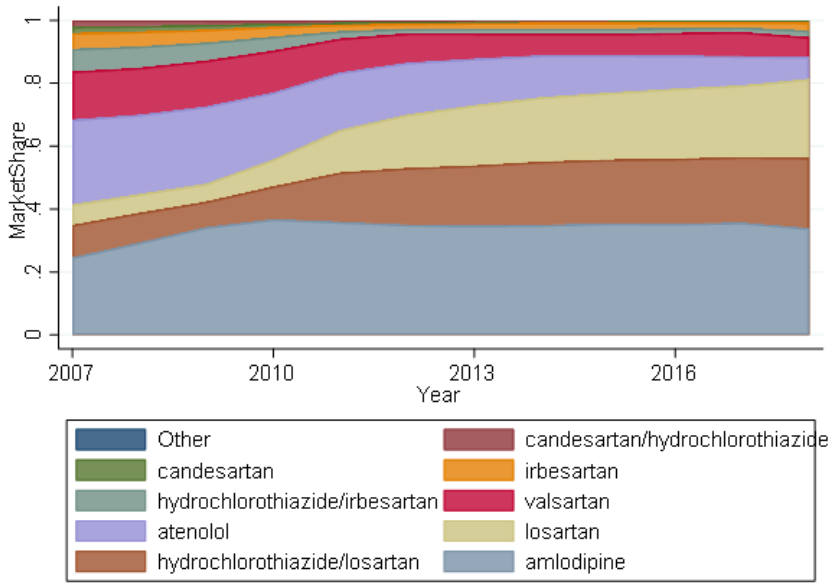
expirations is the main story for hypertension. Quality improvements are minor as there are no new entrants.

Table OA11: QALY Estimates for Hypertension

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
amlodipine	1	0.000
atenolol	0	-0.091
candesartan	0	0.285
candesartan/hydrochlorothiazi	0	0.262
hydrochlorothiazide/irbesartan	0	0.265
hydrochlorothiazide/losartan	0	0.257
irbesartan	0	0.280
losartan	0	0.237
valsartan	0	0.261

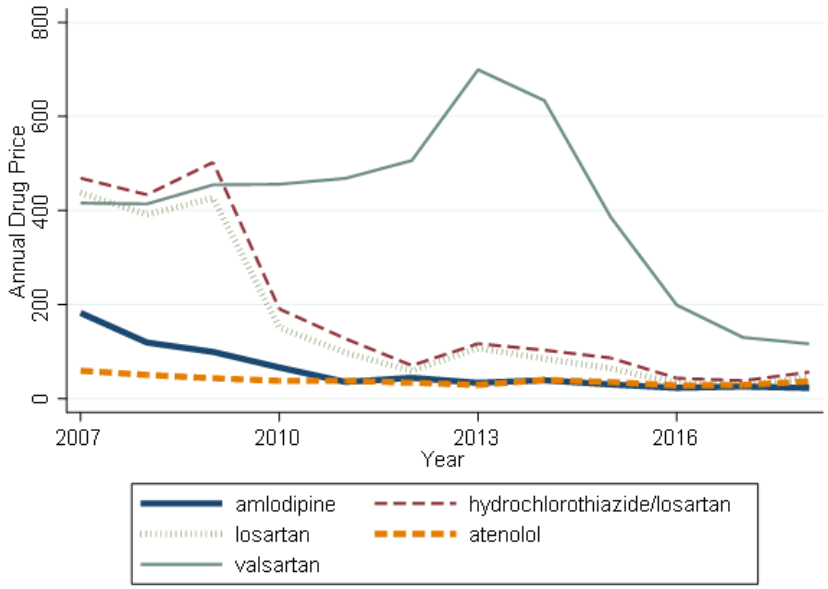
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA16: Market shares for the top Hypertension treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Hypertension across our entire sample period in the MarketScan data.

Figure OA17: Prices for the top Hypertension treatments over time



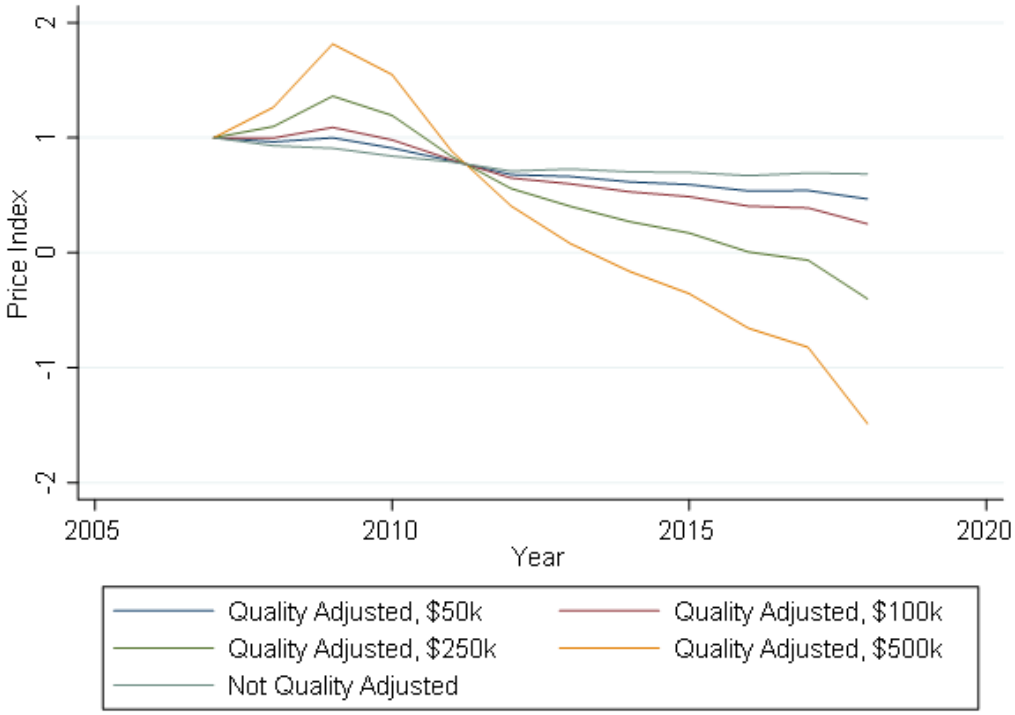
Notes: This figure presents the average price per year of the 5 highest volume drugs for Hypertension in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA12: Price Indexes and Changes in Welfare by Year for Hypertension

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	9	1.000	1.000	0	1.000	0
2008	-0.006	9	0.930	0.997	0	1.264	-2
2009	-0.017	8	0.908	1.090	-1	1.816	-7
2010	-0.013	8	0.840	0.981	0	1.549	-5
2011	-0.002	7	0.790	0.809	2	0.885	1
2012	0.006	7	0.710	0.649	3	0.407	5
2013	0.012	7	0.728	0.598	4	0.079	8
2014	0.016	6	0.703	0.530	4	-0.164	11
2015	0.019	6	0.698	0.487	5	-0.357	12
2016	0.024	6	0.670	0.404	5	-0.658	15
2017	0.028	6	0.693	0.390	6	-0.822	17
2018	0.040	6	0.684	0.250	7	-1.487	23

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA18.

Figure OA18: Price Indexes for Hypertension



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

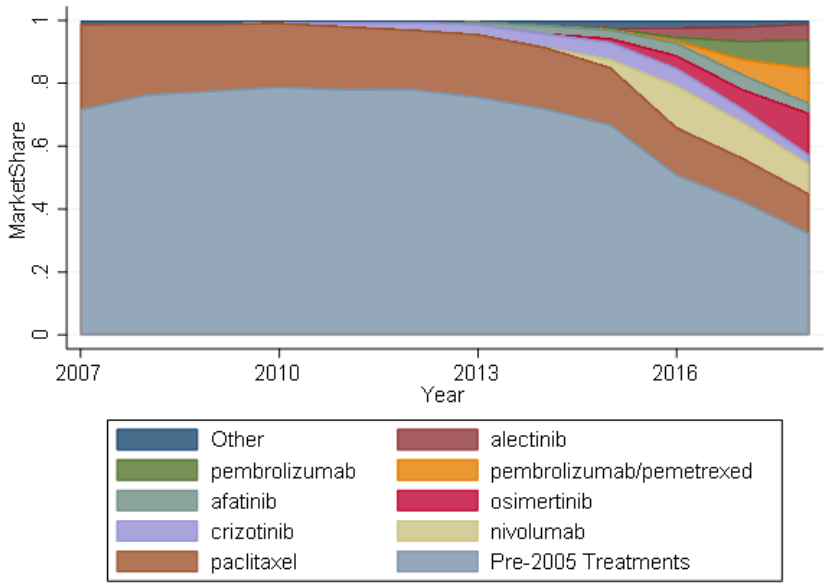
OA.A.7 Lung Cancer

Table OA13: QALY Estimates for Lung Cancer

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
afatinib	0	0.354
alectinib	0	1.102
bevacizumab/paclitaxel	0	0.377
bevacizumab/pemetrexed	0	0.464
ceritinib	0	0.316
crizotinib	0	0.226
docetaxel	0	-0.102
erlotinib	0	-0.026
gemcitabine	0	0.172
gemcitabine/pemetrexed	0	0.254
nivolumab	0	1.165
osimertinib	0	0.922
paclitaxel	1	0.000
pembrolizumab	0	1.019
pembrolizumab/pemetrexed	0	2.674
pemetrexed	0	0.153
vinorelbine	0	0.036

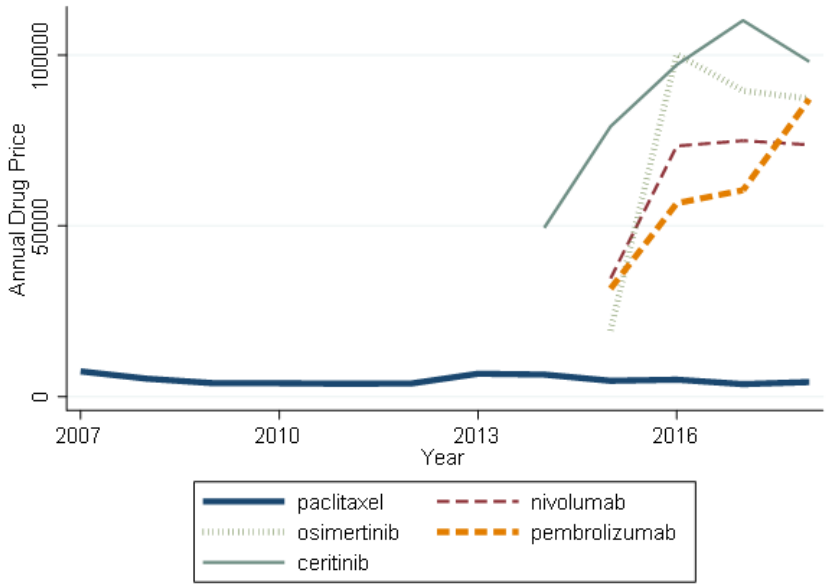
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA19: Market shares for selected treatments for lung cancer over time



Notes: This figure presents the market shares by year for selected treatments for lung cancer across our entire sample period in the MarketScan data.

Figure OA20: Prices for selected lung cancer treatments over time



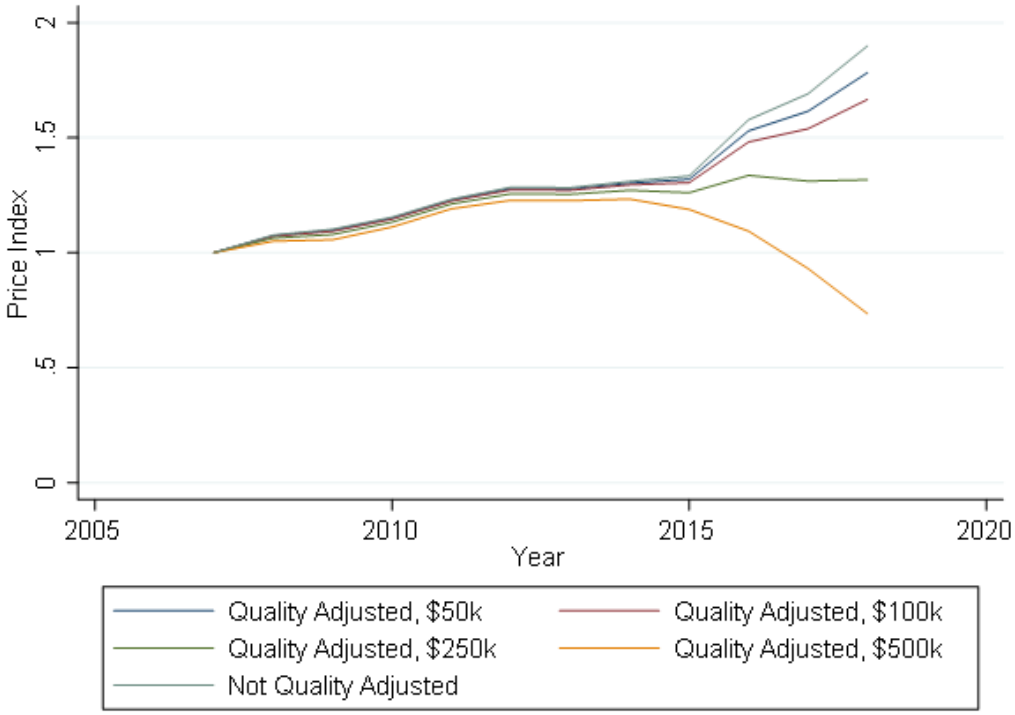
Notes: This figure presents the average price per year selected treatments for lung cancer in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA14: Price Indexes and Changes in Welfare by Year for Lung Cancer

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	281	1.000	1.000	0	1.000	0
2008	0.015	302	1.077	1.071	-20	1.050	-14
2009	0.026	309	1.102	1.093	-26	1.056	-16
2010	0.024	324	1.154	1.146	-41	1.112	-31
2011	0.024	346	1.233	1.224	-63	1.191	-54
2012	0.032	361	1.285	1.273	-77	1.228	-64
2013	0.032	360	1.283	1.271	-76	1.226	-63
2014	0.043	368	1.310	1.295	-83	1.233	-65
2015	0.082	374	1.334	1.305	-86	1.188	-53
2016	0.272	443	1.578	1.481	-135	1.094	-26
2017	0.427	475	1.691	1.539	-151	0.931	19
2018	0.654	533	1.899	1.666	-187	0.734	75

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time. The price indexes are also graphed in Figure OA21.

Figure OA21: Price Indexes for Lung Cancer



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

OA.A.8 Multiple Sclerosis

Multiple sclerosis (MS) is a chronic autoimmune disease that affects the central nervous system, causing damage to the myelin sheath that covers nerve fibers. This can result in a wide range of symptoms, including problems with vision, movement, balance, and sensation. At the beginning of our sample period, multiple sclerosis was typically treated with older "platform" injectable drugs like glatiramer or interferon beta 1a. There has been considerable entry in the market since the beginning of our sample period. Oral therapies like fumarates (dimethyl fumarate), sphingosine 1-phosphate (S1P) receptor modulators (finglomod), and teriflunomide which are viewed as more convenient but lower efficacy than injectable mon-

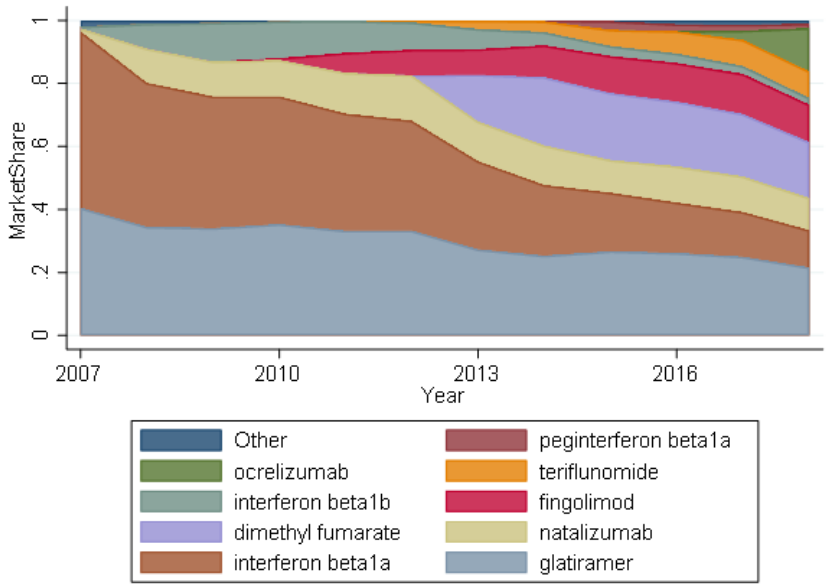
oclonal antibodies such as natalizumab and ocrelizumab (Olek and Mowry, 2022). Both of these sets of treatments are seen as being more effective than the original class of "platform injectable" treatments (Olek and Mowry, 2022). This class difference shows up in our regressions. The oral drugs, dimethyl fumarate, teriflunomide, and finglomod are 0.295, 0.291, and 0.66 QALYs better than interferon beta1a, respectively (See Appendix Table OA15). Injectable monoclonal antibodies such as ocrelizumab and natalizumab are 0.61 and 1.1 QALYs more effective than interferon beta1a. While highly effective, unlike for hepatitis C, none of these drugs are curative and they are only partially effective for reducing the relapse rate. Hence the magnitude of QALY improvements is below that of rheumatoid arthritis and hepatitis C, but still relatively large as they reduce or delay serious disability. While there are real quality improvements, prices are rising rapidly. Interestingly, price increases do not appear to be driven by new drugs entering at high price points, but rather drugs increasing their prices rapidly after entry. As the right panel of Figure **??** shows, nearly all the new entrants in this market more than doubled their prices during our sample period. In Table 4 we see that the cost of multiple sclerosis went up by 303% during our sample period representing an level increase of nearly $464k. While average QALYs increased by a sizeable 0.4, this was not enough to offset the large price increases. Hence, multiple sclerosis has one of the fastest growing quality-adjusted price indexes, even though there were large quality improvements during our sample period.

Table OA15: QALY Estimates for Multiple Sclerosis

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
dimethyl_fumarate	0	0.320
fingolimod	0	0.716
glatiramer	0	-0.262
interferon_beta1a	1	0.000
interferon_beta1b	0	-0.217
natalizumab	0	1.201
ocrelizumab	0	0.666
peginterferon_beta1a	0	0.290
teriflunomide	0	0.315

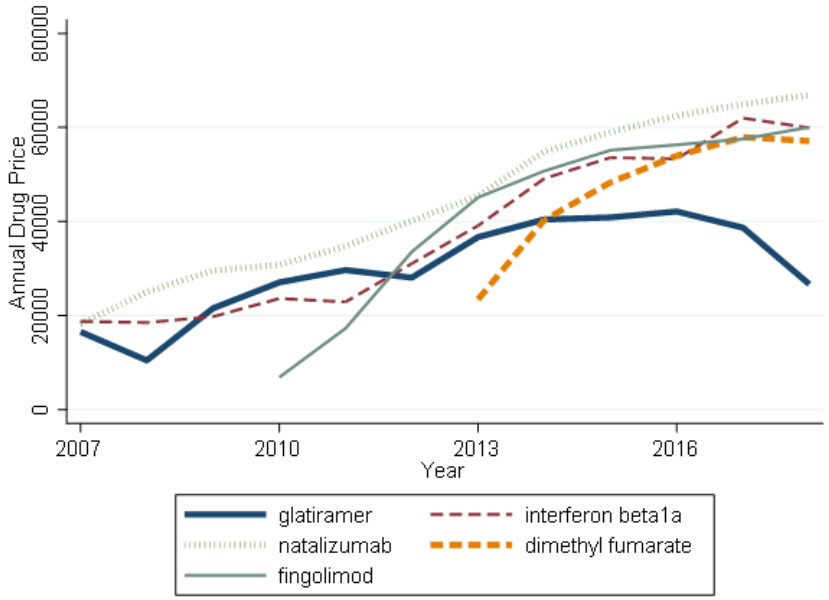
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA22: Market shares for the top Multiple Sclerosis treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Multiple Sclerosis across our entire sample period in the MarketScan data.

Figure OA23: Prices for the top Multiple Sclerosis treatments over time



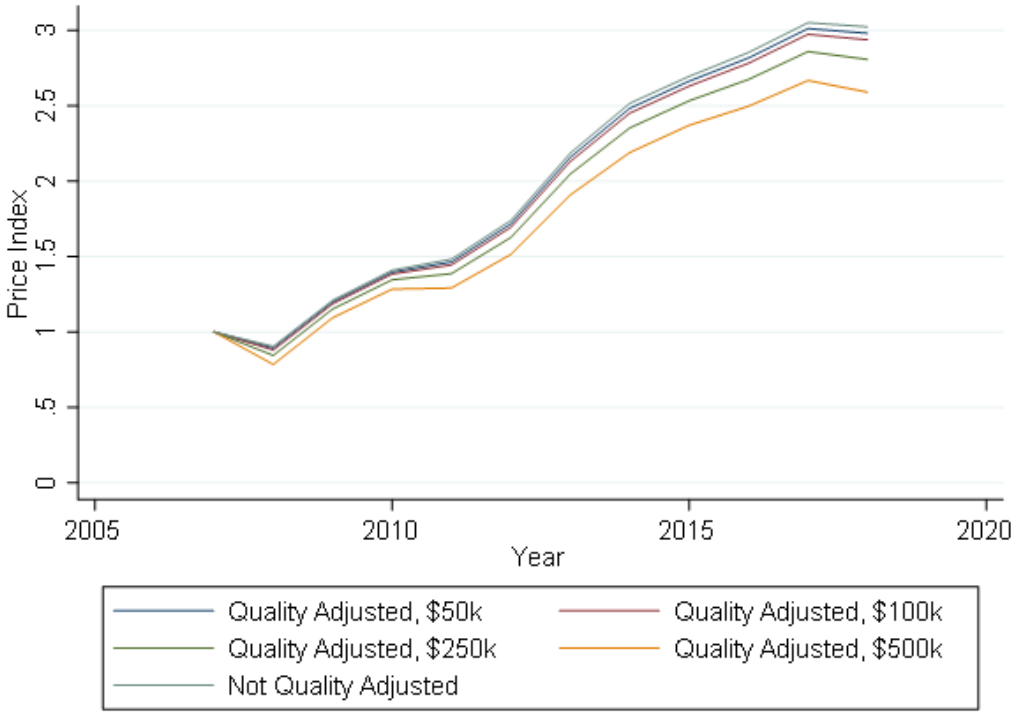
Notes: This figure presents the average price per year of the 5 highest volume drugs for Multiple Sclerosis in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA16: Price Indexes and Changes in Welfare by Year for Multiple Sclerosis

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	507	1.000	1.000	0	1.000	0
2008	0.121	458	0.904	0.880	61	0.784	109
2009	0.118	613	1.210	1.187	-95	1.094	-47
2010	0.126	714	1.408	1.383	-194	1.284	-144
2011	0.194	751	1.482	1.444	-225	1.291	-148
2012	0.228	881	1.740	1.695	-352	1.515	-261
2013	0.282	1,108	2.187	2.131	-573	1.909	-461
2014	0.331	1,275	2.516	2.451	-735	2.190	-603
2015	0.329	1,366	2.696	2.631	-826	2.371	-695
2016	0.362	1,446	2.854	2.783	-903	2.497	-759
2017	0.388	1,546	3.051	2.974	-1,000	2.668	-845
2018	0.440	1,533	3.025	2.938	-982	2.590	-806

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA24.

Figure OA24: Price Indexes for Multiple Sclerosis



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

OA.A.9 Osteoporosis

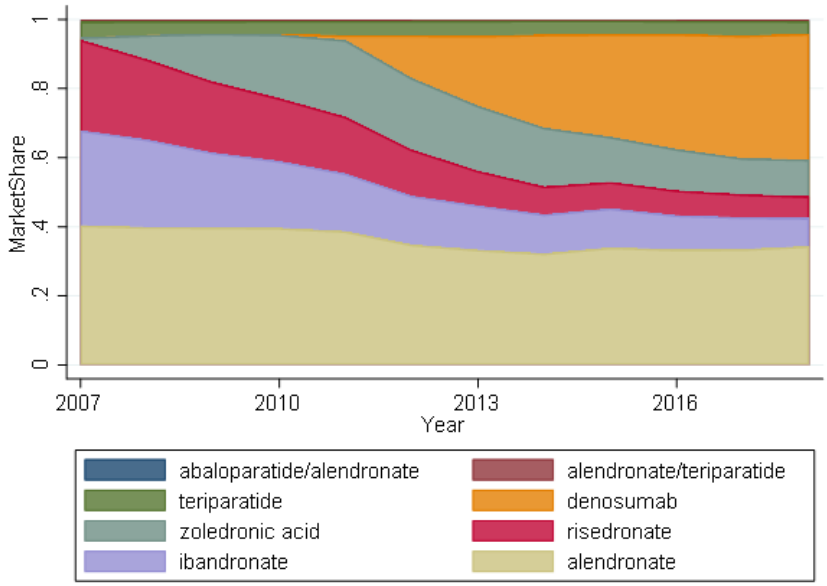
Osteoporosis has the entry of denusumab, which takes considerable market share. Our methodology views denosumab as modestly more effective than its comparators, hence a modest quality improvement for osteoporosis. However, denusumab is also higher cost, leading to sizeable unadjusted price index growth. On net, quality-adjusted prices are rising for VSLY of $100k, but falling for the VSLY of $250k as for this VSLY the quality improvement is sizeable relative to the price increases.

Table OA17: QALY Estimates for Osteoporosis

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
abaloparatide/alendronate	0	0.077
alendronate	1	0.000
alendronate/teriparatide	0	0.060
denosumab	0	0.076
ibandronate	0	-0.056
risedronate	0	0.010
teriparatide	0	0.059
zoledronic_acid	0	-0.011

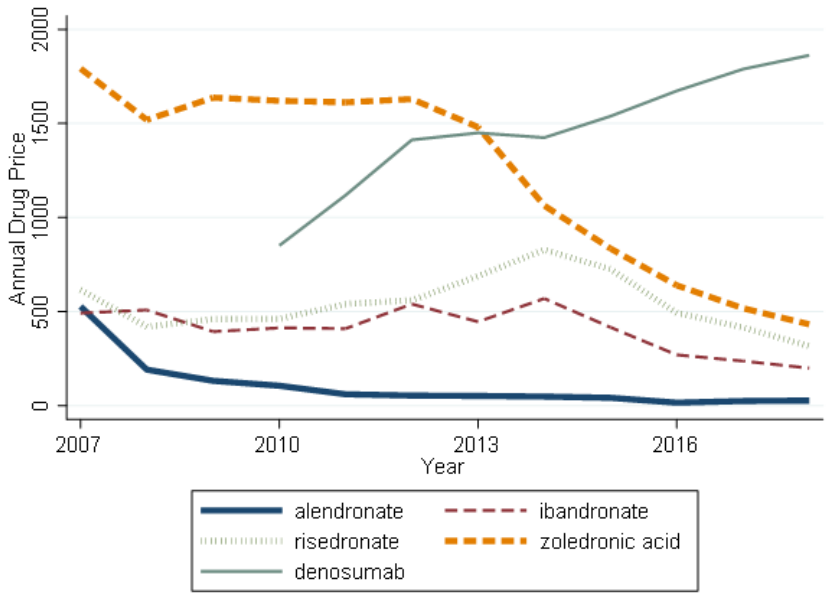
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA25: Market shares for the top Osteoporosis treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Osteoporosis across our entire sample period in the MarketScan data.

Figure OA26: Prices for the top Osteoporosis treatments over time



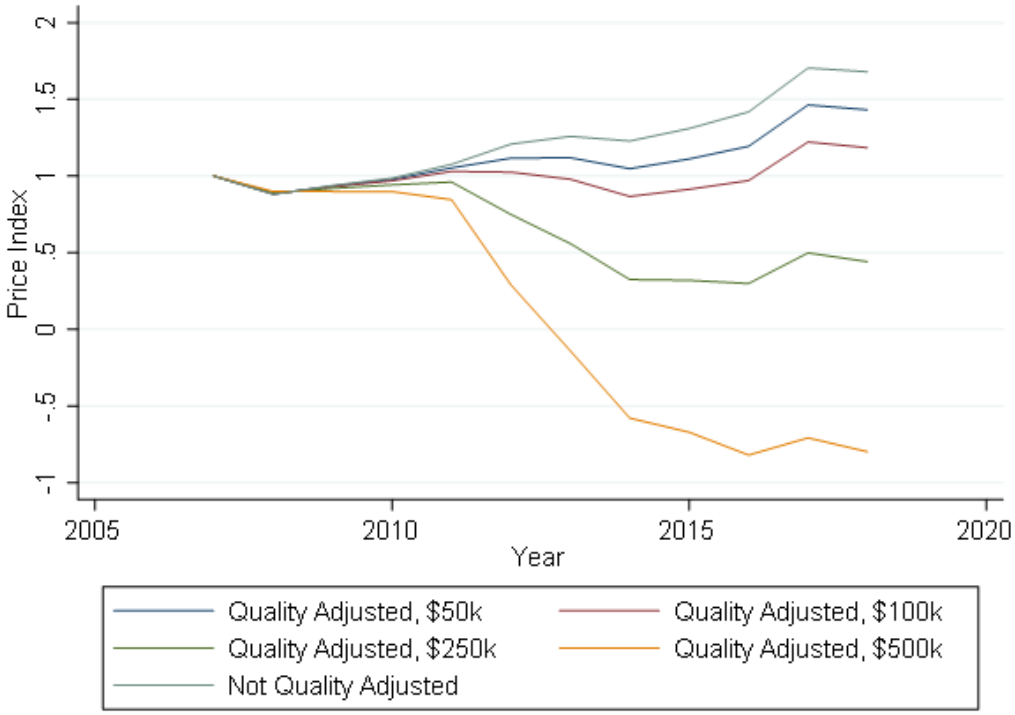
Notes: This figure presents the average price per year of the 5 highest volume drugs for Osteoporosis in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA18: Price Indexes and Changes in Welfare by Year for Osteoporosis

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	7	1.000	1.000	0	1.000	0
2008	-0.000	6	0.881	0.884	1	0.899	1
2009	0.001	7	0.939	0.931	0	0.900	1
2010	0.001	7	0.986	0.969	0	0.898	1
2011	0.003	7	1.076	1.030	0	0.846	1
2012	0.013	8	1.209	1.025	0	0.290	5
2013	0.020	9	1.259	0.979	0	-0.141	8
2014	0.025	9	1.229	0.867	1	-0.580	11
2015	0.028	9	1.310	0.914	1	-0.671	12
2016	0.031	10	1.418	0.970	0	-0.821	13
2017	0.034	12	1.705	1.222	-2	-0.708	12
2018	0.035	12	1.680	1.185	-1	-0.799	13

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA27.

Figure OA27: Price Indexes for Osteoporosis



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

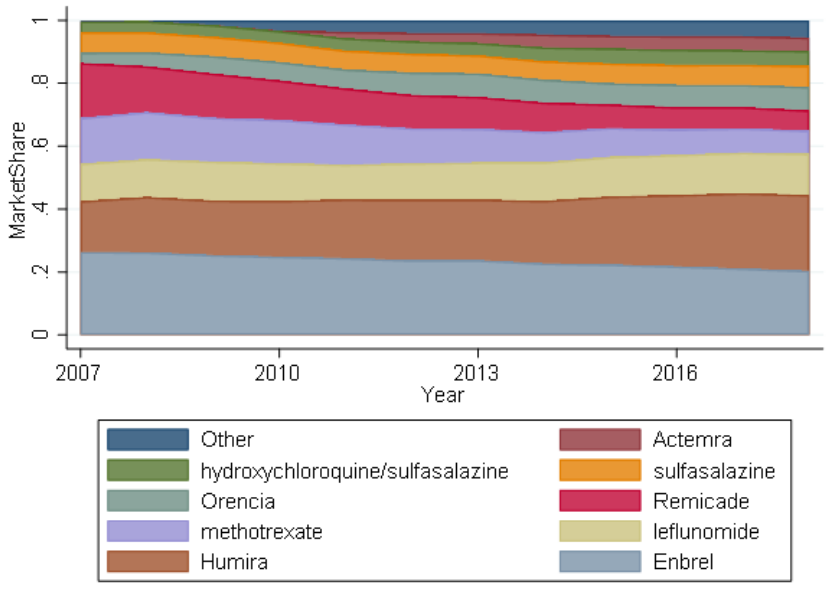
OA.A.10 Rheumatoid Arthritis

Table OA19: QALY Estimates for Rheumatoid Arthritis

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
Actemra	0	3.164
Enbrel	0	2.153
Humira	0	2.295
Orencia	0	2.641
Remicade	0	1.907
hydroxychloroquine/sulfasalaz	0	2.165
leflunomide	0	0.048
methotrexate	1	0.000
sulfasalazine	0	-0.259

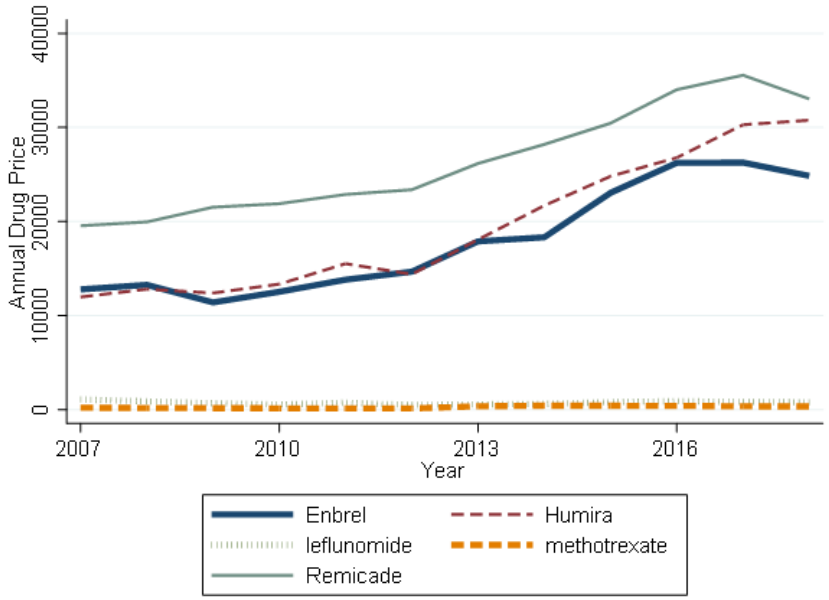
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA28: Market shares for the top Rheumatoid Arthritis treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Rheumatoid Arthritis across our entire sample period in the MarketScan data.

Figure OA29: Prices for the top Rheumatoid Arthritis treatments over time



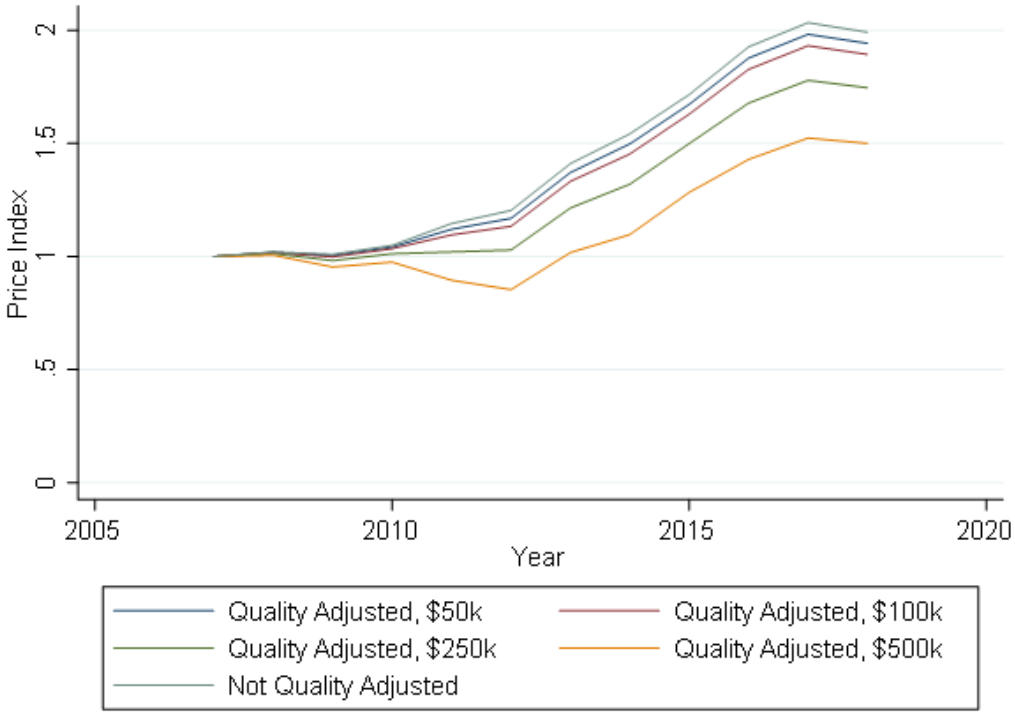
Notes: This figure presents the average price per year of the 5 highest volume drugs for Rheumatoid Arthritis in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA20: Price Indexes and Changes in Welfare by Year for Rheumatoid Arthritis

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	198	1.000	1.000	0	1.000	0
2008	0.006	202	1.020	1.017	-3	1.006	-1
2009	0.022	200	1.010	0.999	0	0.954	9
2010	0.030	208	1.049	1.034	-7	0.975	5
2011	0.099	227	1.146	1.096	-19	0.895	21
2012	0.139	239	1.204	1.134	-27	0.854	29
2013	0.156	280	1.411	1.332	-66	1.018	-3
2014	0.177	306	1.543	1.454	-90	1.097	-19
2015	0.171	340	1.716	1.630	-125	1.284	-56
2016	0.197	382	1.927	1.828	-164	1.429	-85
2017	0.202	403	2.034	1.932	-185	1.524	-104
2018	0.195	395	1.992	1.894	-177	1.500	-99

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA30.

Figure OA30: Price Indexes for Rheumatoid Arthritis



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

OA.A.11 Schizophrenia

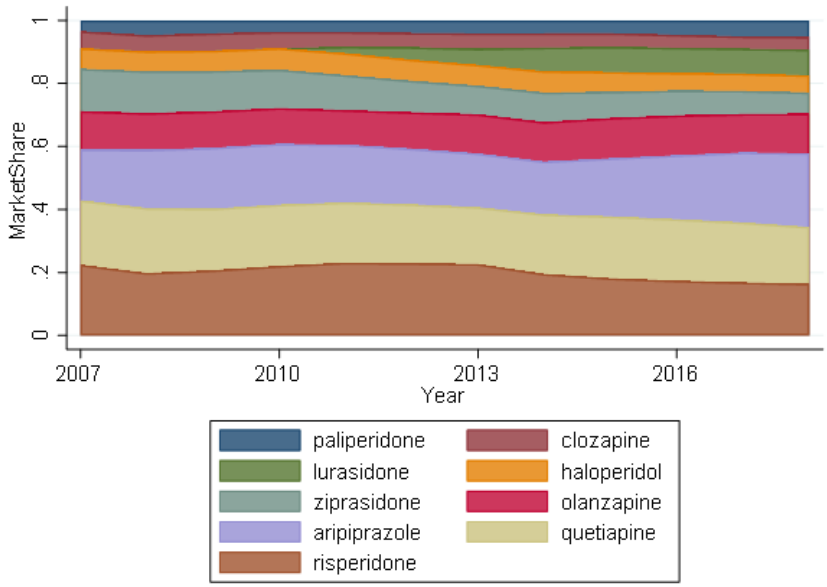
Schizophrenia has relatively stable market shares, except for the entry of lurasidone, which drives some modest quality improvements. Similar to hypertension, schizophrenia also has multiple treatments coming off patent leading to falling price indexes.

Table OA21: QALY Estimates for Schizophrenia

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
aripiprazole	0	0.362
clozapine	0	-0.684
haloperidol	0	-0.192
lurasidone	0	0.371
olanzapine	0	0.429
paliperidone	0	0.851
quetiapine	0	-1.086
risperidone	1	0.000
ziprasidone	0	-0.120

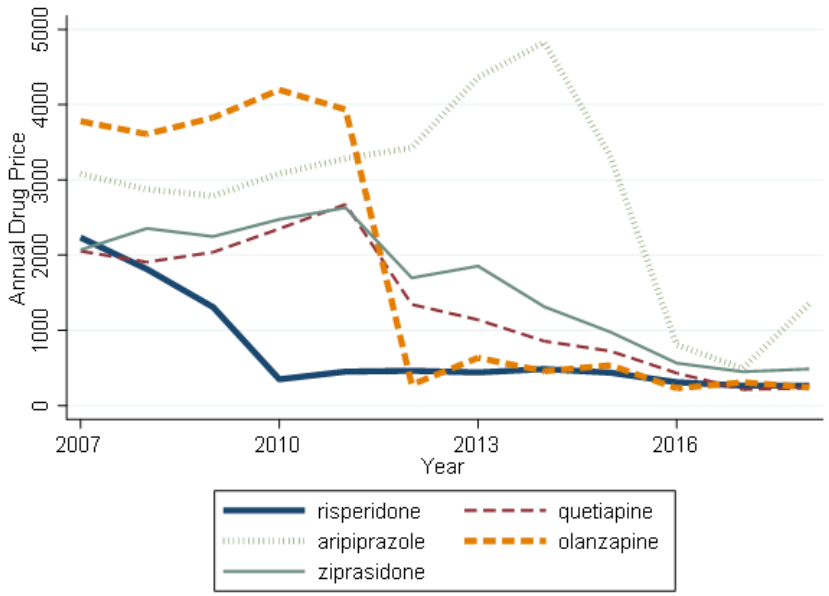
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA31: Market shares for the top Schizophrenia treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Schizophrenia across our entire sample period in the MarketScan data.

Figure OA32: Prices for the top Schizophrenia treatments over time



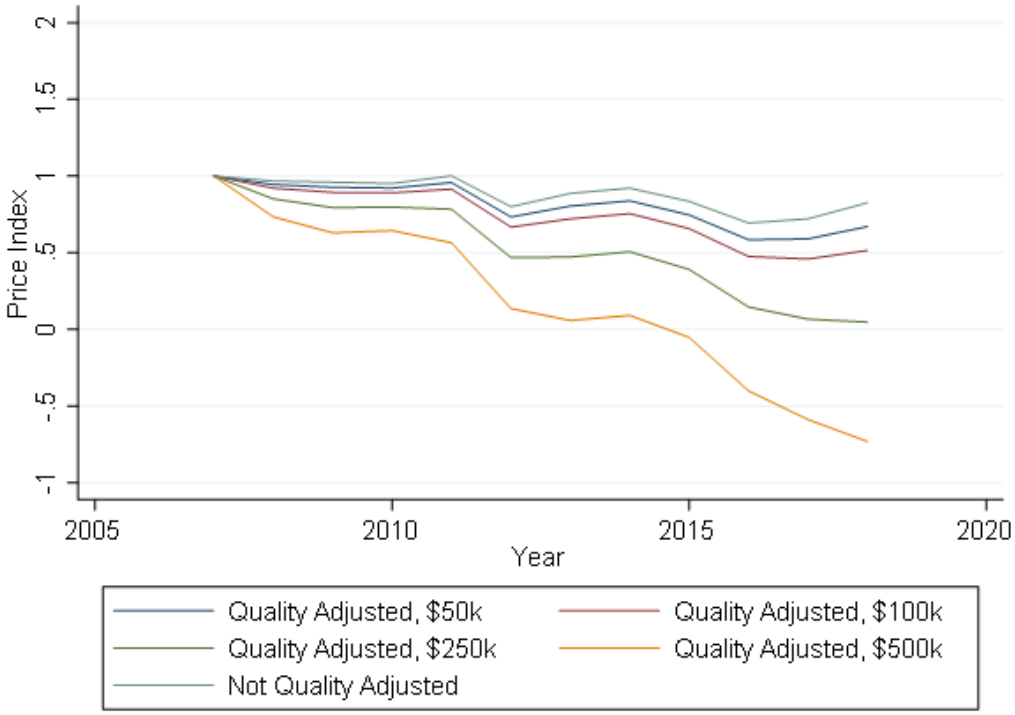
Notes: This figure presents the average price per year of the 5 highest volume drugs for Schizophrenia in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA22: Price Indexes and Changes in Welfare by Year for Schizophrenia

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	39	1.000	1.000	0	1.000	0
2008	0.018	37	0.968	0.921	3	0.734	10
2009	0.025	37	0.959	0.893	4	0.629	14
2010	0.024	37	0.952	0.890	4	0.644	14
2011	0.034	39	1.001	0.914	3	0.565	17
2012	0.051	31	0.800	0.667	13	0.135	33
2013	0.064	34	0.887	0.721	11	0.058	36
2014	0.064	36	0.921	0.755	9	0.091	35
2015	0.069	32	0.835	0.657	13	-0.053	41
2016	0.085	27	0.694	0.475	20	-0.402	54
2017	0.101	28	0.721	0.459	21	-0.588	61
2018	0.120	32	0.826	0.514	19	-0.732	67

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA33.

Figure OA33: Price Indexes for Schizophrenia



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

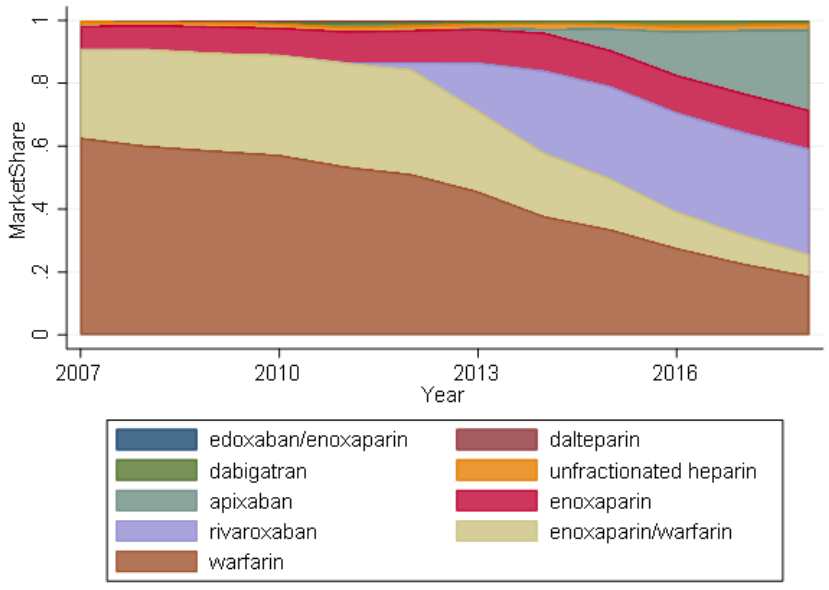
OA.A.12 Venous Thromboembolism

Table OA23: QALY Estimates for Venous Thromboembolism

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
apixaban	0	0.241
dabigatran	0	0.188
dalteparin	0	0.298
edoxaban/enoxaparin	0	0.324
enoxaparin	0	0.195
enoxaparin/warfarin	0	0.172
rivaroxaban	0	0.203
unfractionated_heparin	0	0.285
warfarin	1	0.000

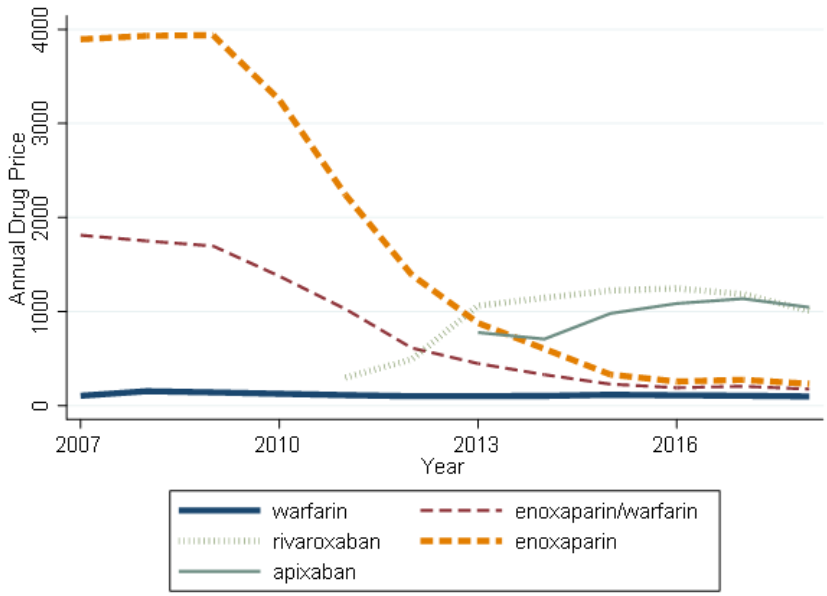
Notes: This table presents the estimated QALYs using the Tufts CEAR data and applying the regression methodology discussed in the text. Column 1 is an indicator for the index treatment for each condition, which all other QALYs are compared to. The second column is the QALY estimate relative to the index drug.

Figure OA34: Market shares for the top Venous Thromboembolism treatments over time



Notes: This figure presents the market shares by year for the 9 highest volume drugs for Venous Thromboembolism across our entire sample period in the MarketScan data.

Figure OA35: Prices for the top Venous Thromboembolism treatments over time



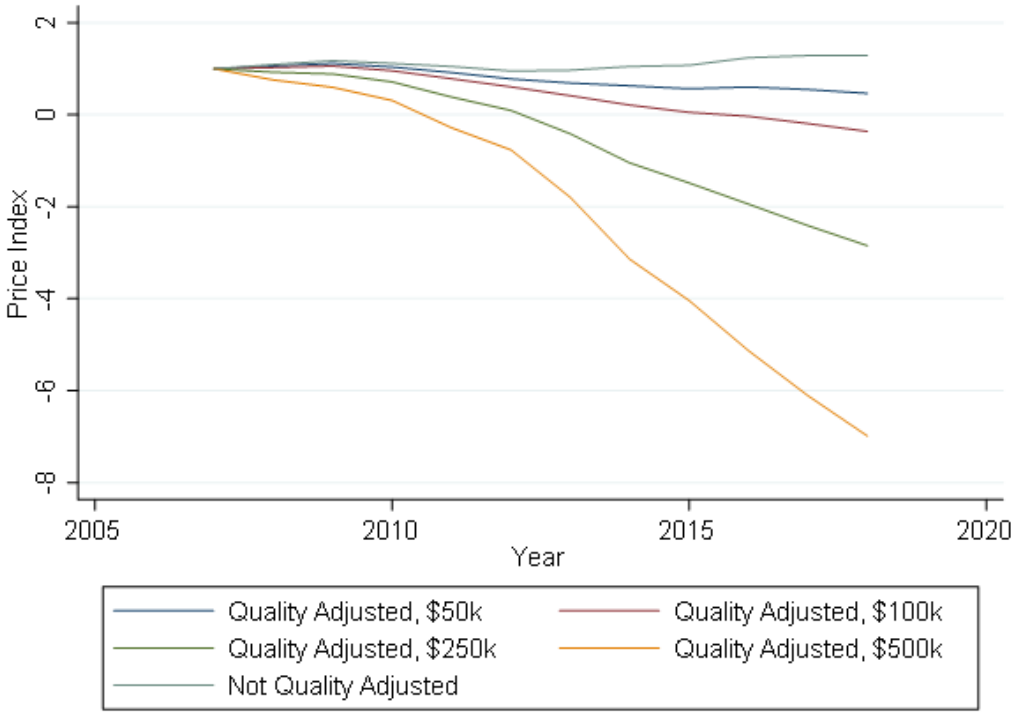
Notes: This figure presents the average price per year of the 5 highest volume drugs for Venous Thromboembolism in our sample. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. Prices are from the MarketScan data and are average costs of that drug for a patient who takes that drug in a calendar year. They are *not* scaled to lifetime costs. The drug prices are deflated to 2018 dollars using the PCE deflator and adjusted for rebates using SSR health data.

Table OA24: Price Indexes and Changes in Welfare by Year for Venous Thromboembolism

	(1) Change in Avg QALYs	(2) Estimated MktScan Lifetime Costs ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Consumer Welfare $500k VSLY ($1,000s)
2007	0.000	6	1.000	1.000	0	1.000	0
2008	0.004	7	1.092	1.025	0	0.756	2
2009	0.007	7	1.174	1.058	0	0.596	3
2010	0.010	7	1.117	0.956	0	0.311	4
2011	0.017	7	1.050	0.784	1	-0.283	8
2012	0.022	6	0.950	0.606	3	-0.766	11
2013	0.035	6	0.967	0.414	4	-1.798	18
2014	0.053	7	1.049	0.210	5	-3.144	26
2015	0.065	7	1.076	0.052	6	-4.043	32
2016	0.081	8	1.241	-0.034	7	-5.132	39
2017	0.094	8	1.287	-0.193	8	-6.110	45
2018	0.105	8	1.292	-0.364	9	-6.989	51

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes and consumer welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3 and 4. The price indexes are also graphed in Figure OA36.

Figure OA36: Price Indexes for Venous Thromboembolism



Notes: This figure presents quality adjusted price indexes using various assumptions about the value of a statistical life year. These results are constructed using data from Tufts, MarketScan, and SSR Health.

Online Appendix OA.B Robustness Checks

OA.B.1 Robustness Checks Referenced in the Main Text

Table OA25: Counterfactual: Price Indexes and Changes in Welfare for Each Condition Between 2007 and 2018, but Simulating Prices After Drugs Go Off Patent

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.002	16	0.999	0.985	0	0.927	0
AtrialFibrillation	0.461	16	2.615	-0.226	20	-11.590	46
ColonCancer	-0.060	403	0.537	0.552	180	0.612	-6
CysticFibrosis	0.233	718	2.047	2.015	-728	1.885	23
HIV	0.188	348	0.425	0.371	219	0.154	19
HepatitisC	2.905	42	0.257	-6.641	322	-34.233	291
Hypertension	0.040	9	0.677	0.243	7	-1.495	4
LungCancer	0.654	281	0.697	0.464	151	-0.468	65
MultipleSclerosis	0.440	507	1.019	0.932	35	0.584	44
Osteoporosis	0.035	7	0.694	0.198	6	-1.785	3
RheumatoidArthritis	0.195	198	0.827	0.729	54	0.335	20
Schizophrenia	0.120	39	0.767	0.455	21	-0.791	12
VenousThromboembolism	0.105	6	0.969	-0.688	11	-7.313	11

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. The results are similar to Table 4, except in 2018 we assume that prices declined by 85% for on-patent drugs. This is meant to simulate a "long-run" outcome where these drugs have lost patent protection. Note that we allow non-drug costs to change between 2007 and 2018, so conditions like atrial fibrillation which have increases in non-drug spending still see unadjusted prices rising.

Table OA26: Price Indexes and Changes in Welfare for Each Condition Multiplying QALYs by Two

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.005	16	1.013	0.984	0	0.869	0
AtrialFibrillation	0.922	16	2.977	-2.705	60	-25.432	92
ColonCancer	-0.121	403	0.543	0.573	172	0.693	-12
CysticFibrosis	0.466	718	4.349	4.284	-2,356	4.024	47
HIV	0.376	348	1.505	1.397	-138	0.965	38
HepatitisC	5.810	42	1.205	-12.591	572	-67.775	581
Hypertension	0.080	9	0.684	-0.184	11	-3.659	8
LungCancer	1.308	281	1.899	1.433	-122	-0.430	131
MultipleSclerosis	0.880	507	3.025	2.851	-938	2.156	88
Osteoporosis	0.069	7	1.680	0.689	2	-3.278	7
RheumatoidArthritis	0.390	198	1.992	1.795	-158	1.008	39
Schizophrenia	0.241	39	0.826	0.203	31	-2.290	24
VenousThromboembolism	0.211	6	1.292	-2.020	19	-15.271	21

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time.

Table OA27: Price Indexes and Changes in Welfare for Each Condition Multiplying QALYs by One-Half

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.001	16	1.013	1.006	0	0.977	0
AtrialFibrillation	0.230	16	2.977	1.557	-9	-4.125	23
ColonCancer	-0.030	403	0.543	0.551	181	0.581	-3
CysticFibrosis	0.116	718	4.349	4.333	-2,391	4.268	12
HIV	0.094	348	1.505	1.478	-166	1.370	9
HepatitisC	1.453	42	1.205	-2.244	137	-16.040	145
Hypertension	0.020	9	0.684	0.467	5	-0.401	2
LungCancer	0.327	281	1.899	1.783	-220	1.317	33
MultipleSclerosis	0.220	507	3.025	2.981	-1,004	2.807	22
Osteoporosis	0.017	7	1.680	1.433	-3	0.441	2
RheumatoidArthritis	0.098	198	1.992	1.943	-187	1.746	10
Schizophrenia	0.060	39	0.826	0.670	13	0.047	6
VenousThromboembolism	0.053	6	1.292	0.464	3	-2.848	5

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time.

Table OA28: Price Indexes and Changes in Welfare for Each Condition Between 2007 and 2018 - Marginal Cost is 20% of the Negotiated Price

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.002	16	1.013	0.999	0	0.941	0
AtrialFibrillation	0.461	16	2.977	0.136	14	-11.228	40
ColonCancer	-0.060	403	0.543	0.558	178	0.618	31
CysticFibrosis	0.233	718	4.349	4.316	-2,380	4.187	-457
HIV	0.188	348	1.505	1.451	-157	1.235	-16
HepatitisC	2.905	42	1.205	-5.693	282	-33.285	289
Hypertension	0.040	9	0.684	0.250	7	-1.487	5
LungCancer	0.654	281	1.899	1.666	-187	0.734	15
MultipleSclerosis	0.440	507	3.025	2.938	-982	2.590	-161
Osteoporosis	0.035	7	1.680	1.185	-1	-0.799	3
RheumatoidArthritis	0.195	198	1.992	1.894	-177	1.500	-20
Schizophrenia	0.120	39	0.826	0.514	19	-0.732	13
VenousThromboembolism	0.105	6	1.292	-0.364	9	-6.989	10

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8.

Table OA29: Price Indexes and Changes in Welfare for Each Condition Increasing Weighting for High Quality Studies

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.003	16	1.013	0.998	0	0.936	0
AtrialFibrillation	0.495	16	2.977	-0.074	17	-12.278	49
ColonCancer	-0.058	403	0.543	0.558	178	0.615	-6
CysticFibrosis	0.234	718	4.349	4.316	-2,380	4.186	23
HIV	0.194	348	1.505	1.450	-156	1.226	19
HepatitisC	2.895	42	1.205	-5.669	281	-33.164	289
Hypertension	0.040	9	0.684	0.250	7	-1.489	4
LungCancer	0.616	281	1.899	1.680	-191	0.802	62
MultipleSclerosis	0.425	507	3.025	2.941	-983	2.605	43
Osteoporosis	0.037	7	1.680	1.147	-1	-0.985	4
RheumatoidArthritis	0.150	198	1.992	1.917	-182	1.614	15
Schizophrenia	0.128	39	0.826	0.495	19	-0.827	13
VenousThromboembolism	0.110	6	1.292	-0.442	9	-7.378	11

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time.

Table OA30: Price Indexes and Changes in Welfare for Each Condition Using Tufts to Estimate Costs

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated Tufts Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.002	86	0.995	0.992	1	0.981	0
AtrialFibrillation	0.461	88	1.292	0.770	20	-1.319	46
ColonCancer	-0.060	257	0.891	0.915	22	1.009	-6
CysticFibrosis	0.233	1,478	2.279	2.264	-1,868	2.201	23
HIV	0.188	1,263	0.977	0.962	48	0.902	19
HepatitisC	2.905	150	1.240	-0.699	255	-8.453	291
Hypertension	0.040	102	1.044	1.005	0	0.849	4
LungCancer	0.654	380	1.166	0.994	2	0.306	65
MultipleSclerosis	0.440	1,760	1.027	1.002	-4	0.902	44
Osteoporosis	0.035	61	0.994	0.937	4	0.710	3
RheumatoidArthritis	0.195	618	1.022	0.991	6	0.864	20
Schizophrenia	0.120	971	0.999	0.987	13	0.937	12
VenousThromboembolism	0.105	33	0.972	0.651	11	-0.632	11

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time.

Table OA31: Counterfactual: Removing All New Drugs - Assuming $50k VSLY

	(1) Baseline Cost Growth 2018 - 2007 ($1,000s)	(2) Cost Growth due to Innovation ($1,000s)	(3) Share of Cost Growth due to Innovation	(4) Change in Consumer Welfare due to Innovation $50k VSLY ($1,000s)	(5) Change in Producer Surplus due to Innovation ($1,000s)	(6) Change in Total Welfare due to Innovation $50k VSLY ($1,000s)
Asthma	0	0	0.316	0	0	0
AtrialFibrillation	32	7	0.209	33	7	40
ColonCancer	-184	0	-0.000	0	0	0
CysticFibrosis	2,403	667	0.277	-657	667	9
HIV	176	89	0.505	-85	89	4
HepatitisC	9	9	1.000	137	9	145
Hypertension	-3	0	0.000	0	0	0
LungCancer	252	295	1.169	-263	295	33
MultipleSclerosis	1,026	204	0.199	-195	204	9
Osteoporosis	5	2	0.517	-1	2	1
RheumatoidArthritis	197	12	0.061	-9	12	3
Schizophrenia	-7	3	-0.421	-1	3	2
VenousThromboembolism	2	2	1.083	1	2	3
Aggregate	276	66	0.239	-60	66	6

Notes: Column 1 presents the cost growth we see without the counterfactual. This can be calculated as Column 2 multiplied by [Column (3) minus 1] in Table 4. Column 2 tells us the amount of cost growth due to innovation. This is calculated by determining the counterfactual where we replace all "new" drugs with "old" drugs in proportion to "old" drug market share in 2018. We then calculate the cost growth between 2007 and the 2018 counterfactual. Column 2 presents the difference between the cost growth we observe and this counterfactual. Column 3 then computes the share of cost growth that is due to innovation (column 2 divided by column 1). Column 4 presents the change in consumer welfare due to innovation. Column 5 presents producer surplus which is the same as column 2 as we assume marginal costs are constant. Column 6 presents the change in total welfare due to innovation, which is just $100k multiplied by the change in QALYs due to innovation (not shown). These numbers are similar to Table 4 because most of the quality improvements are due to innovation.

Table OA32: Counterfactual: Removing All New Drugs - Assuming $500k VSLY

	(1) Baseline Cost Growth 2018 - 2007 ($1,000s)	(2) Cost Growth due to Innovation ($1,000s)	(3) Share of Cost Growth due to Innovation	(4) Change in Consumer Welfare due to Innovation $500k VSLY ($1,000s)	(5) Change in Producer Surplus due to Innovation ($1,000s)	(6) Change in Total Welfare due to Innovation $500k VSLY ($1,000s)
Asthma	0	0	0.316	0	0	0
AtrialFibrillation	32	7	0.209	393	7	400
ColonCancer	-184	0	-0.000	0	0	0
CysticFibrosis	2,403	667	0.277	-573	667	94
HIV	176	89	0.505	-49	89	40
HepatitisC	9	9	1.000	1,445	9	1,453
Hypertension	-3	0	0.000	0	0	0
LungCancer	252	295	1.169	31	295	326
MultipleSclerosis	1,026	204	0.199	-114	204	90
Osteoporosis	5	2	0.517	12	2	15
RheumatoidArthritis	197	12	0.061	23	12	35
Schizophrenia	-7	3	-0.421	15	3	18
VenousThromboembolism	2	2	1.083	32	2	34
Aggregate	276	66	0.239	-6	66	60

Notes: Column 1 presents the cost growth we see without the counterfactual. This can be calculated as Column 2 multiplied by [Column (3) minus 1] in Table 4. Column 2 tells us the amount of cost growth due to innovation. This is calculated by determining the counterfactual where we replace all "new" drugs with "old" drugs in proportion to "old" drug market share in 2018. We then calculate the cost growth between 2007 and the 2018 counterfactual. Column 2 presents the difference between the cost growth we observe and this counterfactual. Column 3 then computes the share of cost growth that is due to innovation (column 2 divided by column 1). Column 4 presents the change in consumer welfare due to innovation. Column 5 presents producer surplus which is the same as column 2 as we assume marginal costs are constant. Column 6 presents the change in total welfare due to innovation, which is just $100k multiplied by the change in QALYs due to innovation (not shown). These numbers are similar to Table 4 because most of the quality improvements are due to innovation.

Table OA33: Counterfactual: Removing All New Drugs - Assuming That Marginal Costs are 20% of Negotiated Prices

	(1) Baseline Cost Growth 2018 - 2007 ($1,000s)	(2) Cost Growth due to Innovation ($1,000s)	(3) Share of Cost Growth due to Innovation	(4) Change in Consumer Welfare due to Innovation $100k VSLY ($1,000s)	(5) Change in Producer Surplus due to Innovation ($1,000s)	(6) Change in Total Welfare due to Innovation $100k VSLY ($1,000s)
Asthma	0	0	0.316	0	0	0
AtrialFibrillation	32	7	0.209	73	5	79
ColonCancer	-184	0	-0.000	0	0	0
CysticFibrosis	2,403	667	0.277	-648	533	-115
HIV	176	89	0.505	-81	71	-10
HepatitisC	9	9	1.000	282	7	289
Hypertension	-3	0	0.000	0	0	0
LungCancer	252	295	1.169	-230	236	6
MultipleSclerosis	1,026	204	0.199	-186	163	-23
Osteoporosis	5	2	0.517	0	2	2
RheumatoidArthritis	197	12	0.061	-5	10	5
Schizophrenia	-7	3	-0.421	1	2	3
VenousThromboembolism	2	2	1.083	5	2	6
Aggregate	276	66	0.239	-54	53	-1

Notes: Column 1 presents the cost growth we see without the counterfactual. This can be calculated as Column 2 multiplied by [Column (3) minus 1] in Table 4. Column 2 tells us the amount of cost growth due to innovation. This is calculated by determining the counterfactual where we replace all "new" drugs with "old" drugs in proportion to "old" drug market share in 2018. We then calculate the cost growth between 2007 and the 2018 counterfactual. Column 2 presents the difference between the cost growth we observe and this counterfactual. Column 3 then computes the share of cost growth that is due to innovation (column 2 divided by column 1). Unlike with constant marginal costs, column 2 no longer represents producer surplus in our framework, producer surplus is .8 multiplied by column 2, which shown in Column 5. Columns 4 and 6 present the change in consumer welfare and total welfare due to innovation.

OA.B.2 Additional Robustness Checks

Multiple Drug Classes — For each condition, in the main results we focus on the most prescribed class of treatments. However, for some conditions we observe are multiple classes of drugs. For example, for rheumatiod arthritis, there are disease-modifying antirheumatic drugs (DMARDs) and non-steroidal anti-inflammatory drugs (NSAIDs). Patients could take medicines in both classes (and often do), as they have distinct purposes. This complicates the regression methodology as not all treatments are directly or indirectly compared to each other. To handle this we create QALY estimates for each class, then weight across classes by quantity. Results are shown in Appendix Table OA34. Results are very similar.

Table OA34: Price Indexes and Changes in Welfare Using Multiple Groups of Treatment

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.001	16	1.013	1.009	0	0.991	0
AtrialFibrillation	0.393	16	2.977	0.555	7	-9.135	39
ColonCancer	-0.051	403	0.543	0.556	179	0.607	-5
CysticFibrosis	0.233	718	4.349	4.316	-2,380	4.187	23
HIV	0.194	348	1.505	1.450	-156	1.227	19
HepatitisC	2.905	42	1.205	-5.693	282	-33.285	291
Hypertension	0.016	9	0.684	0.507	5	-0.202	2
LungCancer	0.638	281	1.899	1.672	-189	0.764	64
MultipleSclerosis	0.440	507	3.025	2.938	-982	2.590	44
Osteoporosis	0.031	7	1.680	1.235	-2	-0.546	3
RheumatoidArthritis	0.106	198	1.992	1.939	-186	1.726	11
Schizophrenia	0.088	39	0.826	0.597	16	-0.317	9
VenousThromboembolism	0.105	6	1.292	-0.364	9	-6.989	11

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time.

Other Prescription Drug Spending — One challenge is that prescription drug claims do not include diagnosis codes. For our main results, we include inpatient and outpatient claims in baseline annual spending, as well as drugs classified in the Tufts. For this robustness check, we include all drug claims were we observe that drug having once been listed as a treatment for that condition in the MEPS (which have diagnosis codes on drug claims). Table OA35 presents results. We note that the MEPS has a lot of treatments which may not actually

mainly focus on the condition at hand, so this likely overstates costs for a given condition. For example, a hepatitis C patient with high cholesterol may be marked as taking a statin on a hepatitis C claim. Hence, we prefer the narrower version of treatments. The unadjusted prices are similar in ordering, but there are some differences in the magnitudes of unadjusted price changes, though they do not systematically overstate or understate cost growth. Quality adjustments are generally smaller as the level of spending with this measure is higher, but overall qualitative results are very similar.

Table OA35: Price Indexes and Changes in Welfare for Each Condition Using Additional Drug Claims

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.002	48	1.180	1.175	-8	1.156	0
AtrialFibrillation	0.461	42	2.016	0.914	4	-3.496	46
ColonCancer	-0.060	434	0.610	0.624	163	0.680	-6
CysticFibrosis	0.233	1,411	6.233	6.217	-7,360	6.151	23
HIV	0.188	885	1.566	1.544	-482	1.459	19
HepatitisC	2.905	76	1.185	-2.634	276	-17.914	291
Hypertension	0.040	48	1.112	1.029	-1	0.696	4
LungCancer	0.654	318	1.877	1.671	-213	0.848	65
MultipleSclerosis	0.440	880	2.814	2.764	-1,552	2.564	44
Osteoporosis	0.035	32	1.574	1.465	-15	1.032	3
RheumatoidArthritis	0.195	317	2.144	2.082	-343	1.836	20
Schizophrenia	0.120	82	0.800	0.653	28	0.068	12
VenousThromboembolism	0.105	24	1.333	0.888	3	-0.891	11

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time.

Alternative QALY Regression Estimates — We also estimate QALYs making different assumptions regarding how the regression is estimated. Tables OA36 and OA37 show results for QALY changes and price indexes assuming $100k VSLY. Note that these results don't change the unadjusted price index. Column 1 is the main results (Table 4). Column 2 does not normalize for heterogeneity in study assumptions, see Appendix Section OA.D.3

for more details of how we normalize for different assumptions that studies make. Column 3 makes the same heterogeneity adjustments as column 1, but does so on the raw QALYs, rather than adjusting the study fixed effects.

In our main specification, we drop studies which say they are a "placebo," "no treatment," "usual care," "standard of care," and "status quo." We do this because it is often unclear what these treatments are, and we worry that these categories will add a lot of noise or biases.[53] However, dropping these categories drops some studies. In our main regression there were 1,923 comparisons. Adding back placebo and no treatment increases that number to 2,179. Adding back standard of care, usual care, and status quo (on top of no treatment) increases that number to 2,210. Columns 4 and 5 present results with these comparisons added back in.

Column 6 estimates the QALY regressions in levels. Overall, results are quite similar across the board. There are some differences in magnitudes (with hepatitis C being a notable case). However, qualitative conclusions like all conditions having increasing quality, the rough ordering of quality changes, and the sign of quality adjusted price indexes are all consistent for all specifications.

[53]For example, there are cases where placebo makes little sense, like rheumatoid arthritis DMARDs compared against a placebo. Likewise, the standard of care can change. Some studies list what the standard of care is (and we classify those treatments), but we can see the standard of care differ across studies.

Table OA36: Change in QALYs with different Tufts regression specifications

	(1) Baseline Result	(2) Don't Normalize Heterogeneity	(3) Normalize Heterogeneity First	(4) Add No Treatment	(5) Add Standard of Care	(6) Estimate Regression in Levels
Asthma	0.002	0.006	0.000	0.003	0.003	0.000
AtrialFibrillation	0.461	0.383	0.516	0.524	0.537	0.453
ColonCancer	-0.060	-0.147	0.099	-0.067	-0.067	0.118
CysticFibrosis	0.233	0.137	0.128	0.253	0.254	0.189
HIV	0.188	0.091	0.106	0.128	0.132	0.152
HepatitisC	2.905	1.514	1.722	3.132	3.144	2.319
Hypertension	0.040	0.034	0.030	0.045	0.043	0.044
LungCancer	0.654	2.787	1.157	0.689	0.706	0.407
MultipleSclerosis	0.440	0.407	0.416	0.552	0.577	0.313
Osteoporosis	0.035	0.027	0.031	0.037	0.037	0.023
RheumatoidArthritis	0.195	0.229	1.356	0.221	0.223	0.165
Schizophrenia	0.120	0.084	0.007	0.135	0.138	0.008
VenousThromboembolism	0.105	0.081	0.064	0.119	0.122	0.047

Notes: This table presents estimated changes in average QALYs between 2007 and 2018 using different specifications in our regressions. Column 1 is the baseline result from Table 4. Column 2 does not normalize study heterogeneity. Column 3 normalizes study heterogeneity on the raw QALYs, rather than the study fixed effects. Columns 4 and 5 add additional studies which are less specific about the treatments in the regressions. Column 6 estimates the regressions in levels rather than logs.

Table OA37: $100k VSLY Price Index Results With Different Tufts Regression Specifications

	(1) Baseline Result	(2) Don't Normalize Heterogeneity	(3) Normalize Heterogeneity First	(4) Add No Treatment	(5) Add Standard of Care	(6) Estimate Regression in Levels
Asthma	0.999	0.976	1.012	0.993	0.993	1.012
AtrialFibrillation	0.136	0.617	-0.204	-0.251	-0.335	0.183
ColonCancer	0.558	0.580	0.519	0.560	0.560	0.514
CysticFibrosis	4.316	4.330	4.331	4.314	4.314	4.322
HIV	1.451	1.479	1.475	1.468	1.467	1.462
HepatitisC	-5.693	-2.390	-2.883	-6.232	-6.260	-4.301
Hypertension	0.250	0.320	0.354	0.199	0.213	0.202
LungCancer	1.666	0.906	1.487	1.654	1.648	1.754
MultipleSclerosis	2.938	2.944	2.942	2.916	2.911	2.963
Osteoporosis	1.185	1.300	1.231	1.143	1.144	1.353
RheumatoidArthritis	1.894	1.877	1.308	1.881	1.880	1.909
Schizophrenia	0.514	0.610	0.807	0.477	0.468	0.805
VenousThromboembolism	-0.364	0.022	0.288	-0.576	-0.623	0.551

Notes: This table presents estimated quality adjusted price indexes assuming a $100k VSLY using different specifications in our regressions. Column 1 is the baseline result from Table 4. Column 2 does not normalize study heterogeneity. Column 3 normalizes study heterogeneity on the raw QALYs, rather than the study fixed effects. Columns 4 and 5 add additional studies which are less specific about the treatments in the regressions. Column 6 estimates the regressions in levels rather than logs.

Heterogeneity in treatment effectiveness — One potential bias, noted by Lucarelli et al. (2022), is if there is heterogeneity in preferences, side effect profiles, or subpopulations where certain treatments work more than others. Our methodology, which compares average quality will understate the benefits if there is heterogeneity.[54] Handling unobserved heterogeneity is a ubiquitous challenge in economics, but especially so when distortions in the market preclude use of revealed preference methods which are typically used to handle unobserved

[54]Many cost-effectiveness studies do account for heterogeneity by focusing on specific subpopulations (i.e. estimating QALYs for populations with a certain stage of cancer or a specific genotype of Hepatitis C), but because we cannot always observe these subpopulations in either the Tufts or MarketScan data, we ignore this information. Therefore, our quality estimates may actually be capturing some of the benefits new drugs provide to heterogenous populations.

heterogeneity. To try to understand how sensitive our results are to heterogeneity, we take a few approaches. First, we add an unobserved idiosyncratic quality shock: For person i the quality of treatment r is: $H_{i,r,d} = \bar{H}_{r,d} + \epsilon_{i,r,d}$, where $\bar{H}_{r,d}$ is our estimate of QALYs in the Tufts. We assume that $\epsilon_{i,r,d}$ has a type 1 extreme value distribution because this distribution has closed form solution for the expected maximum. Then, following Small and Rosen (1981), the expected maximum quality drug for person i is:

$$H_{d,t} = log \sum_{r \in \mathcal{R}_{d,t}} exp(\bar{H}_{r,d}). \tag{A1}$$

Table OA38 shows the results from using Equation A1. Most conditions have higher estimated changes in QALYs, as new entrants unambiguously increase this estimate of consumer welfare. Across all conditions, The unweighted average difference in the change in QALYs across all conditions is .35 QALY, which is a sizeable difference. Colon cancer and hypertension have no change in QALYs as they do not have new entrants in our sample. Among the six conditions where consumer welfare was declining at \$100k VSLY in our main results that difference is .23 QALYs. Still five of these six conditions had declining consumer welfare (osteoporosis has rising consumer welfare with this number). However, at \$500k VSLY, rheumatoid arthritis and HIV have increasing consumer welfare with this measure.

Table OA38: Price Indexes and Changes in Welfare for Each Condition Allowing for Heterogeneity in Quality using Equation A1

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.266	16	1.013	-0.605	26	-7.077	27
AtrialFibrillation	1.784	16	2.977	-8.018	146	-51.997	178
ColonCancer	0.000	403	0.543	0.543	184	0.543	0
CysticFibrosis	0.471	718	4.349	4.283	-2,356	4.021	47
HIV	0.425	348	1.505	1.383	-133	0.894	43
HepatitisC	3.898	42	1.205	-8.052	381	-45.079	390
Hypertension	0.000	9	0.684	0.684	3	0.684	0
LungCancer	1.200	281	1.899	1.472	-132	-0.238	120
MultipleSclerosis	0.254	507	3.025	2.974	-1,001	2.774	25
Osteoporosis	0.304	7	1.680	-2.687	26	-20.159	30
RheumatoidArthritis	0.485	198	1.992	1.748	-148	0.768	49
Schizophrenia	0.151	39	0.826	0.434	22	-1.134	15
VenousThromboembolism	0.608	6	1.292	-8.257	59	-46.456	61

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time.

One issue with this approach to capture heterogeneity is that there is no weighting of the health variables across treatments in equation A1. Specifically, the quality of a drug that has a 90% share would be as important as a drug that has a quality of a 10% share in this specification. While the above specification still captures the expected maximum quality, one might be concerned that this measure does not capture diffusion well.[55] As another back-

[55]If one is willing to assume preferences are revealed by choices, then one could add unobserved quality

of-the-envelope way to address concerns of heterogeneity, we apply insights from Ackerberg and Rysman (2005) to derive a simple functional form for how heterogeneity may increase welfare given a logit functional form.[56] In a logit model where goods are of equal quality (so $\bar{H_{r,d}} = H$), then there is a simple formula for calculating the quality increase as a function of the number of goods:

$$\begin{aligned} H_{d,t} &= log(n_{d,t} \cdot exp(H)) \\ &= log(n_{d,t}) + log(exp(H)) \\ &= log(n_{d,t}) + H \end{aligned} \tag{A2}$$

Where $n_{d,t}$ is the number of drugs in the market in year t. This suggests that the quality change is a function of the number of goods in the market $log(n_{d,t})$, assuming this precise functional form. However, the value of the idiosyncratic error is unknown, so we assign a value that provides a reasonable (upper bound) scaling relative to our health measure.[57] Specifically, let R_d be the range of QALYs for disease d (i.e. it is the difference in QALYs between the highest and lowest QALY drug for that treatment in our sample, across all years). Let γ_d be a scaling factor such that $0.25 = (log(2) \cdot R_d/\gamma_d)$. That is, by setting γ_d we can impose the assumption, in this case, that going from one product to two increases the QALYs by 25% of the range for that condition. We would then allow for an adjusted QALY that adds in an additional QALY of the form $log(n_{d,t}) \cdot \gamma_d$. Specifically, the adjusted amount would be: $H^{alt}_{d,t} = H_{d,t} + log(n_{d,t}) \cdot \gamma_d$, where $H_{d,t}$ is that year's average quality as

that is drug specific to match market shares. However, this term might be picking up features like formulary design, physician learning, or other distortions which are not quality, but do effect choices.

[56] In the context of a choice model Ackerberg and Rysman (2005) they argue that the ϵs in a logit framework may overestimate of the value of heterogeneity from new products. Our conern for this robustenss check is the opposite, and we actually want to add back in this welfare from heterogeneity.

[57] In contrast to Ackerberg and Rysman (2005) attempting to make an adjustment to remove the idiosyncratic error, we are determining how much of the error to add in.

shown in our main tables. This functional form allows us to add additional QALYs for the entry of new drugs, but the QALY increase is dimimishing as more drugs enter the market.

Table OA39 shows the results from using Equation A2. All conditions, besides colon cancer and hypertension have larger QALYs, as all conditions except those two had entry of new treatments. Atrial fibrillation, hepatitis C, and lung cancer have particularly large increases in QALYs (compared to Table 4) as each of these conditions have considerable entry during this time period. Rheumatoid arthritis has a fairly large increase as well because its range of QALYs across conditions is quite large. Again, five conditions have falling consumer welfare assuming \$100k VSLY. At \$500k VSLY, cystic fibrosis and multiple sclerosis have falling consumer welfare. In summary, even with what we think are fairly generous adjustments to account for heterogeneity, we still find that some innovative conditions have declining consumer welfare during our sample period.

Table OA39: Price Indexes and Changes in Welfare for Each Condition Allowing for Heterogeneity in Quality using Equation A2.

	(1) Change in Avg QALYs 2018 - 2007	(2) Estimated MktScan Lifetime Costs in 2007 ($1,000s)	(3) Price Index $0 VSLY	(4) Price Index $100k VSLY	(5) Change in Consumer Welfare $100k VSLY ($1,000s)	(6) Price Index $500k VSLY	(7) Change in Total Welfare $100k VSLY ($1,000s)
Asthma	0.012	16	1.013	0.938	1	0.637	1
AtrialFibrillation	1.627	16	2.977	-7.052	131	-47.171	163
ColonCancer	-0.060	403	0.543	0.558	178	0.618	-6
CysticFibrosis	0.320	718	4.349	4.304	-2,371	4.126	32
HIV	0.354	348	1.505	1.404	-140	0.996	35
HepatitisC	5.087	42	1.205	-10.873	500	-59.184	509
Hypertension	0.040	9	0.684	0.250	7	-1.487	4
LungCancer	1.147	281	1.899	1.491	-138	-0.144	115
MultipleSclerosis	0.949	507	3.025	2.837	-931	2.088	95
Osteoporosis	0.048	7	1.680	0.987	0	-1.787	5
RheumatoidArthritis	0.550	198	1.992	1.715	-142	0.604	55
Schizophrenia	0.203	39	0.826	0.301	27	-1.798	20
VenousThromboembolism	0.174	6	1.292	-1.442	16	-12.381	17

Notes: This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the Tufts, MarketScan, and SSR health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in 2007 for each condition. Column 3 presents unadjusted price index, which is the percentage difference in costs between that year's cost and 2007's cost. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100k, and $500k, respectively. Columns 5 and 7 present changes in consumer and total welfare. All the estimates in columns 3-7 can be calculated directly using the results in columns 1 and 2 and using equations 3, 4 and 8 and assuming marginal costs are constant over time.

Online Appendix OA.C Additional Analyses

OA.C.1 What share of spending growth is due to within-molecule price changes?

In this section we explore what share of spending growth is due to within-molecule price changes. For this counterfactual, we replace the average price of a drug in 2018 with that drug's price in 2007, and recompute a new quantity weighted average cost for that condition in 2018. We leave non-drug spending at the 2018 level. For drugs that were not present in 2007, we leave them at their 2018 price. That is, for this counterfactual everything is the same as in the observed 2018 data, except price growth (or declines) for "old" drugs.

Table OA40 presents results. Column 1 presents the baseline price index (column 3 of Table 4). Column 2 presents the counterfactual index, where the numerator is the 2018 counterfactual without any within-molecule price growth, and the denominator is what we observe in 2007. For rheumatoid arthritis, if we remove all within-molecule inflation, costs would have *fallen* by 26% during our sample period, rather than grown by 99%. On the other hand, hypertension costs would have grown by 10%, rather than declining by 32% if 2007 prices remained constant. This is because prices declined after drugs came off patent for hypertension.

Table OA40: Counterfactual: Removing Within-Molecule Price Changes

	(1) Baseline Price Index $0 VSLY	(2) Counterfactual Price Index with Constant Prices	(3) Share of Cost Growth due to Changing Prices
Asthma	1.013	1.133	-9.050
AtrialFibrillation	2.977	3.009	-0.016
ColonCancer	0.543	0.696	0.334
CysticFibrosis	4.349	3.556	0.237
HIV	1.505	1.404	0.201
HepatitisC	1.205	1.205	0.000
Hypertension	0.684	1.099	-1.313
LungCancer	1.899	1.858	0.045
MultipleSclerosis	3.025	2.356	0.330
Osteoporosis	1.680	1.727	-0.069
RheumatoidArthritis	1.992	0.742	1.260
Schizophrenia	0.826	1.151	-1.868
VenousThromboembolism	1.292	1.628	-1.149
Aggregate	1.700	1.321	0.542

Notes: Column 1 is the baseline unadjusted price index, same as Table 4. Column 2 tells us what the price index would have in our counterfactual with no within-molecule price growth. To compute this counterfactual we replace all "old" drugs 2018 prices with their 2007 prices. We still use the 2018 market share and for "new" drugs we use the 2018 price. Column 3 then computes the share of cost growth that is due to within-molecule price changes. Some prices decline due to patent expiry, for these conditions the share is negative as within-molecule price changes reduce costs.

The third column shows the share of cost growth that is due to within-molecule inflation. Prices for HIV rise by 50% in our data. Our counterfactual suggests HIV would have grown by 40% in the absence of within-molecule price changes. Hence, within-molecule price changes only account for 20% of HIV cost growth.[58] In total, six conditions have negative shares,

[58]Like multiple sclerosis, many new HIV drugs raise their price after entry (see Figure OA14). As our

meaning that they would have been more expensive in the absence of within-molecule price growth. All of these conditions had drugs come off patent. Some of these conditions, like atrial fibrillation had very fast cost growth due to new entrants and within-molecule price declines because other drugs came off patent.[59]

Even though half of our conditions have within-molecule price declines, some of the costliest conditions in our sample have considerable within-molecule price changes. In the aggregate, 60% of the price growth that we see for these 13 conditions is due to within-molecule price growth. However, this finding is extremely sensitive to the inclusion of rheumatoid arthritis in our sample. We find about 15% of cost growth is due to within-molecule inflation when not including rheumatoid arthritis.

Online Appendix OA.D Data and Methods Appendix

OA.D.1 Cleaning and classifying the Tufts data

We chose the 13 conditions which were associated with the most studies in the Tufts and seemed appropriate for our analysis. To determine if a comparison is related to that condition, we search the disease or health intervention variable (this is the variable we use to classify conditions and has names like "hepatitis C" or "rheumatoid arthritis"), the ICD-10 code descriptors, ICD-10 chapter descriptors, and the study title for strings that match our condition names. Most of the observations are classified by the disease or health intervention variable. We intentionally excluded the abstract from the search variables because of its tendency to pick up the effect of treatment on comorbidities rather than the specific condition intended for the treatment (e.g., whether osteoporosis drugs lead to increased risk of acute

results are defining new entry based on a snapshot in time, this counts as cost due to innovation and not within-molecule price growth.

[59] Clopidogrel comes off patent in 2012 and its price drops dramatically (see Figure 6).

myocardial infarction).

The key variable in our data is the treatment variable. These are typically a sentence or two long. We hired multiple research assistants to classify each line in the Tufts data to a specific molecule. Each treatment was classified by two research assistants to ensure accuracy.[60] We ignore variations within a molecule like dose (5 mg vs. 10 mg), form (injectable vs. oral), frequency of treatment (once a day vs once a week) and treatment length (12 weeks vs. 24 weeks) as these can be tricky to map into claims data. Often pharmaceutical treatments are very vague, only listing a drug class (such as DMARDs for rheumatoid arthritis or direct acting antiretrovirals (DAAs) for HIV), these are marked as missing, since they are not specific enough to credibly map to the MarketScan data. Drugs which are given brand names are mapped back to molecule names (such as brand name "Sovaldi" mapped back to molecule "sofosbuvir").

We then merged the Tufts data with the MarketScan data by condition and molecule. To merge by condition we mapped the primary ICD-9 and ICD-10 codes from the Marketscan claims to Clinical Classification System (CCS) categories provided by the Agency for Healthcare and Research Quality (AHRQ) and matched these condition names to those in the Tufts. The only exceptions to using CCS categories were for hepatitis C and atrial fibrillation because the CCS categories for these conditions were too broad. Therefore, instead of using the broad CCS category for "hepatitis," we selected ICD-9 and ICD-10 codes specific to hepatitis C. See Table OA41 below for our mapping of CCS codes or ICD-9/ICD-10 codes to conditions.

To match by molecule we used the treatment names in the Tufts and searched the 2008, 2010, and 2012-17 REDBOOKs for all National Drug Codes (NDCs) associated with these names. Searching across multiple Redbooks ensures that we capture NDCs that enter and

[60] We also spent some time classifying procedures. However, procedure names in the Tufts are not standardized and often hard to match to CPT codes in claims data in an accurate way.

Table OA41: How We Define Conditions

Condition	CCS codes or ICD-9 codes
Asthma	CCS=128
Atrial Fibrillation	ATRFIB related ICDs; 4270, 42731, 42732, 42761, 42781, I480, I481, I482, I483, I484, I4891, I4892, I491
Colon Cancer	CCS=14
Cystic Fibrosis	CCS=56
Hepatitis C	Hep-C related ICDs; 07041, 07044, 07051, 07054, 07070, 07071, B1710, B1711, B1920, B1921, B182
HIV	CCS=5
Hypertension	CCS=98 and 99
Lung Cancer	CCS=19
Multiple Sclerosis	CCS=80
Osteoporosis	CCS=206
Rheumatoid Arthritis	CCS=202
Schizophrenia	CCS=659
Venous Thromboembolism	CCS=118

Notes: This table presents how we map from CCS codes (or ICD codes) to conditions in the MarketScan data.

exit over time. Likewise, we search the HCPCS-NDC crosswalk for all the HCPCS codes associated with a treatment name.[61] Ultimately, this step ensures that we map Tufts treatments to NDC and HCPCS codes.

The Tufts data often compare combinations of molecules with other combinations of molecules. In these cases, we view the Tufts quality estimates as being valid for the combinations, so we use the same combinations in the MarketScan data. We used the Tufts data to identify molecules which patients might take in combination. Once the sample of molecules and combinations of molecules is classified in the Tufts, we search for each patient's condition-specific combinations in the MarketScan data. We look at all drugs that patient took in a given year and create combinations based on what they are observed to take. For example, if a hepatitis C diagnosed patient is observed to have taken Ribavirin, Simeprevir, and Sofosbuvir in 2017, then we identify the following seven treatment possibilities: Ribavirin, Simeprevir, Sofosbuvir, Ribavirin/Simeprevir, Ribavirin/Sofosbuvir, Simeprevir/Sofosbuvir, and Ribavirin/Simeprevir/Sofosbuvir. Among the possibilities, we assign this patient to the maximum combination validated by the Tufts data.[62]

OA.D.2 Tufts Coverage of Spending

To check how well the Tufts data covers the most important treatments, we examine the share of spending we observe in the MEPS data.[63] The MEPS data are useful for this exercise because there are diagnosis codes on pharmaceuticals claims, allowing us to determine the share of MEPS spending we observe in the Tufts. However, the MEPS data do not include

[61] We downloaded the HCPCS-NDC crosswalk from (`https://www.dmepdac.com/palmetto/PDAC.nsf/DID/HJZNZ8E5WD`). It is a simple crosswalk used to identify the HCPCS codes associated with a drug molecule.

[62] While rare, there may be cases, like this example where there are ties in the highest hierarchy. In these cases, a patient can be assigned to multiple combinations Ribavirin/Simeprevir, Ribavirin/Sofosbuvir, and Simeprevir/Sofosbuvir. This does not affect our average cost per year estimates, but impacts our market share calculations.

[63] For this analysis we combine all years of MEPS data from 2007-2017.

5-digit CPT codes, which limits our ability to measure physician administered drugs. The MEPS also masks some NDCs for expensive drugs for confidentiality reasons, so high cost drugs like Sovaldi are not in the MEPS data. This will bias our results towards zero. We also do not include cystic fibrosis in this analysis, as MEPS masks cystic fibrosis in the data after 2009, again due to confidentiality reasons.

Table OA42 provides evidence of how much spending we can classify. The first column shows the percentage of total spending, in the MEPS, that is pharmaceutical spending for a condition. For example, 72% of hepatitis C spending is associated with pharmaceuticals (and in the drug files), though this misses some high cost hepatitis C drugs like Sovaldi. Non-pharmaceutical spending includes physician visits, screenings, and diagnostic imaging. As this spending is counted in costs, we are assuming there is no quality improvement for it, which would bias our QALY change results towards zero. Another example of this is hypertension, which is mostly treated with pharmaceuticals, but we are picking up a lot of doctor's visits where hypertension is the first listed diagnosis. For atrial fibrillation there is considerable spending on ablation procedures. For venous thromboembolism inferior vena cava filters and thrombectomy/embolectomy are important treatments for some patients. For the remainder of conditions we consider, at least 60% of costs are pharmaceuticals.

The second column shows the share of total drug spending in the MEPS data that is captured by the Tufts data. The MEPS data often contain more classes of drugs that treat a condition, for example painkillers or anti-nausea medication, which are symptom aids that treat many conditions. In addition, comorbidities can inflate spending. For example, if a patient has high cholesterol and hepatitis, we may see statins in their hepatitis C claims. To better understand how much coverage we have for each condition, column 3 limits to just drugs that have at least 5% market share over the sample period, which drops many of these other drugs. In column 3, we see that we capture at least 79% of spending on drugs that have at least 5% market share for all conditions except atrial fibrillation (63 percent).

Table OA42: Share of MEPS Spending we Classify

	(1) % of All Spending On RX for Condition	(2) % of RX Spending We Categorize in Tufts	(3) % of RX Spending in Tufts on 5% Market Share Drugs
Asthma	68	69	80
AtrialFibrillation	9	47	63
HIV	73	89	100
HepatitisC	72	91	100
Hypertension	44	41	79
MultipleSclerosis	69	91	100
Osteoporosis	75	84	93
RheumatoidArthritis	66	84	100
Schizophrenia	61	89	100
VenousThromboembolism	15	78	84

Notes: This table presents results for how much drug spending in the MEPS we classify in the Tufts. We use the MEPS 2007-2017 data for this analysis. Column 1 presents the share of all spending in the MEPS is pharmaceutical spending. Column 2 is the amount of all pharmaceutical spending we classify in the Tufts. Some drugs are not in the MEPS as rare/expensive drugs have masked NDC codes. Also, drugs administered by physicians are not included. Both of these will bias are results towards zero. Column 3 is the same as column 2, but only keeps drugs that have 5% market share over the sample period. Cystic fibrosis is not included because MEPS masks that condition to protect anonymity.

OA.D.3 Accounting for Heterogeneity in Cost and Quality

Studies in the Tufts data often make various assumptions in calculating their costs and QALYs. A study may vary in the discount rate used, the time horizon considered, country of interest, etc. In our analysis we include comparison fixed effects which difference out these factors (equation 6). However, because our results are retransformed including the common effect of a comparison, $\gamma_{u,d}$, we standardize the study common effect based on characteristics of each study. To do this, we regress our estimate of each comparison's $\gamma_{u,d}$

on the characteristics of the study and predict what the study common effect would have been under consistent assumptions. For this regression, the unit of observation is a comparison. Our regression equation is:

$$\begin{aligned}\hat{\gamma_{u,d}} = \beta_0 &+ \beta_1 \mathbb{1}(\text{Study uses lifetime time horizon}_i) + \beta_2 \text{time horizon}_i + \beta_3 \text{time horizon}_i^2 \\ &+ \beta_4 \text{time horizon}_i^3 + \beta_5 \mathbb{1}(\text{Study discounts the future}_i) + \beta_6 \text{Discount rate}_i \\ &+ \gamma_g + \gamma_a + \gamma_r + \gamma_c \times \mathbb{1}(\text{Treatment is placebo}_i) + \epsilon_i\end{aligned} \tag{A3}$$

Where $\hat{\gamma_{u,d}}$ is the estimate of the comparison fixed effect from equation 6. $\gamma_g, \gamma_r, \gamma_c$ are gender, country, and condition fixed effects, respectively. Studies also include indicators for the age groups included (i.e. 0-18, 19-40, etc.). If the study includes multiple age groups we divide this indicator by the number of age groups included to get a share of the age groups.[64]

Table OA43 presents results from this regression using out baseline sample and assumptions. The columns vary by the variables included. Column (1) does not include the discount rate variables, country fixed effects or condition fixed effects. Column (2) adds in country fixed effects, column (3) adds country and conditions fixed effects. Column (4) includes the discount rate and an indicator for whether there is time discounting. Results are consistent across columns and the signs of coefficients are as expected. Studies with older populations generally have lower QALYs, likely because these populations have less time for the intervention to impact their patients. We see that studies with a lifetime time horizon have higher QALYs. For those that do not, longer time horizons are associated with higher QALYs. We use column (4) as our preferred specification.

[64]For example, if a study has the indicators for 0-18 and 19-40, then we assign 0.5 for each of those variables, rather than 1 for each indicator. Results do not change much if we use indicators rather than shares.

Table OA43: QALY heterogeneity regressions

	(1)	(2)	(3)	(4)
Share 0-18 years old	0.0453	-0.729	-1.283**	-1.244**
	(0.13)	(-1.39)	(-2.66)	(-2.97)
Share 19-40 years old	0.146	0.0394	-0.0979	0.0747
	(0.79)	(0.22)	(-0.78)	(0.59)
Share 41-64 years old	0.203*	0.185*	0.0462	-0.0350
	(2.15)	(2.00)	(0.69)	(-0.59)
Share 65+	-0.520***	-0.677***	-0.404***	-0.391***
	(-6.31)	(-8.16)	(-5.75)	(-6.36)
Male	-0.402**	-0.265	-0.319**	-0.259*
	(-2.61)	(-1.67)	(-2.71)	(-2.52)
Both Genders	-0.266*	-0.135	-0.148	-0.192
	(-2.32)	(-1.11)	(-1.30)	(-1.94)
Not Specified Gender	-0.414***	-0.320**	-0.209	-0.232*
	(-3.53)	(-2.63)	(-1.82)	(-2.32)
Lifetime Time Horizon	2.955***	2.924***	2.075***	1.353***
	(34.58)	(34.10)	(23.87)	(15.74)
Time Horizon	0.208***	0.215***	0.196***	0.105***
	(17.60)	(17.85)	(19.54)	(10.21)
Time Horizon^2	-0.00409***	-0.00448***	-0.00437***	-0.00194***
	(-9.73)	(-10.91)	(-14.15)	(-6.37)
Time Horizon^3	0.0000244***	0.0000276***	0.0000274***	0.0000107***
	(6.72)	(7.95)	(10.90)	(4.47)
Discount Rate				-0.0664*
				(-2.20)
Constant	-1.905***	-1.874***	-3.199***	-2.396***
	(-13.71)	(-12.31)	(-18.08)	(-11.31)
Country Fixed Effects	No	Yes	Yes	Yes
Condition Fixed Effects	No	No	Yes	Yes
Observations	1057	1057	1057	1057

Notes: This table presents results from different regression specifications from Equation A3. Columns vary by the set of variables included. We use Column 4 as our preferred specification in all other tables.

Using these results, we predict study common effects using standardized assumptions. For the country-specific dummy, we standardize values to the United States. We also specify that the time horizon is a "lifetime" and set the discount rate to be 3 percent. As the demographics change across conditions, we set the demographic variables (age group share and gender indicators) to the mean for that condition in the Tufts data.[65] Tables OA36 and OA37 present QALY estimates and quality adjusted price indexes which check robustness to other assumptions. In particular, we don't adjust for heterogenity, we adjust the raw QALY (rather than the study fixed effects) for heterogeneity, we add in some additional studies with less precise treatment names, and run the regression in levels. Results are qualitatively similar regardless of specification.

OA.D.4 Lifetime Costs and Annual Scaling Factor

To calculate lifetime costs we re-scale annual estimates using a scaling factor. In this section we describe how the scaling factor is determined and how it relates to lifetime costs.

We take into account four factors when calculating the scaling factor: time discounting, the probability of dying, the age distribution for condition d, and how costs progress for an individual. Consider a person at age a. Each year s into the future they have the probability of dying $l_{a,s}$, and if they are alive they have expected costs $C^{\hat{p}}_{s,d}$. Formally, we calculate the estimated lifetime cost for this individual as:

$$LC^{p}_{a,d} = \sum_{s=0}^{100} (1-\rho)^{s} \cdot l_{a,s} \cdot C^{\hat{p}}_{s,d} \tag{A4}$$

where ρ is the interest rate. To be consistent with our standardized QALY estimates, we

[65] The one exception is we set the share over 65 to be equal to zero to be consistent with the MarketScan data. Results do not change much when we leave the share over 65 as its average value in the Tufts data.

assume ρ is 0.03. $l_{a,s}$ is the probability of someone age a dying in s periods into the future, which is calculated using the life tables. We then weight across individuals with treatments using the disease specific distribution of ages in the MarketScan data, $p_{a,d}$.

$$LC_d^p = \sum_{a=0}^{100} p_{a,d} \sum_{s=0}^{100} (1-\rho)^s \cdot l_{a,s} \cdot C_{s,d}^{\hat{p}} \tag{A5}$$

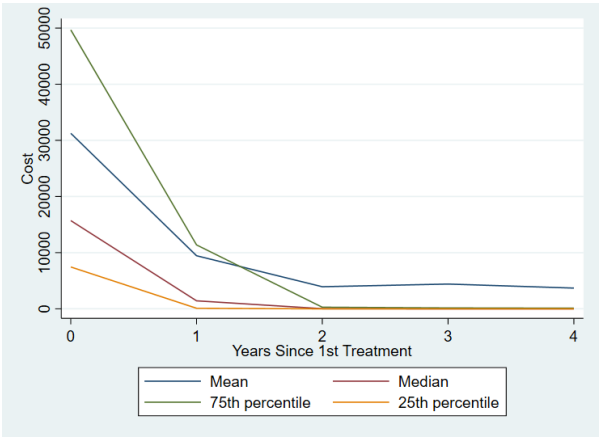
$C_{s,d}^{\hat{p}}$ measures how costs change after an individual with disease d receives treatment. As our goal is to measure lifetime costs, we want to understand how persistent costs are. For example, some treatments might be one-time costs while other treatments may be taken indefinitely. To measure this cost progression, we construct a sample of individuals who are enrolled for four consecutive years after their first treatment and one year prior to treatment (to ensure this is a patient's first treatment). We added the superscript p to indicate this is for our panel of individuals.

Figure OA37 shows how costs evolve for hepatitis C, hypertension, multiple sclerosis, and rheumatoid arthritis. For hepatitis C, in the first year of treatment, the average cost is \$45,000, while in year 3 the average cost is closer to \$5,000. The steep decline in costs after one year of treatment is because the treatments for hepatitis C are typically taken in one course, rather than indefinitely. While our annual cost of treatment measure, used in the rest of the paper focuses on individuals with treatment, our panel measure picks up individuals who are enrolled (which is the condition inclusion in the panel), but may not receive any treatment.

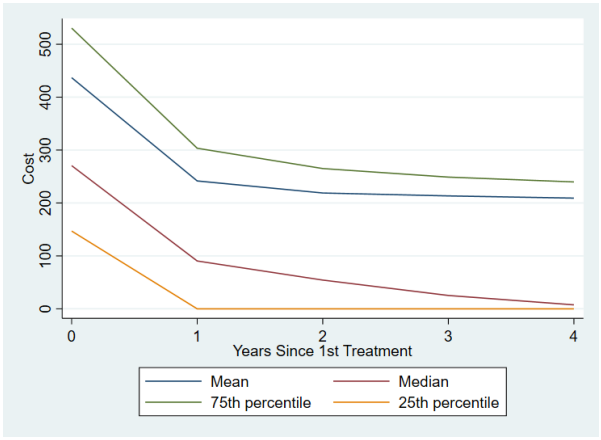
Hypertension has costs which decline from \$600 in year 1 to \$225 in year 3 and 4. Part of this is that in the first year of treatment people may be receiving some additional doctors or diagnostic visits that are not present in years 3 and 4 once their treatment stabilizes. Therefore, the cost progression captures one expensive year and additional moderately expensive years.

For multiple sclerosis and rheumatoid arthritis, treatments are taken indefinitely, so costs do not necessarily decline over time. The increasing slope for these conditions includes the fact that treatments are getting more expensive over time, which is handled by including year fixed effects in our regressions described below.

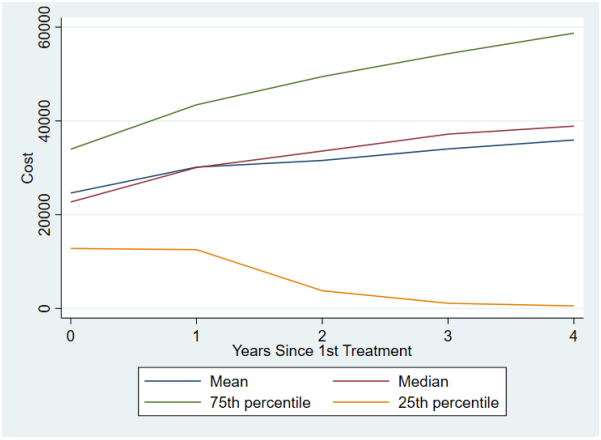
Figure OA37: Cost Progression for Selected Conditions



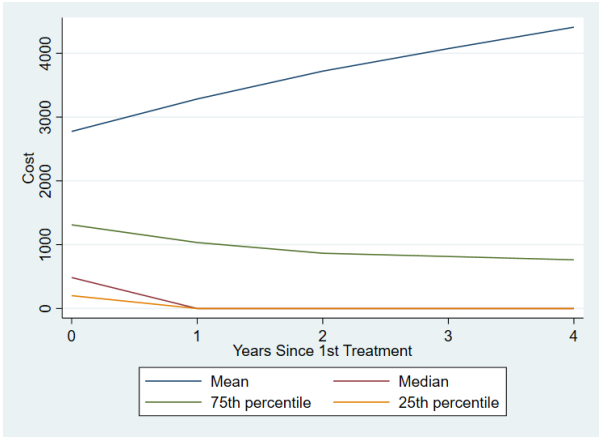
Hepatitis C



Hypertension



Multiple Sclerosis



Rheumatoid Arthritis

Notes: This figure presents the cost progression for an individual with a treatment for the noted disease. Each year is just the sample mean (or sample percentile) of spending for someone X years from their first treatment year. Everyone gets the treatment in year 0. We follow patients for four additional years and take the average of their spending in each year, including patients with no spending.

To approximate this cost progression and extrapolate out over 100 years, we regress costs

on years since first diagnosis fixed effects, up to four years, and calendar-year fixed effects using GLM with a log link.[66] We include calendar-year fixed effects because services are getting more expensive over time which inflates the slopes in Figure OA37. After fitting this regression, we predict costs for each year of having the condition using 2007 as the base year.[67] We then plug these estimates into Equations A4 and A5, to get the lifetime cost estimates for our panel of individuals using 2007 as the base, $LC^{p}_{d,2007}$.

There are two reasons why this lifetime cost estimate differs from the estimates we need. First, costs for individual treatments change over time, whereas this lifetime cost estimate fixes costs in 2007. Second, we need to observe people for a few years to understand how costs evolve, but people who are continuously enrolled for six years or had a year without treatment may have different costs than the average treated person with condition d.

To address the first concern, we multiply $LC^{p}_{d,2007}$ by $\frac{\bar{C}_{d,t}}{\bar{C}_{d,2007}}$, where $\bar{C}_{d,t}$ is just the average spending on disease d in year t. This captures how annual spending evolves over time for the average treated person. For the second issue, we multiply by $\frac{\bar{C}_{d,2007}}{\hat{C}^{p}_{1,d,2007}}$. Where $\hat{C}^{p}_{1,d,2007}$ is the predicted average cost, conditional on treatment, for someone in 2007 who fits our six years of continuous enrollment criteria. This adjusts for the sample selection in using people enrolled for multiple years. That is, our cost estimates are:

$$LifetimeCost_{d,t} = LC^{p}_{d,2007}\frac{\bar{C}_{d,t}}{\bar{C}_{d,2007}}\frac{\bar{C}_{d,2007}}{\hat{C}^{p}{}_{1,d,2007}} = LC^{p}_{d,2007}\frac{\bar{C}_{d,t}}{\hat{C}^{p}{}_{1,d,2007}} \tag{A6}$$

Therefore, throughout the draft we compute $\bar{C}_{d,t}$ and multiply it by our cost multiplier:

[66] We use GLM as it has a better fit than log OLS and retransformation using the smearing estimator in Duan (1983).

[67] In our preferred specification with years since first diagnosis fixed effects, we assume that the year 4 costs remain constant for 96 more years, reflecting a stabilizing in costs. However, we also estimate regressions using a linear trend in years since first treatment. However, this linear trend often goes to zero, which we think understates the persistence of costs.

$$\alpha_d = \frac{LC^p_{d,2007}}{\hat{C}^p{}_{1,d,2007}} \tag{A7}$$

Table OA44 presents the lifetime cost multiplier we estimate for each condition. Column 4 is the version without accounting for the cost slopes (assuming costs are constant). With constant costs, the life tables and time discounting suggest a lifetime cost multiplier of 23-27. The first column is our preferred specification, which assumes that 4th year costs continue indefinitely. For a condition like hepatitis C, our preferred cost multiplier is 3.8. This cost multiplier is much smaller because people mostly only have one expensive year of treatment (i.e. you take one course of Sovaldi). For conditions like rheumatoid arthritis and multiple sclerosis where people continue taking treatments indefinitely, costs are similar to the version without accounting for the cost slope.

Estimated lifetime costs for asthma and hypertension are about half of what they are in the last column, which just takes into account life expectancy. For these conditions, we see some lumpy costs, like doctor's visits and diagnostic tests, which are not paid every year. Likewise, we see that some people stop taking their medications. The annual costs we compute $\bar{C}_{d,t}$ are conditional on having a doctor's visit with an associated diagnosis code and having a treatment, so it likely captures years that are more expensive than the average year. Our lifetime cost estimates, with the panel, accounts for this lumpiness. For these two conditions these cost estimates are telling us an average year is about half as expensive as a year where we observe doctor's visits.

Table OA44: Lifetime estimate cost multipliers for each condition

	(1) Preferred Specification	(2) Uses Years Since Trend Line	(3) No Untreated Prior Year Needed	(4) Constant Costs No Slope
Asthma	12.542	3.646	16.179	27.859
AtrialFibrillation	6.749	2.508	7.473	24.016
ColonCancer	8.254	2.291	8.262	23.735
CysticFibrosis	27.591	20.605	26.802	28.913
HIV	25.959	23.246	24.040	25.905
HepatitisC	3.802	1.551	3.656	24.318
Hypertension	13.507	5.106	15.844	24.296
LungCancer	7.983	3.200	8.941	23.324
MultipleSclerosis	24.307	17.335	21.802	25.422
Osteoporosis	7.293	2.867	7.408	23.249
RheumatoidArthritis	27.337	36.020	29.746	24.872
Schizophrenia	9.335	2.622	14.804	26.703
VenousThromboembolism	3.799	1.473	4.461	24.823

Notes: This table presents lifetime estimate cost multipliers for each condition. All columns account for the age distribution of a condition, life expectancy, and the discount when calculating lifetime costs. Columns 1-3 account for the idea that when someone has treatment in one year that their future costs may not remain constant. Column 1 keeps costs constant at their year four level. Column 2 uses a log-linear trend to predict costs. Column 3 does not condition on having a year without spending prior to the first year of treatment. Column 4 holds treatment costs constant.

The other columns test the robustness of the assumptions we make. Column (2) uses a linear trend for years since treatment rather than assuming the 4th year remains constant. This ultimately predicts costs trend to zero for most conditions, which we think understates the persistence of costs and is why the results in Column (2) are much lower than Column (1).

Column (3) drops the requirement that we observe one year without diagnosis prior to the first year. This increases the multiplier estimates because we have more people who are in the constant cost stage of their treatment, reducing the steepness of the slopes in Figure OA37. Results are fairly similar, suggesting that conditioning on having no spending in the

prior year does not impact results much.

Tables OA45 and OA46 show the $100k and $250 VSLY estimates using all of these specifications and annual costs (assuming the multiplier is 1). With annual costs prices are falling much more quickly than with constant costs. This is a very wide range of assumed values, results are different and should be viewed as very wide bounds on our central estimates.

Table OA45: Price Indexes for Each Condition Using Different Lifetime Cost Assumptions - $100k VSLY

	(1) Preferred Specification	(2) Uses Years Since Trend Line	(3) No Untreated Prior Year Needed	(4) Constant Costs No Slope	(5) Annual Costs
Asthma	0.999	0.964	1.002	1.007	0.833
AtrialFibrillation	0.136	-4.668	0.411	2.179	-16.197
ColonCancer	0.558	0.597	0.558	0.548	0.667
CysticFibrosis	4.316	4.305	4.315	4.318	3.453
HIV	1.451	1.445	1.447	1.451	0.101
HepatitisC	-5.693	-15.702	-5.969	0.127	-25.022
Hypertension	0.250	-0.465	0.314	0.443	-5.182
LungCancer	1.666	1.318	1.691	1.820	0.039
MultipleSclerosis	2.938	2.903	2.928	2.941	0.914
Osteoporosis	1.185	0.419	1.192	1.525	-1.936
RheumatoidArthritis	1.894	1.918	1.902	1.884	-0.699
Schizophrenia	0.514	-0.283	0.630	0.717	-2.083
VenousThromboembolism	-0.364	-2.980	-0.118	1.039	-5.000

Notes: This table presents our quality adjusted price indexes, constructed using the Tufts, MarketScan, and SSR health datasets. Each column assumes that the VSLY is $100k. All columns (except the last) account for the age distribution of a condition, life expectancy, and the discount when calculating lifetime costs. The first three columns account for the idea that when someone has treatment in one year that their future costs may not remain constant. The first column keeps costs constant at their year four level. The second column uses a log-linear trend to predict costs. The third column does not condition on having a year without spending prior to the first year of treatment. The fourth column holds treatment costs constant. Column 5 assumes all costs are in one year, which is clearly unrealistic, but is a clear lower bound.

Table OA46: Price Indexes for Each Condition Using Different Lifetime Cost Assumptions - $500k VSLY

	(1) Preferred Specification	(2) Uses Years Since Trend Line	(3) No Untreated Prior Year Needed	(4) Constant Costs No Slope	(5) Annual Costs
Asthma	0.941	0.766	0.957	0.981	0.111
AtrialFibrillation	-11.228	-35.248	-9.851	-1.015	-92.893
ColonCancer	0.618	0.813	0.618	0.569	1.161
CysticFibrosis	4.187	4.131	4.182	4.194	-0.130
HIV	1.235	1.203	1.213	1.234	-5.514
HepatitisC	-33.285	-83.333	-34.665	-4.187	-129.932
Hypertension	-1.487	-5.060	-1.167	-0.523	-28.648
LungCancer	0.734	-1.007	0.859	1.501	-7.400
MultipleSclerosis	2.590	2.416	2.541	2.609	-7.528
Osteoporosis	-0.799	-4.627	-0.760	0.903	-16.401
RheumatoidArthritis	1.500	1.619	1.540	1.451	-11.462
Schizophrenia	-0.732	-4.721	-0.156	0.281	-13.717
VenousThromboembolism	-6.989	-20.068	-5.759	0.025	-30.169

Notes: This table presents our quality adjusted price indexes, constructed using the Tufts, MarketScan, and SSR health datasets. Each column assumes that the VSLY is $500k. All columns (except the last) account for the age distribution of a condition, life expectancy, and the discount when calculating lifetime costs. The first three columns account for the idea that when someone has treatment in one year that their future costs may not remain constant. The first column keeps costs constant at their year four level. The second column uses a log-linear trend to predict costs. The third column does not condition on having a year without spending prior to the first year of treatment. The fourth column holds treatment costs constant. Column 5 assumes all costs are in one year, which is clearly unrealistic, but is a clear lower bound.

OA.D.5 Rebate Adjustment

To account for manufacturer rebates, we supplement the MarketScan data with data from SSR Health Data. SSR Health, LLC collects data from drug manufacturer SEC filings on revenue net of rebates. They combine the revenue measure with units sold collected by Symphony Health. They then divide net revenues by units sold to estimate a price

net of rebate. They also include the wholesale acquisition cost (WAC), an estimate of the manufacturer's list price. At the brand level, we adjust the level of spending in the MarketScan data by multiplying the payment amounts in the MarketScan data by the ratio of the actual revenue divided by the list price, which removes the rebate amount from our cost estimate.

We aggregate the SSR Health data to the brand-year level. Our SSR Health data includes 1,057 different drugs. To apply this in the MarketScan data, we compute one minus the ratio of net prices to list prices (NET/WAC) which we interpret as the share of revenue which is paid in rebates. Then, at the molecule level, we adjust the level of spending in the MarketScan data by multiplying the payment amounts in the MarketScan data by the NET/WAC ratio for each drug we observe in the SSR Health data. If a molecule is missing in the SSR Health data, which is common (e.g., for most generics), we assume there is no rebate.

Online Appendix OA.E Aggregate Measures

In this section, we aggregate across conditions. However, we caution that we make no claims that these conditions are representative. Indeed, the amount of heterogeneity across conditions suggests that conditions outside of our sample may have very different trends.

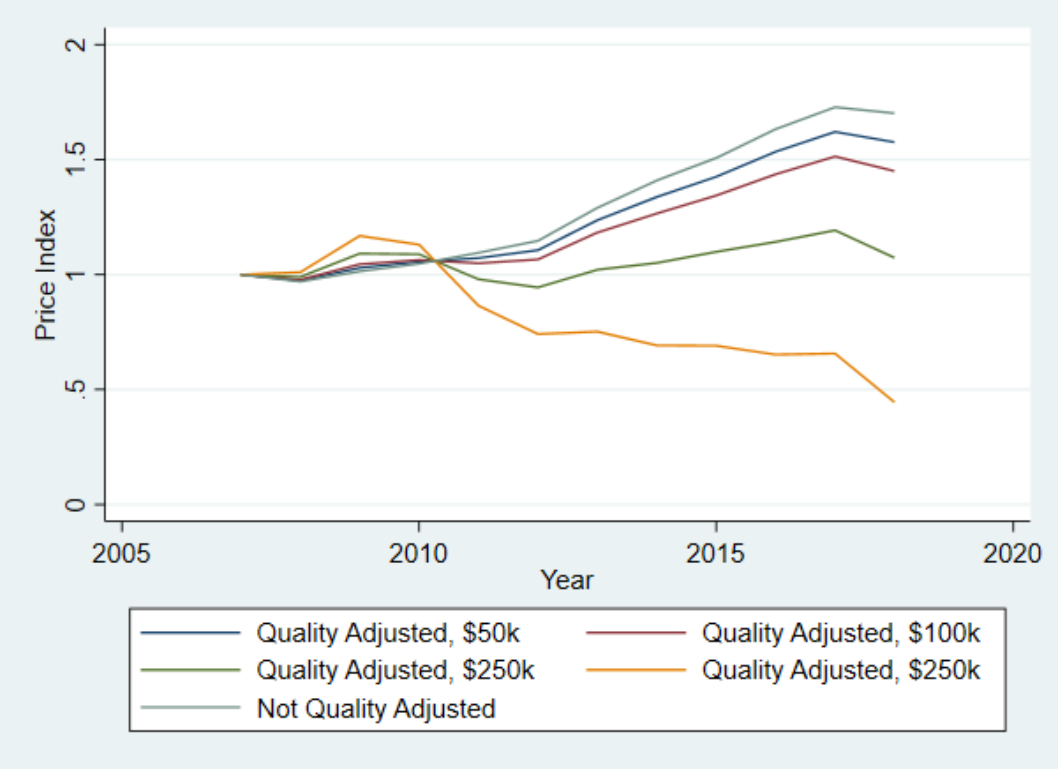
Table OA47 and Figure OA38 present results where we aggregate across all the conditions, weighting by their spending per capita in 2007. The unadjusted price index for these conditions rose by 70% in our sample period. However, quality adjustment reduced this to about 45% assuming a $100k VSLY. At $500k VSLY, quality adjusted prices fell by 55%.

Table OA47: Aggregate Results - Price indexes

	(1) Price Index $0 VSLY	(2) Price Index $100k VSLY	(3) Price Index $500k VSLY
2007	1.000	1.000	1.000
2008	0.970	0.978	1.010
2009	1.014	1.045	1.169
2010	1.047	1.064	1.130
2011	1.095	1.049	0.865
2012	1.147	1.066	0.742
2013	1.291	1.183	0.752
2014	1.409	1.266	0.692
2015	1.507	1.344	0.691
2016	1.633	1.436	0.652
2017	1.728	1.514	0.657
2018	1.702	1.450	0.445

Notes: This table presents quality adjusted price indexes which take the spending weighted average across all 13 conditions we consider. The columns vary by the VSLY assumptions we make. The price indexes are also graphed in Figure OA38. These results are constructed using data from Tufts, MarketScan, and SSR Health.

Figure OA38: Aggregate Price Indexes



Notes: This figure presents quality adjusted price indexes weighted across conditions by spending. These indexes are also shown in Table OA47. These results are constructed using data from Tufts, MarketScan, and SSR Health.