Are Medical Care Prices Still Declining? A Re-examination Based on Cost-effectiveness Studies[1]

Abe Dunn, Bureau of Economic Analysis, (email: abe.dunn@bea.gov)
Anne Hall, U.S. Department of the Treasury, (email: anne.hall@treasury.gov)
Seidu Dauda, World Bank Group, (email: sdauda@worldbank.org)

August 2021

Abstract: More than two decades ago a well-known study on heart attack treatments provided evidence suggesting that, when appropriately adjusted for quality, medical care prices were actually declining (Cutler et al. 1998). Our paper revisits this subject by leveraging estimates from more than 8,000 cost-effectiveness studies across a broad range of conditions and treatments. We find large quality-adjusted price declines associated with treatment innovations. To incorporate these quality-adjusted indexes into an aggregate measure of inflation, we combine an unadjusted medical-care price index, quality-adjusted price indexes from treatment innovations, and proxies for the diffusion rate of new technologies. In contrast to official statistics that suggest medical care prices increased by 0.53 percent per year relative to economy-wide inflation from 2000 to 2017, we find that quality-adjusted medical care prices declined by 1.33 percent per year over the same period.

[1] We would like to thank Ana Aizcorbe, Ben Bridgman, David Byrne, David Cutler, Berthold Herrendorf, Charles Jones, Dennis Fixler, Bronwyn Hall, Bob Hall, Peter Klenow, Eli Liebman, Brett Matsumoto, John Romley, Louise Sheiner, Jon Samuels, Jon Skinner, and the experts on the Bureau of Economic Analysis Academic Advisory panel for comments. We would like to thank seminar and conference participants at the NBER CRIW conference, Society for Economic Measurement in Boston, the Productivity Growth: Past, Present, and Future conference at Arizona State University, the Federal Trade Commission seminar, and NBER's Macroeconomic Perspectives on the Value of Health. We would also like to thank Lasanthi Fernando for research assistance and anonymous referees for several useful comments. The views expressed in this paper are those of the authors and not necessarily the views of the Bureau of Economic Analysis, the U.S. Department of the Treasury or the World Bank.

Introduction

Measuring the output of the medical sector accurately is one of the most important and most challenging tasks in the field of economic measurement. The topic has only increased in importance over the past two decades, as the share of the United States (U.S.) gross domestic product (GDP) devoted to medical care rose from 13 percent in 1998 to nearly 18 percent in 2019 (Martin et al. 2020). Some experts believe that much of the growth in the medical sector's share of GDP is driven by new technologies that improve treatment in the long run (Chernew and Newhouse 2011). Recent work has also shown that innovations have been a key driver of the rapid growth in expenditures over this period for many conditions including rheumatoid arthritis, cancer, hepatitis, and HIV (Dunn et al. 2018). Meanwhile life expectancy at birth has increased by nearly two years over the same period with medical innovations likely playing a significant role (Anderson 2001; Kochanek et al. 2017).

While innovations are a key contributing factor to the growth in spending for medical care, improvements in the quality of medical care that contribute to improved health outcomes are not reflected in U.S. national statistics. Official measures from the Bureau of Labor Statistics (BLS) show negative multifactor productivity growth in the hospital and nursing home industry from 2000 to 2017. It has long been recognized that the lack of quality adjustment leads official statistics to overstate inflation and understate output and productivity growth in the health-care sector (Lebow and Rudd 2003; Groshen et al. 2017).

Some studies have suggested that when adjusted for quality, medical inflation is much lower than the official measures. Cutler et al. (1998) were the first to ask whether medical inflation may even be declining when properly adjusted for quality. They examine this question for the price of treatment for heart attacks and found that after accounting for the value of increased life expectancy following treatment, the price of treatment declined over their period of study, even while the unadjusted price of treatment rose. These findings suggest that the mismeasurement of medical care output and productivity could be

substantial and have major implications for our understanding of individual welfare and economy-wide real output given that medical care is such a large share of the economy.

In this paper, we provide more general evidence that medical inflation has been greatly overstated by official measures. Using a database of thousands of cost-effectiveness studies covering medical innovations for a wide set of medical conditions with information both about the change in price for treatments and the changes in health outcomes delivered by these treatments, we construct quality-adjusted price indexes for each treatment innovation based on the method used by Cutler et al. (1998) which was derived from utility theory.

We find that the entry of new treatments is typically characterized by steep quality-adjusted price declines. The median quality-adjusted price change from a new treatment is a decline of around 20 percent. However, many innovations show quality-adjusted price declines of over 100 percent, indicating that the gains in health outcomes are worth more than the price of the prior treatment. In this case, consumers would not accept the prior technology, even if it were offered for free. These large quality-adjusted price declines are driven by the value of improved health outcomes.

Another contribution of the study is that we incorporate these quality-adjusted price indexes into a simple formula to approximate an aggregate quality-adjusted price index for the health care sector for the period from 2000 to 2017. Price indexes from innovations are not sufficient to measure aggregate inflation rates for the health care sector, as only a portion of the expenditures are spent on new innovations each year. Our method combines unadjusted measures of medical care inflation by condition from the Bureau of Economic Analysis (BEA), quality-adjusted price changes from the cost-effectiveness database, and proxies for the rate of adoption of new technologies. Our proxies for the rate of adoption are based on the diffusion rate of new drugs, shifts in the use of procedure codes, and shifts in hospital process measures. We estimate that the adoption of new technologies accounts for between 1.9 to 3.6 percent of spending per year. We then use this adoption rate to weight both the quality-adjusted index of treatment innovations and the unadjusted medical-care price index, and finally to derive a quality-adjusted index from 2000 to 2017.

Our central estimate shows that quality-adjusted prices are declining by about 1.33 percent per year relative to economy-wide inflation. Over the same period, the official BEA Personal Consumption Expenditure (PCE) price index for health care exceeds economy wide inflation by 0.53 percent per year. Therefore, the difference between the official PCE estimate for health care and our quality-adjusted estimate is 1.86 percentage points per year, indicating that the growth in real output for health care may be understated by this same amount.

While we present evidence that quality-adjusted prices are declining, the exact value of the quality-adjusted price index is sensitive to both the assumed diffusion rate of new technologies and especially to the value placed on a statistical life-year. Our central estimate assumes a value of a statistical life-year of $100,000, but if the value of $150,000 is used, quality-adjusted prices decline by 3.3 percent relative to economy-wide inflation. However, if the value is set to $50,000, we find quality-adjusted prices increase by about 0.7 percent per year, roughly matching the official PCE price index for health care. We explore a variety of assumptions related to the weighting and substitution among technologies, which generally show that quality-adjusted price indexes are declining across a variety of alternative assumptions.

The research literature on quality-adjusted medical price indexes is inconclusive on whether quality-adjusted inflation within health care should be understood as being higher or lower than in other sectors. Several papers that constructed quality-adjusted price indexes for individual medical conditions found steep price decreases after the value of improved health outcomes was taken into account (Cutler et al. 1998; Cutler and McClellan 2001; Shapiro et al. 2001; Berndt et al. 2002; Frank et al. 2004; Lucarelli and Nicholson 2009; Eggleston et al. 2019). However, these case studies of individual conditions found these results in times when new and very beneficial treatments for the medical conditions they studied were diffusing rapidly.

Hult, Jaffe, and Philipson (2018) (henceforth HJP) use the same cost-effectiveness database to study innovations in the health care sector that we use in this study. Their use

of the database for this purpose is novel, but they find that quality-adjusted prices for treatment innovations rose by four percent for the median innovation. The difference between their findings and ours arises from the way the price indexes are formulated. Instead of using an index grounded in utility theory, as we do in this paper, HJP's index is formed as a ratio of price to quality (as measured by life-years added by medical care) or price per quality index. Their approach implicitly assumes that a one percent increase in quality is equivalent to a one percent decrease in price. Their price per quality index is intuitive and does not place a direct value on a statistical life-year, but we show that it places an implicit value on life-years that is implausibly low, leading to an understatement of the value of quality improvement. We find that for the typical innovation the price-per-quality adjustment used in HJP values an additional statistical year of life at around $5,000, which is far below the $100,000 or more typically used in the literature. With an implicit value of just $5,000 per statistical life year, a new treatment that increases life expectancy by 1 year, but costs $5,000, implies no quality adjustment. Meanwhile, the utility-based framework that values a life-year at $100,000 would recognize a large quality-adjustment of $95,000 (=$100,000-$5,000). This low implicit valuation leads to much lower quality adjustment compared to our preferred utility-based index.

In the next section, we provide some background on medical price indexes. We then outline our methodology for constructing price indexes. We then introduce the dataset and show our results. We briefly compare our results to those of HJP and review why their method leads to understatement of the value of quality. We conclude by showing the implications for measured medical inflation, output, and productivity.

Background

Official price indexes at the BLS are often adjusted to reflect quality improvements. A notable example is the case of computers. The quality adjustments in the components of computers (e.g., processor speed, memory, and screen quality) show greater computing power per dollar spent, leading to substantial quality-adjusted price declines. The BLS price index for computers shows a total decline of over 90 percent from 2000 to 2017,

indicating that consumers are getting 10 times more computing power per dollar spent.[2] This sector contrasts with the health care sector where currently prices reported by the BLS are not adjusted for quality.

Price measurement in medical care is complicated by several factors, including the rapid pace of technological change, the presence of third-party payers, and information asymmetries among other factors. Some evidence of quality change may be gleaned from declining mortality rates or individuals living more disability-free years (Cutler, Rosen, and Vijan 2006; Cutler et al. 2017). However, given that non-medical factors may influence health outcomes (e.g., smoking, diet, and obesity), it can be challenging to accurately attribute changes in the health of the population to changes in the medical care sector. Research in this literature uses different methods to adjust for quality without discussing the connections among them (Cutler et al. 1998; Berndt et al. 2002; Frank et al. 2004).[3]

The measurement of medical care prices lies at the center of an important economic question about the forces driving medical expenditures higher. One explanation for the rise in health care spending as a share of the economy is a scenario suggested by Baumol (1967), in which more expenditures shift toward labor-intensive sectors with low measured productivity growth, such as health care. On the other hand, health care has seen significant technological change which has improved health and mortality outcomes over the past 60 years (Cutler, Rosen, and Vijan 2006). If official measures of inflation are not capturing this quality improvement, the reverse scenario could be true: that resources are shifting to health care in response to quality-adjusted prices for health care falling.[4] More accurate measures of price for the health care sector may challenge basic assumptions about price

[2] Bureau of Labor Statistics series CUUS0000SEEE01, personal computers and peripheral equipment in U.S. city average, all urban consumers.

[3] For detailed discussions of this literature and comparisons of methods, see Hall 2016; Sheiner and Malinovskaya 2016; and Dauda, Dunn and Hall 2020.

[4] Chandra et al. (2016) show that consumers prefer and shift toward higher performing hospitals over time. A related economic puzzle is the current slowdown in measured productivity growth in the U.S. (1.6 percentage point lower growth in labor productivity since 2004), which has received considerable attention (Byrne, Fernald, and Reinsdorf 2016; and Syverson, 2017). The full role of the medical care sector contributing to this slowdown is currently unknown given the substantial measurement challenges in this area.

growth in this sector and have implications for understanding across-sector shifts and economy-wide growth.[5]

An important recent contribution in this area is Cutler et al. (2020), who estimate the improvements in medical care productivity based on changes in health outcomes by medical condition for the Medicare population. The approach in Cutler et al. (2020) and the approach in our paper are complementary. Our study exploits cost-effectiveness research which depends on controlled studies to account for changes in health attributable to medical care. In contrast, Cutler et al. (2020) examines changes in population health, due to medical and nonmedical factors, and attempts to isolate changes attributable to the medical care sector using a variety of data sources and assumptions about how medical care impacts health outcomes and how health is allocated across medical conditions in the population. While the method in Cutler et al. (2020) is distinct, their findings are consistent with our findings of large quality-adjusted price declines, as they find large productivity improvements in the medical care sector.

This work also relates to a literature on the value of new goods and their implications for quality-adjusted price indexes outside medical care (Feenstra 1994; Bresnahan and Gordon 1996; Petrin 2002; Aghion et al. 2019; and Redding and Weinstein 2020). Like many papers in this literature, we find that the magnitude of the quality-adjustment is substantial and has important macroeconomic implications.

Methodology

Simple Example

Consider a patient deciding whether a new treatment is worth it or not. The patient will be interested in comparing the price and health benefits of the new treatment to those from other treatments. Determining whether a new treatment represents a price increase or

[5] More generally, accurately measuring the price of health care may be important for understanding growth and sectoral shifts in economies, such as the recent decline in manufacturing, and growth in the service sector (Herrendorf, Rogerson, and Valentinyi 2013; and Duernecker, Herrendorf, and Velentinyi 2019). This recent literature is interested in understanding why resources shift to low productivity service sectors, where health care is typically defined as "low productivity" based on our official measures of inflation.

decrease ultimately depends on the net benefit (benefit minus price) the patient receives from the new treatment relative to the other alternatives.

Let H_{I} be the health benefits of the innovative treatment and let S_{I} be the total price of an innovative treatment (i.e., insurer plus out-of-pocket payments). Let the alternative treatment, the standard of care (SOC) available prior to the new technology, have health benefits H_{SOC} and a treatment price of S_{SOC}. Both H_{I} and H_{SOC} represent the number of healthy life-years accounting for both quality of life and mortality. If the patient values a year of healthy life at an amount of \$VLY, then the new treatment is "worth it" if the health benefit is greater than the additional price of treatment: $\$VLY \cdot (H_{\mathrm{I}} - H_{\mathrm{SOC}}) > S_{\mathrm{I}} - S_{\mathrm{SOC}}$. The difference of the two sides provides the net benefit of the new treatment in dollars:

$$\$VLY \cdot (H_{\mathrm{I}} - H_{\mathrm{SOC}}) - (S_{\mathrm{I}} - S_{\mathrm{SOC}}). \qquad (1)$$

The health benefits, H_{I} and H_{SOC}, are often measured in years of healthy life, generally referred to as quality-adjusted life years (QALYs) in the medical literature (Dolan 2000). Each unit of a QALY is measured as a year of life in perfect health, where quality of life is normalized so that 1 represents a year of life in perfect health while 0 represents death. Typically, the number of QALYs assigned for a particular treatment is the number of QALYs the patient is observed to live following the treatment. The benefits of the innovative treatment in QALYs are then measured relative to the SOC treatment.

The value of life (VLY) is typically measured based on an average estimate in the population and is expressed as the value of a statistical life year (VSLY).[6] In our analysis, we follow Pandya et al. (2015) and Eggleston et al. (2019) in using VSLYs based on three values for a healthy life-year or QALY: \$50,000, \$100,000 and \$150,000 (in 2017 dollars).[7]

[6] For extensive literature on estimating the VSLY, see Viscusi and Aldy (2003).

[7] The \$150,000 amount has been justified as an upper threshold by the World Health Organization (WHO, 2011) because it is approximately three times that of the U.S. GDP per capita (Neumann, Cohen, and Weinstein 2014). We use these values in 2017 dollars, so these estimates are actually a bit conservative, as median household income was \$61,000 in 2017.

To make these ideas a bit more concrete, consider the case of Sovaldi, a hepatitis C treatment that entered the market in 2013 and was viewed as extremely expensive but also very effective (Sanger-Katz 2014). The innovation was particularly important for hepatitis C patients with cirrhosis. A published study of the cost-effectiveness of this treatment found the cost of the Sovaldi treatment to be around $105,488 for patients with cirrhosis (Chhatwal et al. 2015).[8] On the benefits side, patients treated with Sovaldi were observed to live 9.40 QALYs on average following treatment. Prior to the availability of Sovaldi, the standard treatment was with interferon, which had a price of $81,211 and patients treated with it were observed to live 8.28 QALYs on average. The health benefit of Sovaldi relative to standard interferon is therefore 1.12 QALYs.

If we set the VSLY to $100,000, a value often used in the literature, then the relative health benefit of Sovaldi in monetary terms is $\$100k \cdot (9.40 - 8.28) = \$112{,}000$ while the incremental cost is just $\$105{,}488 - \$81{,}211 = \$24{,}277.$ In this case, the benefits outweigh the cost by a large amount, with a net benefit to the patient of $87,723. Even if the VSLY is set to just $50,000, the net benefit is substantial, at $31,723 per patient. In this simple example, a patient would determine that the benefits from Sovaldi are worth the cost relative to the alternative.

Utility-Based Cost-of-Living Index

The formal price index for medical care that we use in this paper was originally presented in Cutler et al. (1998) and further refined by Sheiner and Malinovskaya (2016) and Dauda, Dunn, and Hall (2020). The theoretical justification for the index starts with a utility-based representative consumer model, but the derived index formula closely relates to cost-benefit analysis discussed in the previous simple example. We will be applying the index to measure the price difference between an innovative treatment and the SOC treatment.

[8] The study reviews the cost-effectiveness of the treatment for alternative patient populations. The patient population that we consider here is cirrhotic, implying damage to the liver, and having genotype 1-4, which are the genotypes that Sovaldi is known to treat. All values have been converted to 2017 dollars using the PCE price index for health by function from the Bureau of Economic Analysis.

The index measures the percent change in the quality-adjusted price the consumer receives from the introduction of the innovation, relative to the SOC treatment. The index is constructed as a standard price index that measures the change in the price of a good from one period to the next, with the second-period price in the numerator and the base period price in the denominator. An unadjusted price index would be written, $\frac{S_I}{S_{SOC}}$. To adjust it, we subtract the net change in benefit from the innovation (1) from the price of the SOC treatment in the numerator.[9] This is equivalent to adjusting the innovation price in the numerator with the monetary value of the added health benefit from the new treatment. Given that the index is based on the concept of a cost-of-living index (COLI), we will refer to it as the COLI, to contrast it with the index used by HJP which we will discuss below.

$$COLI = \frac{S_{\text{SOC}} - (S_{\text{SOC}} - S_{\text{I}} + \$\text{VSLY} \cdot (H_I - H_{SOC}))}{S_{\text{SOC}}} = 1 + \frac{S_{\text{I}} - S_{\text{SOC}}}{S_{\text{SOC}}} - \frac{\$\text{VSLY} \cdot (H_I - H_{SOC})}{S_{\text{SOC}}} = \frac{S_{\text{I}} - \$\text{VSLY} \cdot (H_I - H_{SOC}))}{S_{\text{SOC}}}. \quad (2)$$

The price index captures the percent change in the expenditures on treatment, accounting for the treatment benefit measured in improved health. The price index rises with the relative cost of the new treatment, $\frac{S_{\text{I}} - S_{\text{SOC}}}{S_{\text{SOC}}}$, and falls with the relative health benefits of the new treatment, $\frac{\$\text{VSLY} \cdot (H_I - H_{SOC})}{S_{\text{SOC}}}$. On net, it declines if the net benefit of the new treatment is positive and increases if the net benefit is negative.

Returning to the example of Sovaldi and assuming the value of additional QALYs is set at a low value of just $50,000, then the quality-adjusted index is given by applying (2):

$COLI = \frac{\$105{,}488 - \$50{,}000 \cdot (9.40 - 8.28)}{\$81{,}211} = 0.61$. The percent change in the index is found by subtracting the level from one, which indicates that the introduction of Sovaldi represented a 39 percent price decline. If a more central estimate of VSLY of $100,000 is applied, the price index is equal to -0.09. That is, the COLI index indicates a 109 percent decline in

[9] The functional form assumes a constant value for marginal improvements in health following Hall and Jones (2007). They argue that the marginal utility of health is constant with increasing amounts of health and life, unlike the marginal utility of other goods which are generally held to decline with additional units of consumption.

price, making the index fall below zero, because the net benefit is so large that patients would need to be paid to use the SOC technology to make their utility equal across treatments.

The COLI index, which was first formulated by Cutler et al. (1998), is a Laspeyres price index that measures the change in price relative to the base period. We also consider the Paasche equivalent, which measure the price change relative to the second period, where the derivation is parallel to the calculation in Cutler et al. (1998) and is shown in Dunn et al. (2020). The denominator is the reservation price of the innovative good that would make a patient indifferent between the SOC and the innovative good, so we refer to this as the Paasche reservation price (RP) index:

$$RP\ Index = \frac{S_{\mathrm{I}}}{S_{\mathrm{I}}+(S_{\mathrm{SOC}}-S_{\mathrm{I}}+\$\mathrm{VSLY}\cdot(H_{\mathrm{I}}-H_{\mathrm{SOC}}))} = \frac{S_{\mathrm{I}}}{S_{\mathrm{SOC}}+\$\mathrm{VSLY}\cdot(H_{\mathrm{I}}-H_{\mathrm{SOC}})}. \quad (3)$$

Unlike the Laspeyres COLI index, this Paasche RP index (3) is always positive as long as $H_{\mathrm{I}} > H_{\mathrm{SOC}}$. For the case of hepatitis C treatment, using the conservative value of a statistical life-year of $50,000, the RP index equals 0.77, implying that the introduction of Sovaldi entailed a 23 percent price drop. The two utility-based indexes give similar qualitative results but can differ in magnitude for large technological improvements.

Review of the HJP (2018) Methodology

HJP (2018) apply a common method to measure quality-adjusted price indexes in medical care, the goal of which is to measure the price per unit of health produced (Lucarelli and Nicholson 2009; Howard et al. 2015). Specifically, HJP measure the price per QALY (PPQ) added by medical treatment. This index can be written as a simple ratio:

$$PPQ = \frac{S_I/H_I}{S_{SOC}/H_{SOC}}.$$

As discussed elsewhere (Sheiner and Malinovskaya 2016, Dauda, Dunn, and Hall 2020), the PPQ index, while intuitive, does not rest on the same utility formulations as the utility-based indexes and can produce very different results. Sheiner and Malinovskaya (2016) show that the PPQ index is only valid if the patient can purchase as much H as they like at

a constant price per unit of health (i.e., S_{SOC}/H_{SOC}). Both Sheiner and Malinovskaya (2016) and Dauda, Dunn, and Hall (2020) point out that technological constraints could make it impossible to purchase any additional health or there may simply be diminishing returns to additional health care inputs, causing additional units of health to be costlier. In either case, it would be impossible to purchase any further health at the same price.

Returning to the Sovaldi example, if a patient could buy as much health as they wanted at the standard of care price per QALY of $\frac{\$81{,}211}{8.28\ QALY} = \$9{,}808$, then the higher price per QALY of Sovaldi would make them worse off. However, prior to Sovaldi, many patients experienced liver damage from hepatitis C. If it were possible, these patients would certainly have paid $9,808 per QALY to improve their health, as it is well below standard estimates of the monetary value of a QALY. The reason patients did not pay for improved health was therefore clearly due to technological limitations.

The issue with the PPQ index may be seen easily if, following Sheiner and Malinovskaya (2016), we rewrite the PPQ index in a form that parallels the COLI utility-based index:

$$PPQ = \frac{S_I}{S_{SOC}}\left(\frac{H_{SOC}}{H_I}\right) = \frac{S_I}{S_{SOC}}\left(\frac{H_{SOC}}{H_I}\right) + \frac{S_I}{S_{SOC}} - \frac{S_I}{S_{SOC}}\left(\frac{H_I}{H_I}\right) = \frac{S_I - \frac{S_I}{H_I}(H_I - H_{SOC})}{S_{SOC}}. \quad (4)$$

Comparing formulas (2) and (4), it is clear that the PPQ index implicitly values incremental improvements in health at the average price of the improvement for the innovative treatment, $\frac{S_I}{H_I}$, and not at the marginal benefit, as in the utility-based indexes, which is equal to the VSLY. The indexes are only equal if the average price equals the marginal health benefits. This is almost never the case, as a major motivation for innovative treatments is for the marginal benefits to exceed the price, providing a net benefit to patients, while also providing a profit to firms.

Returning to the example of hepatitis C treatment, if we measure the price per QALY delivered by innovative Sovaldi compared with the price per QALY delivered by SOC

interferon, the price of hepatitis C treatment rises by 14 percent, suggesting that patients are actually worse off from the innovation.

$$PPQ = \frac{\$105{,}488/_{9.4\ QALY}}{\$81{,}211/_{8.28\ QALY}} = \frac{\$105{,}488 - \frac{\$105{,}488}{9.4\ QALY}(9.4\ QALY - 8.28\ QALY)}{\$81{,}211} = 1.14$$

Based on this index, patients should not adopt this new technology, even though studies consistently find it to be cost-effective. The reason for the extreme difference between the utility-based indexes and the PPQ index is the implied value placed on incremental improvements in health. In this particular PPQ index, an additional year of healthy life is valued at the price of a QALY from the innovation or $\frac{\$105{,}488}{9.4\ QALY} = \$11{,}221$, which is far lower than the marginal benefit, as there is no accepted VSLY anywhere near it.

Tufts Cost-Effectiveness Database

The main data source used in the analysis is the Tufts Medical Center Cost-Effectiveness Analysis Registry (CEAR) database (Center for the Evaluation of Value and Risk in Health, 2019). The registry is a comprehensive database of original cost-effectiveness studies covering a wide variety of treatments and diseases. The database summarizes and reviews published studies, where each article is screened before inclusion in the registry. To satisfy the criteria for inclusion, the research must be published in English, be an original cost-effectiveness analysis, and measure health benefits as QALYs. Review articles, editorials, and articles missing key features (e.g., quality measures) are excluded. Each article is reviewed by two readers that have been trained in cost-effectiveness and decision analyses. These readers follow a standardized set of forms and instructions and extract over 40 variables for each article, as well as provide specific ratings regarding the quality of the study. The studies vary on numerous dimensions that are recorded in the data: type of intervention (e.g., pharmaceutical), condition treated (e.g., cardiovascular), funding source (e.g., government), as well as numerous other variables. The types of studies vary in the methods that are applied, which are described in the abstract of each paper which is also one of the included data elements.

The CEAR studies present a diverse array of methods applied in the medical literature. Based on a simple word search of the title and abstract, we categorize the type of studies observed in the database. For example, we find that about 34 percent of the articles have the word "random" or "trial."[10] Other types of cost-effectiveness studies include meta-studies or disease-model simulations. The quality of each study is rated by the readers of the study based on a variety of criteria (e.g., health economic methodology, consideration of uncertainty, and transparency). The methods used in forming both cost and QALY estimates vary depending on the study, but they are unified in their goal of estimating the key elements that are necessary to evaluate the cost-effectiveness of treatment, which are the same elements needed to form a quality-adjusted price index from a new treatment.[11]

The version of this database applied in our study contains 8,244 cost-effectiveness studies with 90 percent of the studies coming from the 2004 to 2019 period. Many of the studies in the database contain the critical four elements for understanding the price impact of treatment innovations: (i) the price for the new treatment (i.e., insurer plus patient costs); (ii) the price of treatment for the previous standard of care (SOC); (iii) the QALYs produced by the innovation; and (iv) the QALYs produced by the previous SOC.[12,13] The SOC treatment typically represents the incumbent treatment prior to the arrival of the innovation. Only about 55 percent of the articles in the database include all four of these elements, so not every study may be used to form a quality-adjusted price index. However,

[10] This is based on a simple word search for the terms such as "random" or "trial", which are not classified as meta-studies. A study is classified as a meta study if it contains the word "meta". It is classified as a simulation if it contains the words "simulation" or "markov". All other studies are classified as "other".

[11] While studies have a lot of distinct features, we believe they are generally reflective of the technology being studied. For example, we observe four studies on beta blockers, which are believed to be highly effective treatments for high blood pressure. While the estimates are distinct in each study, all four of them show beta blockers to be highly effective.

[12] The calculation considers only the private cost and benefit. An accurate social cost calculation would consider opportunity costs and remove profits, but this is a challenging task. For example, accounting for the resources involved in prescription drugs and other technologies may be particularly challenging as they involve large fixed R&D costs, but relatively low marginal costs.

[13] Generally, the cost-effectiveness studies in the CEAR database do not report QALYs relative to no treatment. When the CEAR data reports the level of QALYs associated with a treatment, it is the mean level actually observed among patients who received that treatment. This differs from the amount of QALYs relative to no treatment since the number of QALYs a patient lives with no treatment is not normally zero. As noted by Dauda et al. (2020), one of the first application of a PPQ-type index was by Berndt et al. (2002), who argued that the numerator and denominator of the index should both be measured relative to no treatment.

a single article may contain multiple comparisons of treatments, increasing the number of innovations that may be analyzed. We have a total of about 11,000 observations for which we observe the necessary elements to form quality-adjusted price indexes. To clean the data, we remove some of the outlier studies and estimates using the same criteria as HJP.[14]

Descriptive Statistics

The descriptive statistics are reported in Table I. These elements include the innovator QALY (H_I), SOC QALY (H_{SOC}), innovator price (S_I), SOC price (S_{SOC}), innovator price per QALY, SOC price per QALY, and the incremental cost-effectiveness ratio (ICER, defined below). As might be expected, the QALYs associated with innovations are on average greater than QALYs associated with the SOC, with a difference of 0.40 QALYs at the mean, and a slightly larger difference at the median. The mean of the innovator price is higher than the mean of the SOC price, suggesting a higher price generally corresponds to greater medical benefits.

The ICER is a concept often used in cost-effectiveness analysis and is defined as the change in incremental cost divided by the change in incremental effectiveness: $ICER = \frac{S_I - S_{SOC}}{H_I - H_{SOC}}$. The ICER is frequently used as a basis for deciding whether a technology should be adopted. When $S_I - S_{SOC} > 0$, and $H_I - H_{SOC} > 0$, then the ICER is equal to the marginal price of health when moving from the SOC treatment to the innovative treatment. Mechanically, if the VSLY is greater than the ICER, then the innovation leads to a quality-adjusted price decline. To see this, multiply $\$VSLY - ICER$ by the change in health $H_I - H_{SOC}$: $(\$VSLY - ICER) \cdot (H_I - H_{SOC}) = (\$VSLY - \frac{S_I - S_{SOC}}{H_I - H_{SOC}}) \cdot (H_I - H_{SOC}) = \$VSLY \cdot$

[14] The selection rules outlined in HJP: "We omit observations with quality values greater than 100, since it does not make sense for a treatment to add more than 100 years to someone's life. We also omit studies with negative quality values. We omit observations with negative cost for either the innovation or the SOC. We also omit observations where the ICER, price, or price per QALY for the innovation or the SOC is over $10." The negative quality values in many of these studies are because QALYs are estimated relative to some baseline, which is not problematic for the utility-based indexes, but cannot be used for the PPQ index. The negative QALYs account for just half a percent and they have no effect on the main results if they are included or excluded. In order to normalize expenditures in the studies across years to the year 2017, we use a medical care deflator to ensure that the same quantity of medical care may be purchased in 2017 as in the year of the study. We convert medical expenditures into 2017 dollars using the PCE deflator for medical care, rather than the medical CPI, which is only relevant for out-of-pocket costs (Dunn, Grosse, and Zuvekas 2018). However, the main findings do not change by the use of either index. We convert to U.S. dollars using yearly exchange rates.

$(H_I - H_{SOC}) - (S_I - S_{SOC})$. This term is equal to the net welfare benefit, so $\$VSLY - ICER$ determines if the quality-adjusted price change is positive (i.e., welfare reducing) or negative (i.e., welfare improving).

As the ICER is formed as a ratio of differences, it can be both positive or negative, and the interpretation of the ratio changes depending on whether the price and QALY of the innovation is increasing or decreasing, relative to the SOC treatment. Researchers typically divide ICERs into four quadrants: increase in price and increase in QALY (north-east quadrant (NE)); decrease in price and increase in QALY (south-east quadrant (SE)); increase in price and decrease in QALY (north-west quadrant (NW)); and decrease in price and decrease in QALY (south-west quadrant (SW)). A decrease in price and increase in QALY (SE) always implies a welfare improvement, while an increase in price and decrease in QALY (NW) is a welfare reduction. Whether the welfare change from the other two quadrants are a positive or negative depends on the VSLY.

The ICER is available for nearly all studies in the database, while the components of the ICER are only available for around 55 percent of the studies. ICERs for all four quadrants are shown in Table I, with over 85 percent of studies showing QALY improvements. For observations that show an increase in price and an increase in QALYs (NE), the median ICER is just $32,812 indicating that additional health is being purchased for less than its utility value. Even if the value of a statistical-life year is set to just $50,000 the net value per life year gained at the median ICER is $17,188 (=$50,000 – $32,812). The bottom of Table I shows that the distribution of ICERs is very similar in the subsample used to form quality-adjusted price indexes (i.e., the sample where all the components of the ICER are observed), suggesting no systematic differences with the full sample.

Results

Table II reports the distribution of quality-adjusted price indexes constructed according to the utility-based formulas with components taken from the cost-effectiveness studies in the CEAR database. The table shows quality-adjusted prices decline at both the mean and median of the distribution across all assumptions for the value of a life-year for both the

utility-based Laspeyres COLI and Paasche RP indexes. The magnitude of the estimates is highly sensitive to the VSLY, with larger declines for larger VSLY. In fact, the mean change in the Laspeyres COLI is over 100 percent for the VSLYs of $100k and $150k implying that the index level is negative. As mentioned previously, the negative index level is caused by the welfare improvement exceeding the treatment price, which arise from drastic improvements in technology. By construction, the Paasche RP index does not show declines greater than 100 percent and shows fewer extreme changes relative to the Laspeyres COLI.

Choosing between the two equally valid indexes, the COLI and RP, may be difficult when they take on such widely different values. Therefore, we adopt Trajtenberg's (1990) approach by taking a simple average of the two indexes in levels.[15] This distribution of the average of the two utility-based indexes is shown at the bottom of Table II, which similarly shows large quality-adjusted price declines.

Table III shows price changes based on alternative index formulas including the PPQ index. For additional comparisons we also include price index changes based on hedonic regressions (described in detail in the first section of the appendix), which are often used for quality-adjusting prices, but which Berndt et al. (2007), Sheiner and Malinovskaya (2016) and Dauda, Dunn, and Hall (2020) argue are not appropriate in the health care setting where consumers do not face the marginal price of care. All of these methods implicitly place a dollar value on QALYs, and we include the implicit value of the QALY in italics below the estimated price change. While the typical innovation shows a quality-adjusted price decline when using utility-based indexes, these alternative index formulas show quality-adjusted prices increasing at the mean and median. The reason for the price increase is the low implicit value placed on a statistical life year. At the median, the value of a statistical life year for the PPQ index and hedonic indexes is below $20,000 across all methods, which is far below any value of a statistical life-year observed in the literature.

[15] This is similar to a Fisher index that is formed based on the geometric mean of a Laspeyres and Paasche index. The negative values for the COLI index imply the geometric mean is not possible, so we take the arithmetic mean.

The observation that the hedonic methods place an extraordinarily low value on QALYs is consistent with Pakes (2004) who warns that hedonics tend to undervalue quality changes. Although both the PPQ index and utility-based indexes are adjusted for quality, the low implicit valuation placed on health leads to a price measure practically unrelated to the utility-based index. Figure 1 shows a scatterplot of the price change from the average of the utility-based indexes for VSLY of $100k and the PPQ index. The figure shows a red 45-degree line indicating equality between the indexes and the blue line shows the fitted value. About 80 percent of the observations fall below the 45-degree line, indicating the utility-based index is generally lower than the PPQ index. The fitted value shows a positive relationship between the indexes, but the explanatory relationship is weak (R-squared= 0.029). Around 40 percent of the observations fall in the shaded blue region where the PPQ index shows a price increase and the utility-based index shows a price decline.

One notable feature of Table II is that the mean value of the indexes deviates greatly from the median, even after removing the outliers. The reason for this large deviation is that in many cases, the improvements in health can be very large relative to the price, and especially large as the VSLY increases. Figure 2 shows a histogram of the average COLI and RP quality-adjusted price index based on a VSLY of $50k and $150k, where the quality-adjusted price change is winsorized at -5 and 5. The histogram shows a wide distribution of values, both negative and positive, but the distribution is clearly skewed toward price declines. The entire distribution shifts much further to the left when the VSLY increases from $50k to $150k. These large price declines are often from innovations where health care spending is highly efficient relative to the SOC, as measured by the ICER, so that the incremental price for incremental improvements in health is very low. The relationship between the ICER values and quality-adjusted indexes is explored in greater detail in the appendix and in Table AII.[16]

[16] Table AII of the appendix shows a table of quality-adjusted price declines by ranges of ICER values. Efficient health care changes, as reflected by very low ICERs, are associated with extremely large quality-adjusted price declines.

A few examples in Table IV provides the key insights to patterns observed in both Figures 1 and 2. For these studies, the corresponding QALYs and prices are shown along with the utility-based price changes and the PPQ index. The examples show a range of different types of innovations, from new drugs, to management, to screening practices, and a vaccine. In each case, the utility-based COLI and RP indexes show large price declines, assuming a VSLY of $100,000. The first example shows the impact of using the anti-inflammatory drug, Enbrel introduced in 1998, to treat rheumatoid arthritis for patients not responding well to standard treatments such as methotrexate. Enbrel is $100k more expensive, but there is a large gain of more than three QALYs relative to the SOC. This leads to a large quality-adjusted price decline from the utility-based index, but a price increase based on the PPQ index.

The second example, modulated radiation treatment for prostate cancer, similarly shows price declines for the utility-based indexes, while the PPQ index shows a price increase. The third example of diabetes self-management demonstrates that price declines are not exclusive to only high-tech solutions, but also improved management of a condition.

The fourth example is screening for colon cancer, which leads to extremely large declines in utility-based quality-adjusted prices. This decline starkly contrasts with the PPQ index that shows a large price increase for a screening which is widely believed to be beneficial by medical experts. This example helps explain the types of patterns we see in Figures 1 and 2. While this is an old technology, screening for colon cancer remains a highly relevant treatment that has diffused steadily throughout our sample period (Table AIII).

The fifth example is for the human papillomavirus (HPV) vaccine, introduced in 2006, which prevents the HPV virus, a virus that increases the risk of cancer. This technology leads to even more extreme quality-adjusted price declines and is widely viewed as an important and cost-effective technology that has been disseminating among children, teens, and young adults.[17] While the utility-based indexes show extreme quality-adjusted price

[17] One may be concerned that the price decline could affect the mean reported in Table II, but the decline is so large that we mark it as an outlier that does not contribute to the mean, despite its importance to the health of the population. This exclusion likely biases our estimated mean upward.

declines, the PPQ index actually shows a 600 percent price increase. These technologies are beneficial, but another question is whether they are actually diffused and applied in practice. Table AIII presents evidence from the medical care literature for four out of the five examples presented here, with annual diffusion rates ranging from 1.6 to 6.7 percentage points per year.

To address potential concerns that these price declines might come from certain subcategories, Table V reports these quality-adjusted prices change across various dimensions of the data based on the average of the two utility-based indexes shown at the bottom of Table II and using a VSLY of $100,000. The different dimensions examined include, condition categories (e.g., cardiovascular or musculoskeletal), type of intervention (e.g., pharmaceutical or device), the funding sponsor (e.g., government or pharmaceutical firm), and type of study based on a simple word searches of the title and abstract (e.g., randomized or simulation). Table V also shows, in the right three columns, an additional breakout of studies rated as being high quality based on the evaluations of the readers scoring the quality of the research studies along various dimensions. While there are some differences in the mean and median across disease categories, type of intervention, funding sponsor, and type of study, what stands out most is the persistent decline in quality-adjusted prices from treatment innovations. The estimates indicate that price declines from innovation are a prevalent feature of the medical care sector, showing declines at both the mean and median across all dimensions of these data.

With these estimates in mind, it is useful to think again about why the quality-adjustment based on the PPQ is so low relative to the utility-based indexes. Recall that the PPQ quality adjustment is based on the average price per unit of QALY, which is around $5,000 for the median innovation in Table III. Assuming a rational decision maker, this average price per QALY will be far below the QALY benefits as patients receive treatments only if the QALY benefits exceed the price of treatment, as is the case for the typical innovation in Table II. If patients receive multiple treatments for a condition, with diminishing returns to each treatment, then the marginal health benefits per QALY will exceed the marginal price per QALY for all treatments, except perhaps the last treatment where the marginal price

per QALY of the last dollar spent on treatment may be the same as the marginal benefit. In this example, the marginal price per QALY for the last treatment would approximate the marginal benefit to the patient, but the average price per QALY across treatments would not. If the marginal price per QALY is equal to the average price per QALY (i.e., no diminishing returns to treatment or technological constraints), then this would imply the marginal health benefit per QALY of the last treatment is around $5,000, which is implausibly low. Again, the PPQ index assigns an extremely low implicit value to QALYs, which leads to small quality adjustments and substantially higher PPQ indexes, relative to the utility-based indexes.

Implications for Medical Care Inflation, Output and Productivity

In this section, we construct an aggregate quality-adjusted index for the medical care sector from 2000 to 2017. The utility-based estimates in the previous section provide evidence that innovations generally lead to price declines, but additional steps are necessary to incorporate this information into an aggregate quality-adjusted index. While it is a worthwhile goal to match every innovation to every condition and develop quality-adjusted indexes for each one, this is an enormous task that is outside the scope of this paper. Instead, we use proxies for the diffusion rate of new technologies and indexes for groups of conditions to form a more feasible quality-adjusted price index.

To estimate the quality-adjusted index, let S_t be the average expenditure for a condition at time t that is a mixture of both SOC and innovative treatments. If w_t is the share of innovative treatments at time t, then the average expenditure across treated patients is $S_t = w_t \cdot S_I + (1 - w_t) \cdot S_{SOC}$. Next, let H_I be the QALYs delivered by the innovative treatment and H_{SOC} be the QALYs delivered by the SOC, as before. In this case, average health across patients delivered by the composition of treatments at time t is equal to $H_t = w_t \cdot H_I + (1 - w_t) \cdot H_{SOC}$. Using the COLI formula, the quality-adjusted price index (P_t) for the condition in time t is then equal to:

$$P_t = \frac{S_t - \$VSLY \cdot (H_t - H_{t-1})}{S_{t-1}}$$

If $\Delta w = w_t - w_{t-1}$, then $H_t - H_{t-1} = \Delta w \cdot (H_I - H_{SOC})$. We can separate the index into two components.

$$P_t = \frac{S_t \cdot (\Delta w) + S_t \cdot (1 - \Delta w) - \$VSLY \cdot \Delta w \cdot (H_I - H_{SOC})}{S_{t-1}}$$

$$= (1 - \Delta w) \cdot \frac{S_t}{S_{t-1}} + (\Delta w) \cdot \frac{S_t - \$VSLY \cdot (H_I - H_{SOC})}{S_{t-1}}. \quad (5)$$

The first term is an unadjusted medical care treatment index that tracks the cost of treatment over time. The second term is the quality-adjusted component of the index. The index is a weighted average of quality-adjusted and unadjusted indexes, where the weight Δw is based on the diffusion rate of new treatments. This index is similar in spirit to the index in Feenstra (1994) where a quality-adjusted index is formed by combining a more traditional index without quality adjustment with a second term that captures the quality change based on the share of expenditures shifting toward new goods.[18]

For the year in which the innovation enters the market, the quality-adjusted index for the innovation provides an upper bound to the quality-adjusted component of P_t, $\frac{S_t - \$VSLY \cdot (H_I - H_{SOC})}{S_{t-1}}$, as long as, $S_I > S_{SOC}$, which is a pattern we typically see in the data. To see this, consider that in the period before the innovation is introduced, $w_{t-1} = 0$ and $S_{t-1} = S_{SOC}$. Therefore:

$$\frac{S_t - \$VSLY \cdot (H_I - H_{SOC})}{S_{t-1}} = \frac{w_t \cdot S_I + (1 - w_t) \cdot S_{SOC} - \$VSLY \cdot (H_I - H_{SOC})}{S_{SOC}} < \frac{S_I - \$VSLY \cdot (H_I - H_{SOC})}{S_{SOC}}. \quad (6)$$

Based on the equation (6) inequality, the last term, which is the COLI index (equation (2)), can be substituted into equation (5) to create an upper-bound quality-adjusted price index change which we bring to the data:

$$P_t \approx (1 - \Delta w) \cdot \frac{S_t}{S_{t-1}} + (\Delta w) \cdot \frac{S_I - \$VSLY \cdot (H_I - H_{SOC})}{S_{SOC}}. \quad (7)$$

[18] However, the index is distinct from Feenstra (1994) as it uses the cost-effectiveness database to infer quality improvements from innovation, rather than revealed preferences.

The next step is to obtain corresponding empirical estimates to substitute into equation (7). For the component, $\frac{S_I - \$VSLY \cdot (H_I - H_{SOC})}{S_{SOC}}$ in equation (7), we must determine how to aggregate over the thousands of innovations in the data. Without additional information on what specific innovations are adopted, we apply a simple arithmetic mean of the different innovations, which implicitly assumes that treatments gravitate more toward innovations that have larger quality-adjusted price declines, although alternative functional forms are explored in the appendix (e.g., Cobb Douglas and CES functional forms) with qualitatively similar results.[19] As part of the average, we are also including innovations that lead to quality-adjusted prices increasing, which assumes that welfare-reducing technologies are also adopted and is another assumption that we explore in the appendix.

The quality-adjusted index that we apply comes from Table II. As the utility-based COLI index and the RP index are equally valid, we use the average of the two listed on the bottom of Table II to substitute for $\frac{S_I - \$VSLY \cdot (H_I - H_{SOC})}{S_{SOC}}$ in (7).

For $\frac{S_t}{S_{t-1}}$ we want an unadjusted index that includes the full price of treating a condition, including shifts across treatments, such as the use of new and expensive technologies. For this we use the disease-based price index from BEA's Health Care Satellite Account (HCSA), which is estimated to be representative of the U.S. population and tracks the average of the full price of treatment for a comprehensive set of more than 260 conditions (Dunn et al. 2015; Dunn et al. 2018). Following the recommendation of Berndt et al. (2007) and National Research Council (2010), the full price of treatment includes the total expenditures on treatment including the cost of treatment innovations.[20] For example, for diabetes, the HCSA tracks the average expenditure to treat a diabetic patient in 2000 relative to 2017, where the cost includes physician services, hospital services, prescription

[19] The third section of the appendix discusses the implications of using different theoretical specifications of the utility function as the basis for aggregating across innovations.

[20] This type of price index is often referred to as a Medical Care Expenditure Index in the literature, as it is a price that includes all associated expenditures on treatment. The Health Care Satellite Account (HCSA) is available here https://www.bea.gov/data/special-topics/health-care, with a more detailed description available in Dunn et al. (2015). The index aggregates over condition-specific expenditure indexes using a Fisher index formula.

drugs, and other medical care services. The HCSA's disease-based index is not adjusted for quality.

From 2000 to 2017, the aggregate disease-based index grew by 1.96 percent above aggregate economy-wide inflation (as measured by BEA's implicit deflator for GDP) on an annual basis. In contrast, the official PCE index for health care (PCE by function health index from BEA), which only tracks prices of identical products and services (e.g., the price of a 15-minute office visit), grew 0.53 percent above economy-wide inflation . The higher price growth for the disease-based index is partially driven by expenditures shifting to more costly technologies (Dunn et al. 2018), a shift which is not captured in the official PCE index for health care.

For both the CEAR data and the HCSA, we can map innovations, spending, and unadjusted price indexes to 18 broad condition categories that are common across datasets (e.g., circulatory, neoplasms, mental health, etc.). Therefore, in addition to constructing a simple aggregate price index based on equation (7), we also construct an alternative weighted quality-adjusted price index where we estimate the quality-adjusted price change for each of the 18 broad condition categories, then aggregate based on the expenditure share of each category in 2000.

The remaining empirical challenge is estimating Δw, the share of the expenditure growth arising from treatment innovations. There are many ways for treatments to be "innovative" (e.g., entirely new treatments, shifts in practice patterns, or the timing of when services are received) and there is no standard measure of innovation adoption. Therefore, we turn to three alternative proxies for innovation: (i) the share of the increase in drug spending that is on new prescription drug molecules; (ii) the share of the increase in medical service spending on "new" CPT codes; and (iii) shift in hospital adoption of best practices from 2004 to 2009. None of these proxies are perfect but taken together they suggest that Δw between 2 and 4 percent.

The first approach focuses specifically on the introduction of new molecular compounds, which are easier to identify in claims data relative to other forms of innovation (e.g., the

speed at which particular services are performed). For the period from 2002 to 2017, we use estimates from the IBM MarketScan® Research Databases, a private claims data source from Truven Health Analytics, part of the IBM Watson Health business, containing millions of commercial enrollees under the age of 65, as well as retirees over the age of 65. New molecules are defined as molecules that did not appear in the database in the period 2000 or 2001. We report new molecules' share of expenditures in Table AIV of the online appendix. By 2017, we find that approximately 56 percent of spending is on molecules introduced 2002 or later. This implies an average annual growth rate of 3.5 percent in spending on new treatments per year. For those concerned that this estimate may be biased by rebates that are not accounted for in claims databases and are disproportionally used by brand name drugs, we have also done the calculation based on out-of-pocket expenditures, which are unaffected by rebates. Considering only out-of-pocket costs, spending increased by 2.3 percent on an annual basis.

As a second indicator of the rate of technological adoption, we identify rates in which new procedure codes are used. It should be noted that changes in CPT codes do not necessarily represent new technologies, as codes may change for administrative purposes, but they likely reflect the general pace of change in the medical care system.[21] Similar to the calculation for spending on new molecules, for the period from 2002 to 2017, we use estimates from the IBM MarketScan® Research Databases. New codes defined as codes that did not appear in the database in the period 2000 or 2001. The estimates are shown in the right-hand columns of Table AIV. By 2017, we find that approximately 30 percent of spending is on CPT codes introduced in 2002 or later. This averages out to around 1.9 percent increase use of new treatments per year.

As a third indicator of the rate of technological adoption, we use rates at which hospitals adopt recommended treatments, based on process measures of care in the Hospital Compare database from the Centers for Medicare and Medicaid Services. These process measures are not based on newly invented treatments, but measure whether effective

[21] These codes may also miss important innovations. For instance, an innovation may be to apply the "correct" CPT codes to patients, and not simply apply a new CPT code.

treatments are provided to particular patients, such as "receiving an aspirin at arrival" for heart attack patients or "received preventative antibiotic(s) one hour before incision" for surgery. These values are reported in Table AV of the appendix. In 2004, the average score across the 17 observed measures was 77 percent receiving proper care, and by 2009 the average rate was 96 percent, indicating an average percent change in practice patterns per year of 3.6 percent.[22] Chandra et al. (2016) examine acute myocardial infarction (AMI) treatments over the same period when this Hospital Compare data became available. They find improvements in survival over the study period and also find that patients gravitate toward higher quality hospitals. Our three proxy measures align with the diffusion rates from our four specific examples reported in Table AIII, which range from 1.6 to 6.7 percent per year.

To estimate the aggregate quality-adjusted index, P_t, we apply the corresponding estimates to each of the components of equation (7). Table VI reports the quality-adjusted price index based on alternative diffusion rates and alternative assumptions of the VSLY at the bottom of the table. All the estimates in Table VI have been deflated with the GDP deflator so the indexes measure price changes relative to economy-wide inflation. The estimates are sensitive to both assumptions. For each value, we bootstrap our estimates to obtain a 95 percent confidence interval by resampling cost-effectiveness articles with replacement in the CEAR database 1,000 times. Our central estimate, assuming a diffusion rate of 2.8 percent per year and VSLY of $100,000, implies that quality-adjusted prices are falling by about 1.23 percent per year, relative to economy-wide inflation. For a conservative VSLY of $50,000, our central estimate is an inflation rate of just 0.77 percent per year relative to economy-wide inflation, slightly higher than the official PCE health index. In this case, the downward pressure from the benefits of treatment innovations partly offsets the higher

[22] There are several issues with this particular measure, but it is not clear whether it may overstate or understate the true rate of the adoption of new technologies. Hospitals may target improvement in these tasks, suggesting this may overstate changes in treatment practices. On the other hand, as these practices approach 100 percent, there is no possibility for the score to improve along these dimensions, and other changes in treatment patterns will not be reflected in these indicators, suggesting that these estimates may understate practice pattern changes.

price growth in the disease-based index. With a VSLY of $150,000, quality-adjusted prices decline by 3.1 percent per year relative to economy-wide inflation.

The estimates above rely on calculating a single quality-adjusted price index for all of health care. As an alternative, we re-calculate these estimates at the condition category level using a mapping between the BEA HCSA and the CEAR database for 18 broad condition categories. Using this mapping, we estimate quality-adjusted price changes for each condition category separately by estimating equation (7) for each category. These estimates are reported in Table VII where we also include the number of observations in the CEAR database, the spending share from the BEA HCSA in 2000, and the unadjusted annual price growth rate specific to each condition category from the BEA HCSA. In Table VII we report condition-specific quality-adjusted price indexes based on three different VSLY and a central diffusion rate of 2.8 percent. The bottom of Table VII aggregates across these condition categories using the reported spending shares from 2000. Similar to Table VI, we produce estimates based on different VSLYs and diffusion rates and estimate 95 percent confidence intervals for each value using a bootstrap estimate.

Table VII shows that the number of studies generally correlates with the share of spending in 2000. A notable exception is the infectious disease category, that has a very large number of observations from the CEAR database. This is an area where there has been substantial innovation, including treatments for hepatitis C and HIV. Interestingly, it is a category where the unadjusted price of treatment has grown the fastest, but the quality-adjusted price index actually falls faster than the overall average. While condition specific weights are applied to the estimates in Table VII, the estimates correspond quite closely to those in Table VI that are unweighted. Our central estimate with a diffusion rate of 2.8 percent per year and VSLY of $100k shows a quality-adjusted price decline of 1.33 percent per year.[23]

Tables VI and VII show a range of estimates but impose several assumptions. Section 3 of the appendix and Table AVI present alternative estimates to show how these results are

[23] Price indexes are similar if weights from 2017 are applied instead of 2000, with a central estimate showing a price decline of 1.42 percent per year.

affected by key assumptions. For instance, rather than using an arithmetic mean we apply functional forms implied by the Cobb-Douglas and CES utility functions. As another alternative, we exclude innovations that raise quality-adjusted prices. We also explore the robustness of the results based on the country of study and also remove extreme quality-adjusted price changes. Although our estimates are sensitive to alternative assumptions, we find that our central estimates with a VSLY of $100,000 tend to show quality-adjusted prices declining and correspond well with estimates reported in Tables VI and VII.

Our estimated quality-adjusted price indexes are somewhat speculative, as they do not match diffusion in treatment with the associated technology on a condition-by-condition basis. Moreover, health outcomes in a real-world setting may be distinct from those found in clinical data. However, these critiques are partly addressed by other research. Dauda, Dunn, and Hall (2020), examine health outcomes in a real-world setting using claims data for three acute health conditions. When VSLY is set to $50,000, they find even more rapid declines in the quality-adjusted prices of 3.1 percent per year. Also using medical claims data and real-world outcomes, Romley et al. (2019) look at risk-adjusted health outcome measures compared to costs for eight medical conditions for the Medicare population accounting for 20 percent of Medicare A and B spending. Although they do not apply a utility-based framework, they also find large improvements in health outcomes relative to costs. Cutler et al. (2020) apply a different approach that measures quality based on health outcomes for the entire Medicare population across a comprehensive set of conditions and attribute a portion of the observed health outcomes to the medical care sector. Despite their distinct approach, their framework is both comprehensive and utility-based, similar to this paper. The estimates imply a quality-adjusted price decline of 2.4 percent per year for the Medicare population, showing slightly larger price decline than our central estimates.

Implications for Output and Productivity

The estimates in Table VII have implications for health care inflation, output, and productivity. Our central estimate from Table VII shows an annual price decline of 1.33 percent, relative to aggregate GDP inflation, while the PCE health index showed an

increase of 0.53 percent (shown at the top of Table VI). This implies that inflation is overstated by 1.86 percentage points (i.e., 0.53 – (-1.33) = 1.86)) and real output growth is therefore understated by the same amount.

The official estimates of multifactor productivity growth most related to our study from BLS are for Hospitals and Nursing and Residential Care Facilities (North American Industry Classification System (NAICS) industries 622 and 623) and Ambulatory Health Care Services (NAICS industry 621). The official estimates show that multifactor productivity (MFP) for Hospitals and Nursing and Residential Care Facilities declined by 0.2 percent annually from 2000 to 2017 and MFP for Ambulatory Health Care Services increased 0.4 percent annually. For our rough estimate of the productivity change for the health care sector, we use the hospital and nursing homes estimate, which accounts for a larger share of spending.[24]

Multifactor productivity growth is defined as the growth in real output divided by the growth in real inputs. The potential effect of the quality-adjusted index on the measurement of multifactor productivity growth depends on the magnitude of the quality-adjustment bias in the output price index which is used to deflate nominal spending measures to give real output. We define the bias to be the difference between the average annual growth rate of the BLS output price index for hospitals and nursing homes, which equals 0.46 percent per year, relative to economy-wide inflation and the average annual growth rate of the quality-adjusted index (i.e., -1.33 percent), which equals 1.79. This implies a quality-adjusted multifactor productivity growth rate of 1.56 percent a year (i.e., official productivity growth + bias= - 0.23 + 1.79).[25]

This growth rate is faster than the BLS multifactor productivity growth rates for manufacturing, retail, and services (0.6, 0.2, and 0.3 percent per year, respectively), but

[24] The multifactor productivity estimates are from the BLS 1987–2019 Combined Sector and Industry Multifactor Productivity (revised 03/23/2021) https://www.bls.gov/mfp/mprdload.htm#Multifactor%20Productivity%20Tables. Official productivity growth increased 0.4 percent annually for Ambulatory Health Care Services over the 2000-2017 period.

[25] Table AVII of the appendix contains the multifactor productivity estimates and the associated adjustment. Improvements in medical care technology used by the hospital sector (e.g., improved pharmaceuticals) are included in this calculation.

slower than the growth rates in multifactor productivity for the high-tech manufacturing industries computer and electronic products (NAICS 334) and data processing, internet publishing, and other information services (NAICS 518,519) (4.2 and 2.3 percent per year, respectively). In summary, the quality adjustment is substantial enough for the hospital and nursing homes industry to be one of the more productive areas of the economy, rather than being an area of lagging productivity growth.

Conclusion

Using a database containing thousands of cost-effectiveness studies, we find widespread price declines from the introduction of treatment innovations, similar to those found in other high-technology industries. We then calculate a quality-adjusted deflator for the health care sector for years 2000-2017 based on the quality-adjusted prices of innovations and assumptions about how quickly those innovations diffuse. Our estimates suggest that the current PCE index for health care may overstate annual inflation by as much as 1.86 percentage points. Given that health care was found to be nearly 18 percent of the economy in 2019 (Martin et al. 2020), accounting for quality adjustment has profound implications for the measurement of economy-wide output and productivity growth.

Although we find a wide range of potential estimates based on our current framework, there is additional work to be done to obtain a more precise measure of the quality-adjusted price indexes. In this paper, we take a top-down approach and make the strong assumption about the aggregate diffusion rate of technology adoption to construct an aggregate quality-adjusted price index. An alternative, bottom-up approach would weight the importance of each technology by connecting the observed technologies to the actual diffusion rates of those technologies. While this bottom-up approach is an enormous task, it may be feasible for future research given both the availability of the CEAR database and medical claims data sources. For both the top-down and bottom-up approaches, an advantage of using these cost-effectiveness studies is that they distill complex medical care products and services into well understood measures of quality, creating both a rigorous and feasible method for constructing quality-adjusting health care price indexes.

Changes in population health also have macroeconomic implications for welfare (Murphy and Topel 2006; and Jones and Klenow 2016). Recent work by Cutler et al. (2020) have taken steps to attribute observed change in population health and welfare to the medical care sector and find results consistent with estimates in this paper. The estimates in our paper and those in Cutler et al. (2020) take very distinct approaches, but both suggest that current estimates understate medical care output and productivity growth and both papers move us closer to more precise measurement of the medical care sector. Accounting for quality adjustment in the health care sector should make it more comparable to other sectors of our economy, such as computers, where quality adjustment has been applied for decades and price indexes show rapid price declines. However, there is more work to be done to apply consistent quality-adjusted indexes across sectors.

There are many areas that warrant additional research. While we find evidence of quality-adjusted prices declining in health care, this does not imply that U.S. health care is efficient. There may be large productivity gains from removing inefficiencies in the health care system (Glied and Sacarny 2018), regardless of technological progress. There are also important research challenges in determining how health factors that may be influenced by medical care treatment, such as smoking, obesity and drug abuse, are attributed to the medical care sector, which we do not address in this paper. In addition, it is possible that some technologies that seem beneficial may ex-post be harmful (e.g., opioids). Moreover, our findings on declining quality-adjusted prices do not inform health care equity. The declines in quality-adjusted prices may not be realized evenly across geographic, racial, or socioeconomic groups (Fleurbaey and Schokkaert 2011; and Case and Deaton 2020).[26]

[26]A related paper by Romley et al. (2020) adds to this literature, although it focuses on a distinct measure of medical care productivity. Similar to Cutler et al. (2020), Romley et al. (2020) analyze quality based on real-world health outcomes, but the paper focuses specifically on outcomes of well-understood conditions that account for 10% of Medicare A and B spending. Romley et al. (2020) shows medical care is productivity is understated if quality adjustments are not properly accounted for.

References

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li. (2019): "Missing Growth from Creative Destruction," American Economic Review, 109, August 2019, 2795-2822

Anderson, Robert N. (2001): "United States life tables, 1998" *National Vital Statistics Reports*, Center of Disease Control and Prevention, 48(18) 1-40.

Baumol, William J. (1967): "Macroeconomics of unbalanced growth: the anatomy of urban crisis," *The American Economic Review*, 57(3) (1967), 415-426.

Berndt. Ernst R, David M. Cutler, Richard Frank, Zvi Griliches, Joseph P. Newhouse, and Jack E. Triplett. (2007): "Price Indexes for Medical Care Goods and Services -- An Overview of Measurement Issues", in *Medical Care Output and Productivity*, Cutler and Berndt, 141-200.

Berndt, Ernst R., A. Bir, S. H. Busch, R. G. Frank, and S. T. Normand. (2002): "The Medical Treatment of Depression, 1991-1996: Productivity Inefficiency, Expected Outcome Variations, and Price Indexes," *Journal of Health Economics*, 21(3), 373-96.

Bresnahan, Timothy F., and Robert J. Gordon. (1996): "Introduction to "The Economics of New Goods"" in *The Economics of New Goods* ed (The Universirty of Chicago Press).

Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf. (2016): "Does the United States have a productivity slowdown or a measurement problem?" *Brookings Papers on Economic Activity*, 2016(1) 109-182.

Case, A., & Deaton, A. (2020). Deaths of Despair and the Future of Capitalism. Princeton University Press.

Center for the Evaluation of Value and Risk in Health. The Cost-Effectiveness Analysis Registry [Internet]. (Boston), Institute for Clinical Research and Health Policy Studies, Tufts Medical Center. Available from: www.cearegistry.org Accessed on February 2019.

Chandra, Amitahb, Amy Finkelstein, Adam Sacarny and Chad Syverson. (2016): "Healthcare Exceptionalism? Performance and Allocation in the U.S. Healthcare Sector," *American Economic Review*, 106 (8), 2110-2144.

Chhatwal, Jagpreet, Fasiha Kanwal, Mark S Roberts, and Michael A Dunn. (2015): "Cost-effectiveness and budget impact of hepatitis C virus treatment with sofosbuvir and ledipasvir in the United States" *Annals of Internal Medicine*, 162(6), 397-406

Chernew, Michael, and Joseph P. Newhouse. (2011): "Health Care Spending Growth," *Handbook of Health Economics*, 2, 1-43.

Cutler, David M., and Mark McClellan. (2001): "Is Technological Change In Medicine Worth It?" *Health Affairs*, 20(2), 11-29.

Cutler, David M., Michael Chernew, Kaushik Ghosh, and Mary Beth Landrum. (2017): "Understanding the Improvement in Disability Free Life Expectancy in the U.S. Elderly Population," NBER Working Paper No. 22306.

Cutler, David, Kaushik Ghosh, Kassandra Messer, Trivellore Raghunathan, Allison Rosen and Susan Stewert. (2020): "A Satellite Account for Health in the United States," National Bureau of Economic Research, Working Paper 27848.

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

Cutler, David M., Allison B. Rosen, and Sandeep Vijan. (2006): "The Value of Medical Spending in the United States, 1960-2000," *New England Journal of Medicine*, 355(9), 920-927.

Dauda, Seidu, Abe Dunn, Anne Hall, (2020), "Are Medical Care Prices Still Declining? A Systematic Examination of Quality-Adjusted Price Index Alternatives for Medical Care". Hutchins Center on Fiscal and Monetary Policy at Brookings Working Paper #59.

Duernecker, Georg, Berthold Herrendorf, and Akos Valentinyi. "Structural Change within the Services Sector and the Future of Cost Disease." Manuscript, ASU (2019).

Dunn, Abe, Scott D. Grosse, and Samuel H. Zuvekas. (2018): "Adjusting Health Expenditures for Inflation: A Review of Measures for Health Services Research in the United States," *Health Services Research*, 53(1), 175-196.

Dunn, Abe, Lindsey Rittmueller, and Bryn Whitmire. (2015): "Introducing the New BEA Health Care Satellite Account," *Survey of Current Business*, 95(1), 1-21.

Dunn, Abe, Bryn Whitmire, Andrea Batch, Lasanthi Fernando, and Lindsey Rittmueller. (2018): "High Spending Growth Rates for Key Diseases in 2000–14 Were Driven by Technology and Demographic Factors," *Health Affairs*, 37(6), 915-924.

Eggleston, Karen, Brian K. Chen, Chih-Hung Chen, Ying Isabel Chen, Talitha Feenstra, Toshiaki Iizuka, Janet Tinkei Lam, Gabriel M. Leung, Jui-fen Rachel Lu, Beatriz Rodriguez-Sanchez, Jeroen N. Struijs, Jianchao Quan, and Joseph P. Newhouse. (2019): "Are Quality-Adjusted Medical Prices Declining for Chronic Disease? Evidence from Diabetes Care in Four Health Systems", NBER Working Paper No. 25971.

Feenstra, Robert C. (1994): "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 84(1), 157-177.

Fleurbaey, M. and Schokkaert, E., 2011. Equity in health and health care. In Handbook of health economics (Vol. 2, pp. 1003-1092). Elsevier.

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

Glied, S., & Sacarny, A. (2018). Is the US health care system wasteful and inefficient? A review of the evidence. Journal of Health Politics, Policy and Law, 43(5), 739-765.

Groshen, Erica L., Brian Moyer, Ana Aizcorbe, Ralph Bradley, and David Friedman. (2017): "How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches," Journal of Economic Perspectives, 31(2), 187-210.

Hall, Anne E. (2016): "Adjusting the Measurement of the Output of the Medical Sector for Quality: A Review of the Literature," Medical Care Research and Review, 74(6), 639-667.

Hall, Robert E., and Charles I. Jones. (2007): "The value of life and the rise in health spending," *The Quarterly Journal of Economics*, 122(1), 39-72.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi. "Two perspectives on preferences and structural transformation." American Economic Review 103, no. 7 (2013): 2752-89.

Howard, David H., Bach, P. B., Berndt, E. R., and Conti, R. M. (2015): "Pricing in the market for anticancer drugs," *The Journal of Economic Perspectives*, 29(1), 139-162.

Hult, Kristopher J., Sonia Jaffe, and Tomas 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.

Jones, Charles I. and Peter J. Klenow. (2016): "Beyond GDP? Welfare across Countries and Time," *American Economic Review* 2016, 106(9): 2426–2457.

Kochanek, Kenneth D., S.L. Murphy, J. Xu, and E. Arias. (2017): "Mortality in the United States, 2016," NCHS Data Brief #293. The Centers for Disease Control and Prevention.

Lebow, David E., and Jeremy B. Rudd. (2003): "Measurement Error in the Consumer Price Index: Where Do We Stand?" *Journal of Economic Literature*, 41(1), 159-201.

Lucarelli, Claudio, and Sean Nicholson. (2009): "A Quality-Adjusted Price Index for Colorectal Cancer Drugs," NBER Working Paper No. w15174.

Martin, A. B., Hartman, H., Lassman D., and Catlin, A. (2020) "National Health Care Spending in 2019: Steady Growth for the Fourth Consecutive Year." *Health Affairs,* 40(1): 14-24.

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. (2010): "Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement," ed. (The National Academies Press).

Neumann, Peter J., Joshua T. Cohen, and Milton C. Weinstein. (2014): "Updating Cost-Effectiveness: The Curious Resilience of the $50,000-per-QALY Threshold," *The New England Journal of Medicine*, 371(9), 796-797.

The World Health Organization (WHO). (2011): "A System of Health Accounts," OECD.

Pakes, Ariel. (2004): "A Reconsideration of Hedonic Price Indexes with an Application to PC's," *The American Economic Review*, 93(5), 1578-1596.

Pandya, Ankur., S. Sy, S. Cho, M. C. Weinstein, and T. A. Gaziano. (2015): "Cost-effectiveness of 10-Year Risk Thresholds for Initiation of Statin Therapy for Primary Prevention of Cardiovascular Disease," *Journal of American Medical Association* 314(2), 142-150.

Petrin, Amil. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4), 705-729.

Redding, Stephen J., and David E. Weinstein. (2020): "Measuring Aggregate Price Indexes with Demand Shocks: Theory and Evidence from CES Preferences," *Quarterly Journal of Economics*, 135(1), 503–560.

Romley, J. A., Dunn, A., Goldman, D., & Sood, N. (2019). Quantifying Productivity Growth in Health Care Using Insurance Claims and Administrative Data. Big Data for 21st Century Economic Statistics.

Sanger-Katz, Margot. (2014): "Is a $1,000 Pill Really Too Much?", New York Times, August 3, available at https://www.nytimes.com/2014/08/03/upshot/is-a-1000-pill-really-too-much.html.

Sheiner, Louise, and Anna Malinovskaya. (2016): "Measuring Productivity in Healthcare: An Analysis of the Literature," Hutchins Center on Fiscal and Monetary Policy, Working Paper.

Shapiro, Irving, Matthew D. Shapiro, and David Wilcox. (2001): "Measuring the Value of Cataract Surgery," In Medical Care Output and Productivity ed (The Universtity of Chicago Press).

Syverson, Chad. (2017): "Challenges to Mismeasurement Explanations for the US Productivity Slowdown," Journal of Economic Perspectives, 31(2), 165-86.

Trajtenberg, Manuel. (1990): "Product Innovations, Price indices and the (Mis) Measurement of Economic Performance", NBER Working Paper

Viscusi, Kip and Joseph Aldy. (2003): "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World", NBER Working Paper No. 9487

Tables

Table I
Descriptive Statistics from CEAR Database

	Mean	Median	p5	p95	sd	Obs
Innovator QALY	9.84	8.01	0.27	25.56	10.11	11,483
SOC QALY	9.37	7.48	0.16	25.10	9.82	11,439
Innovator Price	$125,252	$25,561	$274	$439,475	$487,269	12,057
SOC Price	$107,004	$20,321	$82	$358,591	$466,716	12,032
Innovator Price per QALY	$25,570	$5,137	$30	$104,988	$156,375	11,362
SOC Price per QALY	$21,775	$4,366	$16	$95,132	$221,384	11,175
ICER	$71,745	$17,086	-$146,946	$429,649	$616,427	19,451
			ICER by Quadrant			
(NE) ↑ in Price & ↑ in QALY	$155,037	$32,812	$1,338	$553,284	$578,724	13,959
(SE) ↓ in Price & ↑ in QALY	-$156,455	-$20,713	-$594,903	-$330	$638,419	3,307
(NW) ↑ in Price & ↓ in QALY	-$244,105	-$40,725	-$1,246,364	-$40	$672,924	1,601
(SW) ↓ in Price & ↓ in QALY	$237,976	$41,981	$1,095	$1,064,275	$580,978	393
			ICER by Quadrant for the Price Index Sub-sample			
(NE) ↑ in Price & ↑ in QALY	$152,117	$33,461	$1,338	$516,176	$593,531	7,202
(SE) ↓ in Price & ↑ in QALY	-$174,227	-$23,821	-$680,445	-$546	$686,701	2,235
(NW) ↑ in Price & ↓ in QALY	-$241,260	-$39,650	-$1,246,364	-$188	$633,776	1,248
(SW) ↓ in Price & ↓ in QALY	$228,497	$38,301	$785	$1,035,877	$610,849	297

Notes. The estimates are the authors calculations based on the Tufts CEAR database. The estimates are converted to 2017 dollars using the PCE index for health by function from the Bureau of Economic Analysis. The Incremental Cost-Effectiveness Ratios (ICERs) are calculated as the difference in the innovator QALY and SOC QALY, divided by the difference in the innovator price and SOC price. The meaning of the ratio changes depending on whether prices or QALYs are increasing or decreasing, so we report all four quadrants that are North-east (NE), South-east (SE), North-west (NW), and South-west (SW). There are many studies that include ICERs that do not report the separate components of the ICER (i.e., the innovator QALY, SOC QALY, the innovator price and SOC price). To understand if the sample used to form quality-adjusted price indexes is different from this larger sample, we also provide descriptive statistics on the ICERs for the select subsample where all the components of the ICER are observed. The distribution of ICER estimates in the subsample are similar to the ICER estimates using the full sample.

Table II

Distribution of Quality-Adjusted Price Index Changes

	Mean	Median	p5	p95	sd	obs
Utility-Based COLI Index						
($50,000 VSLY)	-0.673	-0.050	-4.507	1.504	3.991	10,814
($100,000 VSLY)	-1.915	-0.182	-10.216	1.395	8.272	10,814
($150,000 VSLY)	-3.157	-0.320	-16.009	1.445	12.646	10,814
Utility-Based RP Index						
($50,000 VSLY)	-0.069	-0.061	-0.878	0.790	0.587	10,976
($100,000 VSLY)	-0.213	-0.187	-0.959	0.595	0.557	10,976
($150,000 VSLY)	-0.292	-0.291	-0.983	0.542	0.570	10,976
Average Utility-Based COLI and RP Indexes						
($50,000 VSLY)	-0.402	-0.054	-2.656	1.034	2.105	10,627
($100,000 VSLY)	-1.106	-0.178	-5.685	0.849	4.244	10,611
($150,000 VSLY)	-1.777	-0.300	-8.629	0.827	6.449	10,604

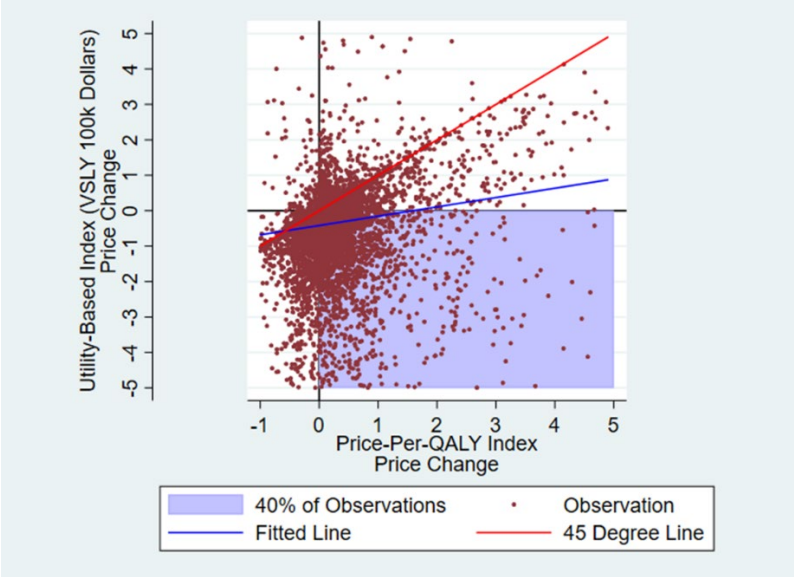
Notes. For each estimate outliers are removed based on the 1st and 99th percentiles for that index. The average utility-based index is calculated at the level of the study and is an average of the COLI Utility-Based Index and the RP, utility-based index. The average utility-based index is marked as an outlier if either the COLI utility-based index is an outlier, or the RP index is an outlier. This leads to a different number of observations across the rows.

Table III
Distribution of Alternative Price Index Changes

	Mean	Median	p5	p95	sd	obs
PPQ Index Change	0.333	0.041	-0.348	1.824	1.062	10,773
Implied VSLY	*$23,795*	*$5,141*	*$39*	*$101,325*	*$94,549*	
Hedonic Index Changes						
Log Regression (Table AI, Col 1)	0.351	0.048	-0.324	1.857	1.079	10,773
Implied VSLY	*$19,207*	*$4,296*	*$32*	*$84,697*	*$54,445*	
Quantile Regression Median (Table AI, Col 2)	0.324	0.036	-0.569	2.004	1.364	10,814
Implied VSLY	*$7,535*	*$7,535*	-	-	*$0*	
Quantile Regression 90th perc. (Table AI, Col 3)	0.056	0.007	-1.377	1.723	1.810	10,814
Implied VSLY	*$19,381*	*$19,381*	-	-	*$0*	

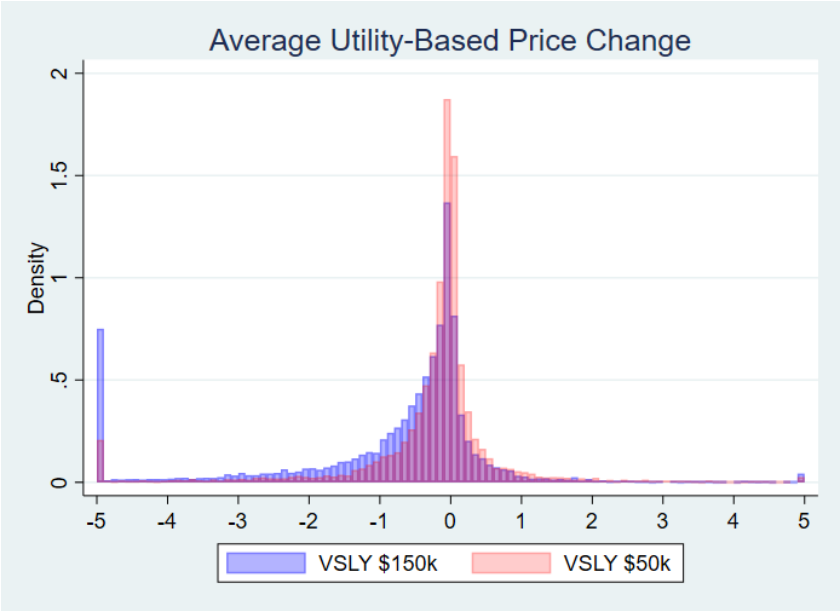
Notes. The PPQ and hedonic indexes show the value of a statistical life year implied by the quality adjustment. The log specification in the first hedonic regression is used to address outliers. The quantile regression is linear so they have only a single value implied by the estimates. Additional details regarding the hedonic regressions are provided in the appendix. The PPQ calculation is the average price per unit of health produced as described in the text and the hedonic is based on the calculation described in footnote 16 of the text. For each estimate outliers are removed based on the 1st and 99th percentiles for each index.

Figure 1. Scatterplot of Prices Change: Average Utility-based Index and PPQ Index



Notes. The figure shows the percent price change for the average of the utility-based COLI and RP indexes and the PPQ index price change. The estimates have been limited to the -5 to 5 range, so excludes some of the larger price changes. The red line is the 45-degree angle where estimates should fall if the price changes agree. The blue line is the fitted value of the price changes, which is only weakly positive. The shaded blue area indicates the quadrant where a plurality of 40 percent of observations fall, indicating the area where the PPQ index shows a price increase and the average utility-based index shows a price decline.

Figure 2. Histogram of Average Quality-Adjusted Price Change for VSLY of $50k and $150k



Notes. The figure shows a histogram of the average quality-adjusted price index change based on two values of a statistical life year: $50k and $150k. The figure has been winsorized at positive 500 percent and negative 500 percent.

Table IV

Select Examples with Price-per-QALY Index and COLI Index assuming VSLY of $100,000

	Price	QALY	COLI Index Price Change	RP Index Price Change	PPQ Index Price Change
1. Disease: Rheumatoid arthritis (includes patients with poor reactions to standard treatments)					
Innovation: Enbrel (Etanercept) plus methotrexate	$256,173	9.515	-1.78	-0.50	0.10
SOC: Methotrexate alone	$143,137	5.833			
Article: Soini et al. (2012)					
2. Disease: Prostate cancer					
Innovation: Modulated radiation therapy	$42,354	6.440	-2.20	-0.57	0.47
SOC: 3D conformal radiation therapy	$25,470	5.710			
Article: Konski et al. (2005)					
3. Disease: Diabetes					
Innovation: Self-management training program	$112,835	8.500	-0.67	-0.37	0.02
SOC: Usual care management	$100,496	7.700			
Article: Gilmer et al. (2007)					
4. Disease: Colon cancer					
Innovation: Colonoscopy every 10 years	$5,441	15.36	-20.12	-0.91	0.80
SOC: No screening	$2,179	14.75			
Article: Wong et al. (2015)					
5. Disease: HPV, cervical cancer					
Innovation: HPV Vaccine and Cancer Screening	$4,295	58.77	-161.19	-0.95	6.97
SOC: Cervical Cancer Screening	$531	57.88			
Article: Sopina and Ashton (2011)					

Notes. The table shows select studies from the CEAR database. All calculations are in 2017 dollars. References to the articles are included in the appendix.

Table V

Distribution of Price Change based on Average Utility-based Index for Innovations ($VSLY $100k): By Condition Category, Type of Intervention, Funding Sponsor, and Type of Study

	All			High Score		
Condition Category	Obs.	Mean	Median	Obs	Mean	Median
Cardiovascular	1,710	-0.86	-0.22	1,235	-0.76	-0.23
Digestive	516	-0.68	-0.08	313	-0.91	-0.11
Endocrine Disorders	873	-1.23	-0.30	634	-1.25	-0.26
Infectious	1,896	-1.98	-0.33	1,345	-1.95	-0.33
Malignant Neoplasms	2,219	-0.76	-0.11	1,543	-0.61	-0.11
Maternal/Child	75	-1.27	-0.07	41	-1.56	-0.08
Musculoskeletal/Rheumatologic	897	-0.62	-0.10	572	-0.67	-0.10
Neuro-Psychiatric/Neurological	829	-0.49	-0.09	584	-0.50	-0.08
Other	1144	-1.46	-0.22	600	-1.13	-0.16
Respiratory	287	-1.02	-0.28	195	-0.75	-0.23
Sense Organ	165	-2.28	-0.22	110	-2.56	-0.31
Intervention						
Care Delivery	357	-1.65	-0.21	207	-1.85	-0.17
Device	382	-1.24	-0.24	261	-1.04	-0.22
Diagnostic	398	-0.50	-0.05	279	-0.42	-0.03
Education	204	-0.61	-0.15	141	-0.36	-0.12
Immunization	300	-1.35	-0.25	225	-0.61	-0.18
Pharmaceutical	5,495	-1.06	-0.20	3,893	-0.96	-0.19
Procedure	1,274	-1.07	-0.21	796	-1.05	-0.22
Screening	1,416	-1.06	-0.04	893	-1.04	-0.03
Surgical	691	-1.62	-0.36	434	-2.09	-0.33
Funding Sponsor						
Foundation	748	-1.68	-0.08	525	-1.74	-0.06
Government	3,004	-0.98	-0.10	1,979	-0.98	-0.09
Health Care Organization	462	-0.78	-0.11	337	-0.82	-0.11
Other	2,956	-1.31	-0.20	1,941	-1.22	-0.18
Pharma or Device Manuf.	3,261	-0.93	-0.27	2,291	-0.79	-0.26
Prof Member Organization	180	-1.46	-0.04	99	-1.11	0.00
Type of Study						
Meta-Analysis	786	-0.46	-0.17	558	-0.43	-0.16
Other	1,887	-0.89	-0.11	980	-0.67	-0.09
Randomized	3,820	-0.92	-0.21	2,662	-0.87	-0.20
Simulation	4,118	-1.49	-0.17	2,972	-1.42	-0.17

Notes: The indexes are reported along four categorical dimensions in this table: disease chapter of the illness being treated, the type of intervention, the funding sponsor for the study, and the type of study based on words observed in the abstract. The reviewers of the medical studies that enter the studies in the CEA database score the quality of the research on various dimensions. An overall rating is included in the database indicating the quality of the study. Following Hult et al. we report overall estimates and estimates based only on those studies with a rating at or above the median of 5, indicating a higher quality study. The indexes at the bottom and top 1 percent of the distribution have been removed for the construction of this table.

Table VI. Estimated Annual Quality-Adjusted Price Index Growth Rate for Health Care Consumption 2000-2017, Relative to Economy-wide Inflation

Official PCE Health Index Growth Rate	0.53%
Disease-Price Growth Rate, HCSA	1.96%
Proxies for Share of Spending on Innovation Per Year	
Share of spending per year on new drugs	3.5%
Share of out-of-pocket spending per year on new drugs	2.3%
Share of hospitals changing practice patterns	3.6%
Share of spending per year on new procedures	1.9%
Average share across proxy measures of innovation	2.8%

Innovation Price Change	$VSLY		
	$50,000	$100,000	$150,000
Quality-Adjusted Price Change (Bottom of Table II)	-40%	-111%	-178%
(95 percent confidence interval)	(-48% -31%)	(-127% -94%)	(-48% -31%)
Annual Quality-Adjusted Disease-Price Growth Rate			
Diffusion Rate Low Estimate 1.9%	1.16%	-0.17%	-1.44%
(95 percent confidence interval)	(1.01% 1.32%)	(-0.48% 0.15%)	(-1.89% -0.98%)
Diffusion Rate Average 2.8%	0.77%	-1.23%	-3.13%
(95 percent confidence interval)	(0.53% 1.00%)	(-1.70% -0.76%)	(-3.82% -2.44%)
Diffusion Rate High Estimate 3.6%	0.43%	-2.12%	-4.55%
(95 percent confidence interval)	(0.19% 0.77%)	(-2.58% -1.42%)	(-5.22% -3.51%)

Notes. The official health care price index for consumption is taken from the PCE health index by function from the BEA. The disease-based index is the price measure for medical spending by disease from the HCSA. The 95 percent confidence interval is constructed using a bootstrap estimate where articles are sampled with replacement 1,000 times and estimates are formed for each of the 1,000 samples. The spending per year on new drugs and spending on new CPT codes is calculated based on the MarketScan claims and encounters database from IBM Watson. The changes in hospital practice patterns were calculated from the Hospital Compare database from the Centers for Medicare and Medicaid Services. The economy-wide GDP deflator has been applied to all estimates, making all estimates relative to economy-wide inflation.

Table VII. Estimated Annual Quality-Adjusted Price Index Growth Rate for Health Care Consumption 2000-2017 Weighted by Condition Categories, Relative to Economy-wide Inflation

				$VSLY		
				$50k	$100k	$150k
Condition Categories	# Obs. CEAR Database	Spending Share 2000	Unad. Ann. Disease Price Growth	Annual Quality-Adjusted Growth	Annual Quality-Adjusted Growth	Annual Quality-Adjusted Growth
Diseases of the circulatory system	1,710	16.2%	1.10%	0.2%	-1.4%	-2.9%
Diseases of the respiratory system	287	10.2%	1.50%	0.6%	-1.4%	-3.5%
Symptoms; signs; and ill-defined conditions	224	9.3%	2.19%	-0.4%	-4.0%	-7.7%
Diseases of the musculoskeletal system	897	8.5%	2.52%	2.0%	0.7%	-0.5%
Injury and poisoning	143	7.4%	2.46%	0.6%	-1.5%	-4.0%
Diseases of the nervous system	580	7.1%	2.94%	2.3%	0.3%	-1.1%
Neoplasms	2,219	6.7%	1.82%	1.3%	-0.4%	-2.1%
Diseases of the genitourinary system	354	6.4%	1.13%	-0.6%	-3.0%	-5.1%
Diseases of the digestive system	516	6.3%	2.36%	1.8%	0.4%	-1.0%
Endocrine; nutritional; and metabolic di	873	5.6%	1.51%	-0.1%	-2.0%	-3.6%
Mental illness	414	4.6%	0.53%	-0.2%	-1.2%	-2.2%
Complications of pregnancy; childbirth;	34	2.8%	1.81%	-1.7%	-4.9%	-7.9%
Infectious and parasitic diseases	1,896	2.8%	4.15%	1.6%	-1.6%	-4.6%
Diseases of the skin and subcutaneous or	121	2.5%	2.73%	2.3%	0.8%	-0.4%
Residual codes; unclassified; all E code	179	1.5%	2.41%	0.2%	-2.4%	-4.9%
Diseases of the blood and blood-forming	91	1.0%	2.58%	1.6%	-0.6%	-2.8%
Certain conditions originating in the pe	21	0.5%	2.49%	2.0%	-0.2%	-2.3%
Congenital anomalies	20	0.5%	1.55%	2.6%	2.1%	1.8%

Annual Quality-Adjusted Disease-Price Growth Rate, Weighted by Condition Category			
$VSLY	$50k	$100k	$150k
Diffusion Rate Low Estimate 1.9%	1.09%	-0.25%	-1.54%
(95 percent confidence interval)	(0.92% 1.23%)	(-0.59% 0.04%)	(-2.10% -1.10%)
Diffusion Rate Average 2.8%	0.67%	-1.33%	-3.26%
(95 percent confidence interval)	(0.42.% 0.89%)	(-1.85% -0.89%)	(-4.07% -2.58%)
Diffusion Rate High Estimate 3.6%	0.33%	-2.23%	-4.70%
(95 percent confidence interval)	(0.00% 0.60%)	(-2.89% -1.67%)	(-5.73% -3.82%)

Notes. The disease-based indexes and spending shares by broad condition category are from the HCSA. The cost-effectiveness studies and BEA Health Care Satellite Account mapped together using broad disease category information from CEAR and the Health Care Satellite Account. The 95 percent confidence interval is constructed using a bootstrap estimate where articles are sampled with replacement 1,000 times and estimates are formed for each of the 1,000 samples. Diffusion rates are based on estimates from Table VI The economy-wide GDP deflator has been applied to all estimates, making all estimates relative to economy-wide inflation, which grew at 1.9 percent per year over the period of study, which grew at 1.9 percent per year over the period of study.

Online Appendix

1. Hedonics

Hedonic regressions are commonly applied to adjust prices for quality changes, especially for high-tech goods, and therefore offer another benchmark to compare to our preferred utility-based indexes. The theory used to justify hedonics relies on the assumption that consumers are responsive to prices, suggesting that hedonic regressions are not well-suited for health care where the full price is not paid for by the patient, although recent evidence suggests that patients are responsive to quality changes (Chandra et al. 2016). We apply a first-difference hedonic model that takes the difference in price and QALY between the innovative and SOC treatments from the same study, which differences out common article-specific factors affecting price and QALYs. To address the skewness in the data, we estimate the first difference using the log functional form:

$$\log\left({}^{S_{Ii}}/_{S_{SOCi}}\right) = \alpha \log\left({}^{H_{Ii}}/_{H_{SOCi}}\right) + \beta X_i + \varepsilon_i$$

where α reflects the change in price associated with the change in quality, and X_i are covariates for other factors that may explain differences in price (e.g., publication year, type of intervention, or health condition).

The first column of Table AI shows the results of this hedonic regression. The α estimate of 0.817 implies that a 10 percent increase in quality yields an 8.2 percent increase in price. At the median SOC price of $20,321 and SOC QALY of 7.48 reported in Table I , this suggests an implicit cost of improving health by one QALY of $2,032, which is implausibly low, and just 2 percent of our central utility value of $100,000.

As an alternative method to address skewness, the second column of Table AI shows the conditional median price of a QALY as estimated by a quantile regression and a linear functional form. The coefficient indicates the value of a QALY is $7,536, again just a small fraction of estimates of the utility value of a QALY. Even when the quantile regression is applied to the upper quartile, shown in column (3), we find an implied value of a QALY

of just $19,381. Regardless of the specification, we find the hedonic method places an extraordinarily low value on QALYs, which is consistent with Pakes (2004) who warns that hedonics tend to undervalue quality changes.[27] For purposes of comparison, we apply hedonic regressions to form adjusted price indexes using standard formulas applied in the literature.[28]

2. Cost-effectiveness and Quality-Adjusted Price Indexes

Table AII shows the distribution of the average of the RP and COLI utility-based index change using a VSLY of $100k for different increments of the ICER. The top portion of the table shows the quality-adjusted price index based on different quadrants of the ICER ratio. A small number of innovations are strictly welfare reducing (NW quadrant). These innovations are associated with quality-adjusted price increases. In contrast, innovations with both decreasing prices and increasing quality (SE quadrant) are highly efficient and account for around 20 percent of the observations and are associated with steep quality-adjusted price declines. The bulk of the observations are the case of both higher cost and higher quality (NE quadrant), which are also generally associated with quality-adjusted prices declining. The magnitude of the decline in the NE quadrant depends on the ICER value, with low ICER values associated with steep price declines. The bottom of the table shows the quality-adjusted prices for different ICER values. The American College of Cardiology (ACC) and the American Heart Association (AHA) consider a cost-effectiveness ratio of $50,000 or less as highly effective, which is associated with quality-adjusted price declines of over 100 percent at the mean and large price declines at the median (Dubois (2015)). This table, then, suggests that standard thresholds for technology adoption are associated with steep quality-adjusted price declines.

---

[27] We tested a variety of alternative functional forms and specifications and the results consistently show that hedonics produce low implicit valuations on QALYs.

[28] Taking the regression estimate $\hat{\alpha}\left(Log(Inn.\,QALY/SOC\;QALY)\right)_i$ we can calculate the hypothetic price absent the quality improvement, where $\hat{s}^0_{1,i} = exp(\log(S_{1,i}) - \hat{\alpha}\left(Log(Inn.\,QALY/SOC\;QALY)\right)_i$. The index formula is then: $Hedonic = \frac{\hat{s}^0_1}{S_0}$, which is the change in price, holding the quality fixed to the level of the SOC treatment.

3. Alternative Quality-Adjusted Indexes

In this appendix, we conduct several robustness checks of our results.

Alternative functional forms We apply alternative functional forms that are commonly used in the price index literature, including the Cobb-Douglas function and Constant Elasticity of Substitution (CES) function. Based on economic theory, the arithmetic mean that we applied to our main estimates is very close to the price index corresponding to the Cobb-Douglas function, assuming each innovation has an equal chance of being adopted. For the Cobb-Douglas function, the associated price index functional form is a geometric mean. The Cobb-Douglas utility function produces conservative estimates, in the sense that the quality-adjustments have a smaller effect, as it assumes that the cross-price elasticity is zero, implying no substitution among treatments, even as technologies improve. In addition to being simple to apply, the arithmetic mean is even more conservative than the geometric mean, as the arithmetic mean places relatively more weight on the larger price increases.

Denote each innovation by i so that the price index formula constructed using the quality-adjusted price for innovation i is P_t^i. The price index formula associated with the Cobb-Douglas function is the geometric mean, where α_i is a parameter that indicates the importance of the technology: $P_t^{agg} = \prod_{i=1}^{N} \left(P_t^i\right)^{\alpha_i}$.

For each innovation i, we construct an index, $P_t^i = (1 - \Delta w) \cdot \frac{S_t}{S_{t-1}} + (\Delta w) \cdot \frac{S_{I,i} - \$VSLY \cdot (H_{I,i} - H_{SOC,i})}{S_{SOC,i}}$, using equation (7). Next, we assign each of the N innovations equal weights, so $\alpha_i = 1/N$. Including these values, the price index from $t-1$ to t is:

$$P_t^{agg} = \prod_{i=1}^{N} \left((1 - \Delta w) \cdot \frac{S_t}{S_{t-1}} + (\Delta w) \cdot \frac{S_{I,i} - \$VSLY \cdot (H_{I,i} - H_{SOC,i})}{S_{SOC,i}} \right)^{\left(\frac{1}{N}\right)}. \quad \text{(A1)}$$

As we will show, the estimates are quite close to those based on the arithmetic mean. In contrast to the Cobb-Douglas function, the CES function allows for positive cross-price elasticities among innovations, allowing for arguably more realistic substitution among alternatives. Relative to the Cobb-Douglas function, the CES will generate more

substitution toward innovations that show larger declines in price, potentially generating larger quality-adjustments and lower price indexes. We calculate the estimate based on a CES price index formula assuming elasticities of substitution of either 2 or 3.[29] The price index associated with the CES-function is: $P_t^{agg,CES} = \prod_{i=1}^{N}\left(\alpha_i (P_t^i)^{(1-\sigma)}\right)^{\frac{1}{1-\sigma}}$.

The parameter σ indicates the elasticity of substitution between products and α_i indicates the importance of the technology, which we assign equal weight, as in the Cobb-Douglas calculation. Substituting in equation (7) to the CES-function (A1) and the assumption of equal weights we get:

$$P_t^{agg,CES} = \prod_{i=1}^{N}\left(\frac{1}{N}\left((1-\Delta w)\cdot\frac{S_t}{S_{t-1}} + (\Delta w)\cdot\frac{S_{I,i} - \$VSLY\cdot(H_{I,i}-H_{SOC,i})}{S_{SOC,i}}\right)^{(1-\sigma)}\right)^{\frac{1}{1-\sigma}}. \quad \text{(A2)}$$

For both the Cobb-Douglas and CES function, only positive values are allowed, so any extremely large price declines are dropped from the analysis. For the unweighted index, we apply this aggregation across all technologies. For the weighted index, we apply the indexes (A1) or (A2) at the condition category level and then aggregate based on the spending share of each condition category.

Table AVI reports these results, where we report estimates assuming a diffusion rate of 2.8 percent across scenarios, and based on different VSLY, as the assumption regarding the VSLY appeared to be more important than the different diffusion rates. Lines 1 and 2 repeat estimates from Tables VI and VII to compare as a baseline. Lines 3 and 4 repeat lines 1 and 2, but apply the index using the Cobb-Douglas functional form. The estimates are quite close to those using the arithmetic mean, as expected, although the arithmetic mean gives greater weight to price increases, relative to the geometric mean. Line 5 and 6 repeats lines 1 and 2, but apply the CES functional form with $\sigma = 2$, which allows a small degree of substitution among innovations. Allowing for some substitution shows much larger quality-adjusted price declines for both the weighted and unweighted estimates, suggesting

[29] The elasticity of substitution across many industries is estimated to be 3 or more based on a CES-type function (Feenstra 1994, Redding and Weinstein 2020; and Aghion et al. 2020).

that the degree of substitution among technologies have an important effect on the price decline. Line 7 is identical to line 6, but allows for a higher elasticity of substitution, $\sigma = 3$. In this case, the quality-adjusted price index falls extremely fast, at a rate of 13 percent per year. As we obtain similar estimates for both the weighted and unweighted values, for the remainder of the robustness checks we only provide the preferrable weighted estimates.

Restricting studies to those from high-income countries

One limitation is that some of these studies are from developing countries which are not necessarily applicable to the U.S., although it is arguable that countries such as China conduct cost-effectiveness studies that are relevant to the U.S.[30] In lines 8, 9 and 10 of Table AVI we repeat the estimates from 4, 6, and 7, respectively, but only use studies from high income countries.[31] The estimates show less of a decline, but the quality-adjusted price decline is still substantial for the estimates using the CES functional form.

Only including innovations that improve welfare

Another assumption is that the estimates assume that those technologies that reduce welfare (i.e., lead to a quality-adjusted price increase) are actually adopted. One reasonable restrictive assumption is that those less efficient technologies are never used. Lines 11, 12, and 13 of Table AVI repeat the estimates from 8, 9, and 10, respectively, but assume that only those technologies that lead to quality-adjusted price declines are adopted and diffused (i.e., only technologies that lead to lower quality-adjusted prices are adopted). These estimates tend to lead to large quality-adjusted price declines for our central estimate when VSLY is set at $100k.

Removing extreme values

[30] For example, studies from China are often advanced comparable treatments to those in the U.S., such as Icotinib, a first line treatment for lung cancer, Apatinib, a treatment for gastric cancer, and Yisaipu, a rheumatoid arthritis treatment.

[31] The grouping of high income countries comes from a list maintained by the World Bank: https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html.

A more general concern may be that the estimates could be affected by large negative or positive values. While we argue that some of those large quality-adjusted price changes are, in fact, very relevant, it is important to check the robustness of the estimates to their exclusion. For the last set of robustness checks, we exclude extreme values, including those technologies that are not welfare improving, but also the 5 percent of technologies that lead to the largest quality-adjusted price declines. Specifically, lines 14, 15, and 16, repeat the estimates from lines 4, 6 and 7, respectively, but removing the more extreme values. We find that across all three specifications, quality-adjusted prices are declining for our central estimate based on the VSLY of $100k, but the magnitude of the decline is smaller at around 0.5 percent per year.

Appendix Tables

Table AI
Hedonic Regressions

	(1)	(2)	(3)
	OLS Regression	Quantile Reg. (Median)	Quantile Reg. (Upper Quartile)
Dependent Variable	Log(Inn. Price) − Log(SOC Price)	Inn. Price − SOC Price	Inn. Price − SOC Price
Log(Innovator QALY) − Log(SOC QALY)	0.817		
	(0.0610)		
Innovator QALY − SOC QALY		7535.5	19380.6
		(494.7)	(1525.1)
Number of Observations	10639	10639	10639
Adjusted R2	0.157		

Notes. Standard errors in parentheses. Standard errors are clustered by condition category for all estimates. The top and bottom 1 percent of outliers are removed based on PPQ index percentiles. The second and third columns apply quantile regressions that predict the conditional median (column 2) and the conditional 75th percentile (column 3). All regressions include controls for the condition category, the publication year and the type of intervention (e.g., pharmaceutical or procedure).

Table AII. Average Utility-Based COLI and RP Index Price Changes by Cost-Effectiveness Type based on the Incremental-Cost-Effectiveness-Ratio (ICER)

		Average Utility-based Index ($100k VSLY)				
Technology Type by ICER Category	obs	mean	median	p5	p95	sd
(NE) ↑ in Price & ↑ in QALY	308	0.408	0.034	-0.393	2.510	1.114
(SE) ↓ in Price & ↑ in QALY	1,010	0.707	0.275	0.009	2.609	1.321
(NW) ↑ in Price & ↓ in QALY	2,279	-1.124	-0.395	-4.134	-0.030	3.112
(SW) ↓ in Price & ↓ in QALY	7,014	-1.428	-0.202	-7.237	0.646	4.809
Total	10,611					
Average Quality-Adjusted Price Change: ↑ in Price & ↑ in QALY:						
ICER<$1,000 & ICER≥0	224	-8.665	-3.288	-37.571	-0.028	12.125
ICER<$10,000 & ICER≥$1,000	1,278	-4.002	-1.086	-17.181	-0.061	7.561
ICER<$50,000 & ICER≥$10,000	2,651	-1.280	-0.476	-4.884	-0.021	2.983
ICER<$100,000 & ICER≥$50,000	1,107	-0.361	-0.089	-0.976	-0.001	1.727
ICER<$200,000 & ICER≥$100,000	727	0.230	0.064	0.001	0.959	0.536
ICER≥$200,000	1,027	0.651	0.234	0.002	2.757	1.139
Total	7,014					

Notes. The table shows the average quality-adjusted price index by the characteristics of the technology. The technology is categorized into four quadrants of either increasing or decreasing quality and cost, as is typical in the cost effectiveness literature. For the quadrant of increasing cost and increasing quality, we break out technologies by the incremental cost-effectiveness ratio.

Table AIII. Diffusion Rates of Specific Treatments

	Time Period	Initial Percentage	Final Percentage	Total Percent Growth	Annual Rate
Disease: Rheumatoid arthritis (includes patients with poor reactions to standard treatments)					
Innovation: Enbrel diffusion rate	2001-2006	0.035	0.217	0.182	0.036
Any biologic disease-modifying antirheumatic drugs	2001-2006	0.035	0.369	0.334	0.067
Source: Krishnan et al. (2012)					
Disease: Prostate cancer	2007-2012	0.286	0.380	0.094	0.019
Innovation: Modulated radiation therapy					
Source: Shahinian et al. (2017)					
Disease: Colon cancer	2000-2018	0.382	0.668	0.286	0.016
Innovation: Colon cancer screening					
Source: https://progressreport.cancer.gov/detection/colorectal_cancer					
Disease: HPV, cancer	2008-2018	0.37	0.70	0.33	0.033
Innovation: HPV Vaccine					
Source: https://progressreport.cancer.gov/prevention/hpv_immunization					

Notes. This table shows diffusion rates based on studies or reports specific to each condition. The selected diffusion rates were selected to correspond to the examples in Table IV in the text. We were able to find corresponding diffusion rates that correspond to the selected examples, except for diabetes management. The diffusion rate ranged from 1.6 percent per year to 6.7 percent per year.

Table AIV. New Molecule and New CPT Code Shares

	New Molecule Spend					CPT Code Spend	
	Pill count share	Spending Share	Difference in Spending Share	Out-of-pocket Spend Share	Difference in Spending Share	Spending Share	Difference in Spending Share
2002	0.47%	0.85%	0.85%	0.99%	0.99%	1.22%	1.22%
2003	1.57%	3.34%	2.48%	3.20%	2.21%	2.99%	1.78%
2004	2.94%	6.51%	3.18%	5.99%	2.79%	5.15%	2.16%
2005	3.67%	8.71%	2.19%	8.04%	2.04%	7.36%	2.21%
2006	5.10%	12.37%	3.67%	11.35%	3.32%	9.93%	2.56%
2007	6.29%	16.10%	3.73%	14.44%	3.08%	12.15%	2.23%
2008	6.76%	18.81%	2.70%	16.74%	2.30%	12.75%	0.59%
2009	7.00%	20.79%	1.98%	18.15%	1.41%	14.78%	2.03%
2010	7.32%	23.97%	3.17%	20.35%	2.20%	16.47%	1.70%
2011	7.72%	27.86%	3.89%	22.57%	2.22%	19.77%	3.30%
2012	7.98%	32.35%	4.49%	24.59%	2.01%	20.72%	0.95%
2013	8.07%	36.22%	3.86%	26.04%	1.45%	23.56%	2.84%
2014	8.40%	41.82%	5.61%	27.98%	1.94%	25.20%	1.63%
2015	9.34%	47.86%	6.04%	32.26%	4.28%	26.89%	1.69%
2016	10.34%	52.28%	4.42%	34.63%	2.37%	28.49%	1.60%
2017	11.71%	56.30%	4.01%	36.81%	2.18%	30.22%	1.73%
Average total change		3.519%		2.300%		1.889%	

Notes. The estimates from this table come from authors calculation based on MarketScan claims data. The calculation is the share of spending on new molecules or new CPT codes, where a new molecule or CPT code is counted if it is not observed in the year 2000 or 2001. The population includes both the commercially insured population under 65, as well as retirees over 65 observed in the database.

Table AV. Average Change in Process Measures of Quality from Hospital Compare Database

Process Measure for Patients Given:	Percent of patients given the following recommended treatment 2004	2009	% Increase	Level Change
Condition: Heart Attack				
ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction	82	96	16.8%	13.8%
Aspirin at Arrival	94	98	4.2%	4.0%
Aspirin at Discharge	94	98	4.2%	4.0%
Beta Blocker at Discharge	93	98	6.0%	5.5%
Smoking Cessation Advice/Counseling	87	99	14.9%	12.9%
Condition: Heart Failure				
ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction	81	94	16.2%	13.1%
Assessment of Left Ventricular Function	87	98	11.7%	10.2%
Discharge Instructions	52	88	68.9%	35.7%
Smoking Cessation Advice/Counseling	74	98	33.5%	24.7%
Condition: Pneumonia				
Patients Assessed and Given Pneumococcal Vaccination	52	93	78.0%	40.6%
Initial Antibiotic(s) within 4 Hours After Arrival	72	95	31.1%	22.5%
Smoking Cessation Advice/Counseling	70	97	38.5%	27.0%
The Most Appropriate Initial Antibiotic(s)	77	91	18.3%	14.1%
Blood Culture Performed Prior to First Antibiotic Received in Hospital	82	95	15.7%	12.9%
Surgical Infection Prevention				
Received Preventative Antibiotic(s) One Hour Before Incision	77	96	24.7%	19.1%
Preventative Antibiotic(s) are Stopped Within 24 hours After Surgery	64	94	45.7%	29.5%
Average Change Across Categories				18.1%
Average Annual Change Across Categories				3.62%

Notes. The estimates from this table come from authors calculation from the Hospital Compare database archives from the Center for Medicare and Medicaid services (https://data.medicare.gov/data/archives/hospital-compare). The estimates are based on a simple weighted average across all hospitals in the database where the weight is determined by the sample size at each hospital. Quality measures that were discontinued or continued in the middle of the sample range are not shown. The year reported in this table is based on the year the information was gathered from the hospital, which is typically lagged one year in the database. For instance, the process measures for 2004 are from the 2005 hospital compare database.

Table AVI. Alternative Quality-Adjusted Price Index Average Annual Growth Rates for Health Care Consumption 2000-2017

	$VSLY			
Alternative Models with diffusion rate of 2.8 percent	$50k	$100k	$150k	Obs
1. Baseline, Unweighted (Table VI)	0.77%	-1.23%	-3.13%	10,611
2. Baseline, Weighted (Table VII)	0.67%	-1.33%	-3.26%	10,579
3. Cobb-Douglas, Unweighted	0.52%	-1.39%	-2.95%	10,562
4. Cobb-Douglas, Weighted	0.44%	-1.70%	-3.03%	10,532
5. CES Model - σ = 2, Unweighted	0.14%	-4.27%	-7.60%	10,562
6. CES Model - σ = 2, Weighted	0.08%	-5.19%	-7.61%	10,532
7. CES Model - σ = 3, Weighted	-0.50%	-13.03%	-17.62%	10,532
8. Cobb-Douglas, Weighted, High Income Countries	0.89%	-0.69%	-2.23%	9,731
9. CES Model - σ = 2, Weighted, High Income Countries	0.69%	-1.73%	-4.67%	9,731
10. CES Model - σ = 3, Weighted, High Income Countries	-0.39%	-7.18%	-13.13%	9,731
11. Cobb-Douglas, Weighted, High Income Countries, Welfare Improving	-0.37%	-2.16%	-3.96%	7,387
12. CES model - σ = 2, Weigh., High Inc. Cntry, Welfare Improving	-0.65%	-3.47%	-6.91%	7,387
13. CES model - σ = 3, Weigh., High Inc. Cntry, Welfare Improving	-0.41%	-9.35%	-15.90%	7,387
14. Cobb-Douglas, Weigh., Drop 5% Largest Declines, Welfare Improving	0.61%	-0.45%	-1.54%	7,241
15. CES model - σ = 2, Weigh., Drop 5% Largest Declines, Welf. Imp.	0.59%	-0.51%	-1.69%	7,241
16. CES model - σ = 3, Weighted, Drop 5% Largest Declines, Welf. Imp.	0.58%	-0.58%	-1.86%	7,241

Notes. Each line corresponds to estimates of quality-adjusted price changes applied under alternative scenarios. We show the quality-adjusted price index based on alternative assumptions of the VSLY, which we found to be more important than the diffusion rate. Across scenarios we assume the diffusion rate is equal to our central diffusion rate of 2.8 percent per year. The first two lines correspond to baseline estimates from Tables VI and VII where the Table VI estimate is unweighted and Table VII is weighted. The baseline models use simple arithmetic averaging, which places more weight on large price increases, leading to less of a quality adjustment. The estimates in rows 3 and 4 use a price index based on the Cobb-Douglas utility function, which applies a geometric mean, rather than an arithmetic mean, and the results are similar to the baseline estimates, but with slightly larger price declines. The Cobb-Douglas utility function allows for no cross-price substitution. A more realistic scenario is to allow for greater substitution among treatments, which is shown using a CES Models in rows 5 (unweighted) and 6 (weighted) where we assume an elasticity of substitution parameter of 2 (allowing for a small amount of substitution to better technologies), leading to much larger price declines relative to the estimates in rows 3 and 4. As we believe the weighted estimates are preferred, we focus on those for the remaining robustness checks. We also focus on estimates grounded in particular utility functions, such as CES or Cobb Douglas, which are commonly used in the literature. Row 7 is the same as row 6, but the elasticity of substitution is set to 3, showing even sharper price declines. Rows 8, 9, and 10 are identical to rows 4, 6 and 7 but include only high income countries. Rows 11, 12 and 13 are identical to rows 8,9 and 10 but only include innovations that lead to quality adjusted price declines. Finally, Rows 14, 15 and 16 eliminates all innovations that show large changes by excluding innovations that are not welfare improving, but also excluding 5 percent of the observations that show the largest declines in quality-adjusted prices.

Table AVII. Hypothetical Adjustment to BLS Multifactor Productivity Estimate for Hospitals and Nursing and Residential Care Facilities (NAICS 622, 623)

	BLS (current)					Adjusted Price, Output, and Productivity			
	Output price index	Real output	Annual output price index growth	Real combined inputs	Multifactor productivity index (real output/real input)	Adj. annual output price index growth	Adj. output price index (rebased)	Adj. real output	Adj. productivity
2000	72.38	62.60	4.4%	61.34	1.02	2.6%	89.54	50.60	0.82
2001	73.78	65.69	1.9%	64.44	1.02	0.1%	89.67	54.05	0.84
2002	75.73	70.15	2.6%	68.79	1.02	0.8%	90.43	58.75	0.85
2003	78.42	72.08	3.5%	70.68	1.02	1.7%	91.99	61.45	0.87
2004	81.47	73.61	3.9%	71.44	1.03	2.1%	93.88	63.88	0.89
2005	84.66	77.59	3.9%	76.34	1.02	2.1%	95.84	68.53	0.90
2006	87.65	79.96	3.5%	78.70	1.02	1.7%	97.49	71.89	0.91
2007	90.17	82.33	2.9%	81.90	1.01	1.1%	98.53	75.35	0.92
2008	93.05	85.17	3.2%	83.25	1.02	1.4%	99.89	79.34	0.95
2009	94.51	89.01	1.6%	85.85	1.04	-0.2%	99.67	84.40	0.98
2010	96.25	92.59	1.8%	90.97	1.02	0.0%	99.72	89.37	0.98
2011	99.02	94.93	2.9%	94.16	1.01	1.1%	100.79	93.26	0.99
2012	100.00	100.00	1.0%	100.00	1.00	-0.8%	100.00	100.00	1.00
2013	101.81	101.69	1.8%	102.27	0.99	0.0%	100.02	103.51	1.01
2014	103.88	103.83	2.0%	104.74	0.99	0.2%	100.26	107.58	1.03
2015	104.81	109.67	0.9%	110.31	0.99	-0.9%	99.38	115.66	1.05
2016	106.03	114.82	1.2%	116.02	0.99	-0.6%	98.77	123.26	1.06
2017	108.06	117.64	1.9%	119.82	0.98	0.1%	98.89	128.54	1.07
Avg. Annual Rate	2.39%	3.78%		4.02%	-0.23%		0.59%	5.64%	1.56%

Notes: The BLS estimates of multifactor productivity taken from the table of productivity for the nonmanufacturing industries (https://www.bls.gov/mfp/mprdload.htm). The output price index, real output, real combined inputs, and multifactor productivity are from the table BLS 1987–2019 Combined Sector and Industry Multifactor Productivity. The annual output price index growth rate is adjusted by the amount of the bias of 1.79 percent per year, which is computed by taking the difference between the growth rate in the output price index of 0.46 per year and the quality-adjusted index of -1.33 percent per year. The adjusted price index affects output and productivity by the amount of this bias, so the multifactor productivity increases by 1.79 percent per year, increasing productivity from -0.23 percent to 1.56 percent.

Appendix References

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., & Li, H. (2019). Missing growth from creative destruction. American Economic Review, 109(8), 2795-2822.

Dubois. (2015): "Cost–effectiveness thresholds in the USA: are they coming? Are they already here?", *Journal of Comparative Effectiveness Research*, 5(1), 9-11

Gilmer, T. P., Roze, S., Valentine, W. J., Emy-Albrecht, K., Ray, J. A., Cobden, D., ... & Palmer, A. J. (2007). Cost-effectiveness of diabetes case management for low-income populations. *Health services research*, 42(5), 1943-1959.

Konski, A., Sherman, E., Krahn, M., Bremner, K., Beck, J. R., Watkins-Bruner, D., & Pilepich, M. (2005). "Economic analysis of a phase III clinical trial evaluating the addition of total androgen suppression to radiation versus radiation alone for locally advanced prostate cancer (Radiation Therapy Oncology Group protocol 86-10)." *International Journal of Radiation Oncology** Biology* Physics, 63(3), 788-794.

Krishnan, E., Lingala, B., Bruce, B., & Fries, J. F. (2012). "Disability in rheumatoid arthritis in the era of biological treatments." *Annals of the Rheumatic Diseases*, 71(2), 213-218.

Shahinian, V. B., Kaufman, S. R., Yan, P., Herrel, L. A., Borza, T., & Hollenbeck, B. K. (2017). "Reimbursement and use of intensity-modulated radiation therapy for prostate cancer." *Medicine*, 96(25).

Soini, E. J., Hallinen, T. A., Puolakka, K., Vihervaara, V., & Kauppi, M. J. (2012). "Cost-effectiveness of adalimumab, etanercept, and tocilizumab as first-line treatments for moderate-to-severe rheumatoid arthritis." *Journal of Medical Economics*, 15(2), 340-351.

Sopina, E., & Ashton, T. (2011). Cost-effectiveness of a cervical screening program with human papillomavirus vaccine. *International journal of technology assessment in health care*, 27(4), 290.

Wong, M. C., Ching, J. Y., Chan, V. C., & Sung, J. J. (2015). The comparative cost-effectiveness of colorectal cancer screening using faecal immunochemical test vs. colonoscopy. *Scientific reports*, 5(1), 1-9.