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

Authors Abe Dunn and Lasanthi Fernando, Bureau of Economic Analysis; Eli Liebman, University of Georgia[1]

Contact abe.dunn@bea.gov; lasanthi.fernando@bea.gov; Eli.Liebman@uga.edu

Date December 2024

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.

Keywords Innovation, price index, quality adjustment, medical spending

JEL Code E31, O30, I10

1. 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.

1. Introduction

U.S. health care spending rose from 5% of GDP in 1960 to 17.7% in 2019 (Hartman et al., 2021). Over this same period, life expectancy in the U.S. 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.[2]

At the heart of the debate is a measurement challenge; 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.[3] 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). This literature has made a substantial contribution towards our understanding of the productivity of medical care spending (Cutler et al., 2022). However, the outcomes-based approach does not identify which technologies are driving the associated changes in outcomes and costs, so it cannot directly measure how innovation impacts these markets. Yet, the pharmaceutical industry spends $120 billion per year on R&D and it is estimated that about 35% of medical spending growth is due to innovation, so understanding how medical innovation impacts welfare is important (Austin and Hayford, 2021; Smith et al., 2022).

Our paper takes a unique approach to measuring 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. We then match these treatments with the medical claims data of millions of commercially-insured individuals to measure how these treatments diffuse. This newly combined dataset contributes a novel and rich source of information for understanding and measuring cost growth and innovation in the health care sector.

Measuring which technologies diffuse and their quality and cost is critical for examining a number of important questions related to medical innovation: Do quality improving innovations diffuse or is the diffusion driven more by the cost-effectiveness of the innovations? Do innovations improve consumer welfare? Societal welfare? And by how much? How much incentive do producers have to innovate? How do these effects change over the life cycle of innovations, especially as prices change due to generic entry? Is the spending on R&D worth the societal benefit? How much do these answers vary across treatments, conditions, and the degree of innovation? This paper is the first that we are aware of that

[2] 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.

[3] These distortions which complicate revealed preference approaches include: insurance coverage insulating patients from risk (moral hazard), insurers distorting demand through formulary design, principal agent problems, and imperfect information (Aizcorbe and Nestoriak, 2012; Dauda et al., 2022).

systematically examines these questions for a large number conditions and innovative treatments.

A cost-effectiveness study typically evaluates a medical innovation against a standard of care treatment. To answer these broader questions, we aggregate across individual studies by classifying treatments to match across studies, estimating average quality for each treatment, and linking treatments to insurance claims data from millions of individuals to track their diffusion. We then apply the framework from Cutler et al. (1998) to translate our cost and quality measures into consumer welfare and quality-adjusted price indexes. In this approach, consumer welfare depends on (1) health outcomes from medical care, measured by combining CEAR quality metrics with diffusion measured in insurance claims data, (2) the value of a statistical life year, for which we present results for a range of assumptions, and (3) the costs of care, which we measure using insurance claims and adjust for rebates. Consistent with the literature, costs represent insurer and patient payments, with "cost" and "spending" used interchangeably in this paper. In this framework, if consumer welfare rises for a condition, the quality-adjusted price of treatment falls, 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 innovations and changes in the quality of treatment. 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 in medical spending annually, or 8%, of total medical expenditure and 14% of pharmaceutical expenditures in 2018.[4]

Our approach yields a number of novel insights into the important questions outlined above. We find that (1) higher quality innovations generally diffuse so measures of inflation are likely overstated; (2) however, in many instances, innovations are priced so high that they are not cost effective and can lower consumer welfare, but still diffuse due to insurance; (3) producers can capture more than 100% of the surplus they create (in the short run); (4) but in the long run, consumers are better off due to generic entry; (5) and it appears that the value created for society is still considerably larger than the R&D costs of innovation. Common theoretical assumptions would rule out some of these findings, which highlights the importance of documenting these empirical facts.

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 $100,000, implying declining quality-adjusted price indexes. If we assume a VSLY of $500,000, 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

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

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 $100,000, rises by 47% — a reduction of 1.6 percentage points in the compound annual growth rate. If we assume a $500,000 VSLY, the quality-adjusted price index *falls* by 65%. 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 in official health care prices from the U.S. Bureau of Labor Statistics (Bosworth et al., 2018).

Surprisingly, several of the markets where consumer welfare *declined* experienced 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: 0.9 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 $150,000 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 $1,500 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).[5]

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 surplus (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. In theory, this could lead to socially inefficient over-investment in costly research and development (R&D). However, in a back-of-the-envelope calculation we find that welfare gains from innovation are an order of magnitude higher than the costs of R&D estimated in DiMasi et al. (2016).

[5] Many drugs for these conditions are not viewed as cost-effective by the literature. As Jena and Philipson (2008) argue, cost-effectiveness is related to consumer welfare. Therefore, the result that consumer welfare is declining is consistent with the cost-effectiveness literature, which finds that many of these drugs that are diffusing are not cost-effective.

While the focus of our paper is on how innovations shape markets, non-innovative markets demonstrate how these markets mature, including eventual patent expiration, providing a more long-run and complete view of how innovations impact 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 in 2018 than in 2007.

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.

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.[6] 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

[6] See Newhouse (1992)Smith et al. (2009), and Smith et al. (2022).

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.[7]

Our paper also relates to recent work by Cutler 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) and abstract from the often non-linear process by which treatment impacts health (Chernew and Newhouse, 2011). On the other hand, the outcome-based approach requires observable outcomes and strong assumptions regarding whether the observed change in health is attributable to improvements in medical care. It also does not link specific innovations to health outcomes, which is essential for our paper.[8] In addition, the outcome-based approach may better capture quality improvements for conditions where outcomes (e.g., mortality or disability) are easier to measure, whereas our approach may better capture conditions where treatments may improve the quality of life, rather than lengthening life. Hence, we view our paper as complementary to these outcomes-based papers. However, our paper is focused on a different question of how specific innovations and the adoption of technologies affects welfare and incentives to both produce new innovations.

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. To produce an aggregate quality-adjusted price change Dunn et al. (2022) imposes strong assumptions , including an assumption that all technologies diffuse at a constant rate. This is an important assumption as the types of technologies adopted (e.g., more or less cost-effective technologies), the size of the market (e.g., heart disease or hemophilia), and the speed and depth by which they are adopted is central to the policy discussion on the cost and benefits to consumers, the incentives of producers to innovate, and the societal value of innovation. Stated differently, those 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, allowing us to systematically examine a number of important questions that would not be possible using a top-down approach.[9]

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

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

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

3. Background

3.1. The Case of Hepatitis C and Rheumatoid Arthritis

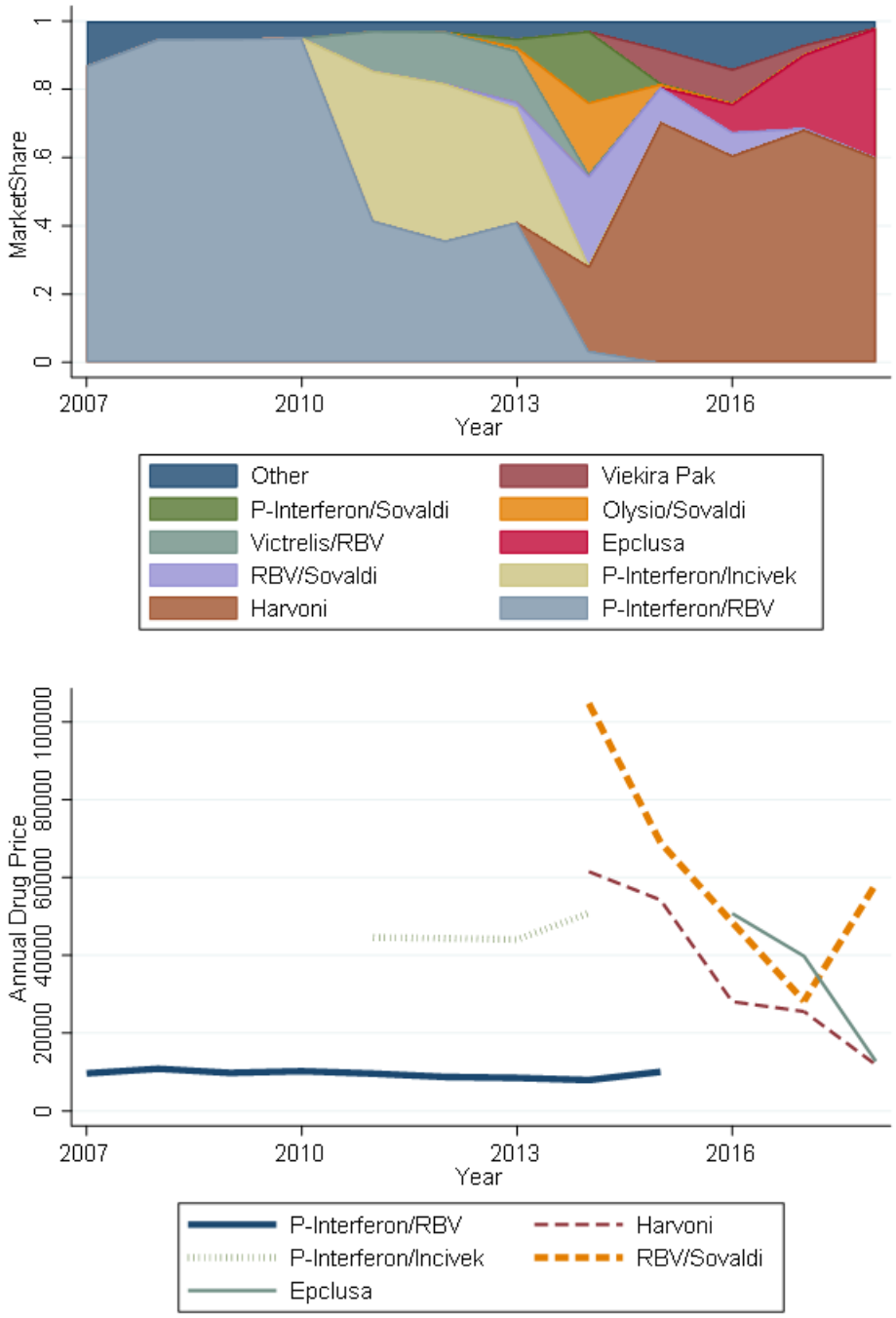
While we construct our index for 13 conditions, we begin our exposition with a focus on hepatitis C and rheumatoid arthritis as they both experienced considerable cost growth during our sample period, but the dynamics in each market are different in ways that help demonstrate how our methodology works and some of the main takeaways.

Hepatitis C is a viral infection that can cause inflammation of the liver. The condition is 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, a biopharmaceutical company, 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 P-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



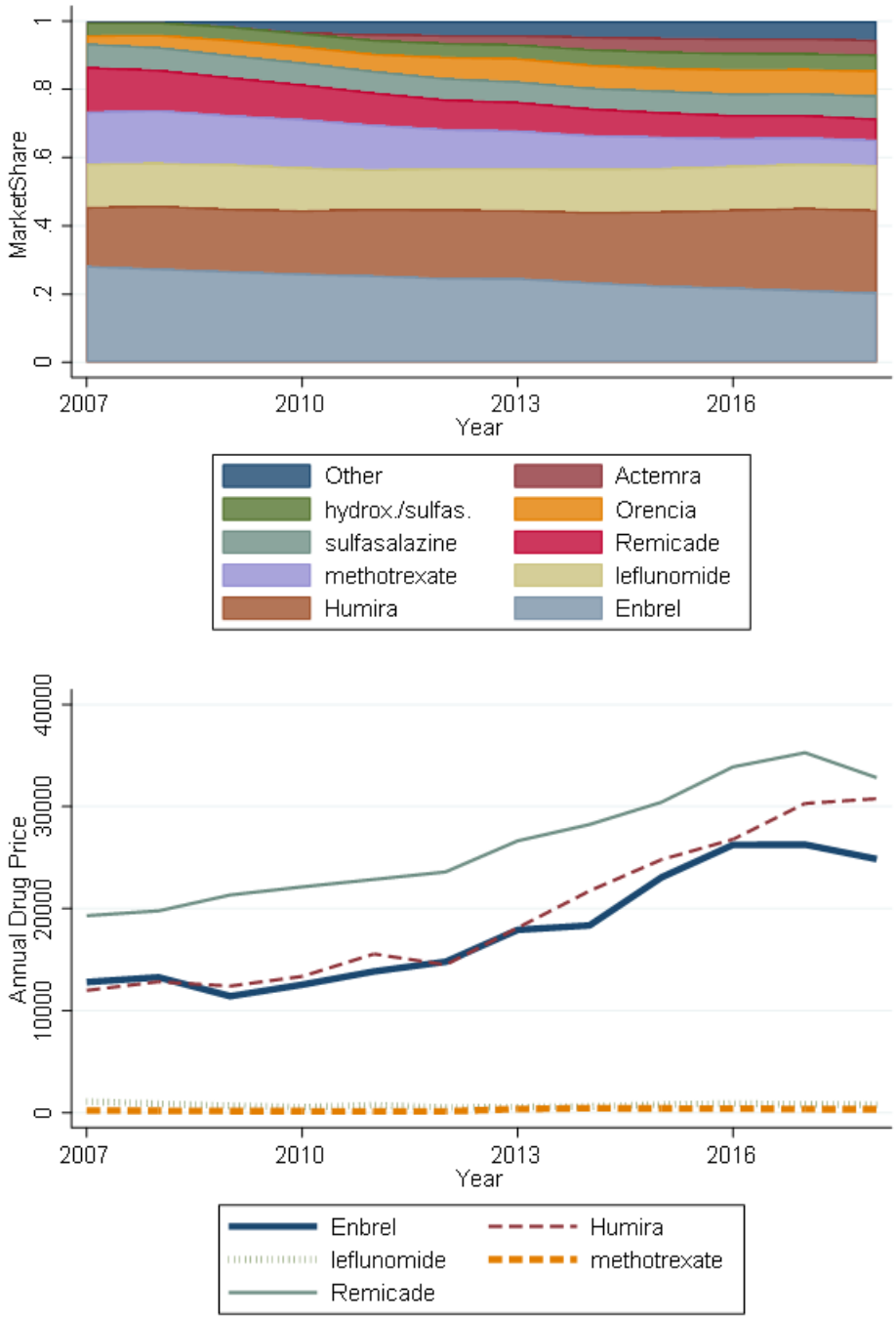
Notes: Estimates are derived from MarketScan claims data described in the data section. The top panel of this figure presents the patient-weighted market shares by year for the nine highest volume drugs for hepatitis C across our entire sample period in the MarketScan data. The bottom panel presents the average price per year of the five 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 Personal Consumption Expenditure (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 deterio-

ration. There are about 1.3 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 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 is 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 received 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 top panel of this figure presents the patient-weighted market shares by year for the nine highest volume drugs for rheumatoid arthritis across our entire sample period in the MarketScan data. The bottom panel presents the average price per year of the five 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.[10] 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 have 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.

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.

[10] The House Committee on Oversight and Reform released a report which highlights a few reasons why prices might rise so rapidly. Potential reasons include executive compensation being tied to revenue targets and firms using the other firms' 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. See "House Committee Takes On Pricing, Patents for Top-Selling Drug Humira," Wall Street Journal, 2021.

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.[11]

As discussed in Fisher and Shell (1972), a utility-based cost-of-living price index measures the relative expenditures needed to maintain the same level of utility across periods, given changes in prices, and in our case, quality. This idea connects directly to the cost-effectiveness discussion in the previous section, but instead of calculating the consumer welfare from switching away from the standard of care treatment, SOC, to the innovative treatment, I, we calculate the consumer welfare of receiving a typical treatment at a point in time, $t-1$, relative to treatments received at time, t.[12] Changing the subscripts in Equation 1 accordingly, we obtain the following equation for the consumer welfare change over time:

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

Δ Consumer Welfare$_{t,t-1}$ accounts for the change in the price and quality of treatment.[13]

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

[11] 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.

[12] 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.

[13] 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.C.2, 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.

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, $\frac{VSLY \cdot (H_{d,t} - H_{d,t-1})}{S_{d,t-1}}$.

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 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. 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, and country the study was performed in, among other characteristics.

Although the CEAR data contain detailed information, they are 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.[14]

To focus our analysis, we concentrate on 13 conditions where pharmaceuticals comprise most of the treatment (or innovation), as pharmaceuticals are much easier to classify in the CEAR and to merge to claims data, relative to procedures or other services.[15] 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.B.2 uses the Medical Expenditure Panel Survey (MEPS) data to explore the share of total drug spending associated with drugs we classify using the CEAR data. The drugs in the CEAR data account for at least 79% of MEPS drug spending for all conditions except atrial fibrillation (60%).[16]

We keep all comparisons where two classified drugs are compared to each other,[17] and both drugs have a non-outlier QALY estimates.[18] In particular, we drop 2,232 comparisons where either cost or QALY information is missing, 197 observations with outlier costs or QALYs, and 1,522 comparisons where one or both of the treatments is not classified. In addition, there are 375 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 1,088 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 at least 360 days. This accounts for 220,658,074 member-years. There are two types of claims files, medical claims and pharmacy claims. Medical claims have information on the diagnosis (characterized by International Classification of Diseases 9th edition (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 National Drug

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

Code (NDC) code (which incorporates a molecule-manufacturer-dose-form). Pharmacy claims do not have diagnosis or procedure codes.

To account for manufacturer rebates, we supplement the MarketScan data with SSR Health data, which has also been used by Kakani et al. (2020) to adjust for rebates.[19] See our data appendix, Section OA.B, 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.[20] 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 (e.g. we reshape the data so there are two observations for each comparison).

Each observation also corresponds to a given treatment r and disease d. Denote the set of treatments used for disease d as $r \in \mathcal{R}_d$. Many different studies may contain a common treatment (e.g., Sovaldi appears in multiple observations), and there are many studies for a given disease (e.g., there are many comparisons and treatments for hepatitis C). To average across quality measures for specific treatments, we use a linear regression model, that allows us to control for the different features of each study. The specific regression is:

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

where the dependent variable is the log of the QALY. The $\gamma_{r,d}$ and $\gamma_{u,d}$ are treatment and comparison-

[19] 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.

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

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.B.4 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. 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.[21] 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, so an innovation that offsets medical care costs will be accounted for in our cost estimates. However, if there are quality improvements in inpatient and outpatient care over time, our

[21] 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. We also try a robustness check where we allocate all drug claims to medical conditions to pick up drugs that are not in the CEAR. Our results do not change much.

results will understate true quality changes.

While we compute average annual costs from the claims data, the QALYs in the Tufts are typically measured in lifetime units. To keep units between costs and benefits consistent, we re-scale 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 costs 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%. 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. See Appendix Section OA.B.5 for a more thorough discussion of this calculation.

7. Results

To demonstrate our methodology 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 drugs' 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.[22] Finally, the newest generation of drugs — Harvoni, Viekira Pak, and Epclusa — are the most effective, each providing 2.7 more QALYs compared to P-Interferon/RBV.

[22] "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.

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

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
Epclusa	0	2.687
Harvoni	0	2.764
Olysio/Sovaldi	0	2.797
P-Interferon/Incivek	0	1.350
P-Interferon/RBV	1	0.000
P-Interferon/Sovaldi	0	2.252
RBV/Sovaldi	0	1.169
Victrelis/RBV	0	0.801
Viekira_Pak	0	2.763

	(1) Is Baseline Treatment	(2) Δ QALYs from Baseline
Actemra	0	3.119
Enbrel	0	2.122
Humira	0	2.262
Orencia	0	2.602
Remicade	0	1.880
hydrox./sulfas.	0	2.134
leflunomide	0	0.048
methotrexate	1	0.000
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 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 appear to be similar in magnitude to the QALY improvements we see from hepatitis C drugs, relative to the baseline treatment.

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

Quality and price index trends for hepatitis C are shown in Table 2. Column 1 of Table 2 calculates how quantity-weighted QALYs are changing over time for the treatment of hepatitis C, relative to 2007. In 2011, when Incivek and Victrelis enter, the average treated hepatitis C patient receives 0.79 more QALYs than they would have in 2007. In 2014, with the emergence of Sovaldi, that number jumps to 2.2 QALYs and in 2018 it is 2.8 QALYs after the entry of Harvoni.[23]

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 $41k. 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,

[23]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.

then they rise to roughly $100,000 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.[24] 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.

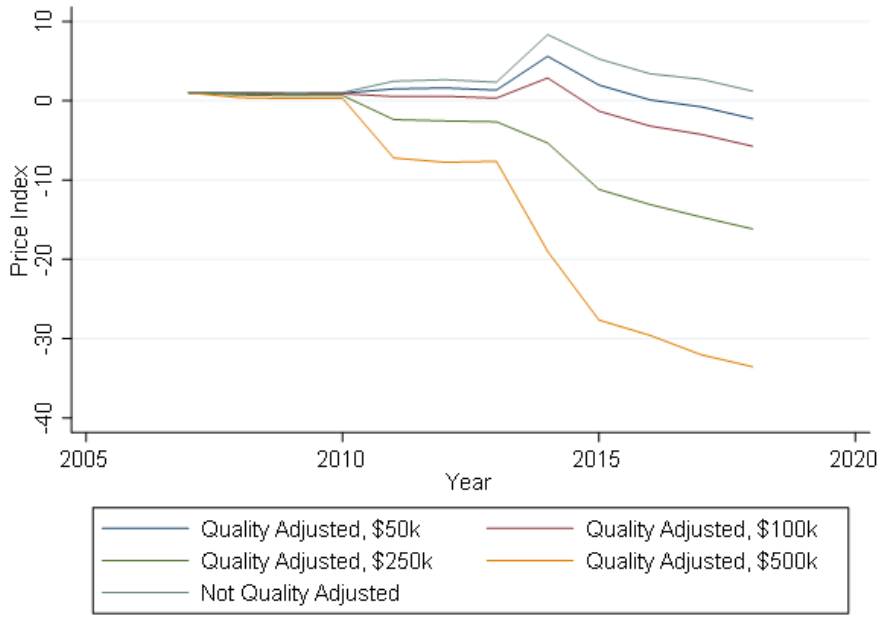
Table 2. Price Indexes and Changes in Welfare by Year for Hepatitis C

	(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	0.056	44	1.075	0.937	3	0.385	25
2009	0.057	40	0.987	0.847	6	0.288	29
2010	0.060	42	1.029	0.882	5	0.295	29
2011	0.791	101	2.467	0.530	19	-7.219	336
2012	0.850	108	2.653	0.572	17	-7.754	357
2013	0.814	95	2.329	0.336	27	-7.636	353
2014	2.231	340	8.337	2.872	-76	-18.987	816
2015	2.689	215	5.277	-1.310	94	-27.658	1,170
2016	2.695	139	3.405	-3.196	171	-29.599	1,249
2017	2.837	111	2.707	-4.243	214	-32.043	1,349
2018	2.838	49	1.204	-5.746	275	-33.548	1,410

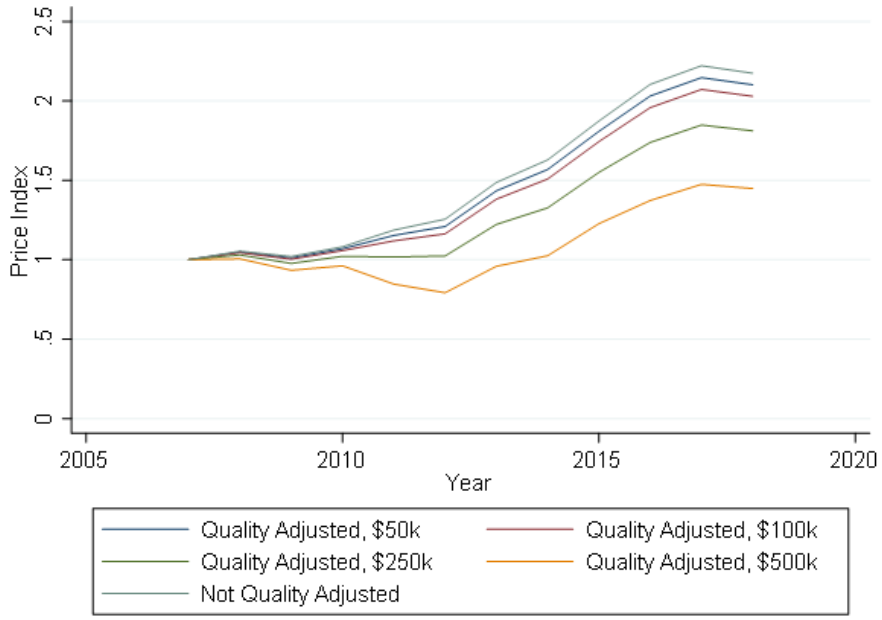
Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes, and consumer welfare, constructed using the CEAR, MarketScan, and SSR Health datasets. Column 1 presents the difference in average QALYs relative to 2007. Column 2 presents estimated lifetime costs in each year. Columns 3, 4, and 6 present price indexes assuming the value of a statistical life year (VSLY) is $0, $100,000, and $500,000, 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.

[24] Our preferred specification averages across all years when determining the lifetime cost multiplier. As a robustness check, we find that allowing for cost dynamics to change over time leads to lower cost growth for hepatitis C treatments. This reflects the fact that newer treatments are curative, so future maintenance costs decline after 2014. We do not find convincing trends for other conditions and worry that making this more flexible specification our preferred specification would add considerable noise. Results of this robustness check are available upon request.

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 VSLY. A subset of these indexes are also shown in Tables 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

$100,000 VSLY. In addition, price indexes for $50,000, $100,000, $250,000, and $500,000 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 2.69 QALYs higher than in 2007 when most patients were receiving interferon based treatments (such as P-Interferon). At a $100,000 VSLY, this represents $269,000 of welfare. Given the $174,000 difference in average costs, this represents a $95,000 gain in consumer welfare. Likewise, the index is $\frac{215-100\cdot 2.69}{41} = -1.31$, 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 $500,000, the index becomes $\frac{215-500\cdot 2.69}{41} = -27.6$. 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 $100,000 VSLY, but at a 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 (Enbrel) and early 2000s (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.[25] As the previous results make clear, this is driven by the lack of diffusion of these newer drugs, rather than the lack of efficacy of these treatments; the newer generation of rheumatoid arthritis drugs have similar relative QALYs as the newest generation of hepatitis C drugs. Column 2 shows that the average lifetime cost in 2007 for a patient with rheumatoid arthritis is $154,000. Lifetime costs doubled during our sample period to $336,000. As discussed in subsection 3.1, this is due to large within-molecule price increases. Drugs like Enbrel and Humira more than doubled their prices during our sample period.

[25]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.

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	0.847	24
2012	0.143	194	1.256	1.163	-25	0.792	32
2013	0.163	230	1.487	1.381	-59	0.959	6
2014	0.186	251	1.629	1.508	-78	1.025	-4
2015	0.199	289	1.873	1.744	-115	1.227	-35
2016	0.226	325	2.104	1.958	-148	1.373	-58
2017	0.231	343	2.222	2.072	-166	1.475	-73
2018	0.224	336	2.174	2.029	-159	1.449	-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 VSLY is $0, $100,000, and $500,000, 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.

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 $100,000 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 for the remaining 11 conditions are available upon request.

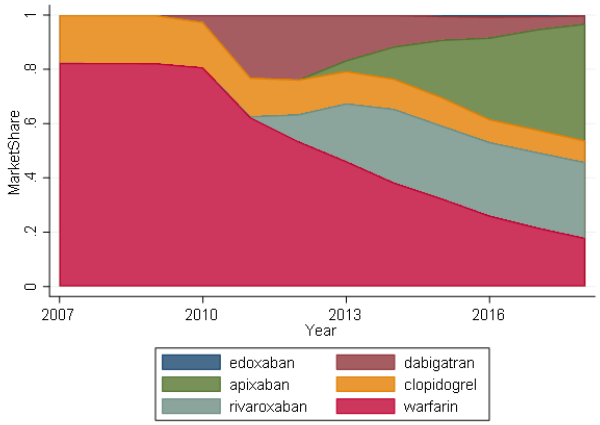
Table 4. Price Indexes and Changes in Welfare for Each Condition Between 2007 and 2018

	(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	1.001	0	0.946	0
Atrial Fibrillation	0.443	14	3.854	0.771	3	-11.564	44
Colon Cancer	-0.040	338	0.607	0.619	129	0.666	-4
Cystic Fibrosis	0.232	622	4.232	4.194	-1,986	4.045	23
HIV	0.167	312	1.505	1.452	-141	1.237	17
Hepatitis C	2.838	41	1.204	-5.746	275	-33.548	284
Hypertension	0.039	9	0.684	0.230	7	-1.587	4
Lung Cancer	0.620	267	2.151	1.919	-246	0.991	62
Multiple Sclerosis	0.419	476	2.998	2.910	-908	2.558	42
Osteoporosis	0.033	7	1.690	1.201	-1	-0.755	3
Rheumatoid Arthritis	0.224	154	2.174	2.029	-159	1.449	22
Schizophrenia	0.113	38	0.823	0.522	18	-0.680	11
Venous Thromboembolism	0.102	6	1.308	-0.316	8	-6.810	10

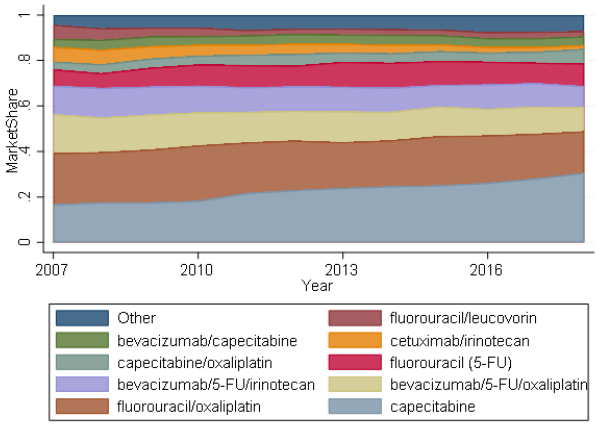
Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes, and consumer and total welfare, constructed using the CEAR, MarketScan, and SSR Health datasets. Columns 1 and 2 presents the difference in average QALYs relative to 2007 and the estimated lifetime costs in 2007 for each condition. Columns 3, 4, and 6 present price indexes assuming theVSLY is $0, $100,000, and $500,000, 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 six selected conditions. Hypertension and colon cancer 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.

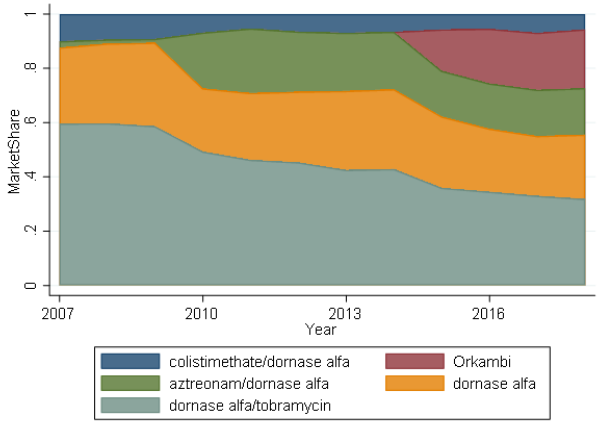
Figure 4. Market share for top treatments over time



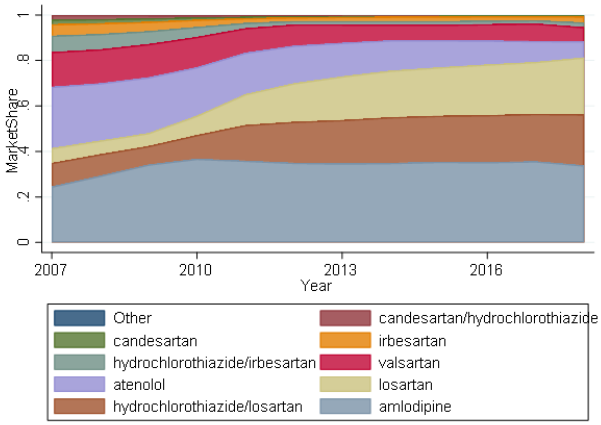
a. Atrial Fibrillation



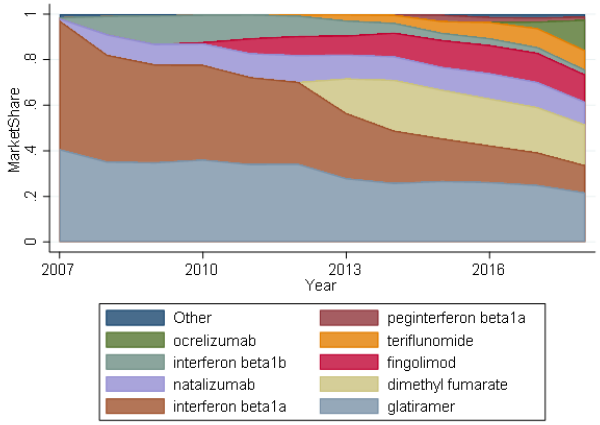
b. Colon Cancer



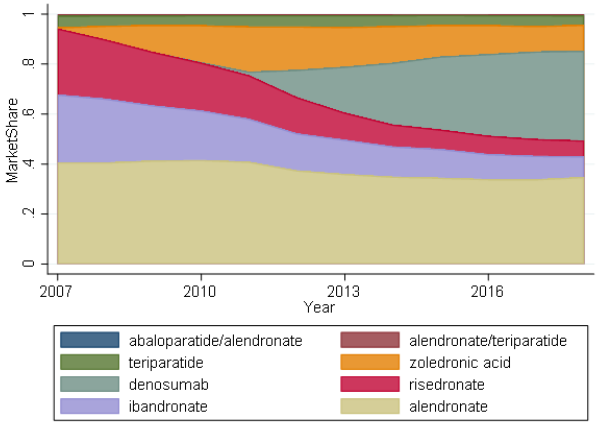
c. Cystic Fibrosis



d. Hypertension



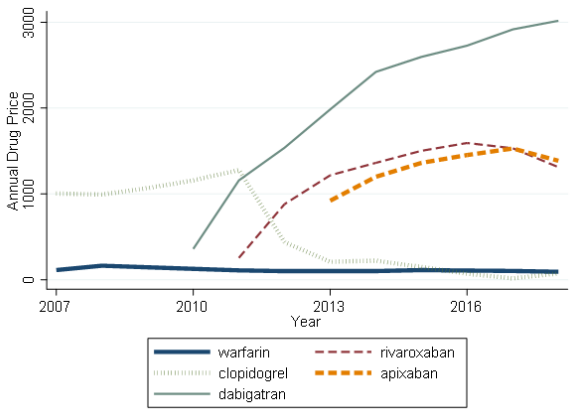
e. Multiple Sclerosis



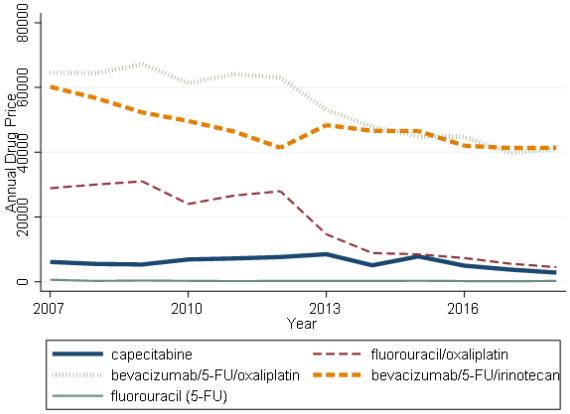
f. Osteoporosis

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.

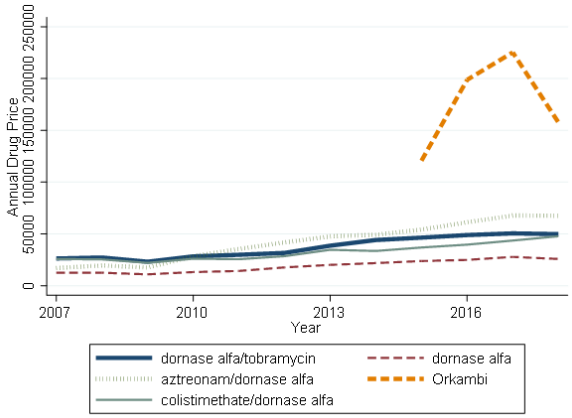
Figure 5. Prices for the Top Five Treatments for Selected Conditions



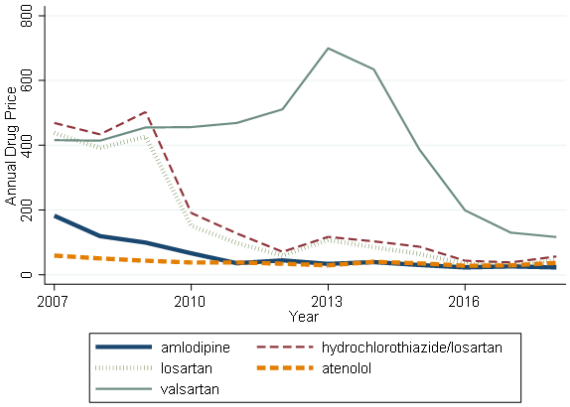
a. Atrial Fibrillation



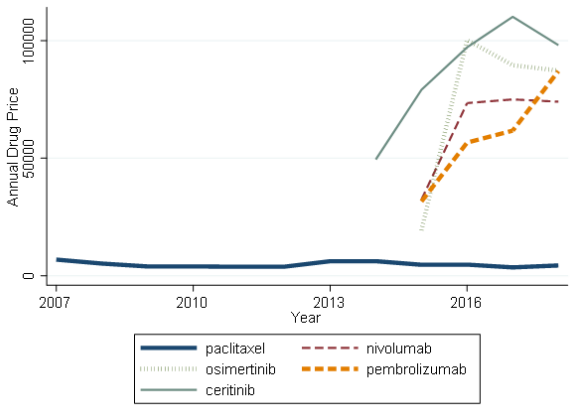
b. Colon Cancer



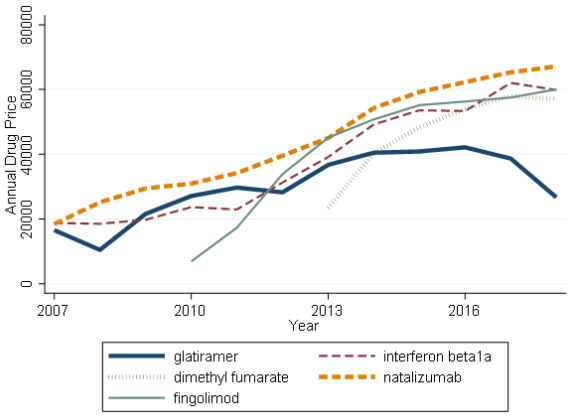
c. Cystic Fibrosis



d. Hypertension



e. Lung Cancer



f. Multiple Sclerosis

Notes: Figure 5 presents the average price per year of the five 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: (i) the condition needs to have new treatments which make large improvements in quality; and (ii) 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).[26]

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 five 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 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 that 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 0.4 QALY improvement over warfarin. Indeed, in 2019, the American College of Cardiology and the American Heart Association recommended these newer anticoagulants as the preferred drug class over warfarin (January et al., 2019), which can also be seen in their large increases in market share (Figure 4). Because these treatments cost $1,000-3,000, 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 $10,000-$40,000 (assuming $100,000 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

[26] 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.

driven by a very high cost entrant, Orkambi. Orkambi was controversially priced at least $150,000 per year and has taken over about 20% market share by 2018.[27] Because of this, costs for cystic fibrosis quadruple in our sample period, where we estimate lifetime costs are over $2.5 million by 2018.[28] However, Orkambi was viewed as a breakthrough therapy and indeed we estimate that it adds 0.89 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. 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.[29]

Lung cancer, multiple sclerosis, and HIV follow a similar pattern to cystic fibrosis.[30] They are clearly innovative with multiple new entrants yielding large quality improvements. However, these new entrants are very expensive. For these cases, despite the quality improvements (rising QALYs), the costs are rising rapidly enough that quality-adjusted prices are still rising and consumer welfare is falling.

While we caution against extrapolating outside the sample, in Appendix Table OA1, we aggregate across conditions, weighted by spending. The unadjusted price index rises by 75%, while the $100,000 VSLY index rises by 47% — a reduction of 1.6 percentage points from the compound annual growth rate. At $500,000 VSLY, the index falls by 65%.

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 the online appendix.

QALY estimates: There are a number of reasons why we may be overstating or understating true quality changes.[31] We think the most sensible approach is to see how our results change if we assume

[27] 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 $150,000 per year.

[28] 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.

[29] 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 0.232. 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. In the introduction, the calculation is specifically focused on Orkambi. 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.

[30] 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.

[31] 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

we are off by a factor of 2. Specifically, we multiply our estimated QALYs by two (Table OA2) or by one half (Table OA3). 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 QALY estimates.

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 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 OA4). While our main estimates focus on the primary treatment class, we include multiple classes of treatments for conditions and results are similar.[32] We also do a number of robustness checks on the CEAR quality regressions (Equation 4) where we run the regression a number of different ways (e.g. in levels, not controlling for heterogeneity, etc.) (Table OA7). 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 375 additional comparisons to the regressions (Table OA7). None of the robustness checks mentioned in this paragraph change our results considerably.

One 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 or high medical bills) and health risk (people get sick, which lowers utility). Entry of new treatments can impact the size of these risks and the overall effect is ambiguous (expensive innovations may reduce health risk, but increase financial risk). However, risk is not accounted for in our baseline model, which is standard in the cost-effectiveness and the quality-adjusted price index literatures. In Appendix Section OA.C.2, 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 due to more choice, it remains the case that innovation appears to be lowering consumer welfare for several conditions.

Another criticism of cost-effectiveness studies, raised by Lucarelli et al. (2022) and others, is that cost-effectiveness studies may understate the value of new treatments if there is heterogeneity in preferences or differences in treatment effectiveness. For example Sovaldi was most effective for only certain genotypes. Many of the studies in the Tufts focus on the subgroups where the new treatments are most effective (e.g. a specific genotype of hepatitis C or those where methotrexate was uneffective for rheumatoid arthritis). If those are the only patients who are taking the new treatments (and pure preference heterogeneity is minimal) then our results should be accurate – the lack of diffusion among subgroups

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.

[32] 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.

which the new drug does not target would be reflected in no change in health for that subgroup. If a broader range of patients are taking the new treatments, we may be overstating consumer welfare gains. If consumer taste is important or the Tufts' studies average over all individuals including less efficacious subpopulations, then we may be understating the consumer welfare gains. We have tried robustness checks (available upon request) where we increase welfare proportional to the number of new entrants, reflecting additional heterogeneity in treatment effectiveness. Even with relatively large assumptions about the welfare gains, we still find consumer welfare falling for many innovative conditions.

Cost estimates: In Appendix Section OA.B.5, we test the sensitivity to our lifetime cost estimates. While we try a number of different modelling choices, the widest range of estimates are observed when 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.[33] The upper bound (constant costs) is 23-28 times the lower bound (one year of costs), so this is a very wide range. In both cases, we still 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.

Another concern is whether new innovations offset other costs, reducing future costs either for that same condition or other conditions.[34] For example, innovations for osteoporosis may reduce costs associated with broken bones. We check for cost offsets in a number of ways and results are available upon request. First, we use the MarketScan data to check whether future costs change over time. For example, has the five year cost of having hepatitis C changed over time? As another alternative, we use a regression methodology from Trogdon et al. (2008) which allocates spending across conditions. The methodology allocates spending across all conditions, regardless of the specific diagnosis code on a claim, so if a drug for osteoporosis reduces the average cost of an osteoporosis patient (e.g., through fewer or less severe broken bones for these patients), then the methodology would suggest the cost of having osteoporosis is falling over time. Finally, we look at studies published by the Institute for Clinical and Economic Review (ICER) for evidence of cost offsets. We find limited evidence of cost offsets for most of the conditions. However, the first methodology and the ICER reports suggest that there are large cost offsets for hepatitis C. These drugs are curative and reduce long-term costs considerably. For the remaining conditions, we do not find compelling evidence of offsets using the first two methodologies. The ICER reports do include some cost offset for osteoporosis drugs and reduced fracture expenditures and reduced expenditures for bleeding associated with replacing warfarin with apixaban and rivaroxaban. These results suggest that our cost estimates for hepatitis C, osteoporosis, and atrial fibrillation may be overstated, but these are mostly conditions where we are finding that consumer welfare is already increasing.

[33] 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.

[34] We include non-drug spending in our cost estimates. Therefore, if a new innovation reduced non-drug spending within the condition, our main estimates would account for that cost offset.

Finally, in Appendix Section OA.C.3 we explore how market expansion impacts our results. For example, if a new drug enters or prices fall due to patent expiration, this may cause more consumers to be treated. We find little evidence of market expansion when drugs lose patent protection. We find more evidence of market expansion for innovative conditions. However, the impact on consumer welfare is mixed. Consumer welfare is lower in markets where highly effective new drugs enter that are not cost-effective. This lowers consumer welfare, though it increases the number of QALYs produced. In other cases consumer welfare increases, such as when new treatments are cost-effective, relative to no treatment. Market expansion has interesting dynamic effects on welfare for hepatitis C where an earlier wave of curative treatments reduces the number of patients treated in 2018. This surprisingly reduces the number of QALYs produced in 2018. Averaging across all these cases, our main results are qualitatively quite similar to results that account for potential expansion effects.

7.4. Total Welfare, Producer Surplus and Long Run Effects

While we find consumer welfare is falling for many innovative conditions, we 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. We 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, the interpretation of cost estimates is that we are measuring the costs of these new treatments after entry. Others, such as DiMasi et al. (2003, 2016), estimate the cost of developing new treatments, which we discuss later.

Adding together consumer welfare from Equation 2 and producer surplus from Equation 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.[35] [36] We think this assumption is a lower bound on marginal

[35] 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.

[36] 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).

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 Table OA5, 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.[37]

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 $100,000 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).[38] 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 figures 4 and 5). While this is exemplified by colon cancer, it is not the only case where this pattern emerges. Hypertension had lower average QALYs from 2008-2011 relative to 2007 because relatively lower quality amlodipine's patent expired in 2007. Its price fell and it gained market share (panel (d) of figures 4 and 5). 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 gained 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 OA6 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.

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

[38] 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. In Appendix Section OA.C.3, we allow for market expansion and we still find negative total welfare for colon cancer. 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).

7.5. Which Treatments Are Driving Our Results?

In this section we ask: which treatments are responsible for the biggest changes in consumer 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.[39] Specifically, for newly obtained market share (diffusion) we replace the 2007 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.

The top panel of Table 5 shows the drugs which contributed to the biggest consumer welfare increases assuming a VSLY of $100,000. 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.

[39] 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 comparison for newly obtained market share.

Table 5. (a) Drugs which account for biggest consumer welfare gains - $100,000 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 - $100,000 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	0.7	12792	24846
5	Multiple Sclerosis	ocrelizumab	0.7		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.

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.[40]

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.

[40]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.

An example of a new innovation whose costs exceed the benefit of quality improvement is Orkambi. As discussed above, the quality improvement for Orkambi is large, which helps explain why insured individuals would use it, but consumer welfare is highly negative. Ocrelizumab, dimethyl fumarate, and Stribild are additional drugs which we estimate as being high quality, but their costs are large enough that our methodology suggests the costs are not worth the benefits for consumers. To explore this more, we isolate the impact of new innovations in the next section.

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, whereas our earlier results included 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", and 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. We keep the old drugs at their 2018 prices.[41]

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.

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

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 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.

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.[42] 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).[43]

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	67	8	75
Colon Cancer	-133	6	-0.048	-5	6	1
Cystic Fibrosis	2,009	560	0.279	-541	560	19
HIV	158	80	0.505	-73	80	7
Hepatitis C	8	8	1.000	276	8	284
Hypertension	-3	0	0.000	0	0	0
Lung Cancer	308	278	0.904	-216	278	62
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	3
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 $100,000 multiplied by the change in QALYs due to innovation (not shown).

[42] 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.

[43] About half of the total cost growth is due to within-molecule price growth, like Enbrel and Humira which raised their prices substantially. Changes in non-drug costs account for the remaining portion of cost growth.

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.[44] Columns 4, 5, and 6 of Table 6 show the change in consumer, producer, and total welfare due to innovation applying the same counterfactual as in Section 8.1 and assuming $100,000 VSLY for each calculation. Given our constant marginal cost assumption, cost growth due to innovation (Column 2) is also the change in producer surplus due to innovation. We present results using other VSLYs and assumptions about marginal costs in the appendix.

Recall that for venous thromboembolism, we estimate that consumer welfare increased by $8,000 over the sample period (Table 4, column 5). In column 4 of Table 8.1 we find that, $5,000 of consumer welfare is due to the entry of new drugs, the other $3,000 is due to compositional changes for old drugs and prices declining due to patent expiration. Producer surplus rose by $2,000 due to innovation so total surplus due to innovation is $7,000 higher. Therefore, we estimate that producers captured about 29% of the surplus from innovations in 2018 for venous thromboembolism.

For cystic fibrosis, consumer welfare falls by $541,000 due to the entry of Orkambi.[45] Total surplus increases by $19k, but producer surplus grows by $560,000 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 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).[46]

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, 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 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 Orkambi diffuses broadly. Even outside of the context of our model (and assumptions), one should expect to find falling consumer

[44] Philipson and Jena (2006); Jena and Philipson (2008); Garrison et al. (2024) also estimate the share of surplus to consumers and producers.

[45] In Column 5 of Table 4 we find that consumer welfare for cystic fibrosis falls by $1.99 million. The difference is that the counterfactual in this section ignores within molecule price growth, and the remainder of the reduction in consumer welfare is because other cystic fibrosis drugs raise their prices considerably.

[46] 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.

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 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.

In theory, producers capturing more than 100% of the surplus from innovations could lead to socially inefficient levels of R&D spending, as profit incentives could cause firms to over-invest in R&D (beyond the socially optimal level). To explore this more thoroughly, in Appendix Section OA.C.1, we do a rough calculation of aggregate (rather than per-patient) welfare gains per new treatment and extend the benefits over the lifecycle of a new treatment. We find our total welfare estimates are an order of magnitude larger than the R&D costs estimated in DiMasi et al. (2016). This suggests that the benefits of these innovations are well worth the costs of R&D, in the aggregate.

Second, our results have important implications for pricing policy that can impact welfare in the U.S. and worldwide (Ho and Pakes, 2024). Some countries in the Organisation for Economic Co-operation and Development and notably the United Kingdom's National Institute for Health and Clinical Excellence 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 (not total) welfare (Jena and Philipson, 2008). 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).[47]

Finally, 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 surplus 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.

[47]This point also makes clear an important caveat 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.

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. First, we find a lot of heterogeneity in spending growth trends, causes of spending growth, and the amount of quality improvements across the 13 conditions. This speaks to the importance of having a scalable framework that can be applied consistently across conditions. Overall, we find quality improvements for 12 of the 13 conditions and for many conditions the quality improvements are large in magnitude. This suggests that price indexes which do not account for quality improvements may overstate price growth.

Overall, the results raise important questions about how health care markets implicitly value quality improvements, amid numerous market distortions. Similar to other sectors of the economy, we provide evidence that innovation has led to sizeable 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

Aizcorbe, A. and N. Nestoriak (2012). Price indexes for prescription drugs: A review of the issues. *The Oxford Handbook of the Economics of the Biopharmaceutical Industry*, 438.

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.

Austin, D. and T. Hayford (2021). Research and development in the pharmaceutical industry. *Congressional Budget Office*.

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.

Bosworth, B., J. Bieler, M. Kleinrock, E. Koepcke, and E. R. Berndt (2018). An evaluation of the cpi indexes for prescription drugs. Technical report, National Bureau of Economic Research.

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. (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., H. G. Grabowski, and R. W. Hansen (2016). Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics 47*, 20–33.

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, et al. (2020). Are quality-adjusted medical prices declining for chronic disease? evidence from diabetes care in four health systems. *The European Journal of Health Economics 21*, 689–702.

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.

Garrison, L., B. Jiao, et al. (2024). Estimating the allocation of the economic value generated by utilization of all-oral direct-acting antivirals for hepatitis c in the united states, 2015 to 2019. *Value in Health 27*(8), 1021–1029.

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).

Ho, K. and A. Pakes (2024). Evaluating pharmaceutical policy options. Technical report, National Bureau of Economic Research.

Hofmeister, M. G., E. M. Rosenthal, et al. (2019). Estimating prevalence of hepatitis c virus infection in the united states, 2013-2016. *Hepatology 69*(3), 1020–1031.

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, et al. (2019). 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation. *Circulation 140*(2).

Jena, A. B. and T. J. Philipson (2008). Cost-effectiveness analysis and innovation. *Journal of health economics 27*(5), 1224–1236.

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.

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.

Philipson, T. J. and A. B. Jena (2006). Who benefits from new medical technologies? estimates of consumer and producer surpluses for hiv/aids drugs. In *Forum for Health Economics & Policy*, Volume 9. De Gruyter.

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.

Trogdon, J. G., E. A. Finkelstein, and T. J. Hoerger (2008). Use of econometric models to estimate expenditure shares. *Health Services Research 43*(4), 1442–1452.

Online Appendix OA.A Robustness Checks Referenced in the Main Text

Table OA1. Aggregate Results - Price indexes

	(1) Δ Avg QALYs 2018 - 2007	(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 Total Welfare $100k VSLY ($1,000s)
2007	0.000	23,812	1.000	1.000	0	1.000	0
2008	-0.002	23,214	0.975	0.984	0	1.019	-1
2009	-0.009	24,196	1.016	1.052	-1	1.196	-1
2010	-0.005	25,176	1.057	1.077	-2	1.155	-0
2011	0.012	26,257	1.103	1.053	-1	0.853	1
2012	0.021	27,540	1.157	1.068	-2	0.716	2
2013	0.028	31,100	1.306	1.188	-5	0.715	3
2014	0.038	34,049	1.430	1.271	-6	0.637	4
2015	0.044	36,962	1.552	1.367	-9	0.628	4
2016	0.052	39,924	1.677	1.457	-11	0.577	5
2017	0.057	42,287	1.776	1.537	-13	0.580	6
2018	0.067	41,599	1.747	1.467	-11	0.347	7

Notes: This table presents changes in QALYs, costs, quality-adjusted price indexes, and consumer welfare aggregated across all conditions. This table is similar to Table 2 except it presents results for all conditions, rather than hepatitis C. See Table 2 for more details.

Table OA2. Price Indexes and Changes in Welfare for Each Condition Multiplying QALYs by Two

	(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	0.005	16	1.014	0.987	0	0.877	0
Atrial Fibrillation	0.886	14	3.854	-2.313	48	-26.982	89
Colon Cancer	-0.080	338	0.607	0.630	125	0.725	-8
Cystic Fibrosis	0.464	622	4.232	4.157	-1,963	3.858	46
HIV	0.335	312	1.505	1.398	-124	0.969	33
Hepatitis C	5.675	41	1.204	-12.697	559	-68.300	568
Hypertension	0.077	9	0.684	-0.224	10	-3.859	8
Lung Cancer	1.240	267	2.151	1.687	-184	-0.169	124
Multiple Sclerosis	0.839	476	2.998	2.822	-867	2.117	84
Osteoporosis	0.066	7	1.690	0.712	2	-3.199	7
Rheumatoid Arthritis	0.448	154	2.174	1.884	-136	0.723	45
Schizophrenia	0.226	38	0.823	0.222	29	-2.184	23
Venous Thromboembolism	0.205	6	1.308	-1.939	19	-14.928	20

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. This table is similar to Table 4 except it multiplies QALYs by two. See Table 4 for more details.

Table OA3. Price Indexes and Changes in Welfare for Each Condition Multiplying QALYs by One-Half

	(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	0.001	16	1.014	1.008	0	0.980	0
Atrial Fibrillation	0.222	14	3.854	2.313	-19	-3.855	22
Colon Cancer	-0.020	338	0.607	0.613	131	0.636	-2
Cystic Fibrosis	0.116	622	4.232	4.213	-1,998	4.138	12
HIV	0.084	312	1.505	1.478	-149	1.371	8
Hepatitis C	1.419	41	1.204	-2.271	134	-16.172	142
Hypertension	0.019	9	0.684	0.457	5	-0.452	2
Lung Cancer	0.310	267	2.151	2.035	-277	1.571	31
Multiple Sclerosis	0.210	476	2.998	2.954	-929	2.778	21
Osteoporosis	0.017	7	1.690	1.445	-3	0.467	2
Rheumatoid Arthritis	0.112	154	2.174	2.102	-170	1.811	11
Schizophrenia	0.057	38	0.823	0.673	12	0.071	6
Venous Thromboembolism	0.051	6	1.308	0.496	3	-2.751	5

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. This table is similar to Table 4 except it multiplies QALYs by one-half. See Table 4 for more details.

Adding weight to high quality studies, reducing weight to industry affiliated studies:

The CEAR data has a 1-7 measure of study quality, as judged by their readers, where the quality measure depends on whether methods and results were communicated clearly, assumptions were reasonable, and whether sensitivity and subgroup analyses were included. In addition, the CEAR contains a variable that indicates if authors have academic or industry affiliations, and whether the study was sponsored by industry. In Table OA4, we set the weight of each study to its quality score. A study rated as a "7" is weighted seven times as much as study rated as a "1." We also add two points for studies with an author with an academic affiliation and subtract two points if the study had an author with industry affiliation or was sponsored by industry. Results are very similar to the equal weighting results, and are not sensitive to changes in the weighting scheme we use or varying which variables we include.

Table OA4. Price Indexes and Changes in Welfare for Each Condition Increasing Weighting for High Quality Studies

	(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	0.940	0
Atrial Fibrillation	0.470	14	3.854	0.583	6	-12.502	47
Colon Cancer	-0.037	338	0.607	0.618	129	0.661	-4
Cystic Fibrosis	0.233	622	4.232	4.194	-1,986	4.045	23
HIV	0.173	312	1.505	1.450	-141	1.228	17
Hepatitis C	2.819	41	1.204	-5.700	274	-33.319	282
Hypertension	0.039	9	0.684	0.230	7	-1.590	4
Lung Cancer	0.586	267	2.151	1.932	-249	1.055	59
Multiple Sclerosis	0.406	476	2.998	2.913	-910	2.572	41
Osteoporosis	0.036	7	1.690	1.164	-1	-0.938	4
Rheumatoid Arthritis	0.189	154	2.174	2.052	-162	1.562	19
Schizophrenia	0.120	38	0.823	0.505	19	-0.767	12
Venous Thromboembolism	0.107	6	1.308	-0.393	9	-7.196	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. This table is similar to Table 4 except it weights by study quality. See Table 4 for more details.

Table OA5. Counterfactual: Removing All New Drugs - Assuming That Marginal Costs are 20% of Negotiated Prices

	(1) Baseline Cost Growth 2018 - 2007 ($1,000s)	(2) Cost Growth due to Innovation ($1,000s)	(3) Share of Cost Growth due to Innovation	(4) 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	67	6	74
Colon Cancer	-133	6	-0.048	-5	5	0
Cystic Fibrosis	2,009	560	0.279	-541	448	-93
HIV	158	80	0.505	-73	64	-9
Hepatitis C	8	8	1.000	276	7	282
Hypertension	-3	0	0.000	0	0	0
Lung Cancer	308	278	0.904	-216	222	6
Multiple Sclerosis	950	192	0.202	-174	154	-21
Osteoporosis	5	2	0.505	0	2	2
Rheumatoid Arthritis	181	10	0.057	-4	8	5
Schizophrenia	-7	3	-0.415	1	2	3
Venous Thromboembolism	2	2	1.049	5	2	6
Aggregate	18	4	0.229	1	3	4

Notes: This table is similar to Table 6 except we assume marginal costs are 20% of the prices we observe. See Table 6 for more details.

Table OA6. Counterfactual: Price Indexes and Changes in Welfare for Each Condition Between 2007 and 2018, but Simulating Prices After Drugs Go Off Patent

	(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	1.001	0.987	0	0.932	0
Atrial Fibrillation	0.443	14	3.363	0.279	10	-12.055	44
Colon Cancer	-0.040	338	0.581	0.593	138	0.640	-4
Cystic Fibrosis	0.232	622	2.001	1.964	-599	1.815	23
HIV	0.167	312	0.425	0.371	196	0.157	17
Hepatitis C	2.838	41	0.257	-6.693	314	-34.495	284
Hypertension	0.039	9	0.677	0.223	7	-1.595	4
Lung Cancer	0.620	267	0.952	0.720	75	-0.208	62
Multiple Sclerosis	0.419	476	1.013	0.925	36	0.572	42
Osteoporosis	0.033	7	0.699	0.210	5	-1.745	3
Rheumatoid Arthritis	0.224	154	0.901	0.755	38	0.175	22
Schizophrenia	0.113	38	0.764	0.463	20	-0.739	11
Venous Thromboembolism	0.102	6	0.978	-0.646	10	-7.140	10

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. The results are similar to Table 4, except in 2018 we assume that prices declined by 85% for on-patent drugs. This is meant to simulate a "long-run" outcome where these drugs have lost patent protection. Note that we allow non-drug costs to change between 2007 and 2018, so conditions like atrial fibrillation, which have increases in non-drug spending, still see unadjusted prices rising.

Alternative QALY regression estimates — We also estimate QALYs making different assumptions regarding how the QALY regression (Equation 4) is estimated. Table OA7 shows results for price indexes assuming $100,000 VSLY. Column 1 is the baseline result from Table 4. Column 2 does not normalize for heterogeneity in study assumptions, see Appendix Section OA.B.4 for more details of how we normalize for different assumptions that studies make. Column 3 makes the same heterogeneity adjustments as column 1, but does so on the QALYs in the raw data, before running the regression in Equation 4, rather than adjusting the study fixed effects.

In our main specification, we drop studies which say they are a "placebo," "no treatment," "usual care," "standard of care," and "status quo." We do this because it is often unclear what these treatments are, and we worry that these categories will add a lot of noise or biases.[48] However, dropping these categories drops some studies. In our main regression there were 1,088 comparisons. Adding back placebo, no treatment, standard of care, usual care, and status quo increases that number to 1,463. Columns 4 and 5 present results with these comparisons added back in. Column 6 estimates the QALY regressions in levels.

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

Table OA7. $100,000 VSLY Price Index Results With Different CEAR Regression Specifications

	(1) Baseline Result	(2) Don't Normalize Heterogeneity	(3) Normalize Heterogeneity First	(4) Add No Treatment	(5) Add Standard of Care	(6) Estimate Regression in Levels
Asthma	1.001	0.977	1.013	0.997	0.997	1.013
Atrial Fibrillation	0.771	1.228	0.345	0.682	0.598	0.787
Colon Cancer	0.619	0.638	0.618	0.624	0.625	0.601
Cystic Fibrosis	4.194	4.210	4.211	4.193	4.193	4.201
HIV	1.452	1.477	1.472	1.386	1.384	1.457
Hepatitis C	-5.746	-2.467	-3.041	-5.756	-5.801	-4.465
Hypertension	0.230	0.289	0.328	0.179	0.183	0.163
Lung Cancer	1.919	1.096	1.718	1.917	1.911	1.997
Multiple Sclerosis	2.910	2.914	2.913	2.854	2.847	2.934
Osteoporosis	1.201	1.306	1.237	1.293	1.292	1.359
Rheumatoid Arthritis	2.029	2.001	1.197	2.018	2.016	2.059
Schizophrenia	0.522	0.607	0.804	0.508	0.498	0.802
Venous Thromboembolism	-0.316	-0.008	0.275	-0.282	-0.327	0.543

Notes: This table presents estimated quality-adjusted price indexes assuming a $100,000 VSLY using different specifications in our regressions. Column 1 is the baseline result from Table 4. Column 2 does not normalize study heterogeneity. Column 3 normalizes study heterogeneity on the raw QALYs, rather than the study fixed effects. Columns 4 and 5 add additional studies which are less specific about the treatments in the regressions. Column 6 estimates the regressions in levels rather than logs.

Overall, results are qualitatively similar, though there are some differences in magnitudes across these results. Except for asthma and one specification for atrial fibrillation, consumer welfare is consistently falling or rising (the index is above or below 1) for each specification.

Online Appendix OA.B Data and Methods Appendix

OA.B.1 Cleaning and Classifying the CEAR data

We chose the 13 conditions which were associated with the most studies in the CEAR and seemed appropriate for our analysis. To determine if a comparison is related to that condition, we search the disease variable (this variable has names like "hepatitis C" or "rheumatoid arthritis"), the ICD-10 code and chapter descriptors, and the study title for strings that match our condition names.

The key variable in our data is the treatment variable. These are typically a sentence or two long. We tasked multiple research assistants to classify each treatment in the CEAR data to a specific molecule. Each treatment was classified by two research assistants to ensure accuracy.[49] We ignore variations within a molecule like dose (5 mg vs. 10 mg), form (injectable vs. oral), frequency of treatment and treatment length as these are not straightforward to map into claims data and are not consistently reported in the CEAR data. Often pharmaceutical treatments are vague, only listing a drug class (such as DMARDs for rheumatoid arthritis). We drop these observations because they are not specific enough to credibly map to the MarketScan data. Brand names are mapped back to molecule names (such as brand name "Sovaldi" mapped back to molecule "sofosbuvir").

[49] We also spent some time classifying procedures. However, procedure names in the CEAR are not standardized and often hard to consistently and accurately match to Current Procedural Terminology (CPT) codes in claims data.

We then merged the CEAR data with the MarketScan data by condition and molecule. To merge by condition we used the Clinical Classification System (CCS) categories provided by the Agency for Healthcare and Research Quality for the MarketScan data and matched the CCS condition names to those in the CEAR. The only exceptions to using CCS categories were for hepatitis C and atrial fibrillation because the CCS categories for these conditions were too broad. Therefore, instead of using the broad CCS category for "hepatitis," we selected ICD-9 and ICD-10 codes specific to hepatitis C.[50]

To match by molecule we used the treatment names in the CEAR and searched the 2008, 2010, and 2012-18 Red Book for all National Drug Codes (NDCs) associated with these names. Searching across multiple Red Book ensures that we capture NDCs that enter and exit over time. Likewise, we search the Healthcare Common Procedure Coding System (HCPCS)-NDC crosswalk for all the HCPCS codes associated with a treatment name.

The CEAR data often compares combinations of molecules with other combinations of molecules. In these cases, we view the CEAR quality estimates as being valid for the combinations, so we use the same combinations in the MarketScan data. We used the CEAR data to identify molecules which patients might take in combination. Once the sample of molecules and combinations of molecules is classified in the CEAR, we search for each patient's condition-specific combinations in the MarketScan data. We look at all drugs that patient took in a given year and create combinations based on what they are observed to take. For example, if a patient diagnosed with hepatitis C is observed to have taken Ribavirin, Simeprevir, and Sofosbuvir in 2018, then we identify the following seven treatment possibilities: Ribavirin, Simeprevir, Sofosbuvir, Ribavirin/Simeprevir, Ribavirin/Sofosbuvir, Simeprevir/Sofosbuvir, and Ribavirin/Simeprevir/Sofosbuvir. Among the possibilities, we assign this patient to the combination with the most drugs that is also in the CEAR data, which in this example would be Ribavirin/Simeprevir/Sofosbuvir.

OA.B.2 CEAR Coverage of Spending

To check how well the CEAR data covers the most important treatments, we examine the share of spending we observe in various datasets. For colon cancer and lung cancer most treatments are physician administered, so the treatments are in medical claims with diagnosis codes. For other conditions, most treatments are in pharmacy claims which do not include diagnosis codes in the MarketScan data. For those conditions, we use the MEPS data.

For colon cancer and lung cancer, we calculate the share of chemotherapy drugs we classify in the Tufts. To do this, we sum all expenditures on any chemotherapy drugs taken by individuals in our colon and lung cancer samples (where chemotherapy drugs are defined by the BETOS category). We also sum up the expenditures on chemotherapy drugs we classify in Tufts for these conditions. We use the years

[50]The exact mapping is as follows: Asthma is CCS code 128. Atrial Fibrillation related ICDs are 4270, 42731, 42732, 42761, 42781, I480, I481, I482, I483, I484, I4891, I4892, I491. Colon Cancer is CCS code 14. Cystic Fibrosis is CCS code 56. Hepatitis C ICDs 07041, 07044, 07051, 07054, 07070, 07071, B1710, B1711, B1920, B1921, B182. HIV is CCS code 5. Hypertension is CCS codes 98 and 99. Lung Cancer is CCS code 19. Multiple Sclerosis is CCS code 80. Osteoporosis is CCS code 206. Rheumatoid Arthritis is CCS code 202. Schizophrenia is CCS code 659. Venous Thromboembolism is CCS code 118.

2007, 2012, and 2017 to capture the coverage across the entire sample period. For lung cancer, at least 92% of all spending on chemotherapy drugs are in the CEAR data. For colon cancer, that number is at least 85%. In both cases, a majority of the missing spending is for drugs which reduce nausea and other chemotherapy side effects. This analysis suggests we are capturing most of the important chemotherapy drugs for these conditions.

For other conditions, we use the 2007-2017 MEPS data to explore CEAR coverage. The MEPS data are useful for this exercise because there are diagnosis codes on pharmaceutical claims, allowing us to determine the share of MEPS spending we observe in the CEAR. However, the MEPS data do not include five-digit CPT codes, which limits our ability to measure physician administered drugs. The MEPS also masks some NDCs for expensive drugs for confidentiality reasons, so high-cost drugs like Sovaldi are not in the MEPS data. This will bias our results towards zero. We also do not include cystic fibrosis in this analysis, as MEPS masks cystic fibrosis in the data after 2009, again due to confidentiality reasons.

Table OA8 provides evidence of how much spending we can classify. The first column shows the percentage of total spending, in the MEPS, that is pharmaceutical spending for a condition (unconditional on whether it is in the Tufts data). For example, 70% of hepatitis C spending is associated with pharmaceuticals (and in the drug files), though this misses some high-cost hepatitis C drugs like Sovaldi. Non-pharmaceutical spending includes hospital stays, physician visits, screenings, diagnostic imaging, and other non-pharmaceutical spending. Another example is hypertension, which is mostly treated with pharmaceuticals, but we are picking up doctor's visits where hypertension is the first listed diagnosis. For atrial fibrillation there is considerable spending on ablation procedures. For venous thromboembolism inferior vena cava filters and thrombectomy/embolectomy are important treatments for some patients. For the remainder of conditions we consider, at least 60% of costs are pharmaceuticals.

Table OA8. Share of MEPS Spending we Classify

	(1) % of All Spending On RX for Condition	(2) % of RX Spending We Categorize in Tufts	(3) % of RX Spending in Tufts on 5% Market Share Drugs
Asthma	65	69	80
Atrial Fibrillation	8	47	60
HIV	71	90	100
Hepatitis C	70	91	100
Hypertension	43	42	79
Multiple Sclerosis	68	92	100
Osteoporosis	70	89	93
Rheumatoid Arthritis	65	85	100
Schizophrenia	52	89	100
Venous Thromboembolism	15	81	85

Notes: This table presents results for how much drug spending in the 2007-2017 MEPS we classify in the CEAR. Column 1 presents the share of all spending in the MEPS is pharmaceutical spending, regardless of whether it is in the CEAR data. Column 2 is the amount of all pharmaceutical spending we classify in the CEAR. Column 3 is the same as column 2, but only keeps drugs that have 5% market share over the sample period. Rare/expensive drugs have masked NDC codes and drugs administered by physicians are not included which will bias our results towards zero. Cystic fibrosis is not included because MEPS masks that condition to protect anonymity. Lung cancer and colon cancer are not included because their treatments are mostly physician administered and the MEPS data do not contain information about these treatments.

The second column shows the share of total drug spending in the MEPS data that is captured by the CEAR data. The MEPS data often contain more classes of drugs that treat a condition, for example painkillers or anti-nausea medication, which are symptom aids that treat many conditions. In addition, comorbidities can inflate spending. For example, if a patient has high cholesterol and hepatitis C, we may see statins in their hepatitis C claims. To better understand how much coverage we have for each condition, column 3 is limited to just drugs that have at least 5% market share over the sample period, which drops many of these other drugs. In column 3, we see that we capture at least 79% of spending on drugs that have at least 5% market share for all conditions except atrial fibrillation (60%).

OA.B.3 Rebate Adjustment

To account for manufacturer rebates, we supplement the MarketScan data with data from SSR Health data. SSR Health, LLC collects data from drug manufacturer SEC filings on revenue net of rebates. They combine the revenue measure with units sold collected by Symphony Health. They then divide net revenues by units sold to estimate a price net of rebates. We aggregate the SSR Health data to the brand-year level. Our SSR Health data includes 1,057 different drugs. To apply this in the MarketScan data, we compute one minus the ratio of net prices to list prices (NET/WAC) which we interpret as the share of revenue which is paid in rebates. Then, at the molecule level, we adjust the level of spending in the MarketScan data by multiplying the payment amounts in the MarketScan data by the NET/WAC ratio for each drug we observe in the SSR Health data. If a molecule is missing in the SSR Health data, which is common (e.g., for most generics), we assume there is no rebate.

OA.B.4 Accounting for Heterogeneity in Cost and Quality

Studies in the CEAR data often make various assumptions in calculating their costs and QALYs. For example, a study may vary in the discount rate used, the time horizon considered, or country of interest. In our analysis we include comparison fixed effects which difference out these factors (Equation 4). However, because our results are retransformed including the common effect of a comparison, $\gamma_{u,d}$, we standardize the study common effect based on the characteristics of each study. To do this, we regress our estimate of each comparison's $\gamma_{u,d}$ on the characteristics of the study and predict what the study common effect would have been under consistent assumptions. For this regression, the unit of observation is a comparison. Our regression equation is:

$$\begin{aligned}\hat{\gamma_{u,d}} = \beta_0 &+ \beta_1 \mathbb{1}(\text{Study uses lifetime time horizon}_i) + \beta_2 \text{time horizon}_i + \beta_3 \text{time horizon}_i^2 \\ &+ \beta_4 \text{time horizon}_i^3 + \beta_5 \mathbb{1}(\text{Study discounts the future}_i) + \beta_6 \text{Discount rate}_i \\ &+ \gamma_g + \gamma_a + \gamma_r + \gamma_c \times \mathbb{1}(\text{Treatment is placebo}_i) + \epsilon_i\end{aligned} \tag{A1}$$

where $\hat{\gamma_{u,d}}$ is the estimate of the comparison fixed effect from Equation 4. $\gamma_g, \gamma_r, \gamma_c$ are gender, country, and condition fixed effects, respectively. Studies also include indicators for the age groups included (i.e. 0-18, 19-40, etc.). If the study includes multiple age groups we divide this indicator by the number of age groups included to get a share of the age groups.[51]

Results for these regressions are available on request. Coefficient estimates are of the expected sign, statistically significant, and robust to inclusion of various additional fixed effects and controls including discount rate variables, indicators for whether there is time discounting, country fixed effects and condition fixed effects.

Using these results, we predict a common effect for all studies using standardized assumptions across studies. For the country-specific dummy, we standardize values to the United States. We also specify that the time horizon is a "lifetime" and set the discount rate to be 3%. As the demographics change across conditions, we set the demographic variables (age group share and gender indicators) to the mean for that condition in the CEAR data.[52] Table OA7 presents quality-adjusted price indexes which check robustness to other assumptions. In particular, we do not adjust for heterogeneity, we adjust the raw QALY (rather than the study fixed effects) for heterogeneity, we add in some additional studies with less precise treatment names, and run the regression in levels. Results are qualitatively similar regardless of specification.

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

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

OA.B.5 Lifetime Costs and Annual Scaling Factor

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

$$LC^{p}_{a,d} = \sum_{s=0}^{100} (1-\rho)^{s} \cdot l_{a,s} \cdot C^{\hat{p}}_{s,d} \tag{A2}$$

where ρ is the interest rate. To be consistent with our standardized QALY estimates, we assume ρ is 0.03. $l_{a,s}$ is the probability of someone age a dying in s periods into the future, which is calculated using the life tables. We then weight across individuals with treatments using the disease-specific distribution of ages in the MarketScan data, $p_{a,d}$.

$$LC^{p}_{d} = \sum_{a=0}^{100} p_{a,d} \sum_{s=0}^{100} (1-\rho)^{s} \cdot l_{a,s} \cdot C^{\hat{p}}_{s,d} \tag{A3}$$

$C^{\hat{p}}_{s,d}$ measures how costs change after an individual with disease d receives treatment. As our goal is to measure lifetime costs, we want to understand how persistent costs are. For example, some conditions might have costs concentrated in one year (e.g., surgery and chemotherapy for cancer typically occurs in one year) while other conditions may have costs that persist indefinitely. To measure this cost progression, we construct a sample of individuals who are enrolled for four consecutive years after their first treatment and one year prior to treatment (to ensure this is a patient's first treatment). We added the superscript p to $C^{\hat{p}}_{s,d}$ to indicate this is for our panel of individuals.

Figure OA1 shows how costs evolve for hepatitis C, hypertension, and rheumatoid arthritis. For hepatitis C, in the first year of treatment (year 0), the average cost is $35,000 while in year 3 the average cost is closer to $5,000. The steep decline in costs after one year of treatment is because the treatments for hepatitis C are typically taken in one course, rather than indefinitely. One can see this in the median and 75th percentile of costs, which go to zero, as we include individuals enrolled but not treated in our panel.[53]

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

[53] While our annual cost of treatment measure, used in the rest of the paper focuses on individuals with treatment (so we drop those with no treatment in a given year), our panel measure picks up individuals who are enrolled (which is the condition for inclusion in the panel), but may not receive any treatment.

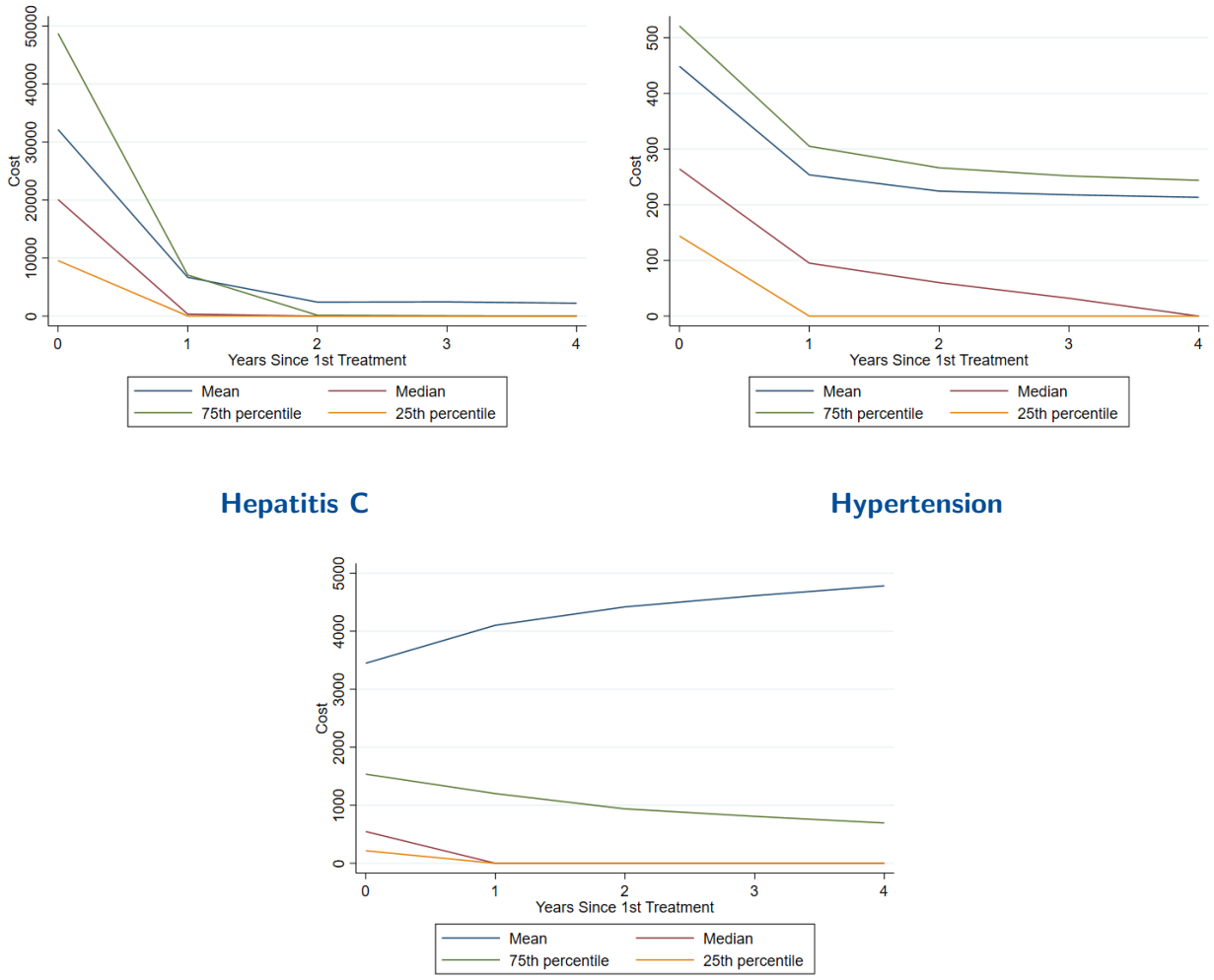
and additional moderately expensive years. For rheumatoid arthritis, treatments are taken indefinitely, so costs do not necessarily decline over time. The increasing slope for these conditions includes the fact that treatments are getting more expensive over time, which is handled by including year fixed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

To approximate this cost progression and extrapolate out over 100 years, we regress costs on years since first diagnosis with fixed effects, up to four years, and calendar-year fixed effects using GLM with a log link.[54] We include calendar-year fixed effects because services are getting more expensive over time which inflates the slopes in Figure OA1. After fitting this regression, we predict costs for each year of having the condition using 2007 as the base year.[55] We then plug these estimates into Equations A2 and A3, to get the lifetime cost estimates for our panel of individuals using 2007 as the base, $LC^{p}_{d,2007}$.

[54] We use Generalized Linear Model (GLM) as it has a better fit than log Ordinary Least Squares (OLS) and then applying a retransformation using the smearing estimator in Duan (1983).

[55] In our preferred specification with years since first diagnosis fixed effects, we assume that the year 4 costs remain constant for 96 more years, reflecting a stabilizing in costs. However, we also estimate regressions using a linear trend in years since first treatment. However, this linear trend often goes to zero, which we think understates the persistence of costs. We have also tried higher order polynomials in these regressions, but these results do not seem credible given how far out of sample we are predicting.

Figure OA1. Cost Progression for Selected Conditions



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

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

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

for the sample selection in using people enrolled for multiple years. That is, our cost estimates are:

$$LifetimeCost_{d,t} = LC^{p}_{d,2007}\frac{\bar{C}_{d,t}}{\bar{C}_{d,2007}}\frac{\bar{C}_{d,2007}}{\hat{C^{p}}_{1,d,2007}} = LC^{p}_{d,2007}\frac{\bar{C}_{d,t}}{\hat{C^{p}}_{1,d,2007}} \tag{A4}$$

This leads to an intuitive cost multiplier $\frac{LC^{p}_{d,2007}}{\hat{C^{p}}_{1,d,2007}}$, which is the lifetime cost of the select sample of people enrolled multiple years, divided by the average annual cost of that select sample. Therefore, throughout the paper we compute $\bar{C}_{d,t}$ and multiply it by our cost multiplier:

$$\alpha_d = \frac{LC^{p}_{d,2007}}{\hat{C^{p}}_{1,d,2007}} \tag{A5}$$

this ratio has an intuitive form, as well, which helps clarify the main assumption we are making. α_d tells us the ratio of lifetime costs to average treatment costs in one year (with treatment) for our continuously enrolled sample. We then assume that this ratio of lifetime costs to one-year costs holds for the main sample.

Table OA9 presents the lifetime cost multiplier we estimate for each condition. Column 4 is the version which assumes costs are constant and the life tables and time discounting suggest a lifetime cost multiplier of 23-27. The first column is our preferred specification, which assumes that 4th year costs continue indefinitely. For a condition like hepatitis C, our preferred cost multiplier is 3.7. This cost multiplier is much smaller because people mostly have only one expensive year of treatment (i.e. you take one course of Sovaldi). For conditions like rheumatoid arthritis and multiple sclerosis where people continue taking treatments indefinitely, costs are similar to the version without accounting for the cost slope.

Table OA9. Lifetime estimate cost multipliers for each condition

	(1) Preferred Specification	(2) Uses Years Since Trend Line	(3) No Untreated Prior Year Needed	(4) Constant Costs No Slope
Asthma	12.571	3.733	15.931	27.859
Atrial Fibrillation	8.103	3.064	8.774	24.016
Colon Cancer	7.885	2.241	7.834	23.735
Cystic Fibrosis	23.167	10.109	25.525	28.913
HIV	23.300	14.772	23.965	25.905
Hepatitis C	3.683	1.507	3.659	24.318
Hypertension	12.491	4.561	15.226	24.296
Lung Cancer	7.468	3.128	8.442	23.324
Multiple Sclerosis	22.658	13.653	21.688	25.422
Osteoporosis	7.167	2.845	7.276	23.249
Rheumatoid Arthritis	23.415	16.870	28.219	24.872
Schizophrenia	9.096	2.636	14.704	26.703
Venous Thromboembolism	3.827	1.483	4.435	24.823

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

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

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

Table OA10. Price Indexes for Each Condition Using Different Lifetime Cost Assumptions - $100,000 VSLY

	(1) Preferred Specification	(2) Uses Years Since Trend Line	(3) No Untreated Prior Year Needed	(4) Constant Costs No Slope	(5) Annual Costs
Asthma	1.001	0.968	1.004	1.008	0.841
Atrial Fibrillation	0.771	-4.300	1.007	2.814	-21.133
Colon Cancer	0.619	0.648	0.619	0.611	0.700
Cystic Fibrosis	4.194	4.146	4.198	4.202	3.367
HIV	1.452	1.421	1.453	1.457	0.257
Hepatitis C	-5.746	-15.786	-5.793	0.151	-24.396
Hypertension	0.230	-0.560	0.312	0.451	-4.991
Lung Cancer	1.919	1.597	1.946	2.077	0.419
Multiple Sclerosis	2.910	2.852	2.906	2.920	1.001
Osteoporosis	1.201	0.458	1.208	1.539	-1.814
Rheumatoid Arthritis	2.029	1.973	2.054	2.038	-1.223
Schizophrenia	0.522	-0.214	0.637	0.721	-1.912
Venous Thromboembolism	-0.316	-2.881	-0.093	1.058	-4.906

Notes: This table presents our quality-adjusted price indexes, constructed using the CEAR, MarketScan, and SSR Health datasets. Each column assumes that the VSLY is $100,000. All columns (except the last) account for the age distribution of a condition, life expectancy, and the discount when calculating lifetime costs. The first three columns account for the idea that when someone has treatment in one year that their future costs may not remain constant. The first column keeps costs constant at their year four level. The second column uses a log-linear trend to predict costs. The third column does not condition on having a year without spending prior to the first year of treatment. The fourth column holds treatment costs constant. Column 5 assumes all costs are in one year, which is clearly unrealistic, but is a clear lower bound.

Column 3 drops the requirement that we observe one year without diagnosis prior to the first year. This increases the multiplier estimates because we have more people who are in the constant cost stage of their treatment, reducing the steepness of the slopes in Figure OA1. Results are fairly similar, suggesting that conditioning on having no spending in the prior year does not impact results much.

Table OA10 shows price index estimates using all of these specifications and annual costs (assuming the

multiplier is 1). With annual costs prices are falling much more quickly than with constant costs. This is a very wide range of assumed values, results are different and should be viewed as very wide bounds on our central estimates.

Online Appendix OA.C Additional Analyses

OA.C.1 Comparing Aggregate Welfare with R&D Costs

In this section, we do a back-of-the-envelope calculation to estimate the aggregate (rather than per-patient) total welfare gains per drug approval. Other studies, such as DiMasi et al. (2003, 2016) have estimated the cost of R&D per drug approval, the latest paper finding costs on the order of $1.4-$2.6 billion per approved drug in 2013 or $1.5-$2.8 billion in 2018 dollars. Our goal in this section is to provide some context for the value we are getting for that R&D expenditure.

We begin with the total welfare estimates from Table 6, which calculates total welfare gained per patient taking new drugs in 2018, versus a 2007 benchmark.[56] These estimates are already in "lifetime" units, for example a change in mean QALYs of 1 means a patient taking an average drug in 2018 gets 1 more QALY than a patient in 2007 over their lifetime. For each condition, we take the per patient estimates, then multiply the number of treated patients in MarketScan. We then scale number of treated patients by the ratio of the number of under-65 privately insured individuals in the U.S. to the number of MarketScan enrollees. To get the per-drug total welfare created by new drugs, we then divide by the number of new drugs in our sample that have been approved between 2007 and 2018.

The first column of Table OA11 presents the results for just 2018. We drop asthma, colon cancer, hypertension, and rheumatoid arthritis because they had few new entrants during our sample period. We calculate that innovations in the treatment of hepatitis C added about $1.9 billion in total welfare per new drug, just for individuals taking hepatitis C drugs in 2018. These estimates depend on the relative health benefit of the innovations and the number of people taking the drugs. Atrial fibrillation drugs have both large health benefits and a lot of people taking them, hence the $13 billion value. Cystic fibrosis has a highly effective entrant, but because it is more rare, it creates only $375 million in welfare in 2018.

To capture the full lifetime value, we need to account for patients that will start taking the drugs in 2019, 2020, etc. To do this, we calculate the share of individuals who take a treatment for the first time in a given year, compared to those continuing treatment. Then we compute the net present value of those additional patients, going out 20 years and assuming a discount rate of 0.97. We call this estimate the lifetime value per drug.

[56] We focus on total welfare for two reasons. First, we think it is a policy-relevant benchmark, though one could argue consumer surplus is just as important or more important. Because total welfare maps directly to health benefits, it is still very much capturing an object of interest. Second, it reduces complications regarding the lifetime split of consumer versus producer surplus which would add another element of uncertainty in this already rough calculation.

Table OA11. Back-of-the-Envelope Aggregate Welfare Calculations

	(1) 2018 Value Per Drug (millions)	(2) Lifetime Value Per Drug (millions)
Atrial Fibrillation	13,089	109,375
Cystic Fibrosis	375	2,197
HIV	191	1,173
Hepatitis C	1,921	22,898
Lung Cancer	389	4,393
Multiple Sclerosis	984	5,738
Osteoporosis	1,335	13,610
Schizophrenia	731	7,198
Venous Thromboembolism	870	8,793

Notes: This table presents results for our back-of-the-envelope aggregate total welfare calculations. The first column shows total welfare due to innovation from Table 6 as described in the preceeding paragraph. Column 2 then takes the net present value of column 1 over 20 years, where we account for the number of people getting treated for a given disease each year. Asthma, colon cancer, hypertension, and rheumatoid arthritis are dropped because they do not have many new entrants.

The second column of Table OA11 presents the lifetime value per drug. Values range from $1.17 billion for HIV to $109 billion for atrial fibrillation. Except for cystic fibrosis and HIV, each condition has lifetime values well above the upper range suggested by DiMasi et al. (2016). For cystic fibrosis this is partially due to the fact that the new drugs are still diffusing, while HIV had its highest value innovations before the sample period. We think this is a lower bound estimate. We are not accounting for market expansion (see Appendix Section OA.C.3) and this calculation is based on the under-65 privately insured population in the United States, so it ignores Medicare, Medicaid, and international markets.[57] Still our estimates suggest that the societal benefits of these new innovations are an order of magnitude larger than the R&D costs of innovation.

OA.C.2 Incorporating Health Risk, Financial Risk, and Insurance

In this section, we begin by deriving our consumer welfare measure which follows Cutler et al. (1998) closely, and is also used in cost-effectiveness studies. We then build on this measure by incorporating the health risk, financial risk, and the value of health insurance, in the spirit of Lakdawalla et al. (2017).

Suppose an individual derives utility from their health and consumption, $u(C, H)$. The standard approach to deriving consumer welfare gains from innovation is to determine how much a consumer would have to pay to be indifferent between states of the world with and without the innovation, denoted 1 and 0, respectively. This value can be implicitly defined using the following expression:

$$u(Y - S_1 - V, H_1) = u(Y - S_0, H_0).$$

[57] According to a recent Global Use of Medicines report by IQVIA, North America accounts for only 8.5% of worldwide drug purchases.

where Y is income, S is the cost of medical care, and H is the health achieved in each state.[58] V is the implicit measure of the consumer welfare generated by the new technology, as it sets utility equal across the two states.

Let u_H and u_C denote the derivatives of the utility function with respect to H and C, respectively. Taking the full derivative, the value of the new innovation can be derived as:

$$V = \frac{u_H}{u_C}(H_1 - H_0) - (S_1 - S_0). \tag{A6}$$

This is the standard formulation of value of innovation in both the cost-effectiveness literature and the quality-adjusted price index literature. This also matches Equation 2, noting that $\frac{u_H}{u_C}$ is the value of health converted into dollars, or the VSLY.

Lakdawalla et al. (2017) note that this "conventional" formulation is missing some important components. First, it does not account for the benefit of reduced risk healthy people face given that they might get sick in the future (health risk): these innovations might improve their welfare if they were to get sick. Second, the conventional formulation does not account for the financial risk that sick people face, namely when someone is sick they also face a cost shock from buying more expensive treatments and potentially lower wages. Finally, the conventional measure treats costs paid by the patient and the insurer equally. This ignores the benefit of insurance where risk is spread from sick to healthy individuals. The bias in the conventional approach's measure of the effect of innovation is ambiguous. Higher quality treatments dampen the health shock of being sick. However, if those treatments also have higher costs, then costly new innovations can increase financial risk. This financial risk can be partly mitigated by health insurance.

To incorporate these factors we consider the ex-ante risk for an individual, prior to knowing whether they will be sick. Let π denote the probability of a individual getting sick. Sick individuals pay P out of pocket for medical care and all individuals pay I for insurance costs. We assume that health status in the healthy state, H_W, does not vary with the innovation. In addition, we let income vary by health status, where Y_W and Y_S denote income for healthy and sick individuals. Then, we can implicitly define the ex-ante value of medical innovation, $V^{ex-ante}$, as:

$$\begin{aligned} \pi u(Y_S - P_1 - I_1 - V^{ex-ante}, H_1) + (1-\pi)u(Y_W - I_1 - V^{ex-ante}, H_W) \\ = \pi u(Y_S - P_0 - I_0, H_0) + (1-\pi)u(Y_W - I_0, H_W). \end{aligned} \tag{A7}$$

Let u_i^S and u_i^W for $i \in \{C, H\}$ denote the derivative of the utility function with respect to i in the sick

[58] The cost of the technology is typically including both what an insurer pays and what a consumer pays out of pocket, under the assumption that either insurer costs will be passed through to consumers as higher premiums or that we are accounting for the "payer" prospective, which is policy relevant if the payer is a government.

and well states, respectively. Then taking the total derivative one can derive the amount of ex-ante consumer welfare the new technology provides:

$$V^{ex-ante} = \frac{\pi u_H^S(H_1 - H_0) - \pi u_C^S(P_1 - P_0 + I_1 - I_0) - (1-\pi)u_C^W(I_1 - I_0)}{\pi u_C^S + (1-\pi)u_C^W}. \tag{A8}$$

This formulation accounts for the three mechanisms discussed above. The first term makes clear that ex-ante there is a benefit to healthy individuals, as they may get sick with probability π, in which case they will get the health benefits of innovation: $H_1 - H_0$. Second, we differentiate between u_c^S and u_c^W, which allows sick and healthy consumers to have different marginal utilities of consumption, which incorporates financial risk. When patients are sick, they have lower wages, so their marginal utility of consumption can be higher. This happens concurrently with their medical expenses being higher. Finally, sick individuals receive the benefit of insurance, as I shifts costs from sick individuals to healthy individuals.

We follow many of the assumptions in Lakdawalla et al. (2017) to parameterize this new formulation for the value of innovation. In particular, we assume that utility takes the additively separable CRRA form:

$$u(C, H) = \frac{C^{1-\sigma} - 1}{1-\sigma} + \frac{H^{1-\sigma} - 1}{1-\sigma}. \tag{A9}$$

This functional form assumption allows us to pin down the marginal utility of consumption terms, u_C^S and u_C^W. To do this, we follow Lakdawalla et al. (2017) and assume that $\sigma = 2$, that full income in a well state is \$120,000 and that income in the sick state is 80% of that in the healthy state. The full income includes the value of non-health consumption, such as the value of leisure. The later assumption is to account for sick individuals' lost wages. We use the observed out-of-pocket costs for each condition to calculate $P_{d,t}$. In our data the out-of-pocket costs are extremely muted by insurance. Conditions like cystic fibrosis, multiple sclerosis, and lung cancer have out-of-pocket costs that are less than 5% of the total amount paid because patients hit their out-of-pocket maximums. For cheaper conditions like hypertension and asthma, insurance covers between 60-80% of the cost. We assume insurance is actuarially fair and higher costs to insurers are fully passed through to all individuals, sick or healthy.[59] Therefore, if we add up the total cost of treatment in the population and divide by the number of individuals with the condition, we obtain the lifetime cost of treatment, $S_{d,t}$, as used in the main specification that ignores insurance. We measure π by calculating the share of individuals in our data with a given disease.[60]

[59] To calculate insurance costs we sum up all the costs paid by insurers (i.e. total costs minus the out-of-pocket amounts, $S_{d,t} - P_{d,t}$). Then, we divide those costs among the total number of enrollees in our sample, including individuals who do not have condition d. For this exercise, $P_{d,t}$ only includes the 1 year annual cost of treatment (rather than the lifetime cost), because the one year costs are deducted from one year of consumption spending.

[60] We calculate the prevalence of a condition, π, in 2007. We then keep this value constant, as changing prevalence would muddy the analysis for how innovation shapes welfare.

To calibrate u_H^S, we assume the value of a life year for a sick individual, $\frac{u_H^S}{u_C^S}$, is $\$100,000$ in 2007. After that, we allow the level of health in the sick state to vary over time in proportion to the average QALY estimate, but do not impose the VSLY to be equal to $100,000 in other years.[61] We calibrate the VSLY in the sick state to be consistent with the conventional approach (and the main approach in our paper), which focuses on sick individuals and therefore are implicitly assuming a VSLY for a sick individual.

The results in the main paper are consumer welfare for just the sick individuals, and it is assumed there is no welfare gain for individuals that are not sick. In this section, we measure welfare gains for both the sick and healthy populations. To make these measures comparable, we divide our ex-ante consumer welfare by the share of the population that is sick, π, $\frac{V^{ex-ante}}{\pi}$. After dividing by π, the ex-ante consumer welfare measure is comparable to our conventional measure as both are measures of consumer welfare per sick individual.

Table OA12 presents results for Hepatitis C. Columns 1-3 reproduce the conventional welfare, health benefit, and cost estimates from Table 2, respectively.[62] Columns 4-6 provide results using $\frac{V^{ex-ante}}{\pi}$. Column 4 presents the full ex-ante value, while column 5 presents just the health benefit (the first term in Equation A8) and column 6 presents the impact the value of the cost change on utility (the second and third term in Equation A8).

Table OA12. Comparing Consumer Welfare With Different Utility Functions: Hepatitis C

	(1) Conventional Change in Consumer Welfare	(2) Conventional Change in Health Benefit	(3) Conventional Change in Cost	(4) New Change in Consumer Welfare	(5) New Change in Health Benefit on Utility	(6) New Change in Cost on Utility
2007	0	0	0	0	0	0
2008	2,559	5,635	3,076	5,862	8,942	3,080
2009	6,243	5,706	-537	9,570	9,056	-514
2010	4,807	5,995	1,188	8,282	9,514	1,232
2011	19,201	79,091	59,890	64,700	124,777	60,077
2012	17,477	84,977	67,500	66,166	133,994	67,828
2013	27,125	81,364	54,239	73,806	128,358	54,552
2014	-76,440	223,112	299,552	47,328	347,191	299,863
2015	94,318	268,925	174,607	243,007	418,135	175,128
2016	171,308	269,489	98,180	320,685	419,556	98,871
2017	214,060	283,747	69,687	371,196	441,561	70,366
2018	275,433	283,763	8,331	433,069	442,062	8,993

Notes: This table presents results for hepatitis C using the conventional measure of welfare and the measure incorporating risk. Columns 1-3 reproduce the conventional welfare, health benefit, and cost estimates from Table 2, respectively. In Table 2 in the main text, these are column 5, column 1 times $100,000, and column 2 minus $41k. Columns 4-6 provide results using $\frac{V^{ex-ante}}{\pi}$. Column 4 presents the full ex-ante value, while column 5 presents just the health benefit (the first term in Equation A8) and column 6 presents the impact the value of the cost change on utility (the second and third term in Equation A8).

[61] If we held the VSLY fixed at $\$100,000$, then new expensive drugs would increase the marginal utility of health, because they increase the marginal utility of consumption and we are holding the ratio constant. This would mean that the health contribution to utility from drugs with the same QALY estimate would be increasing in their cost, which we do not think is reasonable. Instead, in our approach the VSLY will rise when drugs are cheaper, because individuals are able to consume more, which is consistent with Murphy and Topel (2006).

[62] In Table 2 in the main text, these are column 5, column 1 times $100,000, and column 2 minus $41k.

In general, the impact on utility from the costs of these technologies is larger using the new framework than the conventional measures (column 3 versus column 6). This is because sick individuals pay higher health care costs with innovation, which causes the marginal utility of consumption to be higher for sick individuals due to the curvature of the utility function.[63] Sick people also have to pay out-of-pocket costs for their treatments (alongside possible labor market impacts), which means that the utility cost of the price of more expensive treatments is larger than the conventional case with no risk. As discussed above, this is capturing the financial risk created by innovation, so more expensive treatments mean that the financial shock of being sick is larger. However, individuals in our data pay a tiny fraction of the financial cost for these really high cost drugs (for hepatitis C this is less than 5%), which minimizes the financial risk, hence the differences are small. If individuals were uninsured, an $80k annual course of treatment would represent about 80% of income and the utility impact of these costs would be considerably larger.

Column 5 shows the health benefit is larger with the new measure. This is because it dampens the risk of a health shock, which occurs with a consumption shock. Individuals receive a health benefit when their marginal utility is higher because $\pi u_C^S + (1-\pi)u_C^W < u_C^S$, which means the first term in Equation A8 is larger than the first term in Equation A6.

Table OA13. Summary Comparing Consumer Welfare With Different Utility Functions

	(1) Conventional Change in Consumer Welfare	(2) Conventional Change in Health Benefit	(3) Conventional Change in Cost	(4) New Change in Consumer Welfare	(5) New Change in Health Benefit on Utility	(6) New Change in Cost on Utility
Asthma	-11	225	236	76	340	264
Atrial Fibrillation	3,292	44,303	41,011	26,970	68,228	41,258
Colon Cancer	128,812	-3,977	-132,788	126,434	-6,369	-132,803
Cystic Fibrosis	-1,986,003	23,207	2,009,210	-1,973,434	36,670	2,010,103
HIV	-141,116	16,739	157,855	-131,795	26,483	158,278
Hepatitis C	275,433	283,763	8,331	433,069	442,062	8,993
Hypertension	6,546	3,863	-2,683	8,018	5,304	-2,714
Lung Cancer	-245,536	61,975	307,511	-209,907	97,922	307,829
Multiple Sclerosis	-908,444	41,930	950,374	-888,487	62,906	951,393
Osteoporosis	-1,365	3,325	4,690	435	5,163	4,728
Rheumatoid Arthritis	-158,881	22,403	181,284	-148,284	33,412	181,696
Schizophrenia	17,980	11,319	-6,660	24,473	17,829	-6,644
Venous Thromboembolism	8,296	10,237	1,941	13,903	15,989	2,086

Notes: This table presents results for all conditions in 2018 using the conventional measure of welfare and the measure incorporating risk. Columns 1-3 reproduce the conventional welfare, health benefit, and cost estimates from Table 4. In Table 4 in the main paper, these are column 5, column 1 times $100,000, and column 2 times column 3 minus 1. Columns 4-6 provide results using $\frac{V^{ex-ante}}{\pi}$. Column 4 presents the full ex-ante value, while column 5 presents just the health benefit (the first term in Equation A8) and column 6 presents the impact the value of the cost change on utility (the second and third term in Equation A8).

[63] To see this, first consider the full insurance case. In that case, $P_1 = P_0 = 0$ in Equation A8, and I_t is just the total cost of the drugs, spread across the entire population. Because $P_1 = P_0 = 0$, the $I_1 - I_0$ term can be factored out of the second and third terms of Equation A8, the marginal utility terms and πs cancel out. Ex-ante, an individual will have to pay their share of the cost of the drug in either state, so there is no risk. Hence, the insurance value equals the conventional cost. If insurance is not perfect ($P_t \neq 0$), then the costs are shifted towards the sick state, when the marginal utility of consumption is higher.

Table OA13 summarizes results for all conditions in 2018. Results are consistent with the table for hepatitis C. The health benefit tends to be considerably larger. The cost is also larger, but the impact is more muted. On net our measure suggests that benefits of innovation are understated using the conventional measure, and often considerably so. However, even with this new approach, five of the six conditions which had declining consumer welfare with the conventional approach, also have declining consumer welfare using the new measure.

OA.C.3 Market Expansion

In this section we explore how market expansion would impact our results. New drugs may expand the market if they are more effective or treat variations of the condition that older treatments do not treat. In addition, drugs coming off patent could also increase the market if lower prices increase demand.

To quantify the effect of market expansion, we identify the improvement from taking "no treatment" to taking a specific treatment. Most conditions have observations where one of the treatments in the CEAR data is "no treatment," "best supportive care," "placebo," or something similar. We assume these observations are similar to not being treated and include that category as no treatment in the Tufts regressions (Equation 4). Let $H_{d,t}$ be the health benefit of someone taking the bundle average treatment in year t. Likewise, let $H_{d,no}$ be the health benefit associated with no treatment, which is measured using the "no treatment" observations in the Tufts.

We assume the market expansion effect can be measured by the increase in the number of patients taking treatment, relative to 2007, $N_{d,t} - N_{d,2007}$, where $N_{d,t}$ is the number of patients treated in year t. The change in quality is now the sum of the change in quality for those taking the treatment, $H_{d,t} - H_{d,2007}$, and the quality benefit from market expansion, $H_{d,t} - H_{d,no}$. To make these results comparable to the results in the main paper, we define our change in QALYs and welfare in per-patient terms in 2007. Formally, the change in QALYs can be denoted:

$$\frac{N_{d,2007} \times (H_{d,t} - H_{d,2007}) + (N_{d,t} - N_{d,2007}) \times (H_{d,t} - H_{d,no})}{N_{d,2007}}.$$

The numerator is simply the sum of QALYs from those continuing to take the treatment (the first term in the numerator) and the expansion term—the QALYs from the newly treated patients in the market (the second term in the numerator). The denominator divides through by the number of patients in 2007 to keep units consistent with the main results in the paper. The interpretation, both with and without market expansion, corresponds to the number of QALYs gained per patient in 2007, regardless of whether the number of patients changed over time.[64] We define costs of market expansion in a similar way.

[64] To give a concrete example, suppose there are 10 patients in 2007, and the average QALY in 2007 is 1. Now, suppose there are 20 patients in 2018 with an average QALYs of 2. Also, suppose that "no treatment" is associated with 0.5 QALYs.

Without market expansion, like in the main text, we would estimate (2-1=) 1 additional QALYs in 2018. With market expansion, it our estimate would be:

Table OA14 shows results for all conditions in 2018.[65] The first three columns match Table 4, for reference. The last three columns present the results with market expansion. For nearly every condition, we do find that QALY gains are larger when considering market expansion. In general, we find much more market expansion from new innovations, rather than conditions where drugs are going off patent. The gains are considerable for cystic fibrosis, multiple sclerosis, and HIV. The gains in QALYs is driven by two forces, the size of market expansion and the improvement from no treatment. For cystic fibrosis, the share treated only increases by about 6%, but no treatment is considerably worse, leading to a doubling in QALYs produced. HIV and multiple sclerosis both have larger expansions (20-30% increases in the share treated), but no treatment is also significantly worse than any treatment for these populations, hence the disproportionately large increase in QALYs. For hepatitis C, the number being treated in 2018 is slightly smaller than in 2007 (after a spike in 2014-2015), so the change in QALYs in 2018 is somewhat smaller when accounting for market expansion.

However, costs also rise in a proportionate manner, so consumer welfare for cystic fibrosis and multiple sclerosis is lower when accounting for market expansion, even though QALYs rise by so much. These drugs were not cost effective, so having more people taking them does not necessarily improve consumer welfare.[66] However, we do find that total welfare rises considerably when accounting for market expansion. While these drugs may not be cost effective, more health is produced, which improves total welfare.

$$2.5 = \frac{10 \times (2-1) + (20-10) \times 2 - 0.5}{10}.$$

Ten patients get one additional QALY, and 10 patients get 1.5 additional QALYs from switching from no treatment. While technically the average patient is getting 1.25 more QALYs, the 2.5 number expresses the total number of QALYs gain, relative to the number of patients in 2007. This makes it easier to compare across tables. The interpretation is per patient in 2007, how many additional QALYs were gained. Twenty-five additional QALYs were created and there were 10 patients in 2007, so 2.5 QALYs per patient in 2007.

[65] We drop asthma from this calculation because there are no "no treatment" observations in the CEAR. We also do not have a no treatment observation for cystic fibrosis, but a study not in the Tufts, Tice et al. (2020), suggests that Orkambi is 3.6 QALYs better than "best supportive care," so we use that number to calibrate $H_{d,no}$.

[66] This does not necessarily need to be the case, as shown by HIV. These drugs were not cost effective relative to an older generation treatment. However, the relevant QALY and cost comparison for market expansion is no treatment rather than an older drug.

Table OA14. Comparing Consumer and Total Welfare With and Without Market Expansion: All Conditions

	(1) Δ Avg QALYs 2018 - 2007 No Expansion	(2) Δ Consumer Welfare $100k VSLY No Expansion ($1,000s)	(3) Δ Total Welfare $100k VSLY No Expansion ($1,000s)	(4) Δ Avg QALYs 2018 - 2007 With Expansion	(5) Δ Consumer Welfare $100k VSLY With Expansion ($1,000s)	(6) Δ Total Welfare $100k VSLY With Expansion ($1,000s)
Atrial Fibrillation	0.443	3	44	0.680	21	68
Colon Cancer	-0.040	129	-4	-0.012	120	-1
Cystic Fibrosis	0.232	-1,986	23	0.468	-2,123	47
HIV	0.167	-141	17	1.920	-123	192
Hepatitis C	2.838	275	284	2.670	261	267
Hypertension	0.039	7	4	0.027	5	3
Lung Cancer	0.620	-246	62	0.312	-108	31
Multiple Sclerosis	0.419	-908	42	1.066	-1,258	107
Osteoporosis	0.033	-1	3	0.024	-2	2
Rheumatoid Arthritis	0.224	-159	22	0.382	-185	38
Schizophrenia	0.113	18	11	0.216	25	22
Venous Thromboembolism	0.102	8	10	0.115	8	12

Notes: This table presents results for all conditions in 2018 without market expansion (columns 1-3) and with market expansion (columns 4-6). The first two columns match Table 4. Asthma is dropped from this table because we did not have a measure of "no treatment" in the CEAR data. All estimates assume $100,000 VSLY.

For non-innovative conditions with few new entrants, market expansion effects are much more muted. For colon cancer and hypertension we do not find much market expansion. Accordingly, both consumer and total welfare are mostly unchanged or even slightly lower when considering market expansion. This is consistent with Duflos and Lichtenberg (2012); Castanheira et al. (2019) who find generic entry reduces market size, potentially due to reduced advertising.

While not necessarily true in every case, we think these results suggest that our preferred results (without market expansion) may be understated, but in both directions. Entry of drugs which are not cost-effective will lower consumer welfare, and accounting for market expansion will suggest a bigger reduction in consumer welfare if these treatments lead to more people getting treatment. However, these drugs are still improvements over earlier generations or no treatment, so not accounting for market expansion will also understate the health benefits and total welfare gains. While we think these results are informative, we prefer the results without market expansion because these results are heavily influenced by estimates of the relative QALYs of no treatment. There are a few issues with the no treatment category: (1) unlike cases where we are comparing two specific treatments, we are not as confident in what no treatment means in the CEAR data; (2) we are not sure that no treatment is consistently defined across studies in the CEAR data (in fact, we are grouping together different definitions of no treatment); and (3) because of (1) and (2) we are not sure the no treatment definition in CEAR is the same as that in the MarketScan data.

References

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

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.

DiMasi, J. A., H. G. Grabowski, and R. W. Hansen (2016). Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics 47*, 20–33.

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.

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

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

Tice, J., R. Chapman, M. Seidner, D. Rind, S. D. Pearson, K. Kuntz, and K. Wherry (2020). Modulator treatments for cystic fibrosis: Effectiveness and value. *Institute for Clinical and Economic Review*.