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

Abe Dunn, Lasanthi Fernando, and Eli Liebman*†

November 25, 2023

**Abstract**

Medical innovation is a key driver of cost growth and improved life expectancy, but measuring the welfare contribution of innovations is challenging. We leverage thousands of medical studies to estimate the quality of treatments for 13 health conditions and combine these estimates with insurance claims data to quantify how innovations diffuse and their impact on costs and quality. Across nearly all conditions we study, we find higher quality innovations diffuse. Like markets outside of healthcare, we find innovations can improve consumer welfare substantially. However, we also observe a phenomena arguably unique to healthcare, cases where innovation reduces consumer welfare.

*JEL*: E31,O30, I10

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

*Abe Dunn and Lasanthi Fernando: Bureau of Economic Analysis. Eli Liebman: John Munro Godfrey, Sr. 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 Chris Adams, Michael Chernew, David Cutler, Amy Finkelstein, Qing Gong, Scott Grosse, Claudio Lucarelli, Joe Newhouse, Sean Nicholson, Daria Pelech, Julian Reif, David Ridley, Louise Sheiner, Meghan Skira, Keith Teltser, Marcia Weaver, Brett Wendling, and participants in ASSA, Atlanta/Athens Health Economics Research Conference, IHEA, NBER Summer Institute, and SEA conference sessions and seminar participants at CBO and IHME for helpful comments. We would also like to thank Dr. Edward Gorak, D.O., MBA, MS, FACP and Dr. Joseph Nichols, M.D. for helpful review and discussion. 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 a major contributing factor to cost growth (Chernew and Newhouse, 2011). There is also a substantial literature arguing that the prices of new healthcare innovations is excessive, especially in the U.S.[1] At the heart of the debate is a measurement issue; it is difficult to attribute changes in costs and health outcomes to specific innovations.

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 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. We use the Tufts Cost-Effectiveness Analysis Registry (CEAR) database of over 8,000 cost effectiveness studies to estimate the quality of specific treatments. The purpose of a cost-effectiveness study 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 more generally. However, this requires synthesizing the results of thousands of cost-effectiveness studies which we do by: (1) classifying the treatments in each study so they can be matched across studies;[3] (2)

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

[2]Pakes (2003) discusses revealed preference approaches. The distortions in health care markets which complicate revealed preference approaches 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 (Dauda et al., 2022).

[3]The treatments are recorded as text and rarely match across studies.

developing a methodology to estimate average quality for each treatment; and (3) matching these treatments with the medical claims data of millions of commercially-insured individuals to measure how these treatments diffuse. 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 and quality-adjusted price indexes derived in Cutler et al. (1998). In their framework, consumer welfare is a function of (1) the health produced from medical care (which we measure by combining the CEAR quality measures with insurance claims that captures how treatments diffuse); (2) the value of a statistical life year (we present results for a range of assumptions); and (3) the costs of care (which we measure using insurance claims data and a separate dataset which we use to adjust for rebates). Consistent with the literature, the cost we measure refers to the amount insurers and patients pay for a treatment and throughout the draft we use the terms cost and spending interchangeably. In the Cutler et al. (1998) framework, the quality-adjusted price of treatment for a condition is falling if consumer welfare is rising for that condition (and vice-versa).

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 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 CEAR data account for nearly all pharmaceutical spending for that condition. The 13 conditions we study are important in their own right as they account for $191 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: innovative and 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 (e.g. price declines due to patent expiration), as the quality of available products changes little.

[4]These results are based on the BEA Health Care Satellite Account (HCSA) (Dunn et al., 2015).

These two types of markets are exemplified by treatments for hepatitis C and rheumatoid arthritis. During our study period (2007-2018), new innovations for hepatitis C replaced the earlier generations of treatments. In contrast, the major innovations for rheumatoid arthritis occurred in the late 1990s and early 2000s, so we view rheumatoid arthritis as non-innovative market over our sample period. For hepatitis C, there is considerable cost 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, across all the conditions we study, we find considerable evidence of innovation. For all the conditions (except one) 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 to measure the average quality-adjusted price index within our sample. We find that without adjusting for quality, on average prices for these conditions increase by 75% from 2007 to 2018, relative to economy-wide inflation. Meanwhile, our quality-adjusted price index, assuming a VSLY of \$100k, rises by 46% – a reduction of 1.7 percentage points in the compound annual growth rate. If we assume a \$500k VSLY, the quality-adjusted price index *falls* by 68%. This suggests that price indexes which do not account for quality improvements may be overstating price growth. This is important as there is currently no quality adjustment for official health care prices from the Bureau of Labor Statistics (Brown et al., 2020).

Looking at individual conditions, we find a number of novel results. Surprisingly, several of the markets where consumer welfare *declined* had a lot of innovation and large quality improvements. However, those quality improvements were small relative to the cost increase. This result is surprising because without market 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 why this occurs, we take the example of Orkambi, a 2015 breakthrough

therapy for cystic fibrosis. We find a sizeable quality improvement of taking Orkambi: 0.85 quality-adjusted-life years (QALYs), where 1 QALY is a year of life in perfect health. In our data, insurers and patients combine to pay more than $150k per year for Orkambi, or over $3 million in lifetime costs. While Orkambi contributes to major health improvements, in our framework one would need to assume a VSLY of over $3 million for these high costs to be "worth it", which is a value far higher than any in the literature. Hence, when accounting for the total cost this innovation reduces overall consumer welfare as the cost growth overwhelms the quality improvements. Yet, Orkambi had taken roughly 20% market share by 2018. One potential explanation for its diffusion, despite its high cost, is insurance. In our data, on average Orkambi costs $1.5k per year out-of-pocket, suggesting that insured patients using Orkambi are benefiting greatly as most of the cost is paid by insurance (and other consumers if these costs raise premiums). This is not an isolated case. We find that for six of our conditions, consumer welfare in 2018 is lower than it would have been in a counterfactual that assumes no new treatments (i.e., maintains the treatment portfolio in 2007).

While we find several 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 producers are receiving more than 100% of the surplus from their innovations. 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 (such as moral hazard from insurance) lead products with negative consumer surplus to diffuse.

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. We find that non-innovative markets, like colon cancer and hypertension, have falling price indexes and rising consumer welfare because of patent expiration. This highlights how the high prices and reductions in consumer welfare are potentially a short-run phenomenon. To explore this further, we compute a counterfactual where we reduce prices of all on-patent drugs by 85%, simulating the long-run changes in pricing. In this counterfactual, all conditions, except cystic fibrosis, have higher consumer welfare than in 2007.

Another surprising result emerges in non-innovative markets, where we observe two cases

of quality declining after generic entry. For both hypertension and colon cancer, we observe an older, lower quality treatment lose patent protection. Out-of-pocket prices fall for these relatively lower quality drugs and they take market share from relatively higher quality drugs. This reduces the average quality of care, though at a much lower price, so consumer welfare is higher after generic entry. This, in turn, leads to surprising implications for total welfare. If one takes a utilitarian view of welfare and weighs consumer surplus and producer surplus equally, the cost savings for consumers are offset by the profit losses for producers (i.e., the dollar transfer from producers to consumers does not impact the sum of consumer and producer welfare). Under this assumption, the net effect on total welfare is only due to the change in health.[5] Therefore, the only place in our framework where total welfare declines is after generic entry because average quality declines. 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 unique observations about non-innovative markets and patent expiration highlight the importance of looking at the full life-cycle of innovation.[6]

As noted by Bryan and Williams (2021), one of the fundamental challenges of 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. The patterns we observe are consistent with the textbook life-cycle model of innovation, where the benefits first accrue to producers, then to consumers through patent expiration and prices falling. However, we find a number of unconventional (but intuitive) results about how innovation and patent expiration shape welfare in health care markets. These results highlight the importance of empirical evidence in health care markets where multiple distortions can cloud theoretical predictions and conventional wisdom about innovation.

[5]In a standard economic model, generic entry would increase quantity supplied and reduce deadweight loss. While we document some substitution among drugs taken, we do not find any compelling evidence in the MarketScan data that more people are receiving treatment after generic entry. This is line with an empirical literature which has found either no extensive margin effect or even that generic entry actually reduces market shares (possibly due to reduced advertising) (Duflos and Lichtenberg, 2012; Castanheira et al., 2019).

[6]This highlights the value of our detailed yet scaleable methodology. To view the full life-cycle of innovation, one would typically need at least 20 years of data to observe both entry and patent expiration. We overcome this challenge by looking at conditions and treatments in different stages of the life-cycle.

# 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, we find several cases of new treatments that diffuse which are higher quality, but not cost-effective. This is consistent with the findings in Kyle and Williams (2017) who find high-cost drugs diffuse faster in the U.S. than other countries.

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 income), then, following the logic of Solow (1957), the residual is attributed to innovation.[7] Our granular approach allows us to both decompose the share of spending due to innovation, by observing these innovations directly, but also weigh the cost growth against measures of quality improvement.

Having such rich data 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. We view this as an important contribution. While prior case studies have led to advancements in the literature, they vary in 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.[8]

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

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

[8]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.

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.[9] 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 the quality of life, 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 CEAR data to construct quality-adjusted price indexes. Dunn et al. (2022) show that medical innovations typically lead to quality-adjusted prices declining, but Hult et al. (2018) and Dunn et al. (2022) 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 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 information about which technologies are used in practice.[10]

# 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 they both experienced considerable cost growth during our sample period, 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

[9]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.

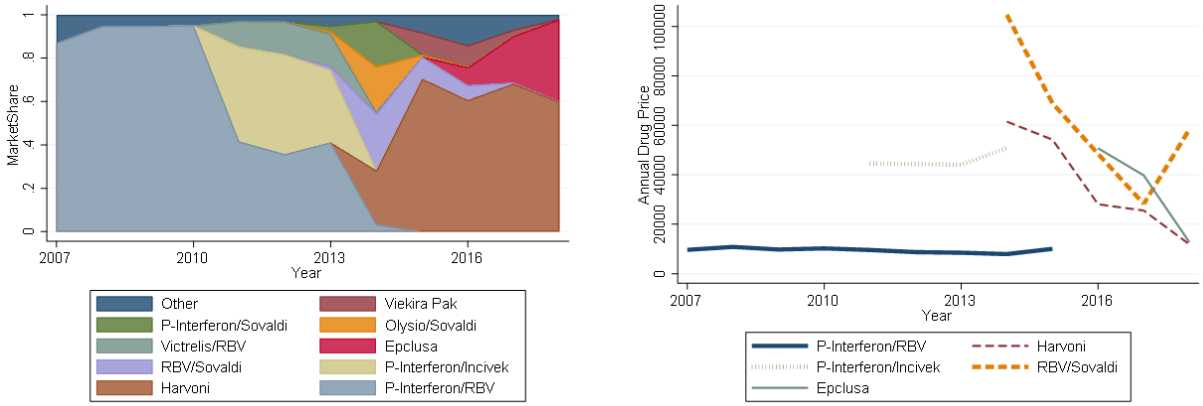
[10]Properly weighting innovations based on their usage is critical for accurately measuring welfare. For example, a major breakthrough innovation for a rare disease may have less of a welfare impact than a marginal innovation which diffuses broadly.

serious, but it can take 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 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 SSR Health data) and market shares for the top hepatitis C treatments evolved during our sample period. At the beginning of our sample in 2007, the 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. Finally, Harvoni, Epclusa, and Viekira Pak entered starting in late 2014. These drugs are more effective than Sovaldi and less expensive, likely due to the greater competition among highly effective treatments upon entry.

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

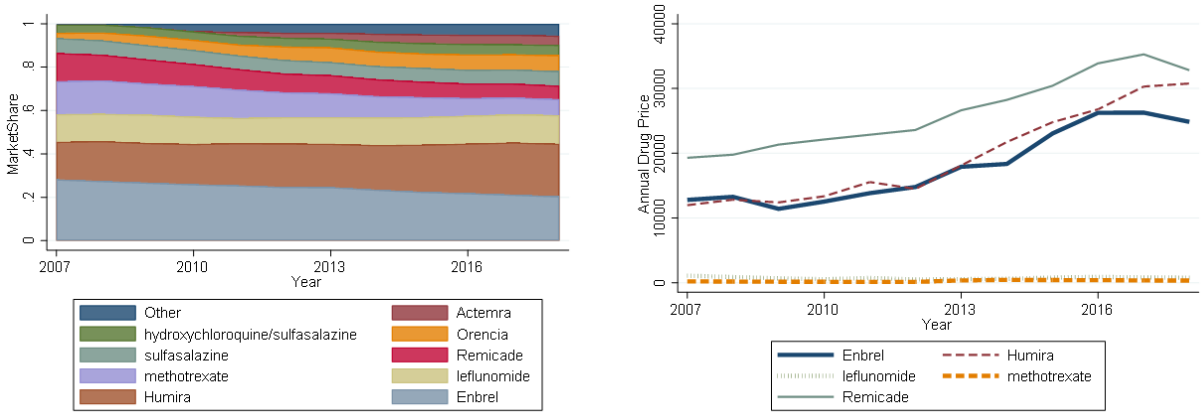



**Notes:** Estimates are derived from MarketScan claims data described in the data section. The left panel of this figure presents the patient-weighted market shares by year for the 9 highest volume drugs for hepatitis C across our entire sample period in the MarketScan data. The right panel presents the average price per year of the 5 highest volume drugs in our sample. They are not scaled to lifetime costs. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. 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. 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. As with hepatitis C, rheumatoid arthritis is typically not fatal with proper treatment.

Figure 2 shows price and market shares for rheumatoid arthritis. The baseline treatment for rheumatoid arthritis is methotrexate, which entered the market in 1947 to treat cancer and was 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 relatively 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 shares and prices for the top rheumatoid arthritis treatments over time




**Notes:** Estimates are derived from MarketScan claims data described in the data section. The left panel of this figure presents the patient-weighted 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. They are *not* scaled to lifetime costs. Drugs do not have prices in all years because either they have not entered the market yet or they stop being used. 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 price increases.[11] 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 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 cost growth does not appear to be due to new innovations.

The goal of this paper is to better understand how these changes in the market translate into quality improvements and cost increases. Our methodology estimates average quality measures for each of these treatments and matches them with their respective market shares to better understand how the quality of treatment for the average patient changes over time. We summarize the consumer 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.

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, such as a new innovation ($I$), with a "comparator" or standard of care ($SOC$) treatment (i.e., a commonly used treatment for a particular condition). Let $S_I$ and $S_{SOC}$ be the costs for the innovation, $I$, and standard of care treatment, $SOC$, respectively. The effectiveness of a treatment, denoted by $H_I$ and $H_{SOC}$, is typically measured in years of life or quality-adjusted-life years (QALYs), where QALYs account for both the mortality and the quality-of-life. One QALY represents one year of life in perfect health.

[11] The House Committee on Oversight and Reform released a report which highlights a few reasons why prices might rise so rapidly (Hopkins, 2021). Potential reasons include executive compensation being tied to revenue targets and firms using the others firm's price increases as political cover to raise prices. This is also consistent with "penetration pricing," where manufacturers may introduce the product with lower prices to induce switching to their product. Then, as patients and doctors gain a taste for the product, manufacturers can raise their prices as current customers become less price sensitive (Lu and Comanor, 1998).

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 measurement of the relative cost and effectiveness of $I$ compared to $SOC$, holding other variables fixed. If a dollar value can be placed on life years gained, then researchers can calculate the net benefit. The dollar value placed on a QALY is often measured as a value of a statistical life year (VSLY), which measures an individual's value of living an additional year.

The elements of a cost-effectiveness study can be used to express the net benefit or consumer welfare from the innovation in a dollar amount.

$$\Delta \text{ Consumer Welfare}_{I,SOC} = VSLY \cdot (H_I - H_{SOC}) - (S_I - S_{SOC}) \tag{1}$$

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.

# 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), and Dauda et al. (2022). We construct price indexes separately by disease, indexed by $d$.[12]

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,

[12]Condition-based inflation measures are recommended in National Research Council (2002) and National Research Council (2011) and are useful when measuring quality changes, which typically affect the treatment of specific conditions.

$t-1$, relative to treatments received at time, $t$.[13] Changing the subscripts in equation (1) accordingly, we obtain the following equation for the consumer welfare change over time:

$$\Delta \text{ Consumer Welfare}_{d,t,t-1} = VSLY \cdot (H_{d,t} - H_{d,t-1}) - (S_{d,t} - S_{d,t-1}). \tag{2}$$

$\Delta$ Consumer Welfare$_{t,t-1}$ accounts for the change in the price and quality of treatment.[14]

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:

$$\begin{aligned}
\text{Price Index}_{d,t,t-1} =& \frac{S_{d,t-1} - \Delta \text{ Consumer Welfare}_{d,t,t-1}}{S_{d,t-1}} \\
=& \frac{S_{d,t-1} - [VSLY \cdot (H_{d,t} - H_{d,t-1}) - (S_{d,t} - S_{d,t-1})]}{S_{d,t-1}} \\
=& \frac{S_{d,t}}{S_{d,t-1}} - \frac{VSLY \cdot (H_{d,t} - H_{d,t-1})}{S_{d,t-1}}.
\end{aligned} \tag{3}$$

The first line of equation (3) shows that the price index falling (being less than 1) means that consumer welfare is rising, and vice versa. The middle line of equation (3) provides an intuitive form of the index. Suppose that the average treatment cost in period $t-1$ is \$50,000. 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 (3) provides a simplified expression that demonstrates how one can separate the unadjusted price change, $\frac{S_{d,t}}{S_{d,t-1}}$, from the quality improvement, $\frac{VSLY \cdot (H_{d,t} - H_{d,t-1})}{S_{d,t-1}}$.

[13] More precisely, 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: $H_{d,t} = \sum_{r \in \mathcal{R}_{d,t}} w_{r,d,t} H_{r,d}$. 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] 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. In Appendix Section OA.E, we follow Lakdawalla et al. (2017) to account for how innovation impact health risks and financial risks. The estimates from this exercise increase the amount of consumer welfare created by innovation, but the main results are qualitatively similar.

# 5 Data

To construct the quality-adjusted indexes, we estimate the cost and QALY of each treatment, as well as determine the share of patients receiving each treatment. 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 compiles more than 8,000 cost-effectiveness analyses which have been published in English and are indexed by Medline, from the years 1976-2019, 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. The data reports the QALY and cost for each treatment, descriptions of treatments, and a disease classification (i.e. asthma, hepatitis C). It also includes information on the journal, author, author affiliation, funding, year published, country the study was performed in, among other characteristics.

Although the CEAR data contain detailed information, it is not in a form that is readily combined with other data sources or across studies within the CEAR data. 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 molecules or combinations of molecules. Accuracy was then verified by an additional review of the independent classifications.[15]

To focus our analysis, we concentrate on 13 conditions where most of the treatment (or innovation) is through pharmaceuticals as these are much easier to classify in the CEAR and to 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 CEAR data. Appendix Section OA.D.2 uses MEPS data to explore the share of total drug spending associated with drugs we classify using the CEAR data. The

[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 CEAR data.

[16]While we tried classifying procedures such as surgeries, the terminology in CEAR did not always map cleanly to procedure codes in claims data.

drugs in the CEAR data account for at least 79% of MEPS drug spending for all conditions except atrial fibrillation (60%).[17]

We keep all comparisons where two classified drugs are compared to each other,[18] and both drugs have a non-outlier QALY estimates.[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 the "no treatment" and "standard of care" 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 and are enrolled for 360 days. 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. 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-

[17] We use MEPS data for this calculation because, unlike Marketscan data, the MEPS data has diagnosis codes on drug claims. For colon cancer and lung cancer we use the Marketscan because chemotherapy drugs are physician administered and include diagnosis codes, so we can link diagnoses and drugs in the medical claims.

[18] Many of the observations in the CEAR 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.

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) to adjust for rebates.[20] See our data appendix, Section OA.D, for more details about how we clean the CEAR data and merge it to the MarketScan data, as well as how we incorporate rebates into estimates.

# 6 Methods

This section discusses how we estimate the variables in Equation 3. This includes how we estimate QALYs using CEAR data and costs using MarketScan claims data.

## 6.1 Estimating QALYs from CEAR Data

Treatment level QALYs can be taken directly from the raw CEAR data for specific studies, but this is not the preferred approach for obtaining QALY estimates for several reasons: (1) the CEAR data makes pairwise comparisons, whereas we need estimates for all treatments; (2) the CEAR data often have multiple observations for each treatment, necessitating some averaging; (3) there is variation in study design, populations, and assumptions which will affect each treatment in a comparison; and (4) there is variation in the drugs that treatments are compared to.[21] We use a regression to address all of these issues.

In the CEAR data, the unit of observation is a comparison, which we subscript with $u$. We reshape the CEAR 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:

[20] SSR Health, LLC collects data from drug manufacturers' 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.

[21] For example, Harvoni is compared to P-Interferon and ribavirin twice, P-Interferon, ribavirin, and Sovaldi seven times, and ribavirin and Sovaldi five times.

$$log(H_{u,c,d}) = \gamma_{r,d} + \gamma_{u,d} + \epsilon_{u,c,d} \tag{4}$$

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 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 in 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, 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 studies (e.g., age, sex or time horizon of the study). See appendix section OA.D.3 for additional details and robustness checks which show that results are robust to different methods of handling comparison specific heterogeneity.

After obtaining regression coefficients, we retransform our estimates into levels using the method proposed in Duan (1983). For disease level estimates, we calculate the average QALY by taking a quantity weighted average across all treatments in a given year.

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 (4). We report results using this approach in Appendix Table OA4, however we strongly prefer using the MarketScan data to estimate costs, as cost-effectiveness studies only reflect costs at a single point in time (e.g., they do not capture prices falling due to patent expiration) and they only reflect costs for a particular setting (e.g., estimates may come from very different health systems).

## 6.2 Estimating Costs from Claims Data

To calculate costs, denoted $C_{d,t}$, we begin by summing over all the expenditures a person has for condition $d$ in a given year $t$.[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, 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 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 3 percent. 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.

[22] Inpatient 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 which condition a prescription treats. In our preferred specification, we include all drugs that we classify using CEAR data, which covers most drug expenditures for our selected conditions. In Appendix Table OA11 we allocate all drug claims to medical conditions to pick up drugs that are not in the CEAR. Our results do not change much.

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

# 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 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 providing 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 | | (1) Is Baseline Treatment | (2) Δ QALYs from Baseline |
|---|---|---|---|---|---|
| Epclusa | 0 | 2.744 | Actemra | 0 | 3.124 |
| Harvoni | 0 | 2.815 | Enbrel | 0 | 2.125 |
| Olysio/Sovaldi | 0 | 2.834 | Humira | 0 | 2.266 |
| P-Interferon/Incivek | 0 | 1.367 | Orencia | 0 | 2.607 |
| P-Interferon/RBV | 1 | 0.000 | Remicade | 0 | 1.883 |
| P-Interferon/Sovaldi | 0 | 2.272 | hydrox./sulfas. | 0 | 2.138 |
| RBV/Sovaldi | 0 | 1.178 | leflunomide | 0 | 0.048 |
| Victrelis/RBV | 0 | 0.807 | methotrexate | 1 | 0.000 |
| Viekira Pak | 0 | 2.797 | sulfasalazine | 0 | -0.256 |

**Notes:** This table presents the estimated QALYs using the 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 2.1 QALYs relative to

[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.

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 compared 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 least 2.1 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 2.3 QALYs and in 2018 it is 2.9 QALYs after the entry of Harvoni.[25]

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 the average lifetime cost of treatment in that year divided by the average lifetime cost in 2007. The change in drug generation is reflected in the costs. Costs are roughly $42k until 2011, then they rise to roughly $100k upon the entry of Incivek and Victrelis. Then in 2014, following the launch of Sovaldi, the average cost jumps to $340k, an 834% 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 $49k, only 20% higher than 2007.

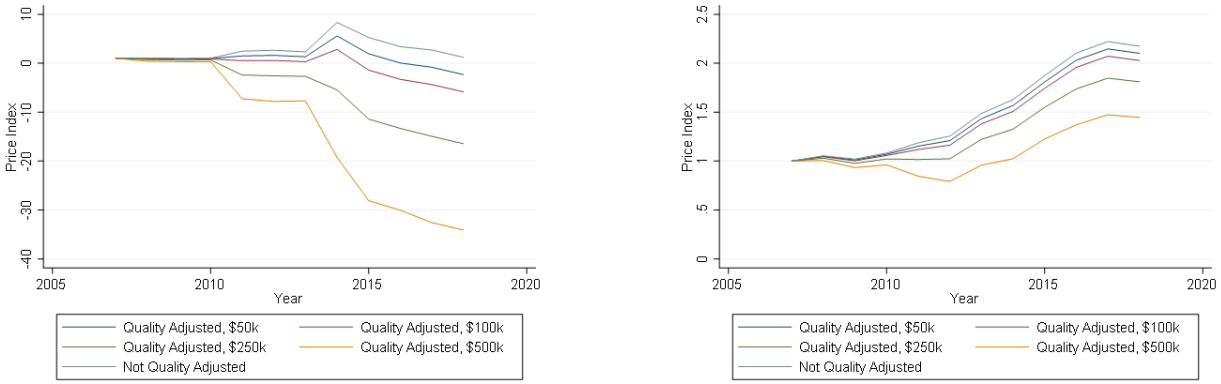
[25]These numbers are somewhat larger than the differences in Table 1. This is because there are other treatments like interferon (rather than P-Interferon) and P-Interferon without ribavirin that have fewer QALYs than the baseline treatment and positive market share in 2007.

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

| | (1) Change in Avg QALYs | (2) MktScan Lifetime Costs ($1,000s) | (3) Price Index $0 VSLY | (4) Price Index $100k VSLY | (5) Δ Consumer Welfare $100k VSLY ($1,000s) | (6) Price Index $500k VSLY | (7) Δ Consumer Welfare $500k VSLY ($1,000s) |
|---|---|---|---|---|---|---|---|
| 2007 | 0.000 | 41 | 1.000 | 1.000 | 0 | 1.000 | 0 |
| 2008 | 0.054 | 44 | 1.075 | 0.944 | 2 | 0.418 | 24 |
| 2009 | 0.054 | 40 | 0.987 | 0.854 | 6 | 0.321 | 28 |
| 2010 | 0.057 | 42 | 1.029 | 0.889 | 5 | 0.330 | 27 |
| 2011 | 0.794 | 101 | 2.467 | 0.523 | 19 | -7.255 | 337 |
| 2012 | 0.853 | 108 | 2.653 | 0.563 | 18 | -7.796 | 359 |
| 2013 | 0.818 | 95 | 2.329 | 0.326 | 28 | -7.685 | 355 |
| 2014 | 2.252 | 340 | 8.337 | 2.822 | -74 | -19.240 | 826 |
| 2015 | 2.725 | 215 | 5.277 | -1.399 | 98 | -28.101 | 1,188 |
| 2016 | 2.731 | 139 | 3.405 | -3.285 | 175 | -30.043 | 1,267 |
| 2017 | 2.881 | 111 | 2.707 | -4.350 | 218 | -32.578 | 1,371 |
| 2018 | 2.883 | 49 | 1.204 | -5.858 | 280 | -34.106 | 1,433 |

**Notes:** This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer welfare, constructed using the CEAR, 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. 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 2 and 3. The price indexes are also graphed in Figure 3.

Figure 3: Price Indexes for Hepatitis C and Rheumatoid Arthritis




Hepatitis C 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 2 and 3. These results are constructed using data from CEAR, MarketScan, and SSR Health.

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, $100k, $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 2 and 3. For example, in 2015, the average QALY was 2.73 QALYs higher than in 2007 when most patients were receiving interferon based treatments. At a \$100k VSLY, this represents \$273k of welfare. Given the \$174k difference in average costs, this represents a \$99k gain in consumer welfare. Likewise, the index is $\frac{215-100\cdot 2.73}{41} = -1.41$, where numbers are slightly different in the table due to rounding. If consumers value life more, the quality adjustment gets larger. If one assumes the VSLY is instead \$500k, the index becomes $\frac{215-500\cdot 2.73}{41} = -28.04$. 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 in order to maintain the same level of utility across periods, individuals would actually need to be paid more than the price of the old technology.

After accounting for QALY differences, prices appear to be declining with each subsequent generation of new drugs. 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. 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 0.22 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, and 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

[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 2011.

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 $154k. Lifetime costs doubled during our sample period to $336k. 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.

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

| | (1) Change in Avg QALYs | (2) MktScan Lifetime Costs ($1,000s) | (3) Price Index $0 VSLY | (4) Price Index $100k VSLY | (5) Δ Consumer Welfare $100k VSLY ($1,000s) | (6) Price Index $500k VSLY | (7) Δ Consumer Welfare $500k VSLY ($1,000s) |
|---|---|---|---|---|---|---|---|
| 2007 | 0.000 | 154 | 1.000 | 1.000 | 0 | 1.000 | 0 |
| 2008 | 0.015 | 163 | 1.054 | 1.044 | -7 | 1.006 | -1 |
| 2009 | 0.027 | 158 | 1.020 | 1.003 | 0 | 0.934 | 10 |
| 2010 | 0.037 | 167 | 1.082 | 1.058 | -9 | 0.962 | 6 |
| 2011 | 0.105 | 183 | 1.187 | 1.119 | -18 | 0.846 | 24 |
| 2012 | 0.143 | 194 | 1.256 | 1.163 | -25 | 0.791 | 32 |
| 2013 | 0.163 | 230 | 1.487 | 1.381 | -59 | 0.958 | 6 |
| 2014 | 0.187 | 251 | 1.629 | 1.508 | -78 | 1.024 | -4 |
| 2015 | 0.200 | 289 | 1.873 | 1.744 | -115 | 1.226 | -35 |
| 2016 | 0.226 | 325 | 2.104 | 1.958 | -148 | 1.371 | -57 |
| 2017 | 0.231 | 343 | 2.222 | 2.072 | -166 | 1.474 | -73 |
| 2018 | 0.224 | 336 | 2.174 | 2.029 | -159 | 1.447 | -69 |

**Notes:** This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer welfare, constructed using the CEAR, 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. 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 2and 3. The price indexes are also graphed in Figure 3.

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 3. 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 doubled during our sample period.

## 7.2 Results for Other Conditions

Table 4 summarizes the 2018 results for all 13 of the conditions. It is the last row of Tables 2 and 3, except it presents the 2007 cost rather than the 2018 cost and we present total welfare, rather than consumer welfare, in column 7, which we discuss in section 7.4. Detailed tables and figures are presented for the remaining 11 conditions in Appendix Section OA.A.

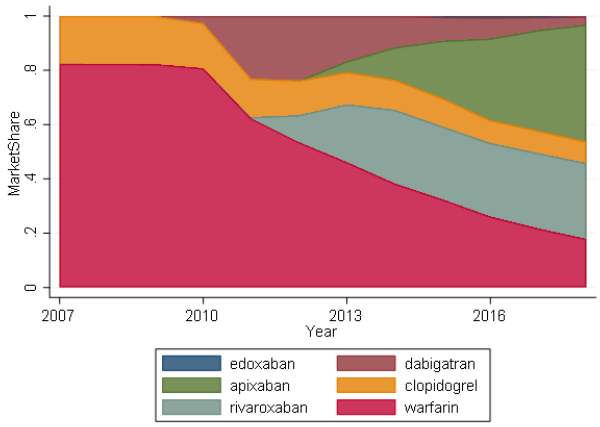
Table 4: Price Indexes and Changes in Welfare for Each Condition Between 2007 and 2018

| | (1) Δ Avg QALYs 2018 - 2007 | (2) MktScan Lifetime Costs in 2007 ($1,000s) | (3) Price Index $0 VSLY | (4) Price Index $100k VSLY | (5) Δ Consumer Welfare $100k VSLY ($1,000s) | (6) Price Index $500k VSLY | (7) Δ Total Welfare $100k VSLY ($1,000s) |
|---|---|---|---|---|---|---|---|
| Asthma | 0.002 | 16 | 1.014 | 1.000 | 0 | 0.944 | 0 |
| Atrial Fibrillation | 0.454 | 14 | 3.854 | 0.695 | 4 | -11.941 | 45 |
| Colon Cancer | -0.040 | 338 | 0.607 | 0.619 | 129 | 0.666 | -4 |
| Cystic Fibrosis | 0.231 | 622 | 4.232 | 4.195 | -1,986 | 4.046 | 23 |
| HIV | 0.186 | 312 | 1.505 | 1.446 | -139 | 1.207 | 19 |
| Hepatitis C | 2.883 | 41 | 1.204 | -5.858 | 280 | -34.106 | 288 |
| Hypertension | 0.039 | 9 | 0.684 | 0.221 | 7 | -1.631 | 4 |
| Lung Cancer | 0.648 | 267 | 2.151 | 1.909 | -243 | 0.939 | 65 |
| Multiple Sclerosis | 0.428 | 476 | 2.998 | 2.909 | -908 | 2.549 | 43 |
| Osteoporosis | 0.033 | 7 | 1.690 | 1.200 | -1 | -0.756 | 3 |
| Rheumatoid Arthritis | 0.224 | 154 | 2.174 | 2.029 | -159 | 1.447 | 22 |
| Schizophrenia | 0.118 | 38 | 0.823 | 0.511 | 18 | -0.738 | 12 |
| Venous Thromboembolism | 0.106 | 6 | 1.308 | -0.377 | 9 | -7.118 | 11 |

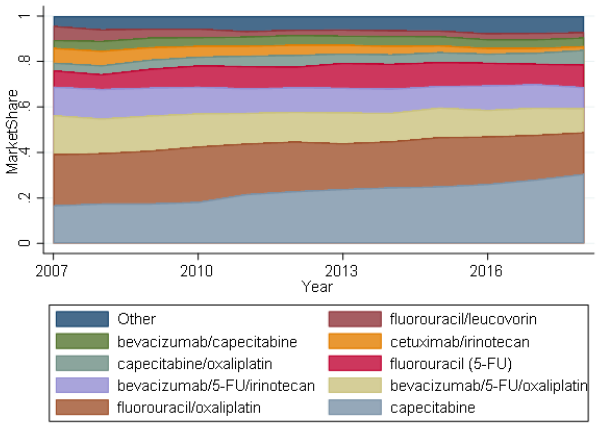
**Notes:** This table presents changes in QALYs, costs, quality adjusted price indexes, and consumer and total welfare, constructed using the CEAR, MarketScan, and SSR health datasets. Columns 1 and 2 presents the difference in average QALYs relative to 2007 and the estimated lifetime costs in 2007 for each condition. 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 4-7 can be calculated directly using the results in columns 1-3 and using equations 2, 3 and 6 and assuming marginal costs are constant over time.

Column 1 shows the change in average QALYs by condition. We find quality improving for all conditions during this time period, except colon cancer (which we explain in Section 7.4). The shift toward higher quality is consistent with Chandra et al. (2016), who find that patients shift toward higher quality hospitals over time. However, we see a lot of heterogeneity in the change in mean QALYs. To highlight where changes in QALYs (and costs) are coming from, Figure 4 shows market shares for 6 selected conditions. Hypertension and colon cancer along with asthma and schizophrenia, (shown in Appendix Figure OA1), have relatively small changes in market shares, and few new entrants over our sample period. Other conditions in Figure 4 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.

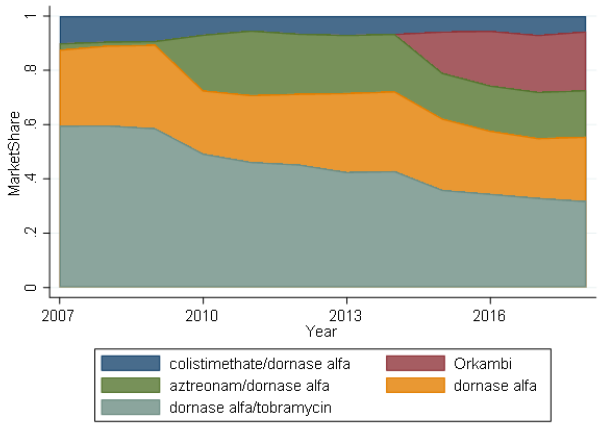
Figure 4: Market share for top treatments over time



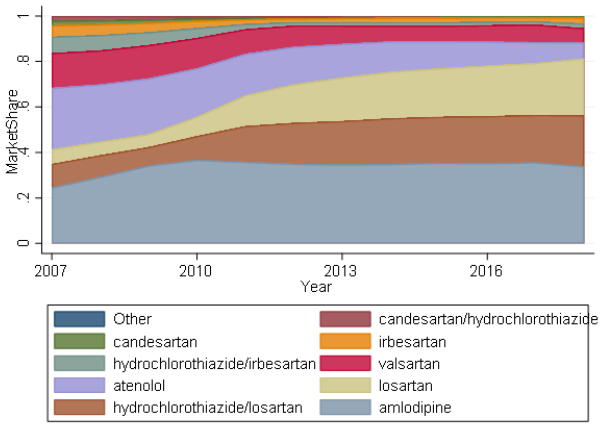

(a) Atrial Fibrillation



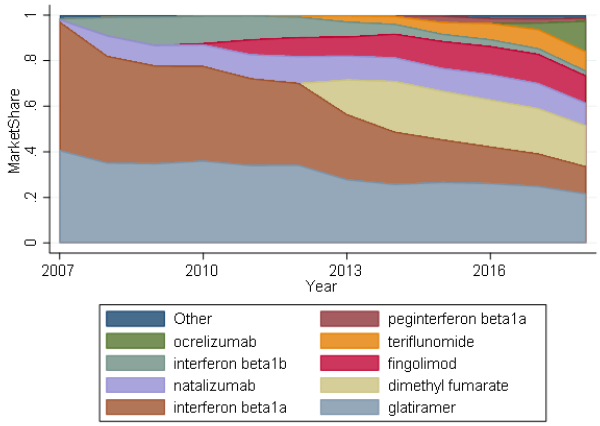

(b) Colon Cancer



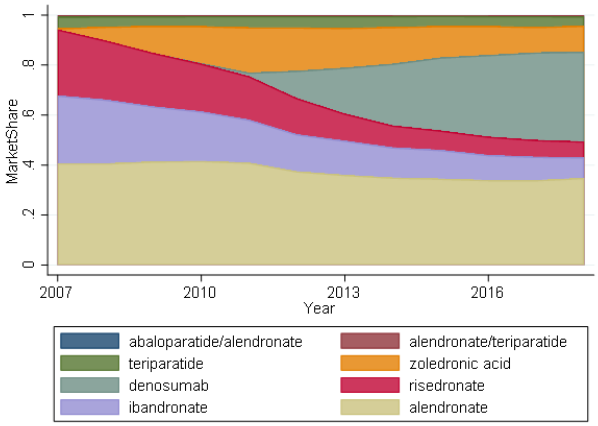

(c) Cystic Fibrosis




(d) Hypertension



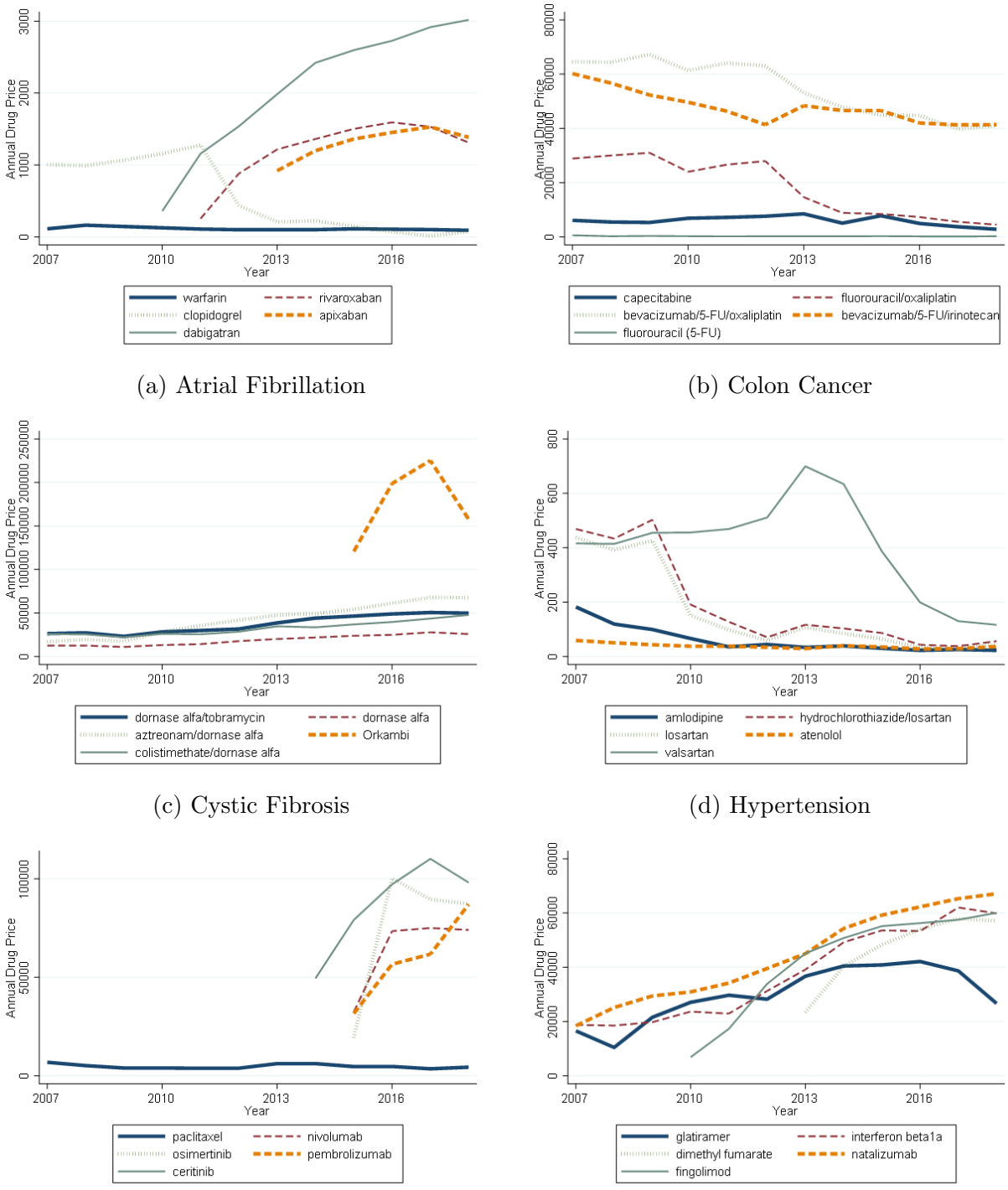

(e) Multiple Sclerosis




(f) Osteoporosis

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

Figure 5: 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.

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 explain these results, Figure 5 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 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 OA4) 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.

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 5). 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

[27]Without merging cost-effectiveness studies to claims data, as in the top-down approach of Dunn et al. (2022), the CEAR data would have suggested that rheumatoid arthritis was highly innovative as the CEAR 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 market share (Figure 4). 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 \$3 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 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, so consumer surplus falls sharply.[31] Indeed, one would need to assume a VSLY of over \$8.7 million for consumer welfare to be improving for cystic fibrosis during our sample period.[32]

Lung cancer, multiple sclerosis, and HIV follow a similar pattern to cystic fibrosis.[33] 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 im-

[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.

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

[32]One can calculate this for every condition using the information in Table 4. For the consumer to be indifferent, the value of health improvements would equal the change in costs: $VSLY \cdot \Delta$ Avg QALYs = $\Delta$ Costs. For cystic fibrosis, the change in QALYs is 0.231. The change in costs is $622k \cdot 4.232 - 622k = 2,010k$. This number is much larger than the number we discuss in the introduction, as that calculation in the introduction is for Orkambi, whereas the number in this section is for cystic fibrosis, generally. The sizeable within-molecule price increases for other drugs increase costs, but not quality, which makes the breakeven VSLY much higher.

[33]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.

provements (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.F, we show results which aggregate across conditions, weighted by spending. The unadjusted price index rises by 75%, while the $100k VSLY index rises by 46% – a reduction of 1.5 percentage points from the compound annual growth rate. At $500k VSLY, the index falls by 68%.

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 Robustness checks

We do a number of other robustness checks that we describe in this section, though we leave many of the details to Appendix Section OA.B.

**QALY estimates:** There are a number of reasons why we may be overstating or understating true quality changes.[34] 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 (Table OA1) or by one half (Table OA2). 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.

We do a number of other robustness checks to test the sensitivity of our QALY estimates to various assumptions and modelling choices. One important robustness check we do is to add weight to studies which an expert Tufts reviewer rates to be of higher quality, and studies conducted by authors with academic affiliations. We also lower the weight on studies authored or sponsored by industry (Table OA3). While our main estimates focus

[34]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.

on the primary treatment class, in Table OA10 we include multiple classes of treatments for conditions.[35] We also do a number of robustness checks on the CEAR quality regressions (Equation 4) where we run the QALY regression a number of different ways (e.g. running the regression in levels, not controlling for heterogeneity, etc.) (Table OA13). We also 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 (Table OA13). None of the robustness checks mentioned in this paragraph change our results considerably.

As Lucarelli et al. (2022) note, cost-effectiveness studies may understate the value of new treatments if there is heterogeneity in preferences, perhaps from differences in treatment effectiveness (e.g., some rheumatoid arthritis patients may respond better than average to Humira, while others respond better to Enbrel). To address this, in a robustness check we add idiosyncratic noise to the QALY, so that some patients get QALYs above or below our estimated average QALY. We then take an expected maximum, implicitly assuming that patients select the best individual treatment. See the discussion around tables OA14 and OA15 of the appendix for details. Even with what we think are fairly large adjustments due to heterogeneity, we still find that many innovative conditions have falling consumer welfare.

Another criticism of cost-effectiveness studies is that they do not account for the insurance value of innovation (Lakdawalla et al., 2017). For example, health shocks can create financial risks (lower earnings, high medical bills which lower other consumption) and health risk (people get sick, which lowers utility). New treatments can impact the size of all of these risks and the overall effect is ambiguous (innovation may reduce health risk, but increase financial risk). However, risk is not accounted for in the model in our paper, which is standard in the cost-effectiveness and the quality-adjusted price index literatures. In Appendix Section OA.E, we allow for these risks (and allow for insurance to mitigate financial risks), following the framework in Lakdawalla et al. (2017). We find that the growth in consumer welfare is higher when we account for these risks. However, even with these higher consumer welfare estimates, it remains the case that innovation appears to be lowering consumer welfare for several conditions.

**Cost estimates:** In Appendix Section OA.D.4 we test the sensitivity to our lifetime cost estimates. While we try a number of different modelling choices, the widest range of

[35]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.

estimates are assuming all treatment costs occur in one year versus assuming treatment costs are constant (only the discount factor and life tables determine the lifetime scaling factor). We observe in our data that neither of these assumptions hold, though the distance from either bound varies by condition.[36] The upper bound (constant costs) is 23-28 times the lower bound (one year of costs), so this is a very wide range. In either case, we see some innovative conditions with consumer welfare falling, and some conditions where price indexes are falling because of quality adjustments. However, with such a wide range of assumptions the number of conditions with falling price indexes and the magnitudes vary considerably.

As an additional robustness check, we bring in additional drug claims. Marketscan drug claims do not include diagnosis codes, so we only use CEAR treatments to account for drug spending. As a robustness check, 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 (Table OA11).[37]

## 7.4 Total Welfare, Producer Surplus and Long Run Effects

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-patient total welfare and producer surplus. 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, $mc_{d,t}$:

$$\Delta\text{Producer Surplus}_{d,t,t-1} = (S_{d,t} - S_{d,t-1}) - (mc_{d,t} - mc_{d,t-1}), \tag{5}$$

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. 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 2 and producer surplus from Equation

[36] One year of costs is an especially extreme assumption for conditions where treatments are taken indefinitely, like hypertension or rheumatoid arthritis. Constant costs is an extreme assumption for conditions where treatments are curative (hepatitis C) or costs are concentrated in one year, like cancers whose costs are mostly surgeries and chemotherapy.

[37] We keep this as a robustness check because the mapping in the MEPS can be noisy. For example, if a person has high cholesterol and hepatitis C, they may have statins on a claim with hepatitis C listed as their diagnosis.

5 provides a measure of per-patient total welfare:

$$\Delta\text{Total Welfare}_{d,t,t-1} = VSLY \cdot (H_{d,t} - H_{d,t-1}) - (mc_{d,t} - mc_{d,t-1}). \tag{6}$$

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 6 to be $VSLY \cdot (H_{d,t} - H_{d,t-1})$, which is the change in the average health benefit.[38] [39] 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, especially relative to the price of the treatments. 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.[40]

Column 7 of Table 4 presents results for total welfare. In our framework, total welfare is simply the health improvement 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 is after generic entry (of a relatively lower quality treatment).[41] 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 4 and panel (b) of Figure 5). While this is exemplified by colon cancer, it is not the only case where this pattern emerges. Hypertension had lower average QALYs from 2008-

[38] This total welfare calculation assumes that consumer and producer surplus are weighed equally. Policymakers likely differ in the weight they place on producer surplus. Our consumer welfare estimates show total welfare if no weight is placed on producer surplus, so readers can average between these estimates to see how the weighting impacts results.

[39] This formulation is similar in spirit to a social planner model where the only good being allocated is health. The social planner can redistribute wealth and ultimately wants to maximize units of health being created. In that framework, total welfare is simply the health produced (after netting out any costs of production).

[40] We may be understating the cost of new innovations if there is considerable spending on advertising.

[41] In a standard economic model, the entry of a lower cost product would increase quantity supplied and reduce deadweight loss, which our measure of per-patient welfare abstracts from. While we document some substitution among drugs taken, we do not find any compelling evidence in the MarketScan data that more people are receiving treatment after generic entry. This is in line with an empirical literature which has found either no extensive margin effect or even that generic entry actually reduces market shares (possibly due to reduced advertising) (Duflos and Lichtenberg, 2012; Castanheira et al., 2019).

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 4 and panel (d) of Figure 5).[42] 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.

Our results for colon cancer, hypertension, and schizophrenia also demonstrate how, in the long run, patents will expire and prices decline. The shift to increasing consumer welfare and a decline in producer surplus when generics enter is part of the product life cycle of pharmaceutical innovations. New innovations are often protected by patents, leading much of the initial surplus to accrue to producers, which later shifts to consumers when patent protection is lost.

To better demonstrate the long-run impacts of these innovations, Table OA9 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.

## 7.5 Which Treatments Are Driving Our Results?

In this section we ask: which treatments are responsible for the biggest changes in welfare? To measure how each treatment impacts welfare, we measure the difference between the observed welfare in 2007, relative to the counterfactual welfare where we introduce a particular treatment at its 2018 levels (i.e., prices, quality and market share), holding all other treatments constant at the 2007 level. We do this for one treatment at a time. We make a distinction between old and new market share, because a drug diffusing changes costs and quality, whereas we assume that a drug that does not diffuse affects welfare through price changes only.[43] Specifically, for newly obtained market share (diffusion) we replace the 2007

[42]Note that we also find that the out-of-pocket prices fall for these drugs after patent expiry.

[43]For a hypothetical individual taking the drug in 2007 and 2018, the relevant welfare counterfactual is simply how the price has changed. This hypothetical consumer should not expect a quality change as they are taking the drug in both periods. Therefore, for retained market share, we only revert the price to 2007 prices. For a hypothetical consumer taking a new treatment in 2018, the counterfactual is not that same drug in 2007 as they were not taking the drug in 2007. Therefore, we use the average bundle in 2007 as the

average bundle with the drug's 2018 price and quality, to capture how diffusion impacts both price and quality. For market share that was retained from 2007, we only replace the price of that drug, as there is no change in the drug taken, so it should not impact quality.[44]

The top panel of 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.

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 for Hepatitis C and rivaroxaban and apixaban, for both atrial fibrillation and venous thromboembolism all provide substantial quality improvements while not being significantly more expensive than their comparators.[45]

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 fluticasone for asthma, aripiprazole for schizophrenia, and losartan for hypertension.

Panel (b) of Table 5 shows the drugs which reduce consumer welfare the most. There are two types of drugs which reduce consumer welfare: high cost drugs which raise their prices and new innovations whose high cost exceeds the quality improvement. Examples of drugs that raise their price considerably include Humira, Enbrel, and interferon beta1a.

An example of a new innovation whose costs exceed the benefit of quality improvement is Orkambi. 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

comparison for newly obtained market share.

[44] As an example, suppose a drug had 10% market share in 2007 and 25% market share in 2018. We would replace its price 2007 price with its 2018 price for 10% of the market share. For the remaining 15% market share we replace the 2007 basket average price and quality with that drug's 2018 price and quality. If the drug instead had 25% market share in 2007, and 10% market share in 2018, then we revert the price for 10% of its market share to 2007 levels. This would capture the effect of the price difference from 2007, but no quality change would be assigned to that treatment whose market share declined.

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

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. healthcare markets, similar to the findings in Kyle and Williams (2017). To explore this more, we isolate the impact of new innovations in the next section.

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

| Rank | Condition | Drug | QALYs | Price in 2007 | Price in 2018 |
|---|---|---|---|---|---|
| 1 | Atrial Fibrillation | apixaban | 0.8 | | 1386 |
| 2 | Hypertension | losartan | 0.1 | 438 | 44 |
| 3 | Hypertension | hydrochlorothiazide/losartan | 0.2 | 469 | 56 |
| 4 | Hepatitis C | Harvoni | 2.9 | | 11881 |
| 5 | Atrial Fibrillation | rivaroxaban | 0.6 | | 1312 |
| 6 | Asthma | fluticasone | 0.0 | 238 | 75 |
| 7 | Hepatitis C | Epclusa | 2.8 | | 12741 |
| 8 | Venous Thromboembolism | rivaroxaban | 0.1 | | 1008 |
| 9 | Venous Thromboembolism | apixaban | 0.2 | | 1045 |
| 10 | Schizophrenia | aripiprazole | 0.5 | 3081 | 1351 |

(b) Drugs which account for biggest consumer welfare reductions - $100k VSLY

| Rank | Condition | Drug | QALYs | Price in 2007 | Price in 2018 |
|---|---|---|---|---|---|
| 1 | Rheumatoid Arthritis | Humira | 0.9 | 11978 | 30773 |
| 2 | HIV | Stribild | 0.1 | | 21271 |
| 3 | Multiple Sclerosis | dimethyl fumarate | 0.4 | | 57127 |
| 4 | Rheumatoid Arthritis | Enbrel | 0.8 | 12792 | 24846 |
| 5 | Multiple Sclerosis | ocrelizumab | 0.8 | | 75285 |
| 6 | Multiple Sclerosis | interferon beta1a | 0.1 | 18767 | 59880 |
| 7 | Cystic Fibrosis | Orkambi | 0.9 | | 157703 |
| 8 | Multiple Sclerosis | fingolimod | 0.8 | | 60068 |
| 9 | Multiple Sclerosis | glatiramer | -0.2 | 16556 | 26667 |
| 10 | Multiple Sclerosis | natalizumab | 1.3 | 18425 | 67121 |

**Notes:** This table presents the 10 drugs which contribute the most to consumer welfare gains and reductions across all the drugs and conditions in our sample. To calculate this, we calculate the difference in consumer welfare between a select counterfactual in 2018 and the 2007 basket average. The counterfactual for 2018 is constructed using the prices, quality and market share of the select treatment at the 2018 level, holding all other treatments to the 2007 level. 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.

# 8 How Does Innovation Affect 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 removing the influence of within-drug price changes which are large in many cases. We begin with an exercise that ignores changes in quality and asks: how much of the growth in costs is due to new entrants? Next, we consider both cost and quality changes to examine 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 innovation on spending growth (Chernew and Newhouse, 2011). Given the difficulty of the problem, researchers often follow Solow (1957) and control for measurable drivers of spending (e.g., age, insurance, price, and income), and the residual is attributed to innovation. Instead, for our subset of conditions, we are able to identify and track new treatments directly.

The thought experiment we have in mind is removing all "new" drugs from the market, where treatments not in the market in 2007 are categorized as "new", while the rest are categorized as "old." 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.[46] We keep the "old" drugs at their 2018 prices.[47]

Table 6 presents results. Column 1 presents our baseline results for how much costs grew in our data without differentiating between innovation and within-drug price growth (Table 4). The average cost of rheumatoid arthritis grew by $181k in our data. Column 2 reports the cost growth due to innovation (i.e., we report the difference between the baseline cost growth and the counterfactual without innovation). For rheumatoid arthritis, only $10k 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 due to within-drug price changes, rather than new drugs entering at higher price points.

There is a lot of heterogeneity across conditions. Hypertension had no new drugs in the CEAR data, so none of the cost growth was driven by innovation. Hepatitis C only has new

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

[47] 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 estimate of any welfare gains due to innovation will be understated.

drugs (no old treatment is used in 2018), so innovation accounts for 100% of its cost growth. Costs for 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 6: 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) Δ Consumer Welfare due to Innovation $100k VSLY ($1,000s) | (5) Δ Producer Surplus due to Innovation ($1,000s) | (6) Δ Total Welfare due to Innovation $100k VSLY ($1,000s) |
|---|---|---|---|---|---|---|
| Asthma | 0 | 0 | 0.292 | 0 | 0 | 0 |
| Atrial Fibrillation | 41 | 8 | 0.196 | 71 | 8 | 79 |
| Colon Cancer | -133 | 6 | -0.048 | -5 | 6 | 1 |
| Cystic Fibrosis | 2,009 | 560 | 0.279 | -541 | 560 | 19 |
| HIV | 158 | 80 | 0.505 | -72 | 80 | 8 |
| Hepatitis C | 8 | 8 | 1.000 | 280 | 8 | 288 |
| Hypertension | -3 | 0 | 0.000 | 0 | 0 | 0 |
| Lung Cancer | 308 | 278 | 0.904 | -213 | 278 | 65 |
| Multiple Sclerosis | 950 | 192 | 0.202 | -174 | 192 | 18 |
| Osteoporosis | 5 | 2 | 0.505 | 0 | 2 | 3 |
| Rheumatoid Arthritis | 181 | 10 | 0.057 | -4 | 10 | 7 |
| Schizophrenia | -7 | 3 | -0.415 | 1 | 3 | 4 |
| Venous Thromboembolism | 2 | 2 | 1.049 | 5 | 2 | 7 |
| Aggregate | 18 | 4 | 0.229 | 1 | 4 | 5 |

**Notes:** Column 1 presents cost growth estimates without the counterfactual. Column 2 is 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 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).

While we make no claim that these conditions are representative, we take a quantity weighted average across conditions to compute aggregate measures reported on the bottom of Table 6. Our average cost growth across all conditions is $18k during our sample period. The counterfactual cost growth due to innovation is $4k. Therefore, we find that about 23% of cost growth during this sample period is due to new innovation. This estimate on the lower end of the range in the literature measuring innovation as the residual of cost growth

that cannot be explained by other factors.[48] 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).

As a lot of cost growth is still unexplained, in Appendix Section OA.C.1, we explore the share of cost growth that is due to within-molecule inflation. We find that more than half of cost growth is due to within-molecule price growth, so within-molecule price growth is more important than innovation in driving higher costs.[49]

## 8.2 What Share of Surplus Goes to Consumers and Producers?

One important question in the innovation literature is what share of surplus is 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. In the last three columns of Table 6, we compute welfare applying the same counterfactual as in Section 8.1 and we assume $100k VSLY for each 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.

Recall that for venous thromboembolism, we estimate that consumer welfare increased by $9k over the sample period (Table 4, column 5). In column 4 of Table 8.1 we find that, $5k of consumer welfare is due to the entry of new drugs, the other $4k is due to compositional changes for old drugs and prices declining due to patent expiration. 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 $541k due to the entry of Orkambi.[50] Total surplus increases by $19k, but producer surplus grows by $560k due to Orkambi. Hence, producers received 2,947% 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

[48] 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 CEAR studies and cost growth by condition to determine the share of growth due to innovation.

[49] Changes in non-drug costs account for the remaining portion of cost growth.

[50] In Column 5 of Table 4 we find that consumer welfare for cystic fibrosis falls by $1.99m. The remainder of the reduction in consumer surplus is because other cystic fibrosis drugs raise their prices considerably during our sample period.

fibrosis. Six of our 13 conditions have lower consumer welfare in 2018 because of those new entrants (and we see this with Sovaldi in 2014, in Table 2 as well).[51]

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, formulary design, uninformed consumers, or provider incentives (or physician detailing), may lead consumers to purchase a drug 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, all of the cost effectiveness studies for Orkambi in the CEAR data find that it is not cost effective at any conventional VSLY, yet Figure 4 shows that Orkambi diffuses broadly. 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 estimate lower consumer welfare have studies that show that they are not cost-effective.

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. 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 100% is not an upper bound, if distortions lead higher quality products with negative consumer surplus to diffuse.

Second, our results have important implications for pricing policy. Some countries in the Organization for Economic Cooperation and Development (OECD) and notably the United Kingdom's National Institute for Health and Clinical Excellence (NICE) often restrict medicines if they 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

[51]It is 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.

(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).[52]

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 received 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 how prices and quality change for colon cancer treatments from 1993-2005 when there are numerous new higher quality 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.

[52]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.

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 quality 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 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


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.

Brown, C., S. Sawyer, and D. Bathgate (2020). A review of hedonic price adjustment techniques for products experiencing. Technical report, Bureau of Labor Statistics.

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.

Castanheira, M., C. Ornaghi, and G. Siotis (2019). The unexpected consequences of generic entry. *Journal of Health Economics 68*, 102243.

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.

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.

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.

Duflos, G. and F. R. Lichtenberg (2012). Does competition stimulate drug utilization? the impact of changes in market structure on us drug prices, marketing and utilization. *International Review of Law and Economics 32*(1), 95–109.

Dunn, A., L. Fernando, and E. Liebman (2023). A direct measure of medical innovation on health care spending: A condition-specific approach. *Health Management Policy and Innovation 8*(2).

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.

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.

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.

Lakdawalla, D., A. Malani, and J. Reif (2017). The insurance value of medical innovation. *Journal of public economics 145*, 94–102.

Lu, Z. J. and W. S. Comanor (1998). Strategic pricing of new pharmaceuticals. *Review of economics and statistics 80*(1), 108–118.

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.

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.

Pakes, A. (2003). A reconsideration of hedonic price indexes with an application to pc's. *American Economic Review 93*(5), 1578–1596.

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.

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.

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.

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.